(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2024-509881
(P2024-509881A)
(43)公表日　令和6年3月5日(2024.3.5)

(51)国際特許分類		FI		テーマコード（参考）
H04N 19/52 (2014.01)		H04N	19/52	5C159
H04N 19/85 (2014.01)		H04N	19/85	

審査請求　未請求　予備審査請求　未請求　（全108頁）

(21)出願番号　特願2023-554362(P2023-554362)
(86)(22)出願日　令和4年1月27日(2022.1.27)
(85)翻訳文提出日　令和5年9月6日(2023.9.6)
(86)国際出願番号　PCT/US2022/014143
(87)国際公開番号　WO2022/191933
(87)国際公開日　令和4年9月15日(2022.9.15)
(31)優先権主張番号　17/198,813
(32)優先日　令和3年3月11日(2021.3.11)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,
最終頁に続く

(71)出願人　507364838
クアルコム，インコーポレイテッド
アメリカ合衆国 カリフォルニア ９２１２１ サン ディエゴ モアハウス ドライブ ５７７５
(74)代理人　100108453
弁理士　村山 靖彦
(74)代理人　100163522
弁理士　黒田 晋平
(72)発明者　レザ・プレッザ
アメリカ合衆国・カリフォルニア・９２１２１・サン・ディエゴ・モアハウス・ドライヴ・５７７５
(72)発明者　タコ・セバスティアーン・コーヘン
アメリカ合衆国・カリフォルニア・９２
最終頁に続く

(54)【発明の名称】　Ｐフレームコーディングシステムを使用する学習型Ｂフレームコーディング

(57)【要約】
単方向コーディングシステムおよび補間された参照フレームを使用する学習型双方向コーディングを実行することなどによって、ビデオデータを処理するための技法が説明される。たとえば、プロセスは、第1の参照フレームおよび第2の参照フレームを取得することを含むことができる。プロセスは、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成することを含むことができる。プロセスは、入力フレームと第3の参照フレームとの間の動きを推定すること、および少なくとも部分的には推定された動きに基づいて第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成することなどによって、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行することを含むことができる。プロセスは、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することを含むことができ、再構成されたフレームは双方向予測フレームを含む。

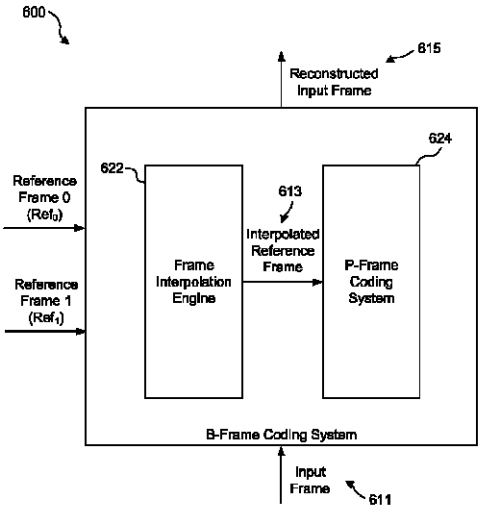


FIG. 6

【特許請求の範囲】
【請求項１】
ビデオデータを処理するためのシステムであって、
補間エンジンと単方向コーディングエンジンとを含む双方向コーディングエンジンを備え、
前記補間エンジンが、
第1の参照フレームおよび第2の参照フレームを取得し、
少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成するように構成され、
前記単方向コーディングエンジンが、
前記第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行するように構成される、
システム。
【請求項２】
前記単方向コーディングエンジンが動きエンジンおよびワーピングエンジンを含み、
前記動きエンジンが、前記入力フレームと前記第3の参照フレームとの間の動きを推定するように構成され、
前記ワーピングエンジンが、少なくとも部分的には前記推定された動きに基づいて前記第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成するように構成される、
請求項1に記載のシステム。
【請求項３】
前記単方向コーディングエンジンが残差エンジンをさらに含み、
前記残差エンジンが、
少なくとも部分的には前記入力フレームと前記ワーピングされたフレームとの間の差分を決定することによって、残差を決定し、
前記残差を使用して予測残差を生成するように構成される、
請求項2に記載のシステム。
【請求項４】
メモリと、前記メモリに結合された少なくとも1つのプロセッサとをさらに備え、前記単方向コーディングエンジンが、
前記ワーピングされたフレームおよび前記予測残差に基づいて、前記入力フレームを表す再構成されたフレームを生成するように構成され、前記再構成されたフレームが双方向予測フレームを含み、
前記少なくとも1つのプロセッサが、前記再構成されたフレームを前記メモリの中に記憶させるように構成される、
請求項3に記載のシステム。
【請求項５】
前記少なくとも1つのプロセッサがニューラル処理ユニット(NPU)を含む、請求項4に記載のシステム。
【請求項６】
前記動きエンジンが第1のニューラルネットワークを含み、前記残差エンジンが第2のニューラルネットワークを含む、請求項3に記載のシステム。
【請求項７】
前記第1のニューラルネットワークが第1のオートエンコーダを含み、前記第2のニューラルネットワークが第2のオートエンコーダを含む、請求項6に記載のシステム。
【請求項８】
前記単方向コーディングエンジンが、
第4の参照フレームを取得し、
追加の入力フレームと前記第4の参照フレームとの間の動きを推定し、

少なくとも部分的には前記推定された動きに基づいて前記第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成し、
前記追加のワーピングされたフレームおよび追加の予測残差に基づいて、前記追加の入力フレームを表す再構成されたフレームを生成するように構成され、前記追加の入力フレームを表す前記再構成されたフレームが単方向予測フレームを含む、
請求項1に記載のシステム。
【請求項９】
少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間を実行することによって、前記第3の参照フレームを生成するために、前記補間エンジンが、
前記第1の参照フレームから前記第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定し、
前記第2の参照フレームから前記第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定し、
少なくとも部分的には動き情報の前記第1のセットを使用して前記第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成し、
少なくとも部分的には動き情報の前記第2のセットを使用して前記第2の参照フレームに対して前記ワーピング関数を実行することによって、第2のワーピング情報を生成し、
前記第1のワーピング情報および前記第2のワーピング情報に基づいて前記第3の参照フレームを生成するように構成される、
請求項1に記載のシステム。
【請求項１０】
動き情報の前記第1のセットおよび動き情報の前記第2のセットが、前記第1の参照フレームと前記第2の参照フレームとの間のピクセル動きに基づいて決定される、請求項9に記載のシステム。
【請求項１１】
動き情報の前記第1のセットが第1のオプティカルフローマップを含み、動き情報の前記第2のセットが第2のオプティカルフローマップを含む、請求項9に記載のシステム。
【請求項１２】
前記ワーピング関数が双線形補間関数を含む、請求項9に記載のシステム。
【請求項１３】
前記入力フレームと前記第3の参照フレームとの間の前記推定された動きがオプティカルフローデータを含む、請求項2に記載のシステム。
【請求項１４】
前記入力フレームと前記第3の参照フレームとの間の前記推定された動きが動的畳み込みデータを含む、請求項2に記載のシステム。
【請求項１５】
前記入力フレームと前記第3の参照フレームとの間の前記推定された動きがブロックベースの動きデータを含む、請求項2に記載のシステム。
【請求項１６】
モバイルデバイス、エクステンデッドリアリティデバイス、またはテレビジョンのうちの1つである、請求項1に記載のシステム。
【請求項１７】
ディスプレイ、および1つまたは複数のビデオフレームをキャプチャするように構成されたカメラのうちの少なくとも1つをさらに備える、請求項1に記載のシステム。
【請求項１８】
ビデオデータを処理するための方法であって、
第1の参照フレームおよび第2の参照フレームを取得するステップと、
少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間

を実行することによって、第3の参照フレームを生成するステップと、
前記第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行するステップと
を備える方法。
【請求項１９】
前記入力フレームと前記第3の参照フレームとの間の動きを推定するステップと、
少なくとも部分的には前記推定された動きに基づいて前記第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成するステップと
をさらに備える、請求項18に記載の方法。
【請求項２０】
少なくとも部分的には前記入力フレームと前記ワーピングされたフレームとの間の差分を決定することによって、残差を決定するステップと、
前記残差を使用して予測残差を生成するステップと
をさらに備える、請求項19に記載の方法。
【請求項２１】
前記ワーピングされたフレームおよび前記予測残差に基づいて、前記入力フレームを表す再構成されたフレームを生成するステップであって、前記再構成されたフレームが双方向予測フレームを含む、ステップと、
前記再構成されたフレームをメモリの中に記憶させるステップと
をさらに備える、請求項20に記載の方法。
【請求項２２】
前記動きが第1のオートエンコーダを使用して推定され、前記予測残差が第2のオートエンコーダを使用して生成される、請求項20に記載の方法。
【請求項２３】
第4の参照フレームを取得するステップと、
追加の入力フレームと前記第4の参照フレームとの間の動きを推定するステップと、
少なくとも部分的には前記推定された動きに基づいて前記第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成するステップと、
前記追加のワーピングされたフレームおよび追加の予測残差に基づいて、前記追加の入力フレームを表す再構成されたフレームを生成するステップと
をさらに備え、前記追加の入力フレームを表す前記再構成されたフレームが単方向予測フレームを含む、
請求項18に記載の方法。
【請求項２４】
前記第1の参照フレームから前記第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定するステップと、
前記第2の参照フレームから前記第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定するステップと、
少なくとも部分的には動き情報の前記第1のセットを使用して前記第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成するステップと、
少なくとも部分的には動き情報の前記第2のセットを使用して前記第2の参照フレームに対して前記ワーピング関数を実行することによって、第2のワーピング情報を生成するステップと、
前記第1のワーピング情報および前記第2のワーピング情報に基づいて前記第3の参照フレームを生成するステップと
をさらに備える、請求項18に記載の方法。
【請求項２５】

動き情報の前記第1のセットおよび動き情報の前記第2のセットが、前記第1の参照フレームと前記第2の参照フレームとの間のピクセル動きに基づいて決定される、請求項24に記載の方法。

【請求項２６】

動き情報の前記第1のセットが第1のオプティカルフローマップを含み、動き情報の前記第2のセットが第2のオプティカルフローマップを含む、請求項24に記載の方法。

【請求項２７】

前記ワーピング関数が双線形補間関数を含む、請求項24に記載の方法。

【請求項２８】

前記入力フレームと前記第3の参照フレームとの間の前記推定された動きがオプティカルフローデータを含む、請求項19に記載の方法。

【請求項２９】

前記入力フレームと前記第3の参照フレームとの間の前記推定された動きが動的畳み込みデータを含む、請求項19に記載の方法。

【請求項３０】

前記入力フレームと前記第3の参照フレームとの間の前記推定された動きがブロックベースの動きデータを含む、請求項19に記載の方法。

【請求項３１】

命令を記憶した非一時的コンピュータ可読媒体であって、前記命令が、1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、

第1の参照フレームおよび第2の参照フレームを取得させ、

少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成させ、

前記第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行させる、

非一時的コンピュータ可読媒体。

【請求項３２】

前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、

入力フレームと前記第3の参照フレームとの間の動きを推定させ、

少なくとも部分的には前記推定された動きに基づいて前記第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成させる命令をさらに備える、

請求項31に記載の非一時的コンピュータ可読媒体。

【請求項３３】

前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、

少なくとも部分的には前記入力フレームと前記ワーピングされたフレームとの間の差分を決定することによって、残差を決定させ、

前記残差を使用して予測残差を生成させる命令をさらに備える、

請求項32に記載の非一時的コンピュータ可読媒体。

【請求項３４】

前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、

前記ワーピングされたフレームおよび前記予測残差に基づいて、前記入力フレームを表す再構成されたフレームを生成することであって、前記再構成されたフレームが双方向予測フレームを含む、生成することと、

前記再構成されたフレームをメモリの中に記憶させることと

を行わせる命令をさらに備える、

請求項33に記載の非一時的コンピュータ可読媒体。

【請求項３５】
前記動きが第1のオートエンコーダを使用して推定され、前記予測残差が第2のオートエンコーダを使用して生成される、請求項33に記載の非一時的コンピュータ可読媒体。
【請求項３６】
前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、
第4の参照フレームを取得させ、
追加の入力フレームと前記第4の参照フレームとの間の動きを推定させ、
少なくとも部分的には前記推定された動きに基づいて前記第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成させ、
前記追加のワーピングされたフレームおよび追加の予測残差に基づいて、前記追加の入力フレームを表す再構成されたフレームを生成させる命令をさらに備え、前記追加の入力フレームを表す前記再構成されたフレームが単方向予測フレームを含む、
請求項31に記載の非一時的コンピュータ可読媒体。
【請求項３７】
前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、
前記第1の参照フレームから前記第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定させ、
前記第2の参照フレームから前記第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定させ、
少なくとも部分的には動き情報の前記第1のセットを使用して前記第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成させ、
少なくとも部分的には動き情報の前記第2のセットを使用して前記第2の参照フレームに対して前記ワーピング関数を実行することによって、第2のワーピング情報を生成させ、
前記第1のワーピング情報および前記第2のワーピング情報に基づいて前記第3の参照フレームを生成させる命令をさらに備える、
請求項31に記載の非一時的コンピュータ可読媒体。
【請求項３８】
ビデオデータを処理するためのシステムであって、
メモリと、
前記メモリに結合された少なくとも1つのプロセッサと
を備え、前記少なくとも1つのプロセッサが、
補間された参照フレーム、および入力フレームと前記補間された参照フレームとの間の動きを表す動き情報を取得し、
前記動き情報に基づいて前記入力フレームと前記補間された参照フレームとの間の前記動きを推定し、
少なくとも部分的には前記補間された参照フレームの1つまたは複数のピクセルを前記推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成し、
前記ワーピングされたフレームおよび予測残差に基づいて、前記入力フレームを表す再構成されたフレームを生成するように構成され、前記再構成されたフレームが双方向予測フレームを含む、
システム。
【請求項３９】
前記少なくとも1つのプロセッサが、
第1の参照フレームおよび第2の参照フレームを取得し、
少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間

を実行することによって、前記補間された参照フレームを生成するように構成される、
請求項38に記載のシステム。
【請求項４０】
前記少なくとも1つのプロセッサが、
ビットストリームから残差を取得し、
前記取得された残差に基づいて前記予測残差を生成するように構成される、
請求項38に記載のシステム。
【請求項４１】
前記動き情報が前記ビットストリームから取得される、請求項40に記載のシステム。
【請求項４２】
前記少なくとも1つのプロセッサが、
第4の参照フレームを取得し、
追加の入力フレームと前記第4の参照フレームとの間の動きを推定し、
少なくとも部分的には前記推定された動きに基づいて前記第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成し、
前記追加のワーピングされたフレームおよび追加の予測残差に基づいて、前記追加の入力フレームを表す再構成されたフレームを生成するように構成され、前記追加の入力フレームを表す前記再構成されたフレームが単方向予測フレームを含む、
請求項38に記載のシステム。
【請求項４３】
ビデオデータを処理するための方法であって、
補間された参照フレーム、および入力フレームと前記補間された参照フレームとの間の動きを表す動き情報を取得するステップと、
前記動き情報に基づいて前記入力フレームと前記補間された参照フレームとの間の前記動きを推定するステップと、
少なくとも部分的には前記補間された参照フレームの1つまたは複数のピクセルを前記推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成するステップと、
前記ワーピングされたフレームおよび予測残差に基づいて、前記入力フレームを表す再構成されたフレームを生成するステップと
を備え、前記再構成されたフレームが双方向予測フレームを含む、
方法。
【請求項４４】
第1の参照フレームおよび第2の参照フレームを取得するステップと、
少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間を実行することによって、前記補間された参照フレームを生成するステップと
をさらに備える、請求項43に記載の方法。
【請求項４５】
ビットストリームから残差を取得するステップと、
前記取得された残差に基づいて前記予測残差を生成するステップと
をさらに備える、請求項43に記載の方法。
【請求項４６】
前記動き情報が前記ビットストリームから取得される、請求項45に記載の方法。
【請求項４７】
第4の参照フレームを取得するステップと、
追加の入力フレームと前記第4の参照フレームとの間の動きを推定するステップと、
少なくとも部分的には前記推定された動きに基づいて前記第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成するステップと、

前記追加のワーピングされたフレームおよび追加の予測残差に基づいて、前記追加の入力フレームを表す再構成されたフレームを生成するステップと
をさらに備え、前記追加の入力フレームを表す前記再構成されたフレームが単方向予測フレームを含む、
請求項43に記載の方法。
【請求項４８】
命令を記憶した非一時的コンピュータ可読媒体であって、前記命令が、1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、
補間された参照フレーム、および入力フレームと前記補間された参照フレームとの間の動きを表す動き情報を取得させ、
前記動き情報に基づいて前記入力フレームと前記補間された参照フレームとの間の前記動きを推定させ、
少なくとも部分的には前記補間された参照フレームの1つまたは複数のピクセルを前記推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成させ、
前記ワーピングされたフレームおよび予測残差に基づいて、前記入力フレームを表す再構成されたフレームを生成させ、前記再構成されたフレームが双方向予測フレームを含む、
非一時的コンピュータ可読媒体。
【請求項４９】
前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、
第1の参照フレームおよび第2の参照フレームを取得させ、
少なくとも部分的には前記第1の参照フレームと前記第2の参照フレームとの間の補間を実行することによって、前記補間された参照フレームを生成させる命令をさらに備える、
請求項48に記載の非一時的コンピュータ可読媒体。
【請求項５０】
前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、
ビットストリームから残差を取得させ、
前記取得された残差に基づいて前記予測残差を生成させる命令をさらに備える、
請求項48に記載の非一時的コンピュータ可読媒体。
【請求項５１】
前記動き情報が前記ビットストリームから取得される、請求項50に記載の非一時的コンピュータ可読媒体。
【請求項５２】
前記1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、
第4の参照フレームを取得させ、
追加の入力フレームと前記第4の参照フレームとの間の動きを推定させ、
少なくとも部分的には前記推定された動きに基づいて前記第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成させ、
前記追加のワーピングされたフレームおよび追加の予測残差に基づいて、前記追加の入力フレームを表す再構成されたフレームを生成させる命令をさらに備え、前記追加の入力フレームを表す前記再構成されたフレームが単方向予測フレームを含む、
請求項48に記載の非一時的コンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】

【０００１】
本開示は、一般に、ビデオコーディング(たとえば、ビデオデータを符号化および/または復号すること)に関する。たとえば、本開示の態様は、単方向コーディングシステム(たとえば、単方向予測フレーム(Pフレーム)エンコーダデコーダまたはコーデック)を使用する学習型双方向予測フレーム(Bフレーム)コーディング(learned bidirectional-predicted frame (B-frame) coding)に関する。

【背景技術】

【０００２】
多くのデバイスおよびシステムは、消費のためにビデオデータが処理および出力されることを可能にする。デジタルビデオデータは、ビデオ品質、性能、および機能において増大する需要を満たすために大量のデータを含む。たとえば、ビデオデータの消費者は、通常、高い忠実度、解像度、フレームレートなどを有する高品質ビデオを所望する。これらの需要を満たすために頻繁に必要とされる大量のビデオデータは、通信ネットワーク、ならびにビデオデータを処理および記憶するデバイスに、著しい負担をかける。ビデオデータを圧縮するためにビデオコーディング技法が使用され得る。ビデオコーディングの1つの例示的な目標は、ビデオ品質の劣化を回避するかまたは最小限に抑えながら、ビデオデータを圧縮してもっと低いビットレートを使用する形態にすることである。絶えず進化するビデオサービスが利用可能になり大量のビデオデータにおける需要が高まるとともに、より良好な性能および効率を有するコーディング技法が必要とされる。

【発明の概要】

【課題を解決するための手段】

【０００３】
参照フレーム間を補間することによって生成された参照フレームを利用できる、単方向(たとえば、Pフレーム)コーディングシステムを使用する学習型双方向(たとえば、Bフレーム)コーディングを含む、システムおよび技法が本明細書で説明される。少なくとも1つの例によれば、ビデオデータを処理するための方法が提供される。方法は、第1の参照フレームおよび第2の参照フレームを取得することと、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成することと、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行することとを含むことができる。

【０００４】
別の例では、ビデオデータを処理するためのシステムが提供される。システムは、補間エンジンと単方向コーディングエンジンとを含む双方向コーディングエンジンを含むことができる。場合によっては、システムは、少なくとも1つのメモリと、メモリに結合された(たとえば、回路構成の中に実装される)1つまたは複数のプロセッサとを含むことができる。補間エンジンは、第1の参照フレームおよび第2の参照フレームを取得し、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成するように構成される。単方向コーディングエンジンは、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行するように構成される。

【０００５】
別の例では、ビデオデータを処理するための非一時的コンピュータ可読媒体が提供される。非一時的コンピュータ可読媒体は、その上に記憶された命令を含むことができ、命令は、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに、第1の参照フレームおよび第2の参照フレームを取得させ、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成させ、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行させる。

【０００６】
別の例によれば、ビデオデータを処理するための装置が提供される。装置は、第1の参

照フレームおよび第2の参照フレームを取得するための手段と、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成するための手段と、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行するための手段とを含むことができる。

【０００７】

いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、入力フレームと第3の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成することとを含むことができる。

【０００８】

いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、少なくとも部分的には入力フレームとワーピングされたフレームとの間の差分を決定することによって、残差を決定することと、残差を使用して予測残差を生成することとを含むことができる。

【０００９】

いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することであって、再構成されたフレームが双方向予測フレームを含むことと、再構成されたフレームをメモリの中に記憶させることとを含むことができる。

【００１０】

いくつかの態様では、動きは第1のオートエンコーダを使用して推定され、予測残差は第2のオートエンコーダを使用して生成される。

【００１１】

いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、第4の参照フレームを取得することと、追加の入力フレームと第4の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成することと、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成することとを含むことができ、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。

【００１２】

いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、第1の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定することと、第2の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定することと、少なくとも部分的には動き情報の第1のセットを使用して第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成することと、少なくとも部分的には動き情報の第2のセットを使用して第2の参照フレームに対してワーピング関数を実行することによって、第2のワーピング情報を生成することと、第1のワーピング情報および第2のワーピング情報に基づいて第3の参照フレームを生成することとを含むことができる。

【００１３】

いくつかの態様では、動き情報の第1のセットおよび動き情報の第2のセットは、第1の参照フレームと第2の参照フレームとの間のピクセル動きに基づいて決定される。

【００１４】

いくつかの態様では、動き情報の第1のセットは第1のオプティカルフローマップを含み、動き情報の第2のセットは第2のオプティカルフローマップを含む。

【００１５】

いくつかの態様では、ワーピング関数は双線形補間関数を含む。

【００１６】

いくつかの態様では、入力フレームと第3の参照フレームとの間の推定された動きは、オプティカルフローデータを含む。

【００１７】

いくつかの態様では、入力フレームと第3の参照フレームとの間の推定された動きは、動的畳み込みデータを含む。

【００１８】

いくつかの態様では、入力フレームと第3の参照フレームとの間の推定された動きは、ブロックベースの動きデータを含む。

【００１９】

少なくとも1つの例によれば、ビデオデータを処理するための方法が提供される。方法は、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得することと、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定することと、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成することと、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することとを含むことができ、再構成されたフレームは双方向予測フレームを含む。

【００２０】

別の例では、ビデオデータを処理するためのシステムが提供される。システムは、少なくとも1つのメモリと、メモリに結合された(たとえば、回路構成の中に実装される)1つまたは複数のプロセッサとを含むことができる。少なくとも1つのプロセッサは、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得し、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定し、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成し、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成するように構成され、再構成されたフレームは双方向予測フレームを含む。

【００２１】

別の例では、ビデオデータを処理するための非一時的コンピュータ可読媒体が提供される。非一時的コンピュータ可読媒体は、その上に記憶された命令を含むことができ、命令は、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得させ、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定させ、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成させ、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成させ、再構成されたフレームは双方向予測フレームを含む。

【００２２】

別の例によれば、ビデオデータを処理するための装置が提供される。装置は、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得するための手段と、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定するための手段と、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成するための手段と、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成するための手段とを含むことができ、再構成されたフレームは双方向予測フレームを含む。

【００２３】
いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、第1の参照フレームおよび第2の参照フレームを取得することと、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、補間された参照フレームを生成することとを含むことができる。
【００２４】
いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、ビットストリームから残差を取得することと、取得された残差に基づいて予測残差を生成することとを含むことができる。
【００２５】
いくつかの態様では、動き情報はビットストリームから取得される。
【００２６】
いくつかの態様では、上記で説明した方法、装置、および非一時的コンピュータ可読媒体は、第4の参照フレームを取得することと、追加の入力フレームと第4の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成することと、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成することとを含むことができ、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。
【００２７】
いくつかの態様では、装置は、モバイルデバイス、ウェアラブルデバイス、エクステンデッドリアリティデバイス(extended reality device)(たとえば、仮想現実(VR)デバイス、拡張現実(AR)デバイス、または複合現実(MR)デバイス)、パーソナルコンピュータ、ラップトップコンピュータ、サーバコンピュータ、テレビジョン、ビデオゲーム機、または他のデバイスを含むことができるか、またはそれらの一部であり得る。いくつかの態様では、装置はモバイルデバイス(たとえば、携帯電話、またはいわゆる「スマートフォン」)を備える。いくつかの態様では、装置は、1つまたは複数の画像またはビデオフレームをキャプチャするための少なくとも1つのカメラをさらに含む。たとえば、装置は、1つもしくは複数の画像および/またはビデオフレームを含む1つもしくは複数のビデオをキャプチャするための1つのカメラ(たとえば、RGBカメラ)または複数のカメラを含むことができる。いくつかの態様では、装置は、1つまたは複数の画像、ビデオ、通知、または他の表示可能なデータを表示するためのディスプレイを含む。いくつかの態様では、装置は、再構成されたビデオフレームを、送信媒体を介して少なくとも1つのデバイスへ送信するように構成された、送信機を含む。いくつかの態様では、プロセッサは、ニューラル処理ユニット(NPU)、中央処理ユニット(CPU)、グラフィックス処理ユニット(GPU)、または他の処理デバイスもしくは構成要素を含む。
【００２８】
本概要は、特許請求される主題の主要なまたは必須の特徴を特定することは意図されず、特許請求される主題の範囲を決定するために独立して使用されることも意図されない。本主題は、この特許の明細書全体、いずれかまたはすべての図面、および各請求項の適切な部分を参照することによって理解されるはずである。
【００２９】
上記のことは、他の特徴および実施形態とともに、以下の明細書、特許請求の範囲、および添付図面を参照すると、より明らかになろう。
【００３０】
本出願の例示的な実施形態が、以下の図を参照しながら以下で詳細に説明される。
【図面の簡単な説明】
【００３１】
【図１】本開示のいくつかの例による、本明細書で説明する様々な技法を実施できる例示

的な画像処理システムを示す図である。
【図２Ａ】本開示のいくつかの例による全結合ニューラルネットワークの一例を示す図である。
【図２Ｂ】本開示のいくつかの例による局所結合ニューラルネットワークの一例を示す図である。
【図２Ｃ】本開示のいくつかの例による畳み込みニューラルネットワークの一例を示す図である。
【図２Ｄ】本開示のいくつかの例による、画像から特徴を認識するように設計された深層畳み込みネットワーク(ＤＣＮ)の詳細な例を示す図である。
【図３】本開示のいくつかの例による別の例示的なＤＣＮを示すブロック図である。
【図４】本開示のいくつかの例による、予測を生成するために単方向の動き圧縮および補償を実行するように構成された、単方向予測フレーム(Ｐフレーム)コーディングシステムの一例を示す図である。
【図５Ａ】本開示のいくつかの例による、予測を生成するために双方向の動き圧縮および補償を実行するように構成された、双方向予測フレーム(Ｂフレーム)コーディングシステムの一例を示す図である。
【図５Ｂ】本開示のいくつかの例による、フレーム補間を使用して予測を生成するように構成された、Ｂフレームコーディングシステムの一例を示す図である。
【図６】本開示のいくつかの例による、Ｐフレームコーディングシステムおよびフレーム補間エンジンを使用する学習型Ｂフレームコーディングを実行するように構成された、Ｂフレームコーディングシステムの一例を示す図である。
【図７Ａ】本開示のいくつかの例による、単方向コーディングおよび/または双方向コーディングを実行するために使用され得るＰフレームコーディングシステムの一例を示す図である。
【図７Ｂ】本開示のいくつかの例による、Ｐフレームコーディングシステムを使用する双方向コーディングを実行するために使用され得るＢフレームコーディングシステムの一例を示す図である。
【図７Ｃ】本開示のいくつかの例による、Ｂフレームコーディングシステムによって使用され得るフレーム補間エンジンの一例を示す図である。
【図８】本開示のいくつかの例による、動き推定技法の例を示す図である。
【図９】本開示のいくつかの例による、Ｉフレーム、Ｂフレーム、Ｉフレーム(ＩＢＩ)コーディング技法の一例を示す図である。
【図１０】本開示のいくつかの例による、Ｉフレーム、Ｂフレーム、Ｐフレーム(ＩＢＰ)コーディング技法の一例を示す図である。
【図１１】本開示のいくつかの例による、双方向フレーム予測のための例示的な逐次方式を示す図である。
【図１２】本開示のいくつかの例による、双方向フレーム予測のための例示的な階層方式を示す図である。
【図１３】図１３Ａおよび図１３Ｂは、本開示のいくつかの例による、本明細書で説明するＢフレームコーディングシステムの実行結果を示すグラフである。
【図１４Ａ】本開示のいくつかの例による、本明細書で説明するＢフレームコーディングシステムのさらなる実行結果を示すグラフである。
【図１４Ｂ】本開示のいくつかの例による、本明細書で説明するＢフレームコーディングシステムのさらなる実行結果を示すグラフである。
【図１４Ｃ】本開示のいくつかの例による、本明細書で説明するＢフレームコーディングシステムのさらなる実行結果を示すグラフである。
【図１４Ｄ】本開示のいくつかの例による、本明細書で説明するＢフレームコーディングシステムのさらなる実行結果を示すグラフである。
【図１４Ｅ】本開示のいくつかの例による、本明細書で説明するＢフレームコーディングシステムのさらなる実行結果を示すグラフである。

【図１４Ｆ】本開示のいくつかの例による、本明細書で説明するBフレームコーディングシステムのさらなる実行結果を示すグラフである。
【図１５】本開示のいくつかの例による、ビデオデータを処理するためのプロセスの一例を示すフローチャートである。
【図１６】本開示のいくつかの例による、ビデオデータを処理するためのプロセスの別の例を示すフローチャートである。
【図１７】本明細書で説明する様々な態様を実施するために使用され得る例示的なコンピューティングシステムを示す図である。
【発明を実施するための形態】
【００３２】
本開示のいくつかの態様および実施形態が、以下で提供される。当業者に明らかになるように、これらの態様および実施形態のうちのいくつかが独立して適用されてよく、それらのうちのいくつかが組合せで適用されてよい。以下の説明では、説明の目的のために、本出願の実施形態の完全な理解をもたらすために具体的な詳細が記載される。しかしながら、様々な実施形態がこれらの具体的な詳細なしに実践され得ることが明らかとなろう。図および説明は限定的であることを意図しない。
【００３３】
以下の説明は、例示的な実施形態を提供するにすぎず、本開示の範囲、適用可能性、または構成を限定することを意図しない。そうではなく、例示的な実施形態の以下の説明は、例示的な実施形態を実施することを可能にする説明を当業者に提供する。添付の特許請求の範囲に記載されるような本出願の範囲から逸脱することなく、要素の機能および構成において様々な変更が加えられてよいことを理解されたい。
【００３４】
デジタルビデオデータは、特に高品質のビデオデータに対する需要が高まり続けているので大量のデータを含むことができる。たとえば、ビデオデータの消費者は、通常、高い忠実度、解像度、フレームレートなどを有する、ますます高い品質のビデオを所望する。しかしながら、そのような需要を満たすために頻繁に必要とされる大量のビデオデータは、通信ネットワーク、ビデオデータを処理および記憶するデバイスにも、著しい負担をかける場合がある。
【００３５】
ビデオデータをコーディングするために様々な技法が使用され得る。場合によっては、ビデオコーディングは、特定のビデオコーディング規格および/または方式に従って実行され得る。例示的なビデオコーディング規格は、特に、高効率ビデオコーディング(ＨＥＶＣ)、高度ビデオコーディング(ＡＶＣ)、ムービングピクチャエキスパートグループ(ＭＰＥＧ)コーディング、多用途ビデオコーディング(ＶＶＣ)を含む。ビデオコーディング技法の1つの目標は、ビデオ品質の劣化を回避するかまたは最小限に抑えながら、ビデオデータを圧縮してもっと低いビットレートを使用する形態にすることである。ビデオサービスに対する需要が高まり新たなビデオサービスが利用可能になるにつれて、より良好な効率および性能を有するコーディング技法が必要とされる。
【００３６】
ビデオコーディングは、ビデオフレームまたは画像もしくはフレームの他のシーケンスの中に存在する冗長性を利用するイントラ予測またはインター予測などの予測方法を使用することができる。イントラ予測は、ビデオの単一のフレーム内のデータを使用して実行され、フレームの空間特性に基づく。イントラ予測を使用してコーディングされたフレームは、Ｉフレームと呼ばれる。フレームのインター予測は、他のフレームに対するフレームの時間特性に基づいて実行される。たとえば、ビデオフレームのインター予測は、そのビデオフレームに対して変化を含む他のビデオフレームの領域、およびそのビデオフレームに対して冗長性を含む領域(たとえば、大部分は変化しないままである背景領域)を識別することによって実行され得る。冗長性は除去することができ、ビデオフレームに対する残差が得られる。残差は(たとえば、エントロピーコーディングを使用して)さらに符号化

することができ、その結果は、記憶されるか、送信されるか、またはさもなければ出力される、ビットストリームの中に含められ得る。

【００３７】

インター予測の例は、単方向予測(単予測)および双方向予測(双予測)を含む。単予測は、フレームのインター予測を実行するときの、単一の参照フレームの使用を含む。単予測を使用してコーディングされたフレームは、Pフレームと呼ばれる。双予測は、フレームのインター予測を実行するときの、2つの参照フレームの使用を伴う。双予測を使用してコーディングされたフレームは、Bフレームと呼ばれる。

【００３８】

場合によっては、ビデオ符号化(圧縮)および復号(解凍)を実行するために機械学習システムが使用され得る。一般に、機械学習(ML)は人工知能(AI)のサブセットである。MLシステムは、パターンおよび推定に依拠することによって様々なタスクを実行するためにコンピュータシステムが使用できる、アルゴリズムおよび統計モデルを含む。MLシステムの一例は、人工ニューロン(たとえば、ニューロンモデル)の相互結合されたグループを含むことができるニューラルネットワーク(人工ニューラルネットワークとも呼ばれる)である。ニューラルネットワークは、特に、画像解析および/またはコンピュータビジョン適用例、インターネットプロトコル(IP)カメラ、モノのインターネット(IoT)デバイス、自律車両、サービスロボットなどの、様々な適用例および/またはデバイスのために使用され得る。

【００３９】

ニューラルネットワークの中の個々のノードは、入力データを取りデータに対して演算を実行することによって、生体ニューロンをエミュレートしてよい。入力データに対して実行された演算の結果が、他のニューロンに選択的に渡される。ネットワークの中の各ベクトルおよびノードに重み値が関連付けられ、これらの値は入力データが出力データにどのように関係するのかを制約する。たとえば、各ノードの入力データは、対応する重み値で乗算されてよく、その積が合計されてよい。積の合計は随意のバイアスによって調整されてよく、その結果に活性化関数が適用されてよく、ノードの出力信号すなわち「出力活性化」(活性化マップまたは特徴マップと呼ばれることがある)を与える。重み値は、ネットワークを通じてデータをトレーニングする反復フローによって最初に決定されてよい。たとえば、特定のクラスをそれらの典型的な入力データ特性によってどのように識別すべきかをネットワークが学習するトレーニングフェーズ中に、重み値が確立されてよい。一例では、ネットワークは、ニューロンに関連するパラメータ(たとえば、活性化パラメータおよび/または重み、バイアスなど)の値を適合させること、ニューロンさらにはニューロンの層を加えることおよび/または除去すること、ニューロン間のエッジを加えることまたは除去することなどによって、特定のタスクを学習するようにトレーニングされてよい。

【００４０】

特に、オートエンコーダ、畳み込みニューラルネットワーク(CNN)、再帰型ニューラルネットワーク(RNN)、多層パーセプトロン(MLP)ニューラルネットワークなどの、様々なタイプのニューラルネットワークが存在する。畳み込みニューラルネットワークは、各々が受容野(たとえば、入力空間の空間的に局在化された領域)を有し、かつ入力空間をまとめてタイリングする、人工ニューロンの集合を含んでよい。畳み込みニューラルネットワークは数多くの適用例を有する。たとえば、CNNはパターン認識および分類のエリアにおいて広く使用され得る。RNNは、層の出力をセーブするとともに層の結果を予測する助けとなるようにこの出力を入力にフィードバックするというやり方で働く。MLPニューラルネットワークでは、データは入力層の中に供給されてよく、1つまたは複数の隠れ層が抽象性のレベルをデータに与える。次いで、抽象化されたデータに基づいて、出力層において予測が行われてよい。MLPは、入力がクラスまたはラベルに割り当てられる分類予測問題に特に適していることがある。

【００４１】

階層ニューラルネットワークアーキテクチャ(複数の隠れ層が存在するとき、深層ニューラルネットワークと呼ばれる)では、人工ニューロンの第1の層の出力が人工ニューロンの第2の層への入力になり、人工ニューロンの第2の層の出力が人工ニューロンの第3の層への入力になり、以下同様である。畳み込みニューラルネットワークは、特徴の階層を認識するようにトレーニングされてよい。畳み込みニューラルネットワークアーキテクチャにおける計算は、1つまたは複数の計算チェーンをなして構成され得る処理ノードの集団にわたって分散されてよい。これらの多層アーキテクチャは、一度に1つの層がトレーニングされてよく、逆伝搬法を使用して微調整されてよい。

【００４２】

いくつかの例では、それぞれ、単予測および双予測を実行するために、機械学習ベースのPフレームおよびBフレームコーディングシステムが使用され得る。場合によっては、そのようなシステムは、ニューラルネットワークアーキテクチャ(たとえば、1つまたは複数のオートエンコーダなどの1つまたは複数の深層ニューラルネットワーク)を含むことができる。機械学習ベースのPフレームコーディングシステムの一例は、現在のフレームと参照フレームとの間の動きの予測を決定するために、現在のフレームおよび参照フレームに対して動き圧縮および動き補償を実行することができる。動き予測は、(たとえば、動き予測の中に含まれる動きベクトルに従って参照フレームのピクセルを移動させることによって)参照フレームのピクセルを修正するために使用され得、現在のフレームに対する予測が得られる。Pフレームコーディングシステムの残差部分は、予測フレームと現在のフレームとの間の差分を表す予測残差を生成することができる。予測残差は、再構成された現在のフレームを生成するために予測と合成され得る。

【００４３】

双予測を実行するために様々なタイプの機械学習ベースのBフレームコーディングシステムが使用され得る。機械学習ベースのBフレームコーディングシステムの一例は上記で説明したPフレームシステムと類似であるが、再構成された現在のフレームを生成するために単一の参照フレームではなく2つの参照フレームを使用する。機械学習ベースのBフレームコーディングシステムの別の例は、2つの参照フレーム間の補間に基づいて予測を生成するフレーム補間エンジンを含む。残差部分は、次いで、補間エンジンからの予測と合成される予測残差を決定する。

【００４４】

単方向(たとえば、Pフレーム)コーディングシステムを使用する学習型双方向(たとえば、Bフレーム)コーディングを実行するためのシステム、方法(プロセスとも呼ばれる)、装置、およびコンピュータ可読媒体(「システムおよび技法」と総称される)が、本明細書で説明される。本明細書で使用するコーディングという用語は、符号化(たとえば、圧縮)、復号(たとえば、解凍)、または符号化と復号の両方を指すことができる。たとえば、少なくとも2つの参照フレームに基づく、補間された参照フレームを生成するためのフレーム補間エンジンを含む、双方向コーディングシステム(Bフレームコーディングシステムとも呼ばれる)が提供される。いくつかの例では、フレーム補間エンジンは、補間された参照フレームを生成するために使用される参照フレームに関連する動き情報(たとえば、オプティカルフローまたは他の動きベースの技法)を使用して、補間された参照フレームを生成することができる。

【００４５】

双方向コーディングシステムは、双方向予測フレーム(Bフレーム)を生成するために単方向コーディングシステム(Pフレームコーディングシステムとも呼ばれる)を使用することができる。たとえば、単方向コーディングシステムは、Bフレームを生成するために、補間された参照フレームを使用することができる。Bフレームは、以前のフレームおよび将来のフレーム(たとえば、表示順序および/または復号順序でBフレームに先行および後続するフレーム)からのデータを参照することによって生成されるタイプのフレームを含むことができる。Bフレームコーディングのために使用されることに加えて、単方向コーディングシステムはまた、特定のフレームに対して単予測モードが実行されるとき、単方

向予測フレーム(Pフレーム)を生成するために使用され得る。たとえば、Pフレームを生成するために、単方向コーディングシステムによって実際の参照フレームが使用され得る。Pフレームは、単一の参照フレームからのデータを参照することによって生成されるタイプのフレームを含むことができ、単一の参照フレームは、以前のフレーム(たとえば、コーディング順序でBフレームに先行するフレーム)または将来のフレーム(たとえば、コーディング順序でBフレームに後続するフレーム)を含むことができる。本明細書で使用するフレーム、ピクチャ、および画像という用語は、互換的に使用され得る。

【００４６】

いくつかの例では、単方向コーディングシステムは、ビデオデータを処理し入力フレームに対する予測を生成するために機械学習を使用できる、ビデオエンコーダおよびビデオデコーダを含むことができる。いくつかの態様では、ビデオエンコーダは、動き予測システムのエンコーダニューラルネットワーク(たとえば、動きオートエンコーダ)を含むことができ、ビデオデコーダは、動き予測システムのデコーダニューラルネットワークを含むことができる。たとえば、ビデオエンコーダ(たとえば、動き予測システムのエンコーダニューラルネットワーク)は、入力フレームと参照フレーム(たとえば、双予測が実行中であるとき、補間された参照フレーム、または単予測が実行中であるとき、補間に基づいて生成されていない再構成されたPフレームもしくはBフレームまたはIフレームなどの、実際の参照フレーム)との間の予測される動きを表す潜在データ(たとえば、ビットストリーム)を生成するために、機械学習を使用することができる。ビデオデコーダ(たとえば、動き予測システムのデコーダニューラルネットワーク)は、潜在データを処理し入力フレームと参照フレームとの間の動きを予測または再構成するために機械学習を使用することができる。参照フレームのそれぞれの部分をワーピングすることによって入力フレームに対する予測を生成するために、予測される動きが単方向コーディングシステムによって使用され得る。たとえば、予測される動きは、動きベクトル(たとえば、入力フレームのピクセルまたはピクセルのブロックごとの動きベクトルを含むオプティカルフローマップ)を含むことができる。ワーピングエンジンは、入力フレーム(ワーピングされるフレームと呼ばれる)に対する予測を生成するために、動きベクトルに従って参照フレームのピクセルを移動させることができる。

【００４７】

いくつかの例では、単方向コーディングシステムのビデオエンコーダおよびビデオデコーダはまた、入力フレームに対する残差を生成するために機械学習を使用することができる。いくつかの態様では、単方向コーディングシステムは、入力フレームに対する残差を生成できる、エンコーダニューラルネットワークおよびデコーダニューラルネットワークを含む残差予測システム(たとえば、残差オートエンコーダ)を含むことができる。たとえば、ビデオエンコーダ(たとえば、残差予測システムのエンコーダニューラルネットワーク)は、入力フレームと参照フレーム(たとえば、双予測が実行中であるとき、補間された参照フレーム、または単予測が実行中であるとき、実際の参照フレーム)との間の差分を含む残差を入力として受信することができる。入力として残差を使用して、ビデオエンコーダは、残差を表す潜在データ(たとえば、ビットストリーム)を生成するために機械学習を使用することができる。ビデオデコーダ(たとえば、残差予測システムのデコーダニューラルネットワーク)は、潜在データを処理し残差を予測または再構成するために機械学習を使用することができる。単方向コーディングシステムは、(たとえば、ワーピングエンジンによって)入力フレームに対して生成された予測に予測残差を加算することによって、入力フレームを表す再構成されたフレームを生成することができる。

【００４８】

いくつかの例では、システムは、動きを表す潜在データおよび残差を表す潜在データを(たとえば、符号化ビデオビットストリームの中で)復号のために別のデバイスへ送信することができる。たとえば、上記で説明したものと同様に(たとえば、動き潜在データを処理して入力フレームと参照フレームとの間の動きを予測すること、予測された動きに従って参照フレームをワーピングすることによって入力フレームに対する予測を生成すること

、残差潜在データを処理して残差を予測すること、および入力フレームに対して生成された予測に予測残差を加算することによって、再構成されたフレームを生成することによって)別のデバイスは、動きを表す潜在データおよび残差を表す潜在データに基づいて入力フレームを再構成するために機械学習を使用できるビデオデコーダ(たとえば、上記で説明したような動き予測システムおよび残差予測システム)を含むことができる。

【００４９】

システムおよび技法のさらなる態様が、図に関して説明される。図1は、場合によっては、本明細書で説明するシステムおよび技法を実施するために使用され得る、画像処理システム100の例示的な実装形態を示す。画像処理システム100は、本明細書で説明する機能のうちの1つまたは複数を実行するように構成された中央処理ユニット(CPU)102またはマルチコアCPUを含むことができる。情報の中でも、パラメータまたは変数(たとえば、ニューラル信号およびシナプス重み)、計算デバイス(たとえば、重みを有するニューラルネットワーク)に関連するシステムパラメータ、遅延、周波数ビン情報、タスク情報、画像データは、ニューラル処理ユニット(NPU)108に関連するメモリブロックの中、CPU102に関連するメモリブロックの中、グラフィックス処理ユニット(GPU)104に関連するメモリブロックの中、デジタル信号プロセッサ(DSP)106に関連するメモリブロックの中、メモリブロック118の中に記憶されてよく、かつ/または複数のブロックにわたって分散されてよい。CPU102において実行される命令は、CPU102に関連するプログラムメモリから、かつ/またはメモリブロック118からロードされてよい。

【００５０】

画像処理システム100はまた、GPU104、DSP106、第5世代(5G)接続性、第4世代ロングタームエボリューション(4G LTE)接続性、Wi-Fi接続性、USB接続性、Bluetooth接続性などを含んでよい接続性ブロック110、および/または、たとえば、画像特徴を検出し得る、マルチメディアプロセッサ112などの、特定の機能を実行するための追加の処理ブロックを含むことができる。いくつかの例では、NPU108は、CPU102、DSP106、および/またはGPU104の中に実装され得る。場合によっては、画像処理システム100はまた、1つもしくは複数のセンサ114、1つもしくは複数の画像信号プロセッサ(ISP)116、および/またはストレージ120を含んでよい。

【００５１】

いくつかの例では、画像処理システム100は、1つまたは複数のプロセッサ用のARM命令セットアーキテクチャを実施することができる。本開示の一態様では、CPU102の中にロードされる命令は、入力値とフィルタ重みとの乗算積に対応するルックアップテーブル(LUT)の中を、記憶された乗算結果を求めて探索するための、コードを含んでよい。CPU102の中にロードされる命令はまた、乗算積のルックアップテーブルヒットが検出されるとき、乗算積の乗算演算の間に乗算器を無効にするための、コードを含んでよい。加えて、CPU102の中にロードされる命令は、乗算積のルックアップテーブルミスが検出されるとき、入力値とフィルタ重みとの算出された乗算積を記憶するためのコードを含んでよい。

【００５２】

画像処理システム100は、1つまたは複数のコンピューティングデバイスの一部であり得る。いくつかの例では、画像処理システム100は、カメラシステム(たとえば、デジタルカメラ、IPカメラ、ビデオカメラ、セキュリティカメラなど)、電話システム(たとえば、スマートフォン、セルラー電話、会議システムなど)、デスクトップコンピュータ、XRデバイス(たとえば、ヘッドマウントディスプレイなど)、スマートウェアラブルデバイス(たとえば、スマートウォッチ、スマートグラスなど)、ラップトップもしくはノートブックコンピュータ、タブレットコンピュータ、セットトップボックス、テレビジョン、ディスプレイデバイス、システムオンチップ(SoC)、デジタルメディアプレーヤ、ゲーム機、ビデオストリーミングデバイス、サーバ、ドローン、自動車の中のコンピュータ、モノのインターネット(IoT)デバイス、または任意の他の好適な電子デバイスなどの、電子デバイス(または、複数のデバイス)の一部であり得る。

【００５３】
いくつかの実装形態では、CPU102、GPU104、DSP106、NPU108、接続性ブロック110、マルチメディアプロセッサ112、1つもしくは複数のセンサ114、ISP116、メモリブロック118、および/またはストレージ120は、同じコンピューティングデバイスの一部であり得る。たとえば、場合によっては、CPU102、GPU104、DSP106、NPU108、接続性ブロック110、マルチメディアプロセッサ112、1つもしくは複数のセンサ114、ISP116、メモリブロック118、および/またはストレージ120は、スマートフォン、ラップトップ、タブレットコンピュータ、スマートウェアラブルデバイス、ビデオゲーミングシステム、サーバ、および/または任意の他のコンピューティングデバイスの中に統合され得る。他の実装形態では、CPU102、GPU104、DSP106、NPU108、接続性ブロック110、マルチメディアプロセッサ112、1つもしくは複数のセンサ114、ISP116、メモリブロック118、および/またはストレージ120は、2つ以上の別個のコンピューティングデバイスの一部であり得る。
【００５４】
画像処理システム100および/またはその構成要素は、本明細書で説明する本開示の態様による技法を使用してビデオ圧縮および/または解凍(ビデオ符号化および/または復号とも呼ばれ、ビデオコーディングと総称される)を実行するように構成され得る。ビデオ圧縮および/または解凍を実行するために本明細書で説明する深層学習アーキテクチャおよび技法を使用することによって、本開示の態様は、デバイスにおけるビデオ圧縮および/もしくは解凍の効率を向上させることができ、かつ/または関連するリソース要件および/もしくはリソース使用量を小さくすることができる。たとえば、本明細書で説明するビデオコーディング技法を使用するデバイスは、より効率的にビデオデータを圧縮することができ、圧縮されたビデオデータの中で宛先デバイスへ送信されるデータの量を減らすことができ、宛先デバイスは、圧縮されたビデオデータを効率的に受信および解凍することができる。いくつかの例では、本明細書で説明する深層学習アーキテクチャおよび技法は、ビデオコンテンツをコーディングするためにエンコーダおよびデコーダなどのコーディングデバイスまたは構成要素の間で交換されるデータの量を減らすことができる。ビデオコーディングのために送信される低減された量のデータは、たとえば、帯域幅、メモリ、ストレージ、電力、計算、ハードウェアなどのコンピューティングリソースにおいて、レイテンシを短縮すること、性能を向上させること、およびコストまたは負担を減らすことができる。
【００５５】
上述のように、ニューラルネットワークは機械学習システムの一例であり、入力層、1つまたは複数の隠れ層、および出力層を含むことができる。入力層の入力ノードからデータが提供され、1つまたは複数の隠れ層の隠れノードによって処理が実行され、出力層の出力ノードを通じて出力が生成される。深層学習ネットワークは、通常、複数の隠れ層を含む。ニューラルネットワークの各層は、人工ニューロン(または、ノード)を含むことができる特徴マップまたは活性化マップを含むことができる。特徴マップは、フィルタ、カーネルなどを含むことができる。ノードは、層のうちの1つまたは複数のノードの重要度を示すために使用される1つまたは複数の重みを含むことができる。場合によっては、深層学習ネットワークは、入力の単純かつ低レベルの特性を決定するために初期の層が使用され、もっと後の層がより複雑かつ抽象的な特性の階層を構築して、一連の多くの隠れ層を有することができる。
【００５６】
深層学習アーキテクチャは、特徴の階層を学習し得る。視覚データが提示される場合、たとえば、第1の層は、入力ストリームの中のエッジなどの特徴を認識することを学習し得る。別の例では、聴覚データが提示される場合、第1の層は、特定の周波数におけるスペクトル電力を認識することを学習し得る。第1の層の出力を入力として取る第2の層は、視覚データに対して形状などの特徴、または聴覚データに対してサウンドの組合せを認識することを学習し得る。たとえば、上位層は、視覚データにおける複雑な形状または聴

覚データにおける単語を表すことを学習し得る。さらに上位の層は、一般的な視覚オブジェクトまたは発話フレーズを認識することを学習し得る。

【００５７】

ニューラルネットワークは、様々な結合性パターンを用いて設計されてよい。フィードフォワードネットワークでは、所与の層の中の各ニューロンが上位層の中のニューロンと通信して、情報が下位層から上位層に渡される。上記で説明したように、フィードフォワードネットワークの連続する層の中で階層表現が構築されてよい。ニューラルネットワークはまた、再帰結合またはフィードバック(トップダウンとも呼ばれる)結合を有してよい。再帰結合では、所与の層の中のニューロンからの出力は、同じ層の中の別のニューロンに伝達されてよい。再帰型アーキテクチャは、シーケンスの中でニューラルネットワークに配信される入力データチャンクのうちの2つ以上にわたるパターンを認識する際に有用であり得る。所与の層の中のニューロンから下位層の中のニューロンへの結合は、フィードバック(または、トップダウン)結合と呼ばれる。多数のフィードバック結合を有するネットワークは、高レベルの概念の認識が入力の特定の低レベルの特徴を区別する助けとなり得るときに有用であり得る。

【００５８】

ニューラルネットワークの層間の結合は、全結合または局所結合されてよい。図2Aは、全結合ニューラルネットワーク202の一例を示す。全結合ニューラルネットワーク202では、第1の隠れ層の中のニューロンは、その出力を第2の隠れ層の中のあらゆるニューロンに伝達してよく、その結果、第2の層の中の各ニューロンは、第1の層の中のあらゆるニューロンから入力を受信する。図2Bは、局所結合ニューラルネットワーク204の一例を示す。局所結合ニューラルネットワーク204では、第1の隠れ層の中のニューロンは、第2の隠れ層の中の限られた数のニューロンに結合されてよい。より一般的には、局所結合ニューラルネットワーク204の局所結合層は、ある層の中の各ニューロンが同じかまたは類似の結合性パターンを、ただし、異なる値(たとえば、210、212、214、および216)を有してよい結合強度を伴って有するように構成されてよい。所与の領域の中の上位層ニューロンは、トレーニングを通じてネットワークへの全入力のうちの制約された部分の特性に調整される入力を受信することがあるので、局所結合の結合性パターンは、上位層の中で空間的に別個の受容野を生じることがある。

【００５９】

局所結合ニューラルネットワークの一例は、畳み込みニューラルネットワークである。図2Cは、畳み込みニューラルネットワーク206の一例を示す。畳み込みニューラルネットワーク206は、第2の層の中の各ニューロンに対する入力に関連する結合強度が共有されるように構成されてよい(たとえば、208)。畳み込みニューラルネットワークは、入力の空間位置が有意味である問題に好適である場合がある。畳み込みニューラルネットワーク206は、本開示の態様に従ってビデオ圧縮および/または解凍の1つまたは複数の態様を実行するために使用され得る。

【００６０】

1つのタイプの畳み込みニューラルネットワークは、深層畳み込みネットワーク(DCN)である。図2Dは、カメラまたは画像センサなどの画像キャプチャデバイス230から入力された画像226から特徴を認識するように設計されたDCN200の一例を示す。いくつかの例では、現在の例のDCN200は、たとえば、画像226の中の1つまたは複数の物体または標識などの、画像226における視覚特徴を識別するようにトレーニングされてよい。

【００６１】

いくつかの例では、DCN200は教師あり学習を用いてトレーニングされてよい。トレーニング中、DCN200は、画像226などの画像が提示されてよく、次いで、出力222を生成するために順方向パスが算出され得る。DCN200は、特徴抽出セクションおよび分類セクションを含んでよい。画像226を受信すると、畳み込み層232は、特徴マップの第1のセット218を生成するために畳み込みカーネル(図示せず)を画像226に適用してよい。一例として、畳み込み層232のための畳み込みカーネルは、28×28の特徴マップを

生成する5×5のカーネルであってよい。本例では、特徴マップの第1のセット218の中で4つの異なる特徴マップが生成されるので、4つの異なる畳み込みカーネルが畳み込み層232において画像226に適用された。畳み込みカーネルは、フィルタまたは畳み込みフィルタと呼ばれることもある。

【００６２】

特徴マップの第1のセット218は、特徴マップの第2のセット220を生成するために最大プーリング層(図示せず)によってサブサンプリングされてよい。最大プーリング層は、特徴マップの第1のセット218のサイズを縮小する。すなわち、14×14などの特徴マップの第2のセット220のサイズは、28×28などの特徴マップの第1のセット218のサイズよりも小さい。縮小されたサイズは、メモリ消費量を低減しながら、類似の情報を後続の層に提供する。特徴マップの第2のセット220は、特徴マップの1つまたは複数の後続のセット(図示せず)を生成するために、1つまたは複数の後続の畳み込み層(図示せず)を介してさらに畳み込まれてよい。

【００６３】

図2Dの例では、第1の特徴ベクトル224を生成するために、特徴マップの第2のセット220が畳み込まれる。さらに、第2の特徴ベクトル228を生成するために、第1の特徴ベクトル224がさらに畳み込まれる。第2の特徴ベクトル228の各特徴は、「標識(sign)」、「60」、および「100」などの画像226の可能な特徴に対応する数を含んでよい。ソフトマックス関数(図示せず)は、第2の特徴ベクトル228における数を確率に変換し得る。したがって、DCN200の出力222は、画像226が1つまたは複数の特徴を含む確率である。

【００６４】

本例では、「標識」および「60」に対する出力222における確率は、「30」、「40」、「50」、「70」、「80」、「90」、および「100」などの、出力222のうちの他のものの確率よりも高い。トレーニングの前に、DCN200によって生成された出力222は、不正確である可能性がある。したがって、出力222とターゲット出力との間で誤差が計算されてよい。ターゲット出力は、画像226のグラウンドトゥルース(たとえば、「標識」および「60」)である。次いで、DCN200の重みが調整されてよく、そのため、DCN200の出力222はターゲット出力とより密に整合される。

【００６５】

重みを調整するために、学習アルゴリズムは重みに対する勾配ベクトルを算出してよい。勾配は、重みが調整された場合に誤差が増加または減少することになる量を示し得る。最上層において、勾配は、最後から2番目の層の中の活性化されたニューロンと出力層の中のニューロンとを結合する重みの値に直接対応し得る。下位層では、勾配は、重みの値、および上位層の算出された誤差勾配に依存し得る。次いで、重みは、誤差を低減するように調整され得る。重みを調整するこの方式は、ニューラルネットワークを通じた「逆方向パス」を伴うので、「逆伝搬法」と呼ばれることがある。

【００６６】

実際には、重みの誤差勾配は、計算された勾配が真の誤差勾配に近似するように、少数の例にわたって計算されてよい。この近似法は、確率的勾配降下法と呼ばれることがある。確率的勾配降下法は、システム全体の達成可能な誤差率が下げ止まるまで、または誤差率がターゲットレベルに達するまで反復されてよい。学習の後、DCNは、新たな画像が提示されてよく、ネットワークを通じた順方向パスが、DCNの推定または予測と見なされてよい出力222を与え得る。

【００６７】

深層信念ネットワーク(DBN)は、隠れノードの複数の層を備える確率的モデルである。DBNは、トレーニングデータセットの階層表現を抽出するために使用され得る。DBNは、制限付きボルツマンマシン(RBM)の層を積層することによって取得され得る。RBMは、入力のセットにわたる確率分布を学習できるタイプの人工ニューラルネットワークである。RBMは各入力がカテゴリー分類されるべき先のクラスについての情報がない状態

で確率分布を学習できるので、RBMは教師なし学習において頻繁に使用される。教師なしと教師ありとのハイブリッドパラダイムを使用して、DBNの下部RBMは、教師なし方式でトレーニングされ得るとともに特徴抽出器として働いてよく、上部RBMは、(前の層およびターゲットクラスからの入力の同時分布上で)教師あり方式でトレーニングされ得るとともに分類器として働いてよい。

【００６８】

深層畳み込みネットワーク(DCN)は、追加のプーリング層および正規化層を用いて構成された、畳み込みネットワークのネットワークである。DCNは、多くのタスクにおいて高性能を達成することができる。DCNは、入力と出力ターゲットの両方が、多数の標本に対して知られており、かつ勾配降下法の使用によってネットワークの重みを修正するために使用される、教師あり学習を使用してトレーニングされ得る。

【００６９】

DCNは、フィードフォワードネットワークであってよい。加えて、上記で説明したように、DCNの第1の層の中のニューロンから次の上位層の中のニューロンのグループへの結合は、第1の層の中のニューロンにわたって共有される。DCNのフィードフォワード結合および共有される結合は、高速処理のために活用され得る。DCNの計算負担は、たとえば、再帰結合またはフィードバック結合を備える、同様にサイズ決定されたニューラルネットワークの計算負担よりも、はるかに小さくてよい。

【００７０】

畳み込みネットワークの各層の処理は、空間的に不変のテンプレートまたは基底射影と見なされてよい。入力が、カラー画像の赤色チャネル、緑色チャネル、および青色チャネルなどの複数のチャネルに最初に分解される場合、その入力に対してトレーニングされる畳み込みネットワークは、画像の軸に沿った2つの空間次元および色情報をキャプチャする第3の次元を有する、3次元と見なされてよい。畳み込み結合の出力は、後続の層の中で特徴マップを形成するものと見なされてよく、特徴マップ(たとえば、220)の各要素が、前の層(たとえば、特徴マップ218)の中のある範囲のニューロンから、かつ複数のチャネルの各々から、入力を受信する。特徴マップの中の値は、整流(rectification)、max(0,x)などの非線形性を用いてさらに処理されてよい。隣接するニューロンからの値はさらに、ダウンサンプリングに相当するプーリングが行われてよく、追加の局所不変性および次元数低減をもたらし得る。

【００７１】

図3は、深層畳み込みネットワーク350の一例を示すブロック図である。深層畳み込みネットワーク350は、結合性および重み共有に基づいて、複数の異なるタイプの層を含んでよい。図3に示すように、深層畳み込みネットワーク350は畳み込みブロック354A、354Bを含む。畳み込みブロック354A、354Bの各々は、畳み込み層(CONV)356、正規化層(LNorm)358、および最大プーリング層(MAX POOL)360を用いて構成されてよい。

【００７２】

畳み込み層356は、特徴マップを生成するために、入力データ352に適用され得る1つまたは複数の畳み込みフィルタを含んでよい。2つの畳み込みブロック354A、354Bしか示されないが、本開示は、そのように限定しておらず、代わりに、設計選好に従って、任意の数の畳み込みブロック(たとえば、ブロック354A、354B)が深層畳み込みネットワーク350の中に含まれてよい。正規化層358は、畳み込みフィルタの出力を正規化し得る。たとえば、正規化層358は、白色化または側方抑制を行ってよい。最大プーリング層360は、局所不変性および次元数低減のために、空間にわたってダウンサンプリング統合を行ってよい。

【００７３】

たとえば、深層畳み込みネットワークの並列フィルタバンクは、高性能および低電力消費を達成するために、画像処理システム100のCPU102またはGPU104上にロードされてよい。いくつかの例では、並列フィルタバンクは、画像処理システム100のDSP106

またはISP116上にロードされてよい。深層畳み込みネットワーク350は、画像処理システム100上に存在することがある他の処理ブロックにアクセスしてよい。

【００７４】

深層畳み込みネットワーク350は、層362A(「FC1」とラベル付けされる)および層362B(「FC2」とラベル付けされる)などの、1つまたは複数の全結合層を含んでよい。深層畳み込みネットワーク350は、ロジスティック回帰(LR)層364を含んでよい。深層畳み込みネットワーク350の各層356、358、360、362、364の間に、更新されることになる重み(図示せず)がある。層(たとえば、356、358、360、362、364)の各々の出力は、畳み込みブロック354Aのうちの第1のものにおいて供給された入力データ352(たとえば、画像、オーディオ、ビデオ、センサデータ、および/または他の入力データ)から階層的な特徴表現を学習するために、深層畳み込みネットワーク350の中の層(たとえば、356、358、360、362、364)のうちの後続の層の入力として働いてよい。深層畳み込みネットワーク350の出力は、入力データ352に対する分類スコア366である。分類スコア366は確率のセットであってよく、ここで、各確率は、入力データが特徴のセットからの特徴を含む確率である。

【００７５】

別のタイプのニューラルネットワークはオートエンコーダである。オートエンコーダは、入力を受信しその出力においてその入力のあるバージョンを生成するように(たとえば、本質的にその入力をその出力にコピーするように)、(たとえば、トレーニングデータおよび1つまたは複数の損失関数を使用して)トレーニングされ得る。オートエンコーダは、教師なし方式で効率的なデータコーディングを学習するようにトレーニングされ得る。たとえば、物体の画像が与えられると、オートエンコーダは、最初に画像を符号化して次元がもっと低い潜在表現にすることができ、次いで、潜在表現を物体の画像に戻して復号することができる。オートエンコーダは、再構成誤差を最小限に抑えながら入力データを圧縮することを(トレーニングを通じて)学習することができる。

【００７６】

上述のように、デジタルビデオデータは、通信ネットワーク、ならびにビデオデータを処理および記憶するデバイスに著しい負担をかける場合がある、大量のデータを含むことができる。たとえば、非圧縮ビデオコンテンツを記録することは、一般に、記録されたビデオコンテンツの解像度が高まるにつれて大幅に増大する大規模なファイルサイズをもたらす。例示的な一例では、1080p/24(たとえば、毎秒24フレームでキャプチャされる、幅が1920ピクセルかつ高さが1080ピクセルという解像度)で記録される、チャネル当り16ビットの非圧縮ビデオは、フレーム当り12.4メガバイト、すなわち毎秒297.6メガバイトを占有し得る。毎秒24フレームで4K解像度で記録される、チャネル当り16ビットの非圧縮ビデオは、フレーム当り49.8メガバイト、すなわち毎秒1195.2メガバイトを占有し得る。

【００７７】

ネットワーク帯域幅は、大きいビデオファイルがそれに対して問題になり得る別の制約である。たとえば、ビデオコンテンツはワイヤレスネットワークを介して(たとえば、LTE、LTEアドバンスト、ニューラジオ(NR)、WiFi(商標)、Bluetooth(商標)、または他のワイヤレスネットワークを介して)頻繁に配送され、消費者インターネットトラフィックの大きい部分を構成する場合がある。したがって、これらのネットワークの中でビデオコンテンツを配送するために使用される帯域幅の量を減らすことが望ましい。

【００７８】

非圧縮ビデオコンテンツは、送信のために物理的な記憶および相当な帯域幅を求めてサイズ決定可能なメモリを伴うことがある、大規模ファイルをもたらす場合があるので、本明細書でさらに説明するように、そのようなビデオコンテンツを圧縮および解凍するためにビデオコーディング技法が利用され得る。ビデオコンテンツのサイズ、したがって、ビデオコンテンツを記憶するために関与するストレージの量、およびビデオコンテンツを配送することに関与する帯域幅の量を縮小するために、特に、HEVC、AVC、MPEG、VV

Cなどの特定のビデオコーディング規格および/または方式に従って、様々なビデオコーディング技法が実行され得る。ビデオコーディングは、ビデオ画像またはビデオシーケンスの中に存在する冗長性を利用する、インター予測またはイントラ予測などの予測方法を使用することができる。ビデオコーディング技法の1つの目標は、ビデオ品質の劣化を回避するかまたは最小限に抑えながら、ビデオデータを圧縮してもっと低いビットレートを使用する形態にすることである。ビデオサービスに対する需要が高まり新たなビデオサービスが利用可能になるにつれて、より良好なコーディング効率、性能、およびレート制御を有するコーディング技法が必要とされる。

【００７９】

符号化デバイスは、ビデオコーディング規格に従ってビデオデータを符号化して符号化ビデオビットストリームを生成することができる。いくつかの例では、符号化ビデオビットストリーム(または、「ビデオビットストリーム」もしくは「ビットストリーム」)は、一連の1つまたは複数のコード化ビデオシーケンスを含むことができる。符号化デバイスは、各ピクチャを複数のスライスに区分することによってピクチャのコード化表現を生成することができる。スライスは、そのスライスの中の情報が同じピクチャ内の他のスライスからのデータへの依存関係なしにコーディングされるように、他のスライスとは無関係である。スライスは、独立スライスセグメント、および存在する場合、前のスライスセグメントに依存する1つまたは複数の従属スライスセグメントを含む、1つまたは複数のスライスセグメントを含む。HEVCでは、スライスは、ルーマサンプルおよびクロマサンプルのコーディングツリーブロック(CTB)に区分される。ルーマサンプルのCTB、およびクロマサンプルの1つまたは複数のCTBは、サンプル用のシンタックスと一緒にコーディングツリーユニット(CTU)と呼ばれる。CTUは、「ツリーブロック」または「最大コーディングユニット」(LCU)と呼ばれることもある。CTUは、HEVC符号化のための基本処理単位である。CTUは、様々なサイズの複数のコーディングユニット(CU)に分割され得る。CUは、コーディングブロック(CB)と呼ばれるルーマサンプルアレイおよびクロマサンプルアレイを含む。

【００８０】

ルーマCBおよびクロマCBは、さらに分割されて予測ブロック(PB)になり得る。PBは、(利用可能であるかまたは使用のために有効化されるとき)インター予測またはイントラブロックコピー(IBC)予測のために同じ動きパラメータを使用するルーマ成分またはクロマ成分のサンプルのブロックである。ルーマPBおよび1つまたは複数のクロマPBは、関連するシンタックスと一緒に、予測ユニット(PU)を形成する。インター予測の場合、動きパラメータのセット(たとえば、1つまたは複数の動きベクトル、参照インデックスなど)は、PUごとにビットストリームの中でシグナリングされ、ルーマPBおよび1つまたは複数のクロマPBのインター予測のために使用される。動きパラメータは動き情報と呼ばれることもある。CBはまた、1つまたは複数の変換ブロック(TB)に区分され得る。TBは、予測残差信号をコーディングするために残差変換(たとえば、場合によっては、同じ2次元変換)がそこに適用される色成分のサンプルの正方形ブロックを表す。変換ユニット(TU)は、ルーマサンプルおよびクロマサンプルのTB、ならびに対応するシンタックス要素を表す。以下でより詳細に変換コーディングが説明される。

【００８１】

HEVC規格によれば、変換はTUを使用して実行され得る。TUは、所与のCU内のPUのサイズに基づいてサイズ決定されてよい。TUは、同じサイズであってよく、またはPUよりも小さくてもよい。いくつかの例では、CUに対応する残差サンプルは、残差4分木(RQT)と呼ばれる4分木構造を使用して、より小さいユニットに再分割され得る。RQTのリーフノードはTUに相当し得る。TUに関連するピクセル差分値は、変換係数を生成するために変換され得る。変換係数は、次いで、符号化デバイスによって量子化され得る。

【００８２】

ビデオデータのピクチャがCUに区分されると、符号化デバイスは予測モードを使用して各PUを予測する。予測ユニットまたは予測ブロックは、次いで、残差(以下で説明する

)を得るために元のビデオデータから減算される。CUごとに、シンタックスデータを使用してビットストリームの内部で予測モードがシグナリングされ得る。予測モードは、イントラ予測(または、イントラピクチャ予測)またはインター予測(または、インターピクチャ予測)を含んでよい。イントラ予測は、ピクチャ内で空間的に隣接するサンプル間の相関を利用する。たとえば、イントラ予測を使用すると、たとえば、PUにとっての平均値を見つけるためのDC予測、平坦面をPUに適合させるための平面予測、隣接データから外挿するための方向予測、または任意の他の好適なタイプの予測を使用して、同じピクチャの中の隣接する画像データから各PUが予測される。インター予測は、画像サンプルのブロックに対する動き補償予測を導出するために、ピクチャ間の時間的な相関を使用する。たとえば、インター予測を使用すると、(出力順序において現在ピクチャの前または後の)1つまたは複数の参照ピクチャの中の画像データからの動き補償予測を使用して、各PUが予測される。ピクチャエリアを、インターピクチャ予測を使用してコーディングすべきか、それともイントラピクチャ予測を使用してコーディングすべきかという決定は、たとえば、CUレベルにおいて行われてよい。

【００８３】

いくつかの例では、ピクチャの1つまたは複数のスライスにスライスタイプが割り当てられる。スライスタイプは、Iスライス、Pスライス、およびBスライスを含むことができる。Iスライス(独立して復号可能なイントラフレーム)は、イントラ予測のみによってコーディングされるピクチャのスライスであり、したがって、スライスの任意の予測ユニットまたは予測ブロックを予測するためにIスライスがフレーム内のデータしか必要としないので独立して復号可能である。Pスライス(単方向予測フレーム)は、イントラ予測を用いてかつ単方向インター予測を用いてコーディングされ得るピクチャのスライスである。Pスライス内の各予測ユニットまたは予測ブロックは、イントラ予測またはインター予測のいずれかを用いてコーディングされる。インター予測が適用されるとき、予測ユニットまたは予測ブロックは、1つの参照ピクチャによって予測され、したがって、参照サンプルは、1つのフレームの1つの参照領域からのものである。Bスライス(双方向予測フレーム)は、イントラ予測およびインター予測(たとえば、双予測または単予測のいずれか)を用いてコーディングされ得るピクチャのスライスである。Bスライスの予測ユニットまたは予測ブロックは、2つの参照ピクチャから双方向予測され得る。各ピクチャは参照領域に寄与することができ、2つの参照領域のサンプルセットが(たとえば、等しい重みを用いて、または異なる重みを用いて)重み付けられて、双方向予測されたブロックの予測信号を生成することができる。上述のように、1つのピクチャのスライスは独立してコーディングされる。場合によっては、ピクチャは、ただ1つのスライスとしてコーディングされ得る。

【００８４】

イントラ予測および/またはインター予測を使用して予測を実行した後、符号化デバイスは変換および量子化を実行することができる。たとえば、予測に続いて、符号化デバイスはPUに対応する残差値を計算し得る。残差値は、コーディングされつつあるピクセルの現在ブロック(PU)と現在ブロックを予測するために使用される予測ブロック(たとえば、現在ブロックの予測されたバージョン)との間のピクセル差分値を含んでよい。たとえば、(たとえば、インター予測またはイントラ予測を使用して)予測ブロックを生成した後、符号化デバイスは、予測ユニットによって生成された予測ブロックを現在ブロックから減算することによって残差ブロックを生成することができる。残差ブロックは、現在ブロックのピクセル値と予測ブロックのピクセル値との間の差分を定量化するピクセル差分値のセットを含む。いくつかの例では、残差ブロックは、2次元ブロックフォーマット(たとえば、ピクセル値の2次元行列またはアレイ)で表され得る。そのような例では、残差ブロックはピクセル値の2次元表現である。

【００８５】

予測が実行された後に残っている場合がある任意の残差データがブロック変換を使用して変換され、ブロック変換は、離散コサイン変換、離散サイン変換、整数変換、ウェーブ

レット変換、他の好適な変換機能、またはそれらの任意の組合せに基づいてよい。場合によっては、1つまたは複数のブロック変換(たとえば、サイズ32×32、16×16、8×8、4×4、または他の好適なサイズ)が、各CUにおける残差データに適用され得る。いくつかの実施形態では、符号化デバイスによって実施される変換プロセスおよび量子化プロセスのためにTUが使用され得る。1つまたは複数のPUを有する所与のCUは、1つまたは複数のTUも含んでよい。以下でさらに詳細に説明するように、残差値は、ブロック変換を使用して変換係数に変換されてよく、次いで、TUを使用して量子化および走査されて、エントロピーコーディングのためのシリアル化変換係数を生成し得る。

【００８６】

符号化デバイスは変換係数の量子化を実行し得る。量子化は、係数を表すために使用されるデータの量を減らすために変換係数を量子化することによって、さらなる圧縮をもたらす。たとえば、量子化は、係数の一部または全部に関連するビット深度を低減し得る。一例では、nビット値を有する係数は、量子化中にmビット値に切り捨てられてよく、nはmよりも大きい。

【００８７】

量子化が実行されると、コード化ビデオビットストリームは、量子化変換係数、予測情報(たとえば、予測モード、動きベクトル、ブロックベクトルなど)、区分情報、および他のシンタックスデータなどの任意の他の好適なデータを含む。コード化ビデオビットストリームの様々な要素が、次いで、符号化デバイスによってエントロピー符号化され得る。いくつかの例では、符号化デバイスは、既定の走査順序を利用して量子化変換係数を走査して、エントロピー符号化され得るシリアル化ベクトルを生成し得る。いくつかの例では、符号化デバイスは適応走査を実行し得る。量子化変換係数を走査してベクトル(たとえば、1次元ベクトル)を形成した後、符号化デバイスはベクトルをエントロピー符号化し得る。たとえば、符号化デバイスは、コンテキスト適応型可変長コーディング、コンテキスト適応型バイナリ算術コーディング、シンタックスベースコンテキスト適応型バイナリ算術コーディング、確率間隔区分エントロピーコーディング、または別の好適なエントロピー符号化技法を使用してよい。

【００８８】

符号化デバイスは、符号化ビデオビットストリームを記憶することができ、かつ/または通信リンクを介して受信デバイスへ符号化ビデオビットストリームデータを送ることができ、受信デバイスは復号デバイスを含むことができる。復号デバイスは、符号化ビデオデータを構成する1つまたは複数のコード化ビデオシーケンスの要素を(たとえば、エントロピーデコーダを使用して)エントロピー復号および抽出することによって、符号化ビデオビットストリームデータを復号し得る。復号デバイスは、次いで、符号化ビデオビットストリームデータを再スケーリングしてよく、符号化ビデオビットストリームデータに対して逆変換を実行してよい。次いで、残差データが復号デバイスの予測ステージに渡される。復号デバイスは、次いで、イントラ予測、インター予測、IBC、および/または他のタイプの予測を使用して、ピクセルのブロック(たとえば、PU)を予測する。いくつかの例では、逆変換の出力(残差データ)に予測が加算される。復号デバイスは、復号ビデオをビデオ宛先デバイスに出力してよく、ビデオ宛先デバイスは、復号ビデオデータをコンテンツの消費者に表示するためのディスプレイまたは他の出力デバイスを含んでよい。

【００８９】

様々なビデオコーディング規格(たとえば、上記で説明したHEVCビデオコーディング技法)によって規定されるビデオコーディングシステムおよび技法は、未加工ビデオコンテンツの中の情報のうちの多くを保持できる場合があり、信号処理および情報理論概念に基づいて事前に規定され得る。しかしながら、様々な圧縮アルゴリズムが一般に(たとえば、任意のタイプのビデオコンテンツに)適用可能であってよいが、これらのビデオコーディング技法は、ネットワークベースの環境または他のタイプの環境において有益ないくつかの特性を欠く場合がある。たとえば、ビデオコーディングシステムは、低レイテンシ動作を実施することが可能であるべきである。しかしながら、既存のビデオコーディング

規格に従って動作するビデオコーディングシステムおよび関係する技法は、いくつかのフレームのバッチ(ピクチャグループすなわちGoPと呼ばれることがある)をなしてビデオを符号化および送信することがあり、送信されるべきバッチ(たとえば、GoP)のフレームの蓄積を待つ必要があり得る。そのようなバッチベースのビデオコーディングは、フレームごとのコーディングおよび送信を使用して達成され得るよりも大きいレイテンシをもたらす。

【００９０】

場合によっては、単予測を実行するために(たとえば、1つまたは複数のニューラルネットワークを含む)機械学習ベースのPフレームコーディングシステムが使用され得る。図4は、機械学習ベースのPフレームコーディングシステム400の一例を示す図である。Pフレームコーディングシステム400の構成要素は、動き圧縮エンジン402、動き補償エンジン404、および残差圧縮エンジン406を含む。Pフレームコーディングシステム400の構成要素は、動き圧縮エンジン402、動き補償エンジン404、および残差圧縮エンジン406を使用して単方向インター予測(単予測)を実行するように構成された、1つまたは複数のニューラルネットワークを含むことができる。

【００９１】

たとえば、動き圧縮エンジン402は、入力フレーム401および参照フレーム403を取得することができる。動き圧縮エンジン402は、入力フレーム401と参照フレーム403との間の動きの予測を決定するために機械学習ベースの動き圧縮を実行することができる。動き圧縮エンジン402は、予測された動き情報を動き補償エンジン404に提供することができる。動き補償エンジン404はまた、入力として参照フレーム403を受信することができる。動き補償エンジン404は、参照フレームのピクセルを修正し入力フレーム403の予測を生成するために、動き予測を使用することができる。たとえば、動き補償エンジン404は、動き予測の中に含まれる動きベクトルに従って参照フレームのピクセルを移動させることができる。残差圧縮エンジン406は、予測と入力フレーム401との間の差分を表す予測残差を生成することができる。予測残差は、次いで、再構成された入力フレームを生成するために予測と合成され得る。再構成された入力フレームが、記憶、表示などのために出力され得る。

【００９２】

場合によっては、双予測を実行するために機械学習ベースのBフレームコーディングシステムが使用され得る。図5Aは、Bフレームコーディングシステム500の一例を示す図である。Bフレームコーディングシステム500の構成要素は、動き圧縮エンジン502、動き補償エンジン504、および残差圧縮エンジン506を含む。Bフレームコーディングシステム500の構成要素は、動き圧縮エンジン502、動き補償エンジン504、および残差圧縮エンジン506を使用して双方向インター予測(双予測)を実行するように構成された、1つまたは複数のニューラルネットワークを含むことができる。双予測は、単一の参照フレームではなく2つの参照フレーム(図5Aでは参照フレーム0(Ref_0)および参照フレーム1(Ref_1)として示す)に基づいて実行される。

【００９３】

たとえば、入力フレーム501ならびに参照フレームRef_0およびRef_1は、動き圧縮エンジン502に入力として提供され得る。動き圧縮エンジン502は、入力フレーム501と第1の参照フレームRef_0との間の動きの第1の予測、および入力フレーム501と第2の参照フレームRef_1との間の動きの第2の予測を決定するために、機械学習ベースの動き圧縮を実行することができる。動き補償エンジン504は、参照フレームRef_0およびRef_1ならびに動き圧縮エンジン502からの予測された動きを入力として受信することができる。動き補償エンジン504は、参照フレームRef_0およびRef_1のピクセルを修正し入力フレーム501の予測を生成するために、2つの動き予測を使用することができる。たとえば、動き補償エンジン504は、第1の動き予測に従って第1の参照フレームRef_0のピクセルを移動させることができ、第2の動き予測に従って第2の参照フレームRef_1のピクセルを移動させることができる。動き補償エンジン504は、2つの予測をマージして入力フレーム501

に対するマージされた予測を生成することができる。

【００９４】

残差圧縮エンジン506は、マージされた予測と入力フレーム501との間の差分を表す予測残差を生成することができる。予測残差は、次いで、マージされた予測と合成されて、再構成された入力フレームを生成することができ、再構成された入力フレームは、記憶または表示のために出力され得る。

【００９５】

図5Bは、Bフレームコーディングシステム510の別の例を示す図である。Bフレームコーディングシステム510の構成要素は、フレーム補間エンジン512および残差圧縮エンジン516を含む。Bフレームコーディングシステム500の構成要素は、フレーム補間エンジン512および残差圧縮エンジン516を使用して双予測を実行するように構成された、1つまたは複数のニューラルネットワークを含むことができる。

【００９６】

図5AのBフレームコーディングシステム500と同様に、Bフレームコーディングシステム510は、2つの参照フレーム(参照フレーム0(Ref_0)および参照フレーム1(Ref_1)として示す)に基づいて双予測を実行することができる。しかしながら、補間エンジン512は、参照フレームRef_0と参照フレームRef_1との間の補間を実行することによって入力フレーム511の予測を生成することができる。予測は、次いで、残差圧縮エンジン516に出力され得る。残差圧縮エンジン506は、予測と入力フレーム511との間の差分を表す予測残差を生成することができる。Bフレームコーディングシステム510は、予測残差をマージされた予測と合成して、再構成された入力フレームを生成することができる。Bフレームコーディングシステム510は、記憶、表示などのために、再構成された入力フレームを出力することができる。

【００９７】

上記で説明した、図5Aおよび図5BのBフレームコーディングシステムは、場合によっては準最適であり得る。たとえば、図5AのBフレームコーディングシステム500は、双方向動き圧縮を実行して、2つの参照フレームRef_0およびRef_1ならびに入力フレーム511に基づいて2つの動き予測を生成する。Bフレームコーディングシステム500は、2つの動き予測に基づいて参照フレームRef_0とRef_1の両方を修正し、次いで、修正されたフレームをマージして入力フレーム511に対する予測を生成する。そのようなプロセスは、トレーニングおよび(機械学習システムがトレーニングされ入力ビデオデータを処理するために使用されていた後の)推定の間に多くの計算を伴う。さらに、(図5Aまたは図5Bの)別個のBフレームコーディングシステムおよび(図4の)Pフレームコーディングシステムのために、(重み、バイアスなどの、異なるパラメータを有する)別個のニューラルネットワークが必要とされる。図5Aおよび図5Bの機械学習システム500および510をトレーニングすることはまた、大量のトレーニングデータ(たとえば、予測ごとに2つの参照フレーム)を必要とする。さらに、図5BのBフレームコーディングシステム510は線形動き想定のもとでしか正確でない補間を決定し、予測が入力フレームと整合されないことがある。場合によっては、Bフレームコーディングシステム510によって非線形フレーム補間器が使用され得るが、そのような場合、依然として不整合が発生することがある。Bフレームコーディングシステム510は、不整合を補償するために単に残差に依拠する。

【００９８】

上述のように、Pフレームコーディングシステムおよび補間された参照フレームを使用する学習型Bフレームコーディングを実行するためのシステムおよび技法が、本明細書で説明される。本明細書で説明するシステムおよび技法は、参照フレームから補間された(たとえば、オプティカルフロー情報などの動き情報を使用して補間された)参照フレームを使用して双方向予測フレーム(Bフレーム)を効率的かつ知的に予測することができる。システムおよび技法は、より小さい計算および帯域幅需要などの、リソース需要がより小さい効率的な高品質ビデオコーディングを可能にすることができる。たとえば、Pフレームコーディングシステムのためにニューラルネットワークが使用されるとき、(Bフレー

ムの)双予測を実行するために、また(Pフレームの)単予測を実行するために、Pフレームコーディングシステムに対してニューラルネットワークパラメータ(たとえば、重み、バイアスなど)の単一のセットが必要とされる。さらに、双予測を実行するために単一の参照フレーム(補間された参照フレーム)しか使用されないことに起因して、各Bフレームの双予測に対して動き情報の単一のセットしか必要とされない。場合によっては、図5BのBフレームコーディングシステム500をトレーニングするために必要とされるトレーニングデータなどと比較して、機械学習システム(たとえば、ニューラルネットワーク)をトレーニングするために、さほど大きいトレーニングデータも必要とされなくてよい。

【００９９】

Bフレームコーディングシステム600はまた、図5AのBフレームコーディングシステム500および図5BのBフレームコーディングシステム510などと比較して、改善されたレートひずみ性能を提供する。たとえば、所与の入力フレームに対して、過去(たとえば、Ref_0)と将来(たとえば、Ref_1)の両方からの参照フレームが利用可能であるとき、たとえば、線形動き想定のもとで、Bフレームコーディングシステム600は、2つの参照フレームを線形に補間することによって入力フレーム611の大まかな予測を決定することができる。2つの参照フレームがすでに利用可能であり、補間を実行するために動きフィールドまたは他の情報が必要とされないので、この予測は0ビットしか消費しない。双方向動き(たとえば、フローまたはワーピング)を使用する図5AのBフレームコーディングシステム500は、そのような情報を利用せず、(たとえば、デコーダ/受信機が物体の動きに気付いていないかのように)両方の参照フレームに関する動き情報(たとえば、オプティカルフロー)を送る。上述のように、図5BのBフレームコーディングシステム510に関して、フレーム補間エンジン512の補間結果は線形動き想定のもとでしか正確でない。そのようなシステムでは、予測は入力フレームと整合されない場合がある。非線形フレーム補間器が採用されるときでさえ、依然として不整合が発生することがある。これらの状況では、コーディングシステムは、不整合を補償するために単に残差に依拠する。その結果、そのようなコーディングシステムのコーディング効率は、割安な副次情報(たとえば、参照フレーム間の補間)を使用して不整合を補正でき、かつ次いで、残差を使用して、再構成された入力フレームを決定できる、図6のBフレームコーディングシステム600と比較して著しく低下する場合がある。

【０１００】

図6は、学習型Bフレームコーディングを実行するように構成された、Bフレームコーディングシステム600の一例を示す図である。Bフレームコーディングシステム600は、ビデオデータを処理するために使用される様々な構成要素を含む。図示のように、Bフレームコーディングシステム600の構成要素は、フレーム補間エンジン622およびPフレームコーディングシステム624を含む。フレーム補間エンジン622およびPフレームコーディングシステム624を使用して、Bフレームコーディングシステム600は、入力フレーム611の双方向インター予測(双予測)を実行して、再構成された入力フレーム615と呼ばれる、再構成されたバージョン(または、復号されたバージョン)の入力フレーム611を生成することができる。フレーム補間エンジン622は、参照フレーム0(Ref_0)および参照フレーム1(Ref_1)を処理することに基づいて、補間された参照フレーム613を生成することができる。フレーム補間エンジン622は任意のタイプの補間技法を利用することができる。いくつかの例では、フレーム補間エンジン622は、参照フレームRef_0およびRef_1に関連する動き情報を使用して、補間された参照フレーム613を生成することができる。たとえば、図7Cを参照しながら説明するように、フレーム補間エンジン622は、参照フレームRef_0および参照フレームRef_1に対して決定された双方向動き情報(たとえば、順方向オプティカルフローおよび逆方向オプティカルフロー)に基づいて、補間された参照フレーム615を決定することができる。場合によっては、補間された参照フレーム615は、1つまたは複数の他の入力フレームによる参照フレームとしての使用などのために記憶され得る。

【０１０１】

Bフレームコーディングシステム600は、Pフレームコーディングシステム624を使用して、再構成された入力フレーム615を(Bフレームとして)生成することができる。たとえば、図6に示すように、Pフレームコーディングシステム624は、補間された参照フレーム613を入力として使用して、再構成された入力フレーム615を生成することができる。任意のタイプのPフレームコーディング技法がPフレームコーディングシステム624によって使用され得る。いくつかの例では、Pフレームコーディングシステム624は、補間された参照フレーム613および入力フレーム611を処理して、再構成された入力フレーム615を生成するために、少なくとも1つの機械学習ベースのビデオエンコーダおよび少なくとも1つの機械学習ベースのビデオデコーダを含むことができる。いくつかの実装形態では、図7Aおよび図7Bに関して以下で説明するように、Pフレームコーディングシステム624は、エンコーダニューラルネットワークおよびデコーダニューラルネットワークを有する動きニューラルネットワークシステムを含むことができる。動きニューラルネットワークシステムは、オートエンコーダ、畳み込みニューラルネットワーク(CNN)、および/または他のニューラルネットワーク構造などの、任意のタイプのニューラルネットワーク構造を含むことができる。動きニューラルネットワークシステムのエンコーダニューラルネットワークは、入力フレーム611に対する動きデータを表す潜在データを生成することができる。動きニューラルネットワークシステムのデコーダニューラルネットワークは、潜在データからの動きを再構成して入力フレーム611に対する予測された動きを生成することができる。Pフレームコーディングシステム624は、入力フレーム611の予測を生成するために、予測された動きを使用して、補間された参照フレーム613をワーピングすることができる。

【０１０２】

Pフレームコーディングシステム624はまた、エンコーダニューラルネットワークおよびデコーダニューラルネットワークを含む残差ニューラルネットワークシステムを含むことができる。残差ニューラルネットワークシステムは、オートエンコーダ、CNN、および/または他のニューラルネットワーク構造などの、任意のタイプのニューラルネットワーク構造を含むことができる。Pフレームコーディングシステム624は、(予測される動きを使用してワーピングされた)ワーピングかつ補間された参照フレーム613と入力フレーム611との間の差分を決定することなどによって、入力フレーム611に対する残差を生成することができる。残差ニューラルネットワークシステムのエンコーダニューラルネットワークは、残差を処理して、残差を表す残差潜在データを生成することができる。残差ニューラルネットワークシステムのデコーダニューラルネットワークは、残差潜在データからの残差を再構成して、入力フレーム611に対する予測残差を生成することができる。Pフレームコーディングシステム624は、ワーピングかつ補間された参照フレーム613に予測残差を加算することによって、再構成された入力フレーム615を生成することができる。

【０１０３】

いくつかの例では、Pフレームコーディングシステム624は、(Bフレームを生成するために)双予測を実行するために使用され得、(Pフレームを生成するために)単予測を実行するためにも使用され得る。たとえば、Pフレームコーディングシステム624は、入力フレームに対して単予測コーディングモードが実行されるべきであると決定されるとき、Pフレームを生成するために使用され得、入力フレームに対して双予測コーディングモードが実行されるべきであると決定されるとき、Bフレームを生成するために使用され得る。いくつかのフレームに対する異なるコーディングモード(たとえば、イントラ予測、単予測、または双予測)の選択に関する詳細が、図9～図12に関して説明される。入力フレームに対して単予測モードが選択されるとき、Pフレームコーディングシステム624は、実際の参照フレーム(たとえば、参照フレームRef_0または参照フレームRef_1)を処理して、再構成されたフレームをPフレームとして生成することができる。単予測に対するPフレームコーディングシステム624の使用に関する詳細が、図7Aに関して以下で説明される。入力フレームに対して双予測モードが選択されるとき、Pフレームコーディングシステ

ム624は、補間された参照フレーム(たとえば、補間された参照フレーム613)を処理して、再構成されたフレームをBフレームとして生成することができる。双予測に対するPフレームコーディングシステム624の使用に関する詳細が、図7Bに関して以下で説明される。

【０１０４】

Bフレームコーディングシステム600はいくつかの構成要素を含むように図示されるが、Bフレームコーディングシステム600が図6に示すものよりも多くの構成要素を含むことができることを、当業者は諒解されよう。場合によっては、Bフレームコーディングシステム600の追加構成要素は、ソフトウェア、ハードウェア、またはソフトウェアとハードウェアとの1つもしくは複数の組合せを含むことができる。たとえば、場合によっては、Bフレームコーディングシステム600は、1つまたは複数のディスプレイデバイス、1つまたは複数の他の処理エンジン、1つまたは複数の画像センサ(たとえば、カメラ)、1つまたは複数の他のセンサ(たとえば、1つまたは複数の慣性測定ユニット(IMU)、レーダー、光検出および測距(LIDAR)センサ、オーディオセンサなど)、および/あるいは(たとえば、ロングタームエボリューション(LTE)プロトコル、5G/ニューラジオ(NR)プロトコル、または他のセルラープロトコルを使用する)セルラーネットワーク通信、(たとえば、802.11Wi-Fiプロトコルを使用する)ワイヤレスローカルエリアネットワーク(WLAN)通信、Bluetoothもしくは他の短距離通信、それらの任意の組合せ、および/または他の通信などの、ワイヤレス通信用の1つまたは複数のワイヤレストランシーバ(または、別個のワイヤレス受信機および送信機)を含むことができる。Bフレームコーディングシステム600が、図6に示されない1つまたは複数の他のソフトウェアおよび/またはハードウェア構成要素を含むことができることを、当業者は諒解されよう。

【０１０５】

いくつかの実装形態では、本明細書で説明する様々な動作を実行するために、Bフレームコーディングシステム600の追加構成要素は、1つまたは複数のプログラマブル電子回路(たとえば、デジタル信号プロセッサ(DSP)、マイクロプロセッサ、マイクロコントローラ、グラフィックス処理ユニット(GPU)、中央処理ユニット(CPU)、それらの任意の組合せ、および/または他の好適な電子回路)を含むことができる、電子回路または他の電子ハードウェアを含むことができ、かつ/またはそれらを使用して実装することができ、かつ/あるいはコンピュータソフトウェア、ファームウェア、またはそれらの任意の組合せを含むことができ、かつ/またはそれらを使用して実装することができる。ソフトウェアおよび/またはファームウェアは、コンピュータ可読記憶媒体上に記憶され、かつBフレームコーディングシステム600を実装する電子デバイスの1つまたは複数のプロセッサによって実行可能な、1つまたは複数の命令を含むことができる。

【０１０６】

Bフレームコーディングシステム600は、コンピューティングデバイスを含むことができるか、またはコンピューティングデバイスの一部であり得る。たとえば、コンピューティングデバイスは、モバイルもしくは固定の電話ハンドセット(たとえば、スマートフォン、セルラー電話など)、エクステンデッドリアリティデバイス(たとえば、仮想現実(VR)デバイス、拡張現実(AR)デバイス、および/または複合現実デバイス)、画像キャプチャデバイス(たとえば、カメラ)、車両、ディスプレイデバイス、(たとえば、車両コンピューティングシステムと通信している)サーバコンピュータ、デスクトップコンピュータ、ラップトップもしくはノートブックコンピュータ、タブレットコンピュータ、セットトップボックス、テレビジョン、デジタルメディアプレーヤ、ビデオストリーミングデバイス、または任意の他の好適なコンピューティングデバイスを含むことができるか、またはそれらの一部であり得る。

【０１０７】

図7Aは、Pフレームの単方向コーディングおよび/またはBフレームの双方向コーディングを実行するために使用され得るPフレームコーディングシステム700の一例を示す図である。Pフレームコーディングシステム700は、図6のPフレームコーディングシステ

ム624の例示的な例である。Pフレームコーディングシステム700は、動き予測システム702、ワーピングエンジン704、および残差予測システム706を含む。動き予測システム702および残差予測システム706は、(たとえば、1つまたは複数のニューラルネットワークを使用する)任意のタイプの機械学習システムを含むことができる。例示的な一例では、動き予測システム702は動きオートエンコーダを含み、残差予測システム706は残差オートエンコーダを含む。いくつかの構成要素を含むように図7AのPフレームコーディングシステム700が示されるが、Pフレームコーディングシステム700が、図7Aに示す構成要素よりも少数または多数の構成要素を含むことができることを、当業者は諒解されよう。

【０１０８】

所与の時間tに対して、Pフレームコーディングシステム700は、入力フレーム(フレームX_tとして示す)および参照フレーム

【０１０９】

【数１】

$$\hat{X}_{ref}$$

【０１１０】

を受信する。参照フレーム

【０１１１】

【数２】

$$\hat{X}_{ref}$$

【０１１２】

は、時間tに先立って生成された(ハットシンボル

【０１１３】

【数３】

“$\hat{*}$”

【０１１４】

によって注記するような)以前に再構成されたフレームである。たとえば、入力フレームX_tは、時間tにおける現在のフレームであり、参照フレーム

【０１１５】

【数４】

$$\hat{X}_{ref}$$

【０１１６】

は、(たとえば、時間tにおける入力フレームX_tの前に時間的または連続的に)時間t-nにおいて再構成または生成された過去のフレームであり得、ただし、nは1以上の整数値である。例示的な一例では、参照フレーム

【０１１７】

【数５】

$$\hat{X}_{ref}$$

【０１１８】

は、時間的または連続的に入力フレームX_tの直前のフレームであり得、その場合、参照

フレーム
【０１１９】
【数６】

$$\hat{X}_{ref}$$

【０１２０】
は、
【０１２１】
【数７】

$$\hat{X}_{t-1}$$

【０１２２】
として示され得る。場合によっては、参照フレーム
【０１２３】
【数８】

$$\hat{X}_{ref}$$

【０１２４】
は、コーディングシステムの復号ピクチャバッファ(DPB)から受信される。
【０１２５】
場合によっては、入力フレームX_tはPフレームであり得、参照フレーム
【０１２６】
【数９】

$$\hat{X}_{ref}$$

【０１２７】
は、Iフレーム、Pフレーム、またはBフレームであり得る。たとえば、参照フレーム
【０１２８】
【数１０】

$$\hat{X}_{ref}$$

【０１２９】
は、(たとえば、Pフレームコーディングシステム700を含むデバイスの一部、またはPフレームコーディングシステム700を含むデバイスとは異なるデバイスであり得る)Iフレームコーディングシステムによって、Pフレームコーディングシステム700(または、Pフレームコーディングシステム700を含むデバイス以外のデバイスのPフレームコーディングシステム)によって、またはBフレームコーディングシステム(たとえば、図6のBフレームコーディングシステム600、図7BのBフレームコーディングシステム720、またはPフレームコーディングシステム700を含むデバイスとは異なるデバイスのBフレームコーディングシステム)によって再構成または生成され得る。場合によっては、参照フレーム
【０１３０】
【数１１】

$$\hat{X}_{ref}$$

【０１３１】
は、図9もしくは図10に示すピクチャグループ(GoP:group of pictures)構造に従って、かつ/または逐次方式(たとえば、図11に示す逐次方式)もしくは階層方式(たとえば、図12に示す階層方式)に従って、あるいは任意の他の方式に基づいて選択され得る。
【０１３２】
動き予測システム702は、参照フレーム
【０１３３】
【数１２】

$$\hat{X}_{ref}$$

【０１３４】
および入力フレームX_tを受信する。動き予測システム702は、参照フレーム
【０１３５】
【数１３】

$$\hat{X}_{ref}$$

【０１３６】
のピクセルと入力フレームX_tのピクセルとの間の(たとえば、オプティカルフロー動きベクトルなどのベクトルによって表される)動きを決定することができる。動き予測システム702は、次いで、入力フレームX_tに対する予測された動き
【０１３７】
【数１４】

$$\hat{f}_t$$

【０１３８】
としてその動きを符号化することができ、場合によってはそれを復号することができる。たとえば、動き予測システム702のエンコーダネットワーク705は、現在のフレームX_tと参照フレーム
【０１３９】
【数１５】

$$\hat{X}_{ref}$$

【０１４０】
との間の動きを決定することができる。エンコーダネットワーク705は、動きを符号化して潜在表現(潜在性z_mとして示す)にすることができる。いくつかの例では、エンコーダネットワーク705は動きを潜在コードにマッピングすることができ、潜在コードは潜在性z_mとして使用され得る。場合によっては、エンコーダネットワーク705は、潜在コードに対してエントロピーコーディングを実行することによって潜在性z_mをビットストリームに変換することができる。いくつかの例では、エンコーダネットワーク705は、(たとえば、潜在コードに対してエントロピーコーディングが実行されるのに先立って)潜在性z_mを量子化することができる。量子化された潜在性z_mは、残差z_mの量子化された表現を含むことができる。場合によっては、潜在性z_mは、量子化されたコードを表すニューラルネットワークデータ(たとえば、ニューラルネットワークノードの活性化マップまたは特徴マップ)を含むことができる。エンコーダネットワーク705は、潜在性z_mを記憶することができ、動き予測システム702のデコーダネットワーク703へ潜在性z_mを送ることができ、かつ/または潜在性z_mを復号できる別のデバイスもしくはシステムへ潜在

性z_mを送ることができる。潜在性z_mを受信すると、デコーダネットワーク703は、潜在性z_mを復号(たとえば、逆エントロピーコーディング、逆量子化、および/または再構成)して参照フレーム

【０１４１】

【数１６】

$$\hat{X}_{ref}$$

【０１４２】

のピクセルと入力フレームX_tのピクセルとの間の予測された動き

【０１４３】

【数１７】

$$\hat{f}_t$$

【０１４４】

を生成することができる。エンコーダネットワーク705およびデコーダネットワーク703は、以下で説明するように、トレーニングデータ(たとえば、トレーニング画像またはトレーニングフレーム)および1つまたは複数の損失関数を使用してトレーニングおよび最適化され得る。

【０１４５】

いくつかの例では、予測される動き

【０１４６】

【数１８】

$$\hat{f}_{\mathrm{t}}$$

【０１４７】

は、図8に関して以下で説明するように、オプティカルフローデータ(たとえば、1つまたは複数の変位ベクトルを含むオプティカルフローマップ)、動的畳み込みデータ(たとえば、データ畳み込み用の行列またはカーネル)、またはブロックベースの動きデータ(たとえば、ブロックごとの動きベクトル)を含むことができる。たとえば、場合によっては、予測される動き

【０１４８】

【数１９】

$$\hat{f}_{\mathrm{t}}$$

【０１４９】

は、オプティカルフローマップを含むことができる。例示的な一例では、オプティカルフローマップは、入力フレームX_tのピクセルごとの動きベクトル(たとえば、第1のピクセルのための第1の動きベクトル、第2のピクセルのための第2の動きベクトルなど)を含むことができる。動きベクトルは、参照フレーム

【０１５０】

【数２０】

$$\hat{X}_{ref}$$

【０１５１】

の中の対応するピクセルに対する現在のフレームX_tの中のピクセルにとっての動きを表

す。
【０１５２】
Ｐフレームコーディングシステムのワーピングエンジン７０４は、動き予測システム７０２によって出力される動き予測
【０１５３】
【数２１】

$$\hat{f}_t$$

【０１５４】
を取得することができる。たとえば、ワーピングエンジン７０４は、ストレージから動き予測
【０１５５】
【数２２】

$$\hat{f}_t$$

【０１５６】
を取り出すことができるか、または動き予測システム７０２から動き予測
【０１５７】
【数２３】

$$\hat{f}_t$$

【０１５８】
を受信することができる。ワーピングエンジン７０４は、(たとえば、動き補償を実行することによって)参照フレーム
【０１５９】
【数２４】

$$\hat{X}_{ref}$$

【０１６０】
のピクセルをワーピングするために、動き予測
【０１６１】
【数２５】

$$\hat{f}_t$$

【０１６２】
を使用することができ、ワーピングされたフレーム
【０１６３】
【数２６】

$$\tilde{X}_t$$

【０１６４】
の生成をもたらす。たとえば、参照フレーム
【０１６５】
【数２７】

$\hat{X}_{ref}$

【０１６６】
のピクセルは、新たな位置に移動され得るか、またはさもなければ再構成された動き推定データ
【０１６７】
【数２８】

$\hat{f}_{\mathrm{t}}$

【０１６８】
に従って修正され得る。参照フレーム
【０１６９】
【数２９】

$\hat{X}_{ref}$

【０１７０】
のピクセルと入力フレームX_tのピクセルとの間の動きを決定するためにオプティカルフローが使用される例示的な一例では、ワーピングエンジン７０４は、参照フレーム
【０１７１】
【数３０】

$\hat{X}_{ref}$

【０１７２】
のピクセルをオプティカルフローマップの動きベクトルによって示される量だけ移動させることができ、ワーピングされたフレーム
【０１７３】
【数３１】

$\tilde{X}_t$

【０１７４】
が得られる。上述のように、他の動きベースの技法も、参照フレーム
【０１７５】
【数３２】

$\hat{X}_{ref}$

【０１７６】
をワーピングするための動き予測
【０１７７】
【数３３】

$\hat{f}_t$

【０１７８】
を決定するために使用され得る(たとえば、動的畳み込みデータ、ブロックベースの動き

データなど)。

【０１７９】

上述のように、ワーピングされたフレーム

【０１８０】

【数３４】

$$\tilde{X}_t$$

【０１８１】

を生成するために、Pフレームコーディングシステム700は、X_tと

【０１８２】

【数３５】

$$\hat{X}_{ref}$$

【０１８３】

との間の動き

【０１８４】

【数３６】

$$\hat{f}_t$$

【０１８５】

(たとえば、オプティカルフローまたは他の動き)を予測すること、および動き予測

【０１８６】

【数３７】

$$\hat{f}_t$$

【０１８７】

を使用して参照フレーム

【０１８８】

【数３８】

$$\hat{X}_{ref}$$

【０１８９】

をワーピングすることによって、動き補償を実行することができる。しかしながら、動き予測

【０１９０】

【数３９】

$$\hat{f}_t$$

【０１９１】

に基づくフレーム予測

【０１９２】

【数４０】

$$\tilde{X}_t$$

【０１９３】
は、再構成されたフレーム
【０１９４】
【数４１】

$$\hat{X}_t$$

【０１９５】
として入力フレームX_tを表すのに十分に正確ではない場合がある。たとえば、入力フレームX_tによって描写されるシーンの中の1つもしくは複数の閉塞したエリア、過度の照明、照明の欠如、または再構成されたフレーム
【０１９６】
【数４２】

$$\hat{X}_t$$

【０１９７】
として使用されるのにフレーム予測
【０１９８】
【数４３】

$$\tilde{X}_t$$

【０１９９】
が十分に正確ではない結果となる他の影響があり得る。
【０２００】
残差予測システム706は、
【０２０１】
【数４４】

$$\tilde{X}_t$$

【０２０２】
の予測を補正するために、またフレームX_tを表す再構成されたフレーム
【０２０３】
【数４５】

$$\hat{X}_t$$

【０２０４】
を生成するために使用され得る。図7Aに示すように、残差r_tは、入力フレームX_tから予測フレーム
【０２０５】
【数４６】

$$\tilde{X}_t$$

【０２０６】
を減算することによって決定(減算演算708を使用して決定)され得る。たとえば、ワーピングエンジン704によって予測フレーム

【０２０７】
【数４７】

$$\tilde{X}_t$$

【０２０８】
が決定されると、Ｐフレームコーディングシステム７００は、予測フレーム
【０２０９】
【数４８】

$$\tilde{X}_t$$

【０２１０】
と入力フレームX_tとの間の差分を(減算演算７０８を使用して)決定することによって残差r_tを決定することができる。残差予測システム７０６のエンコーダネットワーク７０７は、残差r_tを符号化して残差r_tを表す潜在性z_rにすることができる。いくつかの例では、エンコーダネットワーク７０７は、潜在性z_rとして使用され得る潜在コードに残差r_tをマッピングすることができる。場合によっては、エンコーダネットワーク７０７は、潜在コードに対してエントロピーコーディングを実行することによって潜在性z_rをビットストリームに変換することができる。いくつかの例では、エンコーダネットワーク７０７はまた、(たとえば、エントロピーコーディングが実行される前に)潜在性z_rを量子化することができる。量子化された潜在性z_rは、残差r_tの量子化された表現を含むことができる。場合によっては、潜在性z_rは、量子化されたコードを表すニューラルネットワークデータ(たとえば、ニューラルネットワークノードの活性化マップまたは特徴マップ)を含むことができる。エンコーダネットワーク７０７は、潜在性z_rを記憶することができ、残差予測システム７０６のデコーダネットワーク７０９へ潜在性z_rを送ることができ、かつ/または潜在性z_rを復号できる別のデバイスもしくはシステムへ潜在性z_rを送ることができる。潜在性z_rを受信すると、デコーダネットワーク７０９は、潜在性z_rを復号(たとえば、逆エントロピーコーディング、逆量子化、および/または再構成)して予測残差
【０２１１】
【数４９】

$$\hat{r}_t$$

【０２１２】
を生成することができる。エンコーダネットワーク７０７およびデコーダネットワーク７０９は、以下で説明するように、トレーニングデータ(たとえば、トレーニング画像またはトレーニングフレーム)および1つまたは複数の損失関数を使用してトレーニングおよび最適化され得る。
【０２１３】
時間tにおける入力フレームX_tを表す再構成されたフレーム
【０２１４】
【数５０】

$$\hat{X}_t$$

【０２１５】
を生成するために、ワーピングエンジン７０４からのフレーム予測
【０２１６】
【数５１】

$\tilde{X}_t$

【０２１７】
とともに予測残差
【０２１８】
【数５２】

$\hat{r}_t$

【０２１９】
が使用され得る。たとえば、Pフレームコーディングシステム700は、予測残差
【０２２０】
【数５３】

$\hat{r}_t$

【０２２１】
をフレーム予測
【０２２２】
【数５４】

$\tilde{X}_t$

【０２２３】
に(加算演算710を使用して)加算して、再構成されたフレーム
【０２２４】
【数５５】

$\hat{X}_t$

【０２２５】
を生成することができる。場合によっては、残差予測システム706のデコーダネットワーク709が、予測残差
【０２２６】
【数５６】

$\hat{r}_t$

【０２２７】
をフレーム予測
【０２２８】
【数５７】

$\tilde{X}_t$

【０２２９】
に加算することができる。再構成されたフレーム
【０２３０】
【数５８】

$\hat{X}_t$

【０２３１】
は復号フレームと呼ばれる場合もある。再構成された現在のフレーム
【０２３２】
【数５９】

$\hat{X}_t$

【０２３３】
は、(たとえば、復号ピクチャバッファ(ＤＰＢ)または他のストレージの中での)記憶、送信、表示のために、(たとえば、後処理のための、さらなるインター予測における参照フレームなどとして)さらなる処理のために、かつ/または任意の他の使用のために出力され得る。
【０２３４】
上述のように、場合によっては、Ｐフレームコーディングシステム７００は、動きを表す潜在データ(潜在性z_m)および残差を表す潜在データ(潜在性z_r)をビットストリームの中で復号のために別のデバイスへ送信することができる。たとえば、他のデバイスは、潜在性z_mからの予測された動き
【０２３５】
【数６０】

$\hat{f}_t$

【０２３６】
および潜在性z_rからの予測残差
【０２３７】
【数６１】

$\hat{r}_t$

【０２３８】
を復号でき、かつ(たとえば、上記で説明したように、予測された動き
【０２３９】
【数６２】

$\hat{f}_t$

【０２４０】
を使用して生成されたフレーム予測
【０２４１】
【数６３】

$\tilde{X}_t$

【０２４２】
に予測残差
【０２４３】
【数６４】

$\hat{r}_t$

【０２４４】
を加算することによって)予測された動き
【０２４５】
【数６５】

$\hat{f}_t$

【０２４６】
および予測残差
【０２４７】
【数６６】

$\hat{r}_t$

【０２４８】
に基づいて、再構成されたフレーム
【０２４９】
【数６７】

$\hat{X}_t$

【０２５０】
を生成できる、ビデオデコーダ(たとえば、上記で説明したような動き予測システム７０２および残差予測システム７０６、または潜在性z_mおよびz_rを復号するように構成された別のデコーダ)を含むことができる。
【０２５１】
いくつかの例では、動き予測システム７０２および/または残差予測システム７０６は、トレーニングデータおよび1つまたは複数の損失関数を使用してトレーニングおよび最適化され得る。場合によっては、動き予測システム７０２および/または残差予測システム７０６は、(同じトレーニングプロセスの間にすべてのニューラルネットワーク構成要素がトレーニングされる)エンドツーエンド方式でトレーニングされ得る。いくつかの態様では、トレーニングデータはトレーニング画像またはトレーニングフレームを含むことができる。場合によっては、動き予測システム７０２および/または残差予測システム７０６がトレーニング画像またはトレーニングフレームを処理することに基づくトレーニングのために、損失関数Lossが使用され得る。一例では、LossはLoss=D+βRとして定義することができ、ただし、Dは、フレーム(たとえば、入力フレームX_t)とその再構成されたフレームとの間のひずみ(たとえば、D(フレームX_t、フレーム
【０２５２】
【数６８】

$\hat{X}_t$

【０２５３】
))であり、βは、ビットレート(ピクセル当りのビット)を制御するハイパーパラメータであり、Rは、残差(たとえば、残差r_t)を圧縮されたビットストリーム(たとえば、潜在性z_r)に変換するために使用されるビット数である。いくつかの例では、ひずみDは、ピーク信号対雑音比(PSNR)、構造的類似性インデックス尺度(SSIM:structural similarity

index measure)、マルチスケールSSIM(MS-SSIM)などに基づいて計算され得る。トレーニングデータおよび損失関数を使用して、動き予測システム702および/または残差予測システム706のパラメータ(たとえば、重み、バイアスなど)は、所望のコーディング結果がPフレームコーディングシステム700によって達成されるまで調整され得る。

【０２５４】

図7Bは、図7AのPフレームコーディングシステム700を使用する双方向コーディングを実行するために使用され得るBフレームコーディングシステム720の一例を示す図である。Bフレームコーディングシステム720はまた、フレーム補間エンジン722を含む。いくつかの構成要素を含むようにBフレームコーディングシステム720が示されるが、Bフレームコーディングシステム720が、図7Bに示す構成要素よりも少数または多数の構成要素を含むことができることを、当業者は諒解されよう。

【０２５５】

補間エンジン722は、

【０２５６】

【数６９】

$$\hat{X}_{ref_0}$$

【０２５７】

および

【０２５８】

【数７０】

$$\hat{X}_{ref_1}$$

【０２５９】

として示す2つの参照フレームを取得することができる。図示のように、Pフレームコーディングシステム700は、図7Aに関して上記で説明したのと同様に、時間tにおいて入力フレームX_tを取得する。入力フレームX_tは、時間的または連続的に参照フレーム

【０２６０】

【数７１】

$$\hat{X}_{ref_0}$$

【０２６１】

の後かつ参照フレーム

【０２６２】

【数７２】

$$\hat{X}_{ref_1}$$

【０２６３】

の前のフレームであり得る。いくつかの例では、入力フレームX_tは、時間tにおける現在のフレームであり得、参照フレーム

【０２６４】

【数７３】

$$\hat{X}_{ref_0}$$

【０２６５】

は、時間t-nにおける(たとえば、時間的または連続的に時間tにおける入力フレームX_tの前の)過去のフレームであり得、参照フレーム

【０２６６】

【数７４】

$$\hat{X}_{ref_1}$$

【０２６７】

は、時間t+mにおける(たとえば、時間的または連続的に入力フレームX_tの後の)将来のフレームであり得、ただし、nおよびmは同じかまたは異なる値を表すことができる。例示的な一例では、参照フレーム

【０２６８】

【数７５】

$$\hat{X}_{ref_0}$$

【０２６９】

は、時間t-1における入力フレームX_tの直前のフレームであり得、参照フレーム

【０２７０】

【数７６】

$$\hat{X}_{ref_1}$$

【０２７１】

は、時間t+1における入力フレームX_tの直後のフレームであり得る。

【０２７２】

場合によっては、入力フレームXtはBフレームであり得、参照フレーム

【０２７３】

【数７７】

$$\hat{X}_{ref_0}$$

【０２７４】

および

【０２７５】

【数７８】

$$\hat{X}_{ref_1}$$

【０２７６】

の各々は、Iフレーム、Pフレーム、またはBフレームであり得る。たとえば、参照フレーム

【０２７７】

【数７９】

$$\hat{X}_{ref_0}$$

【０２７８】

は、Iフレーム、Pフレーム、またはBフレームであり得、参照フレーム

【０２７９】

【数８０】

$$\hat{X}_{ref_1}$$

【０２８０】
は、Ｉフレーム、Ｐフレーム、またはＢフレームであり得る。場合によっては、参照フレーム
【０２８１】
【数８１】

$$\hat{X}_{ref_0}$$

【０２８２】
と
【０２８３】
【数８２】

$$\hat{X}_{ref_1}$$

【０２８４】
の両方が、Ｂフレーム、Ｐフレーム、またはＩフレームなどの同じタイプのフレームであり得る。他の場合には、参照フレーム
【０２８５】
【数８３】

$$\hat{X}_{ref_0}$$

【０２８６】
および参照フレーム
【０２８７】
【数８４】

$$\hat{X}_{ref_1}$$

【０２８８】
は、異なるタイプのフレームであり得る。たとえば、参照フレーム
【０２８９】
【数８５】

$$\hat{X}_{ref_0}$$

【０２９０】
はＩフレームであり得、参照フレーム
【０２９１】
【数８６】

$$\hat{X}_{ref_1}$$

【０２９２】
はＰフレームであり得る。場合によっては、参照フレーム

【０２９３】
【数８７】

$$\hat{X}_{ref_0}$$

【０２９４】
および
【０２９５】
【数８８】

$$\hat{X}_{ref_1}$$

【０２９６】
は、図9もしくは図10に示すピクチャグループ(GoP)構造に従って、かつ/または逐次方式(たとえば、図11に示す逐次方式)もしくは階層方式(たとえば、図12に示す階層方式)に従って、あるいは任意の他の方式に基づいて選択され得る。
【０２９７】
補間エンジン722は、2つの参照フレーム
【０２９８】
【数８９】

$$\hat{X}_{ref_0}$$

【０２９９】
および
【０３００】
【数９０】

$$\hat{X}_{ref_1}$$

【０３０１】
に対して補間動作を実行して、補間された参照フレーム
【０３０２】
【数９１】

$$\hat{X}_{ref_t}$$

【０３０３】
を生成することができる。たとえば、場合によっては、入力フレームX_tのための参照として使用され得る(たとえば、ピクセルのブロックなどの、フレームの部分だけでなく)参照フレーム全体を生成するために補間が実行される。補間された参照フレーム
【０３０４】
【数９２】

$$\hat{X}_{ref_t}$$

【０３０５】
が記憶され、かつ/またはPフレームコーディングシステム700に出力され得る。フレーム補間エンジン722は、参照フレーム
【０３０６】

【数９３】

$$\hat{X}_{ref_0}$$

【０３０７】
および参照フレーム
【０３０８】
【数９４】

$$\hat{X}_{ref_1}$$

【０３０９】
を使用して任意のタイプの補間を実行して、補間された参照フレーム
【０３１０】
【数９５】

$$\hat{X}_{ref_t}$$

【０３１１】
を生成することができる。フレーム補間エンジン722によって実行され得る補間技法の例は、たとえば、限定はしないが、オプティカルフローに基づくフレーム補間、移動勾配法、位相ベースの方法、カーネルベースの方法、および/または任意の他のフレーム補間技法を含むことができる。例示的な一例では、補間された参照フレーム
【０３１２】
【数９６】

$$\hat{X}_{ref_t}$$

【０３１３】
を生成するために、フレーム補間エンジン722は、参照フレーム
【０３１４】
【数９７】

$$\hat{X}_{ref_0}$$

【０３１５】
および参照フレーム
【０３１６】
【数９８】

$$\hat{X}_{ref_1}$$

【０３１７】
に対して、動き推定情報(たとえば、動きベクトルマップとも呼ばれるオプティカルフローマップなどの、オプティカルフロー情報)を計算することができる。フレーム補間エンジン722は、動き推定情報(たとえば、フローマップ)を補間して、補間された参照フレーム
【０３１８】
【数９９】

$\hat{X}_{ref_t}$

【０３１９】
に関連する時間tに対する動き推定値(たとえば、オプティカルフロー)を決定することができる。フレーム補間エンジン725は、補間された動き推定情報(たとえば、補間されたオプティカルフロー)ならびに参照フレーム
【０３２０】
【数１００】

$\hat{X}_{ref_0}$

【０３２１】
および
【０３２２】
【数１０１】

$\hat{X}_{ref_1}$

【０３２３】
を使用して、(たとえば、少なくとも部分的には参照フレーム
【０３２４】
【数１０２】

$\hat{X}_{ref_0}$

【０３２５】
および
【０３２６】
【数１０３】

$\hat{X}_{ref_1}$

【０３２７】
をワーピングすることによって)補間された参照フレーム
【０３２８】
【数１０４】

$\hat{X}_{ref_t}$

【０３２９】
を生成することができる。
【０３３０】
いくつかの例では、フレーム補間エンジン722は、ピクセルごとの動き推定情報(たとえば、オプティカルフロー情報)を決定することができる。たとえば、フレーム補間エンジン722は、参照フレーム
【０３３１】
【数１０５】

$\hat{X}_{ref_0}$

【０３３２】
の中の各ピクセル(または、参照フレーム
【０３３３】
【数１０６】

$$\hat{X}_{ref_0}$$

【０３３４】
の中のピクセルのサブセット)を参照フレーム
【０３３５】
【数１０７】

$$\hat{X}_{ref_1}$$

【０３３６】
の中の対応するピクセル(または、参照フレーム
【０３３７】
【数１０８】

$$\hat{X}_{ref_1}$$

【０３３８】
の中のピクセルの対応するサブセット)の位置にマッピングすることができ、(たとえば、両方の方向において)逆も同様である。一例では、ピクセルごとの動き推定情報(たとえば、オプティカルフロー情報)は、参照フレーム
【０３３９】
【数１０９】

$$\hat{X}_{ref_0}$$

【０３４０】
と
【０３４１】
【数１１０】

$$\hat{X}_{ref_1}$$

【０３４２】
との間でのピクセルの移動(たとえば、参照フレーム
【０３４３】
【数１１１】

$$\hat{X}_{ref_0}$$

【０３４４】
から参照フレーム
【０３４５】
【数１１２】

$$\hat{X}_{ref_1}$$

【０３４６】
へのピクセルの移動、および/または参照フレーム
【０３４７】
【数１１３】

$$\hat{X}_{ref_1}$$

【０３４８】
から参照フレーム
【０３４９】
【数１１４】

$$\hat{X}_{ref_0}$$

【０３５０】
へのピクセルの移動)を示すベクトルを含むことができる。いくつかの例では、オプティカルフローマップは、参照フレーム
【０３５１】
【数１１５】

$$\hat{X}_{ref_0}$$

【０３５２】
と
【０３５３】
【数１１６】

$$\hat{X}_{ref_1}$$

【０３５４】
との間のオプティカルフローベクトルに基づいて生成され得る。オプティカルフローマップは、フレームの中のピクセルごとのオプティカルフローベクトルを含むことができ、ここで、各ベクトルはフレーム間でのピクセルの移動を示す。
【０３５５】
例示的な一例では、ピクセルに対するオプティカルフローベクトルは、第1のフレームから第2のフレームへのピクセルの移動を示す(たとえば、x変位およびy変位などの水平変位および垂直変位を示す)変位ベクトルであり得る。場合によっては、フレーム補間エンジン722は、参照フレーム
【０３５６】
【数１１７】

$$\hat{X}_{ref_0}$$

【０３５７】
に対する(たとえば、水平変位および垂直変位を示す)2次元(2D)フローマップおよび参照フレーム
【０３５８】
【数１１８】

$$\hat{X}_{ref_1}$$

【０３５９】
に対する2Dフローマップを計算することができる。いくつかの態様では、フレーム補間エンジン722は、2Dフローマップをマージして、補間された参照フレーム
【０３６０】
【数１１９】

$$\hat{X}_{ref_t}$$

【０３６１】
に対する3次元(3D)フローマップを生成することができる。3Dフローマップは、それぞれ、水平変位および垂直変位を示す水平次元および垂直次元、ならびに水平変位および垂直変位が適用されるフレームを識別するフレームインデックスを含むフレーム次元を含むことができる。たとえば、フレーム次元が参照フレーム
【０３６２】
【数１２０】

$$\hat{X}_{ref_0}$$

【０３６３】
を識別する場合、水平次元および垂直次元は、参照フレーム
【０３６４】
【数１２１】

$$\hat{X}_{ref_0}$$

【０３６５】
の中のピクセルに対する水平変位および垂直変位(たとえば、オプティカルフロー情報に基づいて参照フレーム
【０３６６】
【数１２２】

$$\hat{X}_{ref_0}$$

【０３６７】
の中のピクセルがどのくらい移動されるべきであるのか)を示すことができる。
【０３６８】
フレーム補間エンジン722は、補間されたオプティカルフロー(たとえば、3Dフローマップ)を使用して参照フレーム
【０３６９】
【数１２３】

$$\hat{X}_{ref_0}$$

【０３７０】
および参照フレーム
【０３７１】
【数１２４】

$$\hat{X}_{ref_1}$$

【０３７２】
を入力フレームX_tの特定の時間ステップ(たとえば、時間t)にワーピングすることができ、ワーピングされたフレームを融合させて、補間された参照フレーム
【０３７３】
【数１２５】

$$\hat{X}_{ref_t}$$

【０３７４】
を生成することができる。フレーム補間エンジンのそのような例示的な実装形態のさらなる詳細が、図7Cに関して以下で説明される。
【０３７５】
補間された参照フレーム
【０３７６】
【数１２６】

$$\hat{X}_{ref_t}$$

【０３７７】
および入力フレームX_tを入力として使用して、Pフレームコーディングシステム700は、再構成されたフレーム
【０３７８】
【数１２７】

$$\hat{X}_t$$

【０３７９】
を生成するために、図7Aに関して上記で説明したのと同じかまたは類似の動作を実行することができる。たとえば、動き予測システム702のエンコーダネットワーク705は、入力フレームX_tと補間された参照フレーム
【０３８０】
【数１２８】

$$\hat{X}_{ref_t}$$

【０３８１】
との間の動きを決定することができ、その動きを符号化して潜在性z_mにすることができる。動き予測システム702のデコーダネットワーク703は、潜在性z_mを復号して、予測される動き
【０３８２】
【数１２９】

$$\hat{f}_t$$

【０３８３】
を生成することができる。ワーピングエンジン704は、図7Aに関して上記で説明したように、予測された動き
【０３８４】
【数１３０】

$\hat{f}_t$

【０３８５】
を使用して、補間された参照フレーム
【０３８６】
【数１３１】

$\hat{X}_{ref_t}$

【０３８７】
のピクセルを(たとえば、動き補償を実行することによって)修正してフレーム予測
【０３８８】
【数１３２】

$\tilde{X}_t$

【０３８９】
を生成することができる。残差予測システム706のエンコーダネットワーク707は、(フレーム予測
【０３９０】
【数１３３】

$\tilde{X}_t$

【０３９１】
と入力フレームX_tとの間の差分として決定される)残差r_tを取得することができ、残差r_tを符号化して潜在性z_rにすることができる。残差予測システム706のデコーダネットワーク709は、潜在性z_rを復号して予測残差
【０３９２】
【数１３４】

$\hat{r}_t$

【０３９３】
を生成することができる。Bフレームコーディングシステム720は、次いで、フレーム予測
【０３９４】
【数１３５】

$\tilde{X}_t$

【０３９５】
を予測残差
【０３９６】
【数１３６】

$\hat{r}_t$

【０３９７】

に加算して、再構成されたフレーム
【０３９８】
【数１３７】

$$\hat{X}_t$$

【０３９９】
を生成することができる。
【０４００】
図7Cは、Bフレームコーディングシステムによって使用され得る例示的なフレーム補間エンジン725のシステムフローを示す図である。フレーム補間エンジン725は、図7Bのフレーム補間エンジン722の例示的な例である。図7Cの例では、フレーム補間エンジン725は、オプティカルフロー推定層730、オプティカルフロー予測層732、2Dワーピング層734、改善層736、および3Dワーピング層738を含む。
【０４０１】
図7Cに示すように、参照フレーム
【０４０２】
【数１３８】

$$\hat{X}_{ref_0}$$

【０４０３】
および参照フレーム
【０４０４】
【数１３９】

$$\hat{X}_{ref_1}$$

【０４０５】
がオプティカルフロー推定層730に入力される。いくつかの例では、オプティカルフロー推定層730は、1つまたは複数のニューラルネットワークまたは他の機械学習システムを含むことができる。たとえば、オプティカルフロー推定層730は、1つまたは複数のオプティカルフロー推定ニューラルネットワークを使用して実装され得る。オプティカルフロー推定層730は、参照フレーム
【０４０６】
【数１４０】

$$\hat{X}_{ref_0}$$

【０４０７】
および参照フレーム
【０４０８】
【数１４１】

$$\hat{X}_{ref_1}$$

【０４０９】
を使用して、参照フレーム
【０４１０】
【数１４２】

$\hat{X}_{ref_0}$

【０４１１】
および参照フレーム
【０４１２】
【数１４３】

$\hat{X}_{ref_1}$

【０４１３】
に対する双方向動き情報７４０(たとえば、順方向オプティカルフローおよび逆方向オプティカルフロー)を推定することができる。双方向動き情報７４０は、参照フレーム
【０４１４】
【数１４４】

$\hat{X}_{ref_0}$

【０４１５】
から参照フレーム
【０４１６】
【数１４５】

$\hat{X}_{ref_1}$

【０４１７】
へのオプティカルフローまたはオプティカルフローマップ($f_{0\to1}$として示す)、および参照フレーム
【０４１８】
【数１４６】

$\hat{X}_{ref_1}$

【０４１９】
から参照フレーム
【０４２０】
【数１４７】

$\hat{X}_{ref_0}$

【０４２１】
へのオプティカルフローまたはオプティカルフローマップ($f_{1\to0}$として示す)を含むことができる。
【０４２２】
たとえば、双方向動き情報７４０は、参照フレーム
【０４２３】
【数１４８】

$\hat{X}_{ref_0}$

【０４２４】
の中の各ピクセル(または、参照フレーム
【０４２５】
【数１４９】

$$\hat{X}_{ref_0}$$

【０４２６】
の中のピクセルのサブセット)の位置を参照フレーム
【０４２７】
【数１５０】

$$\hat{X}_{ref_1}$$

【０４２８】
の中の対応するピクセル(または、参照フレーム
【０４２９】
【数１５１】

$$\hat{X}_{ref_1}$$

【０４３０】
の中のピクセルの対応するサブセット)の位置にマッピングするオプティカルフローマップ$f_{0\to 1}$を含むことができる。双方向動き情報740はまた、参照フレーム
【０４３１】
【数１５２】

$$\hat{X}_{ref_1}$$

【０４３２】
の中の各ピクセル(または、参照フレーム
【０４３３】
【数１５３】

$$\hat{X}_{ref_1}$$

【０４３４】
の中のピクセルのサブセット)の位置を参照フレーム
【０４３５】
【数１５４】

$$\hat{X}_{ref_0}$$

【０４３６】
の中の対応するピクセル(または、参照フレーム
【０４３７】
【数１５５】

$$\hat{X}_{ref_0}$$

【０４３８】
の中のピクセルの対応するサブセット)の位置にマッピングするオプティカルフローマップ$f_{1\to 0}$を含むことができる。場合によっては、ピクセルごとの動き情報７４０は、参照フレーム
【０４３９】
【数１５６】

$$\hat{X}_{ref_0}$$

【０４４０】
と
【０４４１】
【数１５７】

$$\hat{X}_{ref_1}$$

【０４４２】
との間でのピクセルの移動を示すベクトルを含むことができる。いくつかの例では、オプティカルフローマップ$f_{0\to 1}$および$f_{1\to 0}$は、参照フレーム
【０４４３】
【数１５８】

$$\hat{X}_{ref_0}$$

【０４４４】
と
【０４４５】
【数１５９】

$$\hat{X}_{ref_1}$$

【０４４６】
とのピクセル間のオプティカルフローベクトルに基づいて生成され得る。オプティカルフローマップは、フレームの中のピクセルごとのオプティカルフローベクトルを含むことができ、ここで、各ベクトルはフレーム間でのピクセルの移動を示す。
【０４４７】
オプティカルフロー推定層７３０は、双方向動き情報７４０(たとえば、$f_{0\to 1}$および$f_{1\to 0}$)をオプティカルフロー予測層７３２に提供することができる。オプティカルフロー予測層７３２は、区間０～１の中の時間tを示す時間情報７４２を取得することができる(たとえば、$t\in(0,1)$)、ただし、０は参照フレーム
【０４４８】
【数１６０】

$$\hat{X}_{ref_0}$$

【０４４９】
の時間ステップを表し、１は参照フレーム
【０４５０】
【数１６１】

$\hat{X}_{ref_1}$

【０４５１】
の時間ステップを表し、時間tは、補間されるべきフレーム(たとえば、補間フレーム
【０４５２】
【数１６２】

$\hat{X}_{ref_t}$

【０４５３】
)の時間ステップに対応し、入力フレーム(たとえば、図7Bの入力フレームX_t)の時間ステップにも対応する。オプティカルフロー予測層732は、双方向動き情報740および時間情報742を使用して2Dオプティカルフローマップ744を予測することができる。いくつかの例では、オプティカルフロー推定層730は、参照フレーム
【０４５４】
【数１６３】

$\hat{X}_{ref_0}$

【０４５５】
および参照フレーム
【０４５６】
【数１６４】

$\hat{X}_{ref_1}$

【０４５７】
に対する双方向動き情報740(たとえば、$f_{0\to1}$および$f_{1\to0}$)を補間して、時間tにおけるオプティカルフローを決定することができる。場合によっては、時間tにおけるオプティカルフローは、$f_{0\to1}$と$f_{1\to0}$との間の線形補間であり得る。
【０４５８】
いくつかの例では、2Dオプティカルフローマップ744は、時間tから参照フレーム
【０４５９】
【数１６５】

$\hat{X}_{ref_0}$

【０４６０】
への(たとえば、参照フレーム
【０４６１】
【数１６６】

$\hat{X}_{ref_0}$

【０４６２】
の時間ステップ0への)オプティカルフローマップ(たとえば、
【０４６３】
【数１６７】

$\hat{f}_{t\to 0}^{2D}$

【０４６４】
)、および時間tから参照フレーム
【０４６５】
【数１６８】

$\hat{X}_{ref_1}$

【０４６６】
への(たとえば、参照フレーム
【０４６７】
【数１６９】

$\hat{X}_{ref_1}$

【０４６８】
の時間ステップ1への)オプティカルフローマップ(たとえば、
【０４６９】
【数１７０】

$\hat{f}_{t\to 1}^{2D}$

【０４７０】
)を含むことができる。たとえば、2Dオプティカルフローマップ744は、参照フレーム
【０４７１】
【数１７１】

$\hat{X}_{ref_0}$

【０４７２】
の時間tからある時系列(たとえば、図7Cに示す例では時系列0)までの動き情報を有する逆方向2Dフローマップ
【０４７３】
【数１７２】

$\hat{f}_{t\to 0}^{2D}$

【０４７４】
および参照フレーム
【０４７５】
【数１７３】

$\hat{X}_{ref_1}$

【０４７６】
の時間tからある時系列(たとえば、図7Cに示す例では時系列1)までの動き情報を有する順方向2Dフローマップ
【０４７７】

【数１７４】

$\hat{f}_{t\to 1}^{2D}$

【０４７８】
を含むことができる。
【０４７９】
例示的な一例では、オプティカルフロー予測層７３２は、次のように式１および式２に従って、２Ｄオプティカルフローマップ７４４を予測することができる。
【０４８０】
【数１７５】

$$\hat{f}_{t\to 0}^{2D} = -1(1-t)tf_{0\to 1} + t^2 f_{1\to 0} \qquad \text{式}(1)$$

$$\hat{f}_{t\to 1}^{2D} = 1(1-t)^2 f_{0\to 1} - t(1-t)f_{1\to 0} \qquad \text{式}(2)$$

【０４８１】
上の式１および式２において、ｔは、参照フレーム
【０４８２】
【数１７６】

$\hat{X}_{ref_0}$

【０４８３】
の時間ステップと参照フレーム
【０４８４】
【数１７７】

$\hat{X}_{ref_1}$

【０４８５】
の時間ステップ(たとえば、区間０～１の中の時間ステップ)との間の(入力フレームX_tに対応する)補間フレーム
【０４８６】
【数１７８】

$\hat{X}_{ref_t}$

【０４８７】
の時間ステップを表し、$f_{0\to 1}$は、参照フレーム
【０４８８】
【数１７９】

$\hat{X}_{ref_0}$

【０４８９】
から参照フレーム
【０４９０】

【数１８０】

$$\hat{X}_{ref_1}$$

【０４９１】
へのオプティカルフローを表し、$f_{1\to 0}$は、参照フレーム
【０４９２】
【数１８１】

$$\hat{X}_{ref_1}$$

【０４９３】
から参照フレーム
【０４９４】
【数１８２】

$$\hat{X}_{ref_0}$$

【０４９５】
へのオプティカルフローを表し、
【０４９６】
【数１８３】

$$\hat{f}_{t\to 0}^{2D}$$

【０４９７】
は、tから参照フレーム
【０４９８】
【数１８４】

$$\hat{X}_{ref_0}$$

【０４９９】
に関連する時間ステップ(たとえば、図7Cに示す例では時間ステップ0)への逆方向の動きを識別する2Dオプティカルフローマップを表し、
【０５００】
【数１８５】

$$\hat{f}_{t\to 1}^{2D}$$

【０５０１】
は、tから参照フレーム
【０５０２】
【数１８６】

$$\hat{X}_{ref_1}$$

【０５０３】
に関連する時間ステップ(たとえば、図7Cに示す例では時間ステップ1)への順方向の動きを識別する2Dオプティカルフローマップを表す。

【０５０４】
2Dワーピング層734は、2Dオプティカルフローマップ744、参照フレーム
【０５０５】
【数１８７】

$$\hat{X}_{ref_0}$$

【０５０６】
、および参照フレーム
【０５０７】
【数１８８】

$$\hat{X}_{ref_1}$$

【０５０８】
を取得することができ、ワーピングされたフレーム746を生成することができる。ワーピングされたフレーム746は、参照フレーム
【０５０９】
【数１８９】

$$\hat{X}_{ref_0}$$

【０５１０】
に対応するワーピングされたフレーム(W_{x0})を含むことができる。2Dワーピング層734は、2Dオプティカルフローマップ
【０５１１】
【数１９０】

$$\hat{f}_{t\to 0}^{2D}$$

【０５１２】
を参照フレーム
【０５１３】
【数１９１】

$$\hat{X}_{ref_0}$$

【０５１４】
に適用することによって、ワーピングされたフレームW_{x0}を生成することができる。ワーピングされたフレーム746はまた、参照フレーム
【０５１５】
【数１９２】

$$\hat{X}_{ref_1}$$

【０５１６】
に対応するワーピングされたフレーム(W_{x1})を含むことができる。2Dワーピング層734は、2Dオプティカルフローマップ
【０５１７】
【数１９３】

$\hat{f}_{t\to 1}^{2D}$

【０５１８】
を参照フレーム
【０５１９】
【数１９４】

$\hat{X}_{ref_1}$

【０５２０】
に適用することによって、ワーピングされたフレームW_{x1}を生成することができる。２Ｄワーピング層７３４は、参照フレーム
【０５２１】
【数１９５】

$\hat{X}_{ref_0}$

【０５２２】
を時間ステップｔにワーピングするとともに、参照フレーム
【０５２３】
【数１９６】

$\hat{X}_{ref_1}$

【０５２４】
を時間ステップｔにワーピングするために、１つまたは複数のワーピング関数を実施することができる。
【０５２５】
いくつかの例では、ワーピングされたフレームW_{x0}を生成するために、２Ｄワーピング層７３４は、参照フレーム
【０５２６】
【数１９７】

$\hat{X}_{ref_0}$

【０５２７】
に対して計算された２Ｄオプティカルフローマップ
【０５２８】
【数１９８】

$\hat{f}_{t\to 0}^{2D}$

【０５２９】
を使用して、参照フレーム
【０５３０】
【数１９９】

$\hat{X}_{ref_0}$

【０５３１】
の中のピクセルを、時間tに対応しかつ/または時間tに対する、それぞれの位置に移動させることができる。同様に、ワーピングされたフレームW_{x1}を生成するために、2Dワーピング層734は、参照フレーム
【０５３２】
【数２００】

$$\hat{X}_{ref_1}$$

【０５３３】
に対して計算された2Dオプティカルフローマップ
【０５３４】
【数２０１】

$$\hat{f}_{t\to 1}^{2D}$$

【０５３５】
を使用して、参照フレーム
【０５３６】
【数２０２】

$$\hat{X}_{ref_1}$$

【０５３７】
の中のピクセルを、時間tに対応しかつ/または時間tに対する、それぞれの位置に移動させることができる。場合によっては、ワーピングされたフレーム746は、時間tにおけるピクセルの垂直変位および/または水平変位(および/または、ピクセル位置)を考慮に入れる、参照フレーム
【０５３８】
【数２０３】

$$\hat{X}_{ref_0}$$

【０５３９】
に対する2Dフレーム、ならびに時間tにおけるピクセルの垂直変位および/または水平変位(および/または、ピクセル位置)を考慮に入れる、参照フレーム
【０５４０】
【数２０４】

$$\hat{X}_{ref_1}$$

【０５４１】
に対する2Dフレームを含むことができる。
【０５４２】
改善層736は、2Dオプティカルフローマップ744、ワーピングされたフレーム746、ならびに参照フレーム
【０５４３】
【数２０５】

$$\hat{X}_{ref_0}$$

【０５４４】
および
【０５４５】
【数２０６】

$$\hat{X}_{ref_1}$$

【０５４６】
を入力として取得することができ、入力データを使用して改善動作を実行することができる。改善層７３６は、いくつかの例ではフレーム補間エンジン７２５の随意の構成要素であり得る。場合によっては、改善層７３６は、１つまたは複数のニューラルネットワークまたは他の機械学習システムを含むことができる。例示的な一例では、改善層７３６は、１つまたは複数の補間畳み込みニューラルネットワーク(ＣＮＮ)を含むことができる。いくつかの例では、改善層７３６によって実行される改善動作は、２Ｄオプティカルフローマップ７４４の中の値を改善および/または補正することを含むことができる。たとえば、場合によっては、参照フレーム
【０５４７】
【数２０７】

$$\hat{X}_{ref_0}$$

【０５４８】
に対応する時間ステップ０と参照フレーム
【０５４９】
【数２０８】

$$\hat{X}_{ref_1}$$

【０５５０】
に対応する時間ステップ１との間の移動の一部または全部が線形でない場合がある。改善層７３６は、そのような非線形の移動を検出および補正することができる。改善層７３６は、そのような非線形性を考慮に入れる改善された２Ｄオプティカルフローを生成することができる。追加または代替として、改善層７３６は、ワーピングされたフレーム７４６のうちのいずれか(ならびに/または、参照フレーム
【０５５１】
【数２０９】

$$\hat{X}_{ref_0}$$

【０５５２】
および
【０５５３】
【数２１０】

$$\hat{X}_{ref_1}$$

【０５５４】
のうちのいずれか)の中のアーティファクトを改善または補正することができ、エッジなどの、ワーピングされたフレーム７４６のうちのいずれか(ならびに/または、参照フレーム
【０５５５】

【数２１１】

$$\hat{X}_{ref_0}$$

【０５５６】
および
【０５５７】
【数２１２】

$$\hat{X}_{ref_1}$$

【０５５８】
のうちのいずれか)の部分を改善または補正することができ、ワーピングされたフレーム746のうちのいずれか(ならびに/または、参照フレーム
【０５５９】
【数２１３】

$$\hat{X}_{ref_0}$$

【０５６０】
および
【０５６１】
【数２１４】

$$\hat{X}_{ref_1}$$

【０５６２】
のうちのいずれか)の中の閉塞を考慮に入れるように改善または補正を加えることができ、かつ/あるいはワーピングされたフレーム746ならびに/または参照フレーム
【０５６３】
【数２１５】

$$\hat{X}_{ref_0}$$

【０５６４】
および
【０５６５】
【数２１６】

$$\hat{X}_{ref_1}$$

【０５６６】
に関連する画像データおよび/またはオプティカルフロー情報の任意の他の改善を実行することができる。
【０５６７】
いくつかの例では、改善動作はまた、補間フレーム
【０５６８】
【数２１７】

$$\hat{X}_{ref_t}$$

【０５６９】
に対して2Dオプティカルフローマップ744を3Dオプティカルフローマップ748にマージすることを含むことができる。たとえば、前に言及したように、2Dオプティカルフローマップ744は、参照フレーム
【０５７０】
【数２１８】

$$\hat{X}_{ref_0}$$

【０５７１】
に対する2Dオプティカルフローマップ(たとえば、
【０５７２】
【数２１９】

$$\hat{f}_{t\to 0}^{2D}$$

【０５７３】
)、および参照フレーム
【０５７４】
【数２２０】

$$\hat{X}_{ref_1}$$

【０５７５】
に対する2Dオプティカルフローマップ(たとえば、
【０５７６】
【数２２１】

$$\hat{f}_{t\to 1}^{2D}$$

【０５７７】
)を含むことができる。改善層736は、2Dオプティカルフローマップ744(たとえば、2Dオプティカルフローマップ
【０５７８】
【数２２２】

$$\hat{f}_{t\to 0}^{2D}$$

【０５７９】
と2Dオプティカルフローマップ
【０５８０】
【数２２３】

$$\hat{f}_{t\to 1}^{2D}$$

【０５８１】
)とをマージして3Dオプティカルフローマップ748を生成することができる。3Dオプティカルフローマップ748は、特定のフレーム(たとえば、参照フレーム
【０５８２】
【数２２４】

$\hat{X}_{ref_0}$

【０５８３】
および参照フレーム
【０５８４】
【数２２５】

$\hat{X}_{ref_1}$

【０５８５】
)を、対応するオプティカルフロー情報にマッピングすることができる。たとえば、3Dオプティカルフローマップ748は、2つの空間次元(たとえば、X軸およびY軸または垂直次元および水平次元)、およびフレーム情報を含む第3の次元(たとえば、フレーム次元)を含むことができる。第3の次元の中のフレーム情報は、2つの空間次元の中の動き情報(たとえば、垂直変位および/または水平変位)が適用されるフレーム(たとえば、参照フレーム
【０５８６】
【数２２６】

$\hat{X}_{ref_0}$

【０５８７】
または参照フレーム
【０５８８】
【数２２７】

$\hat{X}_{ref_1}$

【０５８９】
)を識別することができる。
【０５９０】
たとえば、第3の次元(たとえば、フレーム次元)が参照フレーム
【０５９１】
【数２２８】

$\hat{X}_{ref_0}$

【０５９２】
を識別する場合、空間次元の中の(たとえば、X軸およびY軸または垂直次元および水平次元における)対応する値は、参照フレーム
【０５９３】
【数２２９】

$\hat{X}_{ref_0}$

【０５９４】
の中のピクセルにとっての水平変位および/または垂直変位(たとえば、空間次元の中のオプティカルフロー情報に基づいて参照フレーム
【０５９５】

【数２３０】

$$\hat{X}_{ref_0}$$

【０５９６】
の中のピクセルがどのくらい移動されるべきであるのか)を示すことができる。同様に、第3の次元が参照フレーム
【０５９７】
【数２３１】

$$\hat{X}_{ref_1}$$

【０５９８】
を識別する場合、空間次元の中の対応する値は、参照フレーム
【０５９９】
【数２３２】

$$\hat{X}_{ref_1}$$

【０６００】
の中のピクセルにとっての水平変位および/または垂直変位(たとえば、空間次元の中のオプティカルフロー情報に基づいて参照フレーム
【０６０１】
【数２３３】

$$\hat{X}_{ref_1}$$

【０６０２】
の中のピクセルがどのくらい移動されるべきであるのか)を示すことができる。
【０６０３】
場合によっては、改善層736は、2Dオプティカルフローマップ744を使用して2Dオプティカルフローマップ744を3Dオプティカルフローマップ748にマージすることができる。いくつかの例では、改善層736は、2Dオプティカルフローマップ744を改善することによって生成される改善された2Dオプティカルフローを使用して、2Dオプティカルフローマップ744を3Dオプティカルフローマップ748にマージすることができる。場合によっては、改善層736は、参照フレーム
【０６０４】
【数２３４】

$$\hat{X}_{ref_0}$$

【０６０５】
および
【０６０６】
【数２３５】

$$\hat{X}_{ref_1}$$

【０６０７】
、ワーピングされたフレーム746、ならびに/または2Dオプティカルフローマップ744(

または、改善された2Dオプティカルフロー)を入力として使用して、3Dオプティカルフローマップ748を生成することができる。

【０６０８】

3Dワーピング層738は、3Dオプティカルフローマップ748ならびに参照フレーム

【０６０９】

【数２３６】

$$\hat{X}_{ref_0}$$

【０６１０】

および

【０６１１】

【数２３７】

$$\hat{X}_{ref_1}$$

【０６１２】

(および/または、ワーピングされたフレーム746)を使用して、(入力フレームX_tの時間ステップに対応する)時間ステップtに対応する補間された参照フレーム

【０６１３】

【数２３８】

$$\hat{X}_{ref_t}$$

【０６１４】

を生成することができる。たとえば、3Dワーピング層738は、3Dオプティカルフローマップ748を使用して参照フレーム

【０６１５】

【数２３９】

$$\hat{X}_{ref_0}$$

【０６１６】

および

【０６１７】

【数２４０】

$$\hat{X}_{ref_1}$$

【０６１８】

(または、ワーピングされたフレーム746)をワーピングすることができ、ワーピングされたフレーム(たとえば、ワーピングされた参照フレーム

【０６１９】

【数２４１】

$$\hat{X}_{ref_0}$$

【０６２０】

および

【０６２１】

【数２４２】

$\hat{X}_{ref_1}$

【０６２２】
またはワーピングされたフレーム７４６)を融合させて、補間された参照フレーム
【０６２３】
【数２４３】

$\hat{X}_{ref_t}$

【０６２４】
を生成することができる。場合によっては、３Ｄワーピング層７３８は１つまたは複数のワーピング関数を実施することができる。例示的な一例では、補間フレームを生成するために以下のことが使用され得る。
【０６２５】
【数２４４】

$$\hat{I}_t = \propto_0 \odot\ g(I_0, F_{t\to 0}) + (1 - \propto_0) \odot\ g(I_1, F_{t\to 1}),$$

式(3)

【０６２６】
ただし、g(・,・)は、(場合によっては双線形補間を使用して実施され得る)逆方向ワーピング関数であり、パラメータ$\propto_0$は、２つの入力画像I_0およびI_1の寄与を制御し、時間的な一貫性および閉塞推論に依存し、項
【０６２７】
【数２４５】

$\odot$

【０６２８】
は、($\propto_0$パラメータに基づく、入力画像の内容を意識した重み付け(content-aware weighting)をもたらす)要素ごとの乗算を示す。
【０６２９】
場合によっては、オプティカルフロー推定層７３０、オプティカルフロー予測層７３２、２Ｄワーピング層７３４、改善層７３６、および/または３Ｄワーピング層７３８は、１つまたは複数の損失関数を使用してトレーニングおよび/または最適化され得る。場合によっては、オプティカルフロー推定層７３０、オプティカルフロー予測層７３２、２Ｄワーピング層７３４、改善層７３６、および３Ｄワーピング層７３８は、１つまたは複数の損失関数を使用してエンドツーエンドでトレーニングされ得る。例示的な一例では、上記で説明した損失(Loss=D+βR)が使用され得る。
【０６３０】
図７Ｃはオプティカルフロー補間を使用することに関して説明されるが、本明細書ではオプティカルフロー補間が説明の目的で例示的な例として使用されることに留意されたい。図７Ｂのフレーム補間エンジン７２２および/または図７Ｃのフレーム補間エンジン７２５が任意の他のフレーム補間技法を実施できることを、当業者は認識されよう。
【０６３１】
図８は、参照フレーム間の(たとえば、参照フレーム

【０６３２】
【数２４６】

$$\hat{X}_{ref_0}$$

【０６３３】
から参照フレーム
【０６３４】
【数２４７】

$$\hat{X}_{ref_1}$$

【０６３５】
への、またはその逆の)動き情報を決定するために実行され得る異なるタイプの動き推定を示す図である。図８における動き推定は、(たとえば、図７Ｂおよび図７Ｃに関して説明したような)補間された参照フレームを決定するために、かつ/または(たとえば、図７Ａおよび図７Ｂのワーピングエンジン７０４によって)参照フレームをワーピングするために使用され得る。図８において、項ｘは、動きがそこから推定され得る参照フレームを示し、項ｆは動き推定値を示し、項ｙは、次のように、すなわち、ｙ＝ｆ(ｘ)のように算出され得るワーピングされたフレームを示す。
【０６３６】
１つのタイプの動き推定は、ブロックベースの動き推定技法８０２である。ブロックベースの動き推定はブロックごとに実行され得る。たとえば、フレームｙの中のブロックごとに、動き推定値ｆは、フレームｘの中の対応するブロックの位置を規定する。例示的な一例では、動き推定値ｆは、フレームｘの中の対応するブロックに対するフレームｙの中のブロックの変位(たとえば、水平変位および垂直変位)を示す動きベクトルを含むことができる。フレームｘからのブロックは、ブロック間の類似度(たとえば、ピクセル値での類似度)を決定することによって、フレームｙの中のブロックに対応するものと決定され得る。
【０６３７】
実行され得る別のタイプの動き推定は、オプティカルフロー動き推定技法８０４である。オプティカルフロー動き推定はピクセルごとに実行され得る。たとえば、フレームｙの中のピクセルごとに、動き推定値ｆは、フレームｘの中の対応するピクセルの位置を規定する。ピクセルごとの動き推定値ｆは、フレーム間でのピクセルの移動を示すベクトルを含むことができる。場合によっては、フレーム間のオプティカルフローベクトルの計算に基づいてオプティカルフローマップ(動きベクトルマップとも呼ばれる)が生成され得る。オプティカルフローマップは、フレームの中のピクセルごとのオプティカルフローベクトルを含むことができ、ここで、各ベクトルは、フレーム間でのピクセルの移動を示す。例示的な一例では、ピクセルに対するオプティカルフローベクトルは、第１のフレームから第２のフレームへのピクセルの移動を示す(たとえば、ｘ変位およびｙ変位などの水平変位および垂直変位を示す)変位ベクトルであり得る。
【０６３８】
場合によっては、オプティカルフローマップは、フレームの中のすべてに満たないピクセルに対するベクトルを含むことができる。たとえば、高密度オプティカルフローは、フレームの中のピクセルごとのオプティカルフローベクトルを生成するためにフレーム間で算出することができ、高密度オプティカルフローマップの中に含めることができる。いくつかの例では、各オプティカルフローマップは、各ベクトルが第１のフレームから第２のフレームへの点の移動を示す変位ベクトルである、２Ｄベクトルフィールドを含むことができる。
【０６３９】
上述のように、オプティカルフローベクトルまたはオプティカルフローマップは、フレ

ームのシーケンスのフレーム間で算出され得る。2つのフレームは、連続的にキャプチャされたフレームである直接隣接する2つのフレーム、またはフレームのシーケンスの中のいくつかの距離(たとえば、互いの2フレーム内、互いの3フレーム内、または任意の他の好適な距離)だけ離れている2つのフレームを含むことができる。例示的な一例では、フレームxの中のピクセルI(x,y,t)は、フレームyの中で距離すなわち変位(Δx,Δy)だけ移動することができる。

【０６４０】

実行され得る別のタイプの動き推定は、動的畳み込み(適応畳み込みとも呼ばれる)動き推定技法806である。動的畳み込み動き推定はピクセルごとに実行され得る。たとえば、フレームyの中のピクセルごとに、動き推定値fは、フレームxの中のピクセルの近傍の線形結合を規定する。図8に示すように、フレームyの中のピクセル位置におけるピクセルに対して、フレームxの中の同じピクセル位置におけるピクセルを囲むピクセルの近傍の周辺でカーネルまたは行列が規定される。図8における例示的な動的畳み込み動き推定は、フレームyの中の位置(2,7)((行番号,列番号)に対応する)におけるピクセルに対して実行され、ここで、フレームyの中の左上のピクセル位置が位置(0,0)である。フレームxの中の位置(2,7)におけるピクセルの周辺でカーネルが規定される。たとえば、図示のように、サイズ5×5のカーネルが、フレームxの中の位置(2,7)におけるピクセルを囲むピクセルの近傍に適用される。

【０６４１】

カーネルまたは行列の中の各エントリは、数(たとえば、図8の例では5×5個の数のセット)を含むことができる。カーネルの中の数のセットは、フレームyの中でマッピングおよび/または再構成されつつあるピクセルの同じピクセル位置に対応する、フレームxの中のピクセル位置におけるピクセルを囲むピクセルの近傍からのピクセルを、どのように合成すべきかを表すことができる。

【０６４２】

図9および図10は、(GoPの中のフレームの順序に対応する)ピクチャグループ(GoP)構造の様々な例を示す図である。図9は、Iフレーム、Bフレーム、Iフレーム(IBI)コーディング技法の一例を示す。IBI技法は、たとえば、(たとえば、HEVCまたはVVCなどの1つまたは複数のビデオコーディング規格に従って動作する)従来のコーディングシステムではRandomAccessモードにおいて使用され得る。図示のように、第1のGoP(GoP1として示す)に対して、コーディングシステム(たとえば、図6のコーディングシステム600)は、最初および最後のフレームをIフレームとしてコーディング(たとえば、符号化/圧縮および/または復号/解凍)することができ、中間フレームをBフレームとして圧縮することができる。「参照」とラベル付けされた行の中の矢印によって示すように、Bフレーム圧縮のための参照フレームとして2つのIフレームが使用される。「順序」とラベル付けされた行は、Iフレームが最初にコーディングされ中間のBフレームが次にコーディングされることを示す。第2のGoP(GoP2として示す)の中で、Bフレームは、GoP1からの参照フレームを使用してコーディングすることができ、そのため、GoP2の中の最初のフレームはIフレームとしてコーディングされる必要がない。GoP2の中のBフレームは、GoP1の中の最後のフレーム(Iフレーム)を使用して、かつ参照フレームとして最後のフレーム(同じくIフレーム)を使用してコーディングされ得る。図6のBフレームコーディングシステム600または図7BのBフレームコーディングシステム720は、Bフレームをコーディングするために使用され得る。たとえば、図9のGoP1の中の最初のIフレームは、図7Bの中の参照フレーム

【０６４３】

【数２４８】

$$\hat{X}_{ref_0}$$

【０６４４】

であり得、図9のGoP1の中の最後のIフレームは、図7Bの中の参照フレーム
【０６４５】
【数２４９】

$$\hat{X}_{ref_1}$$

【０６４６】
であり得る。図9のGoP1の中のBフレームは、図7Bの中の再構成されたフレーム
【０６４７】
【数２５０】

$$\hat{X}_t$$

【０６４８】
であり得、その場合、GoP1の中の2つのIフレームを使用して補間フレーム
【０６４９】
【数２５１】

$$(\hat{X}_{ref_t})$$

【０６５０】
が生成され得る。
【０６５１】
図10は、Iフレーム、Bフレーム、Pフレーム(IBP)コーディング技法の一例を示す図である。たとえば、本明細書で説明するBフレームコーディングシステム(たとえば、図7BのBフレームコーディングシステム720)はPフレームコーディングシステム(たとえば、図7Aおよび図7BのPフレームコーディングシステム700)を使用するので、単予測用かつ双予測用にPフレームコーディングシステムが使用されるときにパラメータ(たとえば、重み、バイアスなど)が共通である場合、Pフレームは、追加の複雑度を加えることなく同じ構成要素を使用して生成され得る。図10の例では、各GoPの最後のフレームは、(図9のIBI GoP構造において事実であったように)Iフレームである必要がなく、代わりにPフレームとしてコーディングされ得る。図10におけるコーディング順序はまた、図9のコーディング順序と比較して異なる。たとえば、図1の中のGoP1の最初のフレームは、Iフレームとして最初にコーディング(たとえば、符号化/圧縮および/または復号/解凍)される。参照フレームとしてIフレームを使用すると、GoP1の中の最後のフレームは、(たとえば、図7AのPフレームコーディングシステム700によって)Pフレームとしてコーディングされ得る。GoP1の中のIフレームおよびPフレームは、次いで、(たとえば、図6のBフレームコーディングシステム600または図7BのBフレームコーディングシステム720を使用して)中間フレームをBフレームとしてコーディングするために使用され得る。GoP2の場合、GoP1の最後のフレームが、PフレームとしてのGoP2の最後のフレームのコーディングのための参照フレームとして(たとえば、図7AのPフレームコーディングシステム700によって)使用され得る。GoP1の最後のフレームおよびGoP2の最後のフレームは、BフレームとしてのGoP2の中間フレームのコーディングのための参照フレームとして(たとえば、図6のBフレームコーディングシステム600または図7BのBフレームコーディングシステム720によって)使用され得る。
【０６５２】
図11および図12は、GoP内でのフレームコーディング順序を示す図である。たとえば、コーディングシステム(たとえば、Iフレームコーディングシステムおよび/またはPフレームコーディングシステム)がGoPに対して2つの参照フレーム(たとえば、図11および図12のGoPの中の最初および最後のフレーム)を決定すると、(たとえば、図6のBフレ

ームコーディングシステム600または図7BのBフレームコーディングシステム720による)Bフレームコーディングシステムは、GoPの中間フレームをBフレームとしてコーディングし始めることができる。図11および図12の例示的なコーディング順序は、それらのBフレームがコーディング(たとえば、符号化/圧縮および/または復号/解凍)される順序の例を規定する。例示的な一例では、図11および図12のGoPの最初のフレームおよび最後のフレームは、図9に示すものと類似のIフレームを含むことができる。別の例示的な例では、図11および図12のGoPの最初のフレームは、Iフレームを含むことができ、最後のフレームは、図10に示すものと類似のPフレームを含むことができる。

【0653】

図11は、双予測のための逐次方式の一例を示す。図11の逐次方式では、Bフレームは(たとえば、図6のBフレームコーディングシステム600または図7BのBフレームコーディングシステム720によって)左から右に連続的にコーディングされる。図示のように、順序番号2を有するBフレームが、参照フレームとして使用されるGoPの最初および最後のフレームを用いた双予測を使用して最初にコーディングされる。順序番号3を有するBフレームが、前のBフレーム(順序番号2を有する)および参照フレームとして使用されるGoPの最後のフレームを用いた双予測を使用して次にコーディングされる。順序番号4を有するBフレームは、順序番号3を有するBフレームの後にコーディングされ、順序番号5を有するBフレームが後続し、以下同様である。

【0654】

図12は、双方向フレーム予測のための階層方式の一例を示す。図12の階層方式では、2つの初期参照フレーム(GoPの最初および最後のフレーム)が利用可能になると、コーディングシステム(たとえば、図6のBフレームコーディングシステム600または図7BのBフレームコーディングシステム720)は、双予測における2つの初期参照フレームを使用して第5のフレーム(順序番号2を有する中間フレーム)を最初のBフレームとしてコーディングすることができる。コーディングシステムは、次いで、図12の中の「参照」行によって示される参照フレームを使用してGoPの中の(両方とも順序番号3を有する)第3のフレームまたは第7のフレームのいずれかをコーディングすることができる。たとえば、(順序番号3を有する)第3のフレームは、参照フレームとして(Iフレームであり得る)GoPの第1のフレームおよび(Bフレームである)GoPの第5のフレームを使用してコーディングされ得、(順序番号3を有する)第7のフレームは、参照フレームとして(IフレームまたはPフレームであり得る)GoPの最後のフレームおよびGoPの第5のフレーム(Bフレーム)を使用してコーディングされ得る。順序番号4とラベル付けされたフレームが、次いで、図12の中の「参照」行によって示される参照フレームを使用して任意の順序でコーディングされ得る。

【0655】

図13Aおよび図13Bは、補間された参照フレームを使用しないベースラインコーディングシステムと比較した、(たとえば、図7BのBフレームコーディングシステム720を使用して実施される)図6のBフレームコーディングシステム600の性能を示すグラフである。図13Aおよび図13Bのグラフに示す、すべての3つのコーディングシステムは、同じトレーニング条件を使用して(たとえば、同じトレーニングデータセット、同じ損失関数、および同じ回数のトレーニング反復などを使用して)トレーニングされたニューラルネットワークベースのコーディングシステムである。グラフは、ピーク信号対雑音比(PSNR)の関数として(ピクセル当りのビットとして表される)レートをプロットすることによって性能を表す。ライン1302は、図12の階層的なコーディング順序が使用されるときの、図7BのBフレームコーディングシステム720の性能を示す。ライン1304は、図11の逐次的なコーディング順序が使用されるときの、Bフレームコーディングシステム720の性能を示す。ライン1306は、ベースラインコーディングシステムの性能を示す。グラフの左上コーナー(PSNRがより大きくて、レートがより低い)が、より良好な性能を示す。図13Aと図13Bの両方に示すように、ライン1302(Bフレームコーディングシステム720が、階層的なコーディングを使用する)が、最も高い性能を示し、ライン1304(Bフレー

ムコーディングシステム720が、逐次的なコーディングを使用する)が後続し、それにライン1306(ベースラインコーディングシステム)が後続する。

【０６５６】

図14A～図14Fは、(レート対フレームインデックスとラベル付けされたグラフによって示すような)フレーム当りのレート、(PSNR対フレームインデックスとラベル付けされたグラフによって示すような)フレーム当りのPSNR、および(MS-SSIM対フレームインデックスとラベル付けされたグラフによって示すような)フレーム当りのマルチスケール構造的類似度(MS-SSIM)に関して、(たとえば、図7BのBフレームコーディングシステム720を使用して実施される)図6のBフレームコーディングシステム600の性能を示す、グラフである。グラフのx軸はフレーム順序を表すフレームインデックスを含み、y軸は様々な性能メトリック(たとえば、レート、PSNR、およびMS-SSIM)を含む。図14A～図14Fの例では、各GoPの中のフレームの数は12個であり、グラフは、結果を生成するために使用されるビデオシーケンスの中のすべてのGoPの平均を示す。

【０６５７】

各GoPシーケンスの中の(フレームインデックス0を有する)最初のフレームは、(たとえば、図9および図10のGoP構造によって示されるように)イントラ予測を使用してIフレームとして生成される。図14Aの中のドット1402、図14Bの中のドット1404、図14Cの中のドット1406、図14Dの中のドット1408、図14Eの中のドット1410、および図14Eの中のドット1412は、Iフレームのメトリック(レート、PSNR、またはMS-SSIM)を表す。たとえば、図14Aの中のドット1402は、そのIフレームの圧縮のために使用されるレート(ビット数)を示す。各GoPシーケンスの中の(フレームインデックス11を有する)最後のフレームは、(たとえば、図7AのPフレームコーディングシステム700によって)単予測を使用してPフレームとして生成される。図14Aの中のドット1414、図14Bの中のドット1416、図14Cの中のドット1418、図14Dの中のドット1420、図14Eの中のドット1422、および図14Eの中のドット1424は、Pフレームのメトリック(レート、PSNR、またはMS-SSIM)を表す。たとえば、図14Aに示すように、Iフレームのために使用されるビットよりも少ないビットが、Pフレームのために使用される。

【０６５８】

最初のフレームと最後のフレームとの間のフレームは、(たとえば、図7BのBフレームコーディングシステム720によって)Bフレームとしてコーディングされる。Bフレームに対して結果の2つのセットが示され、一方は図11の逐次的なコーディング順序を使用し、他方は図12の階層的なコーディング順序を使用する。一例として図14Aを参照すると、ライン1426は逐次的なコーディング順序を使用する結果を表し、ここで、1というフレームインデックスを有するフレームが最初にBフレームとしてコーディングされ、2というフレームインデックスを有するフレームが後続し、以下同様である。図14Aの中のライン1428は階層的なコーディング順序を使用する結果を表し、ここで、5というフレームインデックスを有するフレームが最初にBフレームとしてコーディングされ、2というフレームインデックスを有するフレームが後続し、8というフレームインデックスを有するフレームが後続し、以下同様である。図12に示すように、後でコーディングされるBフレームにとって、Bフレームまでの参照フレームの距離がより短いことに基づいて、図14Aに示すように、後でコーディングされるBフレームに対してより少数のビットが使用される。たとえば、5というフレームインデックスを有するフレームは、参照フレームとしてIフレーム(GoPの最初のフレーム)およびPフレーム(GoPの最後のフレーム)を使用して双予測されるが、1というフレームインデックスを有するフレームは、参照フレームとしてIフレームおよび2というフレームインデックスを有するBフレームを使用して双予測される。図14B～図14Fの中の滑らかなラインは、図12の階層的なコーディング順序を使用するそれぞれの結果を表し、図14B～図14Fの中のぎざぎざのラインは、図11の逐次的なコーディング順序を使用するそれぞれの結果を表す。

【０６５９】

図15は、本明細書で説明する技法を使用してフレームの双方向予測を実行するための

プロセス1500の一例を示すフローチャートである。ブロック1502において、プロセス1500は、第1の参照フレームおよび第2の参照フレームを取得することを含む。例示的な一例では、第1の参照フレームは図6の参照フレーム0(Ref_0)を含むことができ、第2の参照フレームは図6の参照フレーム1(Ref_1)を含むことができる。別の例示的な例では、第1の参照フレームは図7Bおよび/または図7Cの参照フレーム0

【0660】

【数252】

$$(\hat{X}_{ref0})$$

【0661】

を含むことができ、第2の参照フレームは図7Bおよび/または図7Cの参照フレーム1

【0662】

【数253】

$$(\hat{X}_{ref1})$$

【0663】

を含むことができる。

【0664】

ブロック1504において、プロセス1500は、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成することを含む。たとえば、プロセス1500は、第1の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定することを含むことができる。プロセス1500は、第2の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定することをさらに含むことができる。いくつかの態様では、動き情報の第1のセットおよび動き情報の第2のセットは、第1の参照フレームと第2の参照フレームとの間のピクセル動きに基づいて決定される。場合によっては、動き情報の第1のセットは第1のオプティカルフローマップを含み、動き情報の第2のセットは第2のオプティカルフローマップを含む。例示的な一例として図7Cの説明を参照すると、第1の動き情報は、式(1)を使用して上記で定義された

【0665】

【数254】

$$\hat{f}_{t\to0}^{2D}$$

【0666】

を含むことができ、第2の動き情報は、式(2)を使用して上記で定義された

【0667】

【数255】

$$\hat{f}_{t\to1}^{2D}$$

【0668】

を含むことができる。プロセス1500は、少なくとも部分的には動き情報の第1のセットを使用して第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成することを含むことができる。いくつかの態様では、ワーピング関数は、図7Cに関して上記で説明したものなどの双線形補間関数を含む。プロセス1500は、少なくとも部分的には動き情報の第2のセットを使用して第2の参照フレームに対してワーピング関数を実行することによって、第2のワーピング情報を生成することをさら

に含むことができる。プロセス1500は、第1のワーピング情報および第2のワーピング情報に基づいて第3の参照フレームを生成することを含むことができる。

【０６６９】

ブロック1506において、プロセス1500は、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行することを含む。たとえば、単方向インター予測を実行するために、プロセス1500は、入力フレームと第3の参照フレームとの間の動きを推定することを含むことができる。例示的な一例では、予測された動きは、図7Bの予測された動き

【０６７０】

【数２５６】

$$\hat{f}_t$$

【０６７１】

を含むことができる。場合によっては、入力フレームと第3の参照フレームとの間の推定される動きは、(たとえば、図8に示すオプティカルフロー動き推定技法804を使用して決定される)オプティカルフローデータを含む。いくつかの態様では、入力フレームと第3の参照フレームとの間の推定された動きは、(たとえば、図8に示す動的畳み込み動き推定技法806を使用して決定される)動的畳み込みデータを含む。いくつかの態様では、入力フレームと第3の参照フレームとの間の推定された動きは、(たとえば、図8に示すブロックベースの動き推定技法802を使用して決定される)ブロックベースの動きデータを含む。プロセス1500は、少なくとも部分的には推定された動きに基づいて第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成することをさらに含むことができる。たとえば、例示的な例として図7Bを参照すると、ワーピングエンジン704は、予測される動き

【０６７２】

【数２５７】

$$\hat{f}_t$$

【０６７３】

の中に含まれる動きベクトル(または、他の動きデータ)に基づいて、補間された参照フレーム

【０６７４】

【数２５８】

$$\hat{X}_{ref_t}$$

【０６７５】

のピクセルをワーピングすることができる。

【０６７６】

いくつかの例では、単方向インター予測をさらに実行するために、プロセス1500は、少なくとも部分的には入力フレームとワーピングされたフレームとの間の差分を決定することによって、残差を決定することを含むことができる。プロセス1500は、残差を使用して予測残差を生成することをさらに含むことができる。例示的な例として図7Bを参照すると、Pフレームコーディングシステム700は、フレームt予測

【０６７７】

【数２５９】

$$\tilde{X}_t$$

【０６７８】
と入力フレームX_tとの間の差分として残差r_tを生成することができる。残差予測システム７０６は、上記で説明した技法を使用して予測残差
【０６７９】
【数２６０】

$$\hat{r}_t$$

【０６８０】
を生成することができる。
【０６８１】
いくつかの態様では、動きは第１のニューラルネットワーク(たとえば、第１のオートエンコーダまたは他のタイプのニューラルネットワーク)を使用して推定され、予測残差は第２のニューラルネットワーク(たとえば、第２のオートエンコーダまたは他のタイプのニューラルネットワーク)を使用して生成される。例示的な一例では、第１のニューラルネットワークは図７Ｂの動き予測システム７０２を含むことができ、第２のニューラルネットワークは図７Ｂの残差予測システム７０６を含むことができる。いくつかの例では、プロセス１５００は、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することを含むことができる。再構成されたフレームは双方向予測フレームを含む。例示的な例として再び図７Ｂを参照すると、Ｐフレームコーディングシステム７２０は、(補間された参照フレーム
【０６８２】
【数２６１】

$$\hat{X}_{ref_t}$$

【０６８３】
をワーピングすることによって生成される)フレームｔ予測
【０６８４】
【数２６２】

$$\tilde{X}_t$$

【０６８５】
を予測残差
【０６８６】
【数２６３】

$$\hat{r}_t$$

【０６８７】
に加算することによって、再構成されたフレームｔ
【０６８８】
【数２６４】

$$(\hat{X}_t)$$

【０６８９】
を生成することができる。場合によっては、プロセス１５００は、再構成されたフレームをメモリ(たとえば、復号ピクチャバッファ(ＤＰＢ)または他のメモリ)の中に記憶させること

をさらに含むことができる。いくつかの例では、記憶される再構成されたフレームは、他のフレームのインター予測のために使用され得る。

【０６９０】

上記で説明したように、本明細書で説明するPフレームコーディングシステム(たとえば、図7AのPフレームコーディングシステム700)は、補間された参照フレーム(たとえば、補間された参照フレーム

【０６９１】

【数２６５】

$$\hat{X}_{ref_t}$$

【０６９２】

)を使用して双予測を実行することができ、単予測も実行することができる。たとえば、単予測を実行するために、プロセス1500は、第4の参照フレーム(たとえば、参照フレーム

【０６９３】

【数２６６】

$$\hat{X}_{ref}$$

【０６９４】

)を取得することを含むことができる。プロセス1500は、追加の入力フレームと第4の参照フレームとの間の動きを(たとえば、動き予測システム702を使用して)推定することをさらに含むことができる。例示的な一例では、推定された動きは、図7Aに示す予測された動き

【０６９５】

【数２６７】

$$\hat{f}_t$$

【０６９６】

を含むことができる。プロセス1500は、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを(たとえば、ワーピングエンジン704を使用して)生成することを含むことができる。例示的な例として再び図7Aを参照すると、追加のワーピングされたフレームは、フレーム予測

【０６９７】

【数２６８】

$$\tilde{X}_t$$

【０６９８】

を含むことができる。プロセス1500は、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成することを含むことができ、その場合、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。例示的な例として図7Aを参照すると、Pフレームコーディングシステム720は、(参照フレーム

【０６９９】

【数２６９】

$\hat{X}_{ref}$

【０７００】
をワーピングすることによって生成される)フレームt予測
【０７０１】
【数２７０】

$\tilde{X}_t$

【０７０２】
を予測残差
【０７０３】
【数２７１】

$\hat{r}_t$

【０７０４】
に加算することによって、再構成されたフレームt
【０７０５】
【数２７２】

$(\hat{X}_t)$

【０７０６】
を生成することができる。場合によっては、プロセス1500は、追加の入力フレームを表す再構成されたフレームをメモリ(たとえば、復号ピクチャバッファ(DPB)または他のメモリ)の中に記憶させることを含むことができる。いくつかの例では、追加の入力フレームを表す、記憶される再構成されたフレームは、他のフレームのインター予測のために使用され得る。
【０７０７】
図16は、本明細書で説明する技法を使用してフレームの双方向予測を実行するためのプロセス1600の別の例を示すフローチャートである。場合によっては、プロセス1600の少なくとも一部は、動き予測システム702のデコーダネットワーク703によって、また残差予測システム706のデコーダネットワーク709によって実行され得る。ブロック1602において、プロセス1600は、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得することを含む。例示的な一例では、補間された参照フレームは、図7Bの補間された参照フレーム
【０７０８】
【数２７３】

$\hat{X}_{ref_t}$

【０７０９】
を含むことができる。いくつかの態様では、動き情報はビットストリームから取得される。たとえば、動き情報は、図7Bに示す潜在性z_mを含むことができる。ブロック1604において、プロセス1600は、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定することを含む。たとえば、動き予測システム702のデコーダネットワーク703は、潜在性z_mに基づいて入力フレームX_tと補間された参照フレーム
【０７１０】

【数２７４】

$\hat{X}_{ref_t}$

【０７１１】
との間の動きを推定することができる。推定された動きは、図７Ｂの予測される動き
【０７１２】
【数２７５】

$\hat{f}_t$

【０７１３】
を含むことができる。
【０７１４】
ブロック１６０６において、プロセス１６００は、少なくとも部分的には補間された参照フレームの１つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成することを含む。たとえば、例示的な例として図７Ｂを参照すると、ワーピングエンジン７０４は、予測された動き
【０７１５】
【数２７６】

$\hat{f}_t$

【０７１６】
の中に含まれる動きベクトル(または、他の動きデータ)に基づいて、補間された参照フレーム
【０７１７】
【数２７７】

$\hat{X}_{ref_t}$

【０７１８】
のピクセルをワーピングすることができる。
【０７１９】
ブロック１６０８において、プロセス１６００は、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することを含む。再構成されたフレームは双方向予測フレームを含む。例示的な例として再び図７Ｂを参照すると、Ｐフレームコーディングシステム７２０は、(補間された参照フレーム
【０７２０】
【数２７８】

$\hat{X}_{ref_t}$

【０７２１】
をワーピングすることによって生成される)フレームｔ予測
【０７２２】
【数２７９】

$\tilde{X}_t$

【０７２３】
を予測残差
【０７２４】
【数２８０】

$$\hat{r}_t$$

【０７２５】
に加算することによって、再構成されたフレームt
【０７２６】
【数２８１】

$$(\hat{X}_t)$$

【０７２７】
を生成することができる。場合によっては、プロセス1500は、再構成されたフレームをメモリ(たとえば、復号ピクチャバッファ(DPB)または他のメモリ)の中に記憶させることをさらに含むことができる。いくつかの例では、記憶される再構成されたフレームは、他のフレームのインター予測のために使用され得る。
【０７２８】
いくつかの例では、プロセス1600は、第1の参照フレームおよび第2の参照フレームを取得することを含むことができる。プロセス1600は、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、補間された参照フレームを生成することをさらに含むことができる。本明細書で説明する(たとえば、図7Cおよび/または図15に関して説明する)技法が、補間された参照フレームを生成するために使用され得る。
【０７２９】
いくつかの例では、プロセス1600は、ビットストリームから残差を取得することを含むことができる。たとえば、残差予測システム706のデコーダネットワーク709は、ビットストリームから潜在性z_rを取得することができる。潜在性z_rは残差r_tを表す。プロセス1600は、取得された残差に基づいて予測残差を生成することをさらに含むことができる。たとえば、例示的な例として図7Bを参照すると、デコーダネットワーク709は、潜在性z_rを入力として使用して予測残差
【０７３０】
【数２８２】

$$\hat{r}_t$$

【０７３１】
を生成することができる。
【０７３２】
いくつかの態様では、プロセス1500に関して上記で説明したのと同様に、プロセス1600は、第4の参照フレームを取得することと、追加の入力フレームと第4の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成することとを含むことができる。プロセス1600は、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成することを含むことができ、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。
【０７３３】

いくつかの例では、プロセス1500および/またはプロセス1600は、1つまたは複数のトレーニング反復に対して、損失関数(たとえば、Loss)を使用してニューラルネットワークシステムをトレーニングすることを含むことができる。場合によっては、損失関数は、1つまたは複数の再構成されたBフレームと1つまたは複数の再構成されたBフレームによって表される1つまたは複数のフレームとの間の損失を決定することができる。いくつかの例では、第1の損失関数は、前に説明したようにD+βRに等しい場合がある。

【０７３４】

いくつかの例では、本明細書で説明するプロセス(たとえば、プロセス1500、プロセス1600、および/または本明細書で説明する他のプロセス)は、図17に示すコンピューティングデバイスアーキテクチャ1700を有するコンピューティングデバイスなどの、コンピューティングデバイスまたは装置によって実行されてよい。コンピューティングデバイスは、自律車両コンピュータ、ロボティックデバイス、モバイルデバイス(たとえば、モバイルフォン)、デスクトップコンピューティングデバイス、タブレットコンピューティングデバイス、ウェアラブルデバイス、サーバデバイス、ビデオゲームデバイス、エクステンデッドリアリティデバイス(たとえば、仮想現実(VR)デバイス、拡張現実(AR)デバイス、または複合現実(MR)デバイス)、カメラデバイス、セットトップボックスデバイス、および/またはプロセス800を含む本明細書で説明するプロセスを実行するためのリソース能力を有する任意の他のコンピューティングデバイスなどの、任意の好適なデバイスを含むことができる。いくつかの例では、コンピューティングデバイスは、モバイルデバイス、ウェアラブルデバイス、XRデバイス、パーソナルコンピュータ、ラップトップコンピュータ、ビデオサーバ、テレビジョン、カメラ、セットトップボックス、ビデオゲーム機、または他のデバイスを含むことができる。いくつかの例では、プロセス1500および/またはプロセス1600は、図6のBフレームコーディングシステム600および/または図7BのBフレームコーディングシステム720を実装するコンピューティングデバイスアーキテクチャ1700を有するコンピューティングデバイスによって実行され得る。

【０７３５】

場合によっては、コンピューティングデバイスまたは装置は、1つもしくは複数の入力デバイス、1つもしくは複数の出力デバイス、1つもしくは複数のプロセッサ、1つもしくは複数のマイクロプロセッサ、1つもしくは複数のマイクロコンピュータ、1つもしくは複数の送信機、受信機、もしくは複合送信機受信機(たとえば、トランシーバと呼ばれる)、1つもしくは複数のカメラ、1つもしくは複数のセンサ、および/または本明細書で説明するプロセスのステップを実行するように構成される他の構成要素などの、様々な構成要素を含んでよい。いくつかの例では、コンピューティングデバイスは、ディスプレイ、データを通信および/もしくは受信するように構成されたネットワークインターフェース、それらの任意の組合せ、ならびに/または他の構成要素を含んでよい。ネットワークインターフェースは、インターネットプロトコル(IP)ベースのデータまたは他のタイプのデータを通信および/または受信するように構成され得る。

【０７３６】

コンピューティングデバイスの構成要素は、回路構成の中に実装され得る。たとえば、構成要素は、1つまたは複数のプログラマブル電子回路(たとえば、マイクロプロセッサ、グラフィックス処理ユニット(GPU)、デジタル信号プロセッサ(DSP)、中央処理ユニット(CPU)、ニューラル処理ユニット(NPU)、特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)、および/または他の好適な電子回路)を含み得る、電子回路または他の電子ハードウェアを含むことができ、かつ/またはそれらを使用して実装することができ、かつ/あるいは本明細書で説明する様々な動作を実行するために、コンピュータソフトウェア、ファームウェア、またはこれらの任意の組合せを含むことができ、かつ/またはそれらを使用して実装することができる。

【０７３７】

プロセス1500およびプロセス1600は、論理フロー図として図示され、その動作は、ハードウェア、コンピュータ命令、またはそれらの組合せで実装され得る動作のシーケン

スを表す。コンピュータ命令のコンテキストでは、動作は、1つまたは複数のプロセッサによって実行されたとき、記載された動作を実行する、1つまたは複数のコンピュータ可読記憶媒体上に記憶されたコンピュータ実行可能命令を表す。一般に、コンピュータ実行可能命令は、特定の機能を実行するかまたは特定のデータタイプを実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。動作が説明される順序は、限定として解釈されることを意図せず、任意の数の説明する動作は、プロセスを実施するために任意の順序で、かつ/または並列に組み合わせられ得る。

【０７３８】

追加として、(プロセス1500およびプロセス1600を含む)本明細書で説明するプロセスは、実行可能命令を用いて構成された1つまたは複数のコンピュータシステムの制御下で実行されてよく、ハードウェアまたはそれらの組合せによって1つまたは複数のプロセッサ上で集合的に実行するコード(たとえば、実行可能命令、1つもしくは複数のコンピュータプログラム、または1つもしくは複数のアプリケーション)として実装されてよい。上述のように、コードは、たとえば、1つまたは複数のプロセッサによって実行可能な複数の命令を備えるコンピュータプログラムの形態で、コンピュータ可読記憶媒体または機械可読記憶媒体上に記憶されてよい。コンピュータ可読記憶媒体または機械可読記憶媒体は、非一時的であってよい。

【０７３９】

図17は、本明細書で説明する様々な技法を実施できる例示的なコンピューティングデバイスの例示的なコンピューティングデバイスアーキテクチャ1700を示す。いくつかの例では、コンピューティングデバイスは、モバイルデバイス、ウェアラブルデバイス、XRデバイス、パーソナルコンピュータ、ラップトップコンピュータ、ビデオサーバ、ビデオゲーム機、ロボティックデバイス、セットトップボックス、テレビジョン、カメラ、サーバ、または他のデバイスを含むことができる。たとえば、コンピューティングデバイスアーキテクチャ1700は、図6のBフレームコーディングシステム600、図7AのPフレームコーディングシステム700、および/または図7BのBフレームコーディングシステム720を実装することができる。コンピューティングデバイスアーキテクチャ1700の構成要素は、バスなどの接続部1705を使用して互いに電気通信しているものとして示される。例示的なコンピューティングデバイスアーキテクチャ1700は、処理ユニット(CPUまたはプロセッサ)1710、ならびに読取り専用メモリ(ROM)1720およびランダムアクセスメモリ(RAM)1725などのコンピューティングデバイスメモリ1715を含む様々なコンピューティングデバイス構成要素をプロセッサ1710に結合する、コンピューティングデバイス接続部1705を含む。

【０７４０】

コンピューティングデバイスアーキテクチャ1700は、プロセッサ1710に直接接続されるか、その極近傍にあるか、またはその一部として統合される、高速メモリのキャッシュを含むことができる。コンピューティングデバイスアーキテクチャ1700は、プロセッサ1710による高速アクセスのために、メモリ1715および/または記憶デバイス1730からキャッシュ1712にデータをコピーすることができる。このようにして、キャッシュは、データを待つ間のプロセッサ1710の遅延を回避する、性能向上をもたらすことができる。これらおよび他のモジュールは、様々なアクションを実行するようにプロセッサ1710を制御することができ、またはそれを制御するように構成され得る。他のコンピューティングデバイスメモリ1715も、使用のために利用可能であってよい。メモリ1715は、様々な性能特性を有する様々なタイプの複数のメモリを含むことができる。プロセッサ1710は、任意の汎用プロセッサおよびハードウェア、またはプロセッサ1710を制御するように構成された、記憶デバイス1730の中に記憶されるサービス1 1732、サービス2 1734、およびサービス3 1736などのソフトウェアサービス、ならびにソフトウェア命令がプロセッサ設計の中に組み込まれる専用プロセッサを含むことができる。プロセッサ1710は、複数のコアまたはプロセッサ、バス、メモリコントローラ、キャッシュなどを含む、自立型のシステムであってよい。マルチコアプロセッサは、対称であっても非対

称であってもよい。

【０７４１】

コンピューティングデバイスアーキテクチャ1700とのユーザ対話を可能にするために、入力デバイス1745は、音声用のマイクロフォン、ジェスチャーまたはグラフィカル入力用のタッチセンシティブスクリーン、キーボード、マウス、モーション入力、音声などの、任意の数の入力メカニズムを表すことができる。出力デバイス1735も、ディスプレイ、プロジェクタ、テレビジョン、スピーカーデバイスなどの、当業者に知られているいくつかの出力メカニズムのうちの1つまたは複数であり得る。いくつかの事例では、マルチモードコンピューティングデバイスは、コンピューティングデバイスアーキテクチャ1700と通信するためにユーザが複数のタイプの入力を与えることを可能にすることができる。通信インターフェース1740は、一般に、ユーザ入力およびコンピューティングデバイス出力を統率および管理することができる。いかなる特定のハードウェア構成上で動作することに対しても制約はなく、したがって、改善されたハードウェアまたはファームウェア構成が開発されるにつれて、ここでの基本的な特徴がそれらに対して容易に置き換えられ得る。

【０７４２】

記憶デバイス1730は不揮発性メモリであり、ハードディスク、または磁気カセット、フラッシュメモリカード、ソリッドステートメモリデバイス、デジタル多用途ディスク、カートリッジ、ランダムアクセスメモリ(RAM)1725、読取り専用メモリ(ROM)1720、およびそれらのハイブリッドなどの、コンピュータによってアクセス可能なデータを記憶できる他のタイプのコンピュータ可読媒体であり得る。記憶デバイス1730は、プロセッサ1710を制御するためのサービス1732、1734、1736を含むことができる。他のハードウェアまたはソフトウェアモジュールが企図される。記憶デバイス1730は、コンピューティングデバイス接続部1705に接続され得る。一態様では、特定の機能を実行するハードウェアモジュールは、機能を実行するために、プロセッサ1710、接続部1705、出力デバイス1735などの必要なハードウェア構成要素と接続されたコンピュータ可読媒体の中に記憶されるソフトウェア構成要素を含むことができる。

【０７４３】

「デバイス」という用語は、(1つのスマートフォン、1つのコントローラ、1つの処理システムなどの)1つまたは特定の個数の物理的な物体に限定されない。本明細書で使用するとき、デバイスは、本開示の少なくともいくつかの部分を実装し得る1つまたは複数の部分を有する任意の電子デバイスを含むことができる。説明および例は、本開示の様々な態様を説明するために「デバイス」という用語を使用するが、「デバイス」という用語は、物体の特定の構成、タイプ、または個数に限定されない。追加として、「システム」という用語は複数の構成要素または特定の実施形態に限定されない。たとえば、システムは、1つまたは複数のプリント回路板または他の基板上に実装されてよく、移動可能または静的な構成要素を有してよい。説明および例は、本開示の様々な態様を説明するために「システム」という用語を使用するが、「システム」という用語は、物体の特定の構成、タイプ、または個数に限定されない。

【０７４４】

本明細書で提供される実施形態および例の完全な理解を与えるために、本説明の中で具体的な詳細が提供される。しかしながら、本実施形態がこれらの具体的な詳細なしに実践され得ることが当業者によって理解されよう。説明を明快にするために、いくつかの事例では、本技術は、デバイス、デバイス構成要素、ソフトウェアの中で具現される方法におけるステップもしくはルーチン、またはハードウェアとソフトウェアとの組合せを備える機能ブロックを含む、個々の機能ブロックを含むものとして提示されることがある。図の中に示されかつ/または本明細書で説明されるもの以外の、追加の構成要素が使用されてよい。たとえば、不必要な詳細で本実施形態を不明瞭にしないように、回路、システム、ネットワーク、プロセス、および他の構成要素がブロック図の形態で構成要素として示されることがある。他の事例では、本実施形態を不明瞭にすることを避けるために、よく知

られている回路、プロセス、アルゴリズム、構造、および技法は、不必要な詳細なしに示されることがある。

【０７４５】

個々の実施形態が、フローチャート、フロー図、データフロー図、構造図、またはブロック図として示されるプロセスまたは方法として上記で説明されることがある。フローチャートは動作を逐次プロセスとして説明することがあるが、動作の多くは並行してまたは同時に実行され得る。加えて、動作の順序は並べ替えられてよい。プロセスは、その動作が完了するときに終了するが、図の中に含まれない追加のステップを有する場合がある。プロセスは、メソッド、関数、プロシージャ、サブルーチン、サブプログラムなどに相当し得る。プロセスが関数に相当するとき、その終了は、その関数が呼出し関数またはメイン関数に戻ることに相当し得る。

【０７４６】

上記で説明した例によるプロセスおよび方法は、記憶されるかまたは別のやり方でコンピュータ可読媒体から利用可能なコンピュータ実行可能命令を使用して実施され得る。そのような命令は、たとえば、汎用コンピュータ、専用コンピュータ、もしくは処理デバイスにいくつかの機能もしくは機能のグループを実行させるか、またはさもなければ、いくつかの機能もしくは機能のグループを実行するように汎用コンピュータ、専用コンピュータ、もしくは処理デバイスを構成する、命令およびデータを含むことができる。使用されるコンピュータリソースの部分は、ネットワークを介してアクセス可能であり得る。コンピュータ実行可能命令は、たとえば、バイナリであってよく、アセンブリ言語、ファームウェア、ソースコードなどの中間フォーマット命令であってもよい。

【０７４７】

「コンピュータ可読媒体」という用語は、限定はしないが、ポータブルまたは非ポータブルの記憶デバイス、光記憶デバイス、ならびに命令および/またはデータを記憶、格納、または搬送できる様々な他の媒体を含む。コンピュータ可読媒体は、データがそこに記憶され得るとともに、ワイヤレスにまたは有線接続を介して伝搬する搬送波および/または一時的な電子信号を含まない、非一時的媒体を含んでよい。非一時的媒体の例は、限定はしないが、特に、磁気ディスクまたは磁気テープ、フラッシュメモリなどの光記憶媒体、メモリまたはメモリデバイス、磁気ディスクまたは光ディスク、フラッシュメモリ、不揮発性メモリを用いて設けられるUSBデバイス、ネットワーク化された記憶デバイス、コンパクトディスク(CD)またはデジタル多用途ディスク(DVD)、それらの任意の好適な組合せを含んでよい。コンピュータ可読媒体は、プロシージャ、関数、サブプログラム、プログラム、ルーチン、サブルーチン、モジュール、ソフトウェアパッケージ、クラス、または命令、データ構造、もしくはプログラムステートメントの任意の組合せを表し得る、その上に記憶されたコードおよび/または機械実行可能命令を有してよい。コードセグメントは、情報、データ、引数、パラメータ、またはメモリ内容を渡すことおよび/または受けることによって、別のコードセグメントまたはハードウェア回路に結合されてよい。情報、引数、パラメータ、データなどは、メモリ共有、メッセージパッシング、トークンパッシング、ネットワーク送信などを含む、任意の好適な手段を介して渡され、転送され、または送信されてよい。

【０７４８】

いくつかの実施形態では、コンピュータ可読記憶デバイス、媒体、およびメモリは、ビットストリームなどを含むケーブル信号またはワイヤレス信号を含むことができる。しかしながら、述べられるとき、非一時的コンピュータ可読記憶媒体は、エネルギー、キャリア信号、電磁波、および信号自体などの媒体を明確に除外する。

【０７４９】

これらの開示に従ってプロセスおよび方法を実施するデバイスは、ハードウェア、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語、またはそれらの任意の組合せを含むことができ、様々なフォームファクタのうちのいずれかをとることができる。ソフトウェア、ファームウェア、ミドルウェア、またはマイクロコ

ードで実装されるとき、必要なタスクを実行するためのプログラムコードまたはコードセグメント(たとえば、コンピュータプログラム製品)は、コンピュータ可読媒体または機械可読媒体の中に記憶されてよい。プロセッサは必要なタスクを実行してよい。フォームファクタの典型的な例は、ラップトップ、スマートフォン、モバイルフォン、タブレットデバイスまたは他の小型フォームファクタパーソナルコンピュータ、携帯情報端末、ラックマウントデバイス、スタンドアロンデバイスなどを含む。本明細書で説明した機能性はまた、周辺装置またはアドインカードの中で具現され得る。そのような機能性はまた、さらなる例として、異なるチップのうちの回路基板上、または単一のデバイスの中で実行する異なるプロセス上で実施され得る。

【０７５０】

命令、そのような命令を伝えるための媒体、命令を実行するためのコンピューティングリソース、およびそのようなコンピューティングリソースをサポートするための他の構造は、本開示で説明した機能を提供するための例示的な手段である。

【０７５１】

上記の説明では、本出願の態様はそれらの特定の実施形態を参照しながら説明されるが、本出願がそれらに限定されないことを当業者は認識されよう。したがって、本出願の例示的な実施形態が本明細書で詳細に説明されているが、本発明の概念が別のやり方で様々に具現および採用され得ること、ならびに従来技術によって限定される場合を除き、添付の特許請求の範囲がそのような変形を含むものと解釈されることが意図されることを理解されたい。上記で説明した本出願の様々な特徴および態様は、個別にまたは共同で使用され得る。さらに、実施形態は、本明細書の範囲から逸脱することなく、本明細書で説明したもの以外の任意の数の環境および適用例において利用され得る。したがって、本明細書および図面は限定的ではなく例示的と見なされるべきである。例示の目的のために、方法は特定の順序で説明された。代替実施形態では、方法は説明された順序とは異なる順序で実行され得ることを諒解されたい。

【０７５２】

本明細書で使用する、よりも小さい(「〈」)およびよりも大きい(「〉」)というシンボルまたは用語が、本説明の範囲から逸脱することなく、それぞれ、以下(「≦」)および以上(「≧」)というシンボルで置き換えられ得ることを、当業者は諒解されよう。

【０７５３】

構成要素がいくつかの動作を実行する「ように構成される」ものとして説明される場合、そのような構成は、たとえば、動作を実行するように電子回路もしくは他のハードウェアを設計することによって、動作を実行するようにプログラマブル電子回路(たとえば、マイクロプロセッサ、または他の好適な電子回路)をプログラムすることによって、またはそれらの任意の組合せで達成され得る。

【０７５４】

「に結合される」という句は、直接もしくは間接的のいずれかで別の構成要素に物理的に接続される任意の構成要素、および/または直接もしくは間接的のいずれかで別の構成要素と通信している(たとえば、有線接続もしくはワイヤレス接続および/または他の好適な通信インターフェースを介して他の構成要素に接続される)任意の構成要素を指す。

【０７５５】

集合「のうちの少なくとも1つ」および/または集合のうちの「1つまたは複数」と記載する請求項の文言または他の文言は、集合の1つのメンバーまたは集合の(任意の組合せでの)複数のメンバーが請求項を満たすことを示す。たとえば、「AおよびBのうちの少なくとも1つ」または「AまたはBのうちの少なくとも1つ」と記載する請求項の文言は、A、B、またはAおよびBを意味する。別の例において、「A、B、およびCのうちの少なくとも1つ」または「A、B、またはCのうちの少なくとも1つ」と記載する請求項の文言は、A、B、C、またはAおよびB、またはAおよびC、またはBおよびC、またはAおよびBおよびCを意味する。集合「のうちの少なくとも1つ」および/または集合のうちの「1つまたは複数」という文言は、集合の中で列挙される項目にその集合を限定しない。たとえば

、「AおよびBのうちの少なくとも1つ」または「AまたはBのうちの少なくとも1つ」と記載する請求項の文言は、A、B、またはAおよびBを意味することができ、追加として、AおよびBの集合の中で列挙されない項目を含むことができる。

【０７５６】

本明細書で開示する実施形態に関して説明した様々な例示的な論理ブロック、モジュール、回路、およびアルゴリズムステップは、電子ハードウェア、コンピュータソフトウェア、ファームウェア、またはそれらの組合せとして実装され得る。ハードウェアとソフトウェアとのこの互換性を明確に示すために、様々な例示的な構成要素、ブロック、モジュール、回路、およびステップが、概してそれらの機能性に関して上記で説明されている。そのような機能性がハードウェアとして実装されるのかそれともソフトウェアとして実装されるのかは、特定の適用例および全体的なシステムに課される設計制約に依存する。当業者は、説明した機能性を特定の適用例ごとに様々な方法で実施し得るが、そのような実装決定は、本出願の範囲からの逸脱を引き起こすものと解釈されるべきではない。

【０７５７】

本明細書で説明した技法はまた、電子ハードウェア、コンピュータソフトウェア、ファームウェア、またはそれらの任意の組合せで実装され得る。そのような技法は、汎用コンピュータ、ワイヤレス通信デバイスハンドセット、またはワイヤレス通信デバイスハンドセットおよび他のデバイスにおける適用例を含む複数の用途を有する集積回路デバイスなどの、様々なデバイスのうちのいずれかにおいて実装され得る。モジュールまたは構成要素として説明した任意の特徴は、集積論理デバイスの中で一緒に、または個別であるが相互動作可能な論理デバイスとして別々に実装され得る。ソフトウェアで実装される場合、技法は、実行されたとき、上記で説明した方法のうちの1つまたは複数を実行する命令を含むプログラムコードを備えるコンピュータ可読データ記憶媒体によって、少なくとも部分的に実現され得る。コンピュータ可読データ記憶媒体は、パッケージング材料を含み得るコンピュータプログラム製品の一部を形成し得る。コンピュータ可読媒体は、シンクロナスダイナミックランダムアクセスメモリ(SDRAM)などのランダムアクセスメモリ(RAM)、読取り専用メモリ(ROM)、不揮発性ランダムアクセスメモリ(NVRAM)、電気的消去可能プログラマブル読取り専用メモリ(EEPROM)、FLASHメモリ、磁気または光データ記憶媒体などの、メモリまたはデータ記憶媒体を備えてよい。技法は、追加または代替として、伝搬される信号または波などの、命令またはデータ構造の形態でのプログラムコードを搬送または通信するとともに、コンピュータによってアクセスされ、読み取られ、かつ/または実行され得る、コンピュータ可読通信媒体によって少なくとも部分的に実現されてよい。

【０７５８】

プログラムコードは、1つまたは複数のデジタル信号プロセッサ(DSP)、汎用マイクロプロセッサ、特定用途向け集積回路(ASIC)、フィールドプログラマブル論理アレイ(FPGA)、または他の均等な集積論理回路構成もしくは個別論理回路構成などの1つまたは複数のプロセッサを含み得る、プロセッサによって実行されてよい。そのようなプロセッサは、本開示で説明した技法のうちのいずれかを実行するように構成され得る。汎用プロセッサはマイクロプロセッサであってよいが、代替として、プロセッサは、任意の従来のプロセッサ、コントローラ、マイクロコントローラ、またはステートマシンであってよい。プロセッサはまた、コンピューティングデバイスの組合せ、たとえば、DSPとマイクロプロセッサとの組合せ、複数のマイクロプロセッサ、DSPコアと連携した1つもしくは複数のマイクロプロセッサ、または任意の他のそのような構成として実装されてよい。したがって、本明細書で使用する「プロセッサ」という用語は、上記の構造、上記の構造の任意の組合せ、または本明細書で説明した技法の実装に適した任意の他の構造もしくは装置のうちのいずれかを指すことがある。

【０７５９】

本開示の例示的な例は、以下を含む。

【０７６０】

態様1:ビデオデータを処理するためのシステムであって、システムは、補間エンジンと単方向コーディングエンジンとを含む双方向コーディングエンジンを備え、補間エンジンは、第1の参照フレームおよび第2の参照フレームを取得し、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成するように構成され、単方向コーディングエンジンは、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行するように構成される。

【０７６１】

態様2:態様1のシステムであって、単方向コーディングエンジンは動きエンジンおよびワーピングエンジンを含み、動きエンジンは、入力フレームと第3の参照フレームとの間の動きを推定するように構成され、ワーピングエンジンは、少なくとも部分的には推定された動きに基づいて第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成するように構成される。

【０７６２】

態様3:態様2のシステムであって、単方向コーディングエンジンは残差エンジンをさらに含み、残差エンジンは、少なくとも部分的には入力フレームとワーピングされたフレームとの間の差分を決定することによって、残差を決定し、残差を使用して予測残差を生成するように構成される。

【０７６３】

態様4:態様3のシステムであって、メモリと、メモリに結合された少なくとも1つのプロセッサとをさらに備え、単方向コーディングエンジンは、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成するように構成され、再構成されたフレームは双方向予測フレームを含み、少なくとも1つのプロセッサは、再構成されたフレームをメモリの中に記憶させるように構成される。

【０７６４】

態様5:態様4のシステムであって、少なくとも1つのプロセッサはニューラル処理ユニット(ＮＰＵ)を含む。

【０７６５】

態様6:態様3～5のうちのいずれかのシステムであって、動きエンジンは第1のニューラルネットワークを含み、残差エンジンは第2のニューラルネットワークを含む。

【０７６６】

態様7:態様6のシステムであって、第1のニューラルネットワークは第1のオートエンコーダを含み、第2のニューラルネットワークは第2のオートエンコーダを含む。

【０７６７】

態様8:態様1～7のうちのいずれかのシステムであって、単方向コーディングエンジンは、第4の参照フレームを取得し、追加の入力フレームと第4の参照フレームとの間の動きを推定し、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成し、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成するように構成され、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。

【０７６８】

態様9:態様1～8のうちのいずれかのシステムであって、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成するために、補間エンジンは、第1の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定し、第2の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定し、少なくとも部分的には動き情報の第1のセットを使用して第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成し、少なくとも部分的には動き情報の第2のセットを使用して第2の参照フレームに対してワーピング関数を実行することによって、第2のワーピング情報を生成し、第1のワーピング情報および第2のワーピン

グ情報に基づいて第3の参照フレームを生成するように構成される。

【０７６９】

態様10:態様9のシステムであって、動き情報の第1のセットおよび動き情報の第2のセットは、第1の参照フレームと第2の参照フレームとの間のピクセル動きに基づいて決定される。

【０７７０】

態様11:態様9または10のうちのいずれかのシステムであって、動き情報の第1のセットは第1のオプティカルフローマップを含み、動き情報の第2のセットは第2のオプティカルフローマップを含む。

【０７７１】

態様12:態様9～11のうちのいずれかのシステムであって、ワーピング関数は双線形補間関数を含む。

【０７７２】

態様13:態様2～12のうちのいずれかのシステムであって、入力フレームと第3の参照フレームとの間の推定された動きは、オプティカルフローデータを含む。

【０７７３】

態様14:態様2～12のうちのいずれかのシステムであって、入力フレームと第3の参照フレームとの間の推定された動きは、動的畳み込みデータを含む。

【０７７４】

態様15:態様2～12のうちのいずれかのシステムであって、入力フレームと第3の参照フレームとの間の推定された動きは、ブロックベースの動きデータを含む。

【０７７５】

態様16:態様1～15のうちのいずれかのシステムであって、システムは、モバイルデバイス、エクステンデッドリアリティデバイス、またはテレビジョンのうちの1つである。

【０７７６】

態様17:態様1～16のうちのいずれかのシステムであって、ディスプレイ、および1つまたは複数のビデオフレームをキャプチャするように構成されたカメラのうちの少なくとも1つをさらに備える。

【０７７７】

態様18:ビデオデータを処理するための方法であって、方法は、第1の参照フレームおよび第2の参照フレームを取得することと、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成することと、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行することとを備える。

【０７７８】

態様19:態様18の方法であって、入力フレームと第3の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第3の参照フレームの1つまたは複数のピクセルをワーピングすることによって、ワーピングされたフレームを生成することとをさらに備える。

【０７７９】

態様20:態様19の方法であって、少なくとも部分的には入力フレームとワーピングされたフレームとの間の差分を決定することによって、残差を決定することと、残差を使用して予測残差を生成することとをさらに備える。

【０７８０】

態様21:態様20の方法であって、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することであって、再構成されたフレームが双方向予測フレームを含むことと、再構成されたフレームをメモリの中に記憶させることとをさらに備える。

【０７８１】

態様22:態様20または21のうちのいずれかの方法であって、動きは第1のオートエン

コーダを使用して推定され、予測残差は第2のオートエンコーダを使用して生成される。

【０７８２】

態様23:態様18～22のうちのいずれかの方法であって、第4の参照フレームを取得することと、追加の入力フレームと第4の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成することと、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成することとをさらに備え、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。

【０７８３】

態様24:態様18～23のうちのいずれかの方法であって、第1の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第1のセットを決定することと、第2の参照フレームから第3の参照フレームへのピクセル動きを表す動き情報の第2のセットを決定することと、少なくとも部分的には動き情報の第1のセットを使用して第1の参照フレームに対してワーピング関数を実行することによって、第1のワーピング情報を生成することと、少なくとも部分的には動き情報の第2のセットを使用して第2の参照フレームに対してワーピング関数を実行することによって、第2のワーピング情報を生成することと、第1のワーピング情報および第2のワーピング情報に基づいて第3の参照フレームを生成することとをさらに備える。

【０７８４】

態様25:態様24の方法であって、動き情報の第1のセットおよび動き情報の第2のセットは、第1の参照フレームと第2の参照フレームとの間のピクセル動きに基づいて決定される。

【０７８５】

態様26:態様24または25のうちのいずれかの方法であって、動き情報の第1のセットは第1のオプティカルフローマップを含み、動き情報の第2のセットは第2のオプティカルフローマップを含む。

【０７８６】

態様27:態様24～26のうちのいずれかの方法であって、ワーピング関数は双線形補間関数を含む。

【０７８７】

態様28:態様19～27のうちのいずれかの方法であって、入力フレームと第3の参照フレームとの間の推定された動きは、オプティカルフローデータを含む。

【０７８８】

態様29:態様19～27のうちのいずれかの方法であって、入力フレームと第3の参照フレームとの間の推定された動きは、動的畳み込みデータを含む。

【０７８９】

態様30:態様19～27のうちのいずれかの方法であって、入力フレームと第3の参照フレームとの間の推定された動きは、ブロックベースの動きデータを含む。

【０７９０】

態様31:その上に記憶された命令を有する非一時的コンピュータ可読媒体であって、命令は、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに、第1の参照フレームおよび第2の参照フレームを取得させ、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成させ、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行させる。

【０７９１】

態様32:態様31の非一時的コンピュータ可読媒体であって、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに態様19～30のうちのいずれかによる動作を実行させる命令をさらに備える。

【０７９２】
態様33:ビデオデータを処理するための装置であって、第1の参照フレームおよび第2の参照フレームを取得するための手段と、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、第3の参照フレームを生成するための手段と、第3の参照フレームに基づいて入力フレームに対して単方向インター予測を実行するための手段とを備える。
【０７９３】
態様34:態様33の装置であって、態様19～30のうちのいずれかによる動作を実行するための手段をさらに備える。
【０７９４】
態様35:ビデオデータを処理するためのシステムであって、システムは、メモリと、メモリに結合された少なくとも1つのプロセッサとを備え、少なくとも1つのプロセッサは、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得し、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定し、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成し、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成するように構成され、再構成されたフレームは双方向予測フレームを含む。
【０７９５】
態様36:態様35のシステムであって、少なくとも1つのプロセッサは、第1の参照フレームおよび第2の参照フレームを取得し、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、補間された参照フレームを生成するように構成される。
【０７９６】
態様37:態様35のシステムであって、少なくとも1つのプロセッサは、ビットストリームから残差を取得し、取得された残差に基づいて予測残差を生成するように構成される。
【０７９７】
態様38:態様37のシステムであって、動き情報はビットストリームから取得される。
【０７９８】
態様39:態様35のシステムであって、少なくとも1つのプロセッサは、第4の参照フレームを取得し、追加の入力フレームと第4の参照フレームとの間の動きを推定し、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成し、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成するように構成され、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。
【０７９９】
態様40:ビデオデータを処理するための方法であって、方法は、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得することと、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定することと、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成することと、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成することとを備え、再構成されたフレームは双方向予測フレームを含む。
【０８００】
態様41:態様40の方法であって、第1の参照フレームおよび第2の参照フレームを取得することと、少なくとも部分的には第1の参照フレームと第2の参照フレームとの間の補間を実行することによって、補間された参照フレームを生成することとをさらに備える。

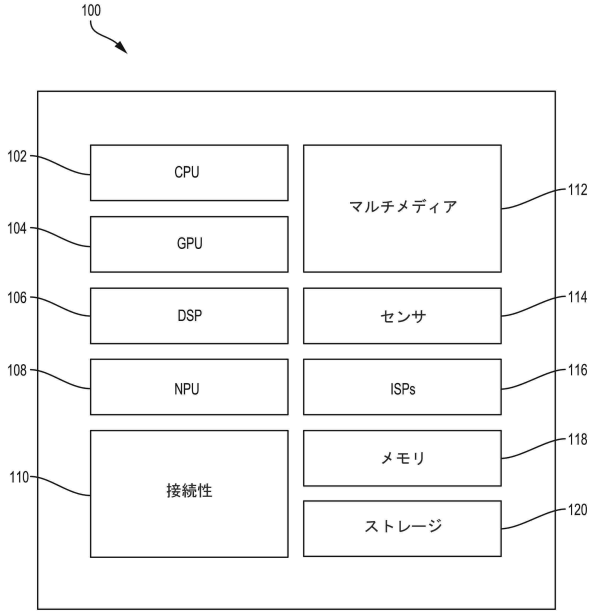
【０８０１】
態様42:態様40の方法であって、ビットストリームから残差を取得することと、取得された残差に基づいて予測残差を生成することとをさらに備える。
【０８０２】
態様43:態様42の方法であって、動き情報はビットストリームから取得される。
【０８０３】
態様44:態様40の方法であって、第4の参照フレームを取得することと、追加の入力フレームと第4の参照フレームとの間の動きを推定することと、少なくとも部分的には推定された動きに基づいて第4の参照フレームの1つまたは複数のピクセルをワーピングすることによって、追加のワーピングされたフレームを生成することと、追加のワーピングされたフレームおよび追加の予測残差に基づいて、追加の入力フレームを表す再構成されたフレームを生成することとをさらに備え、追加の入力フレームを表す再構成されたフレームは単方向予測フレームを含む。
【０８０４】
態様45:その上に記憶された命令を有する非一時的コンピュータ可読媒体であって、命令は、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得させ、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定させ、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成させ、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成させ、再構成されたフレームは双方向予測フレームを含む。
【０８０５】
態様46:態様45の非一時的コンピュータ可読媒体であって、1つまたは複数のプロセッサによって実行されたとき、1つまたは複数のプロセッサに態様41～44のうちのいずれかによる動作を実行させる命令をさらに備える。
【０８０６】
態様47:ビデオデータを処理するための装置であって、補間された参照フレーム、および入力フレームと補間された参照フレームとの間の動きを表す動き情報を取得するための手段と、動き情報に基づいて入力フレームと補間された参照フレームとの間の動きを推定するための手段と、少なくとも部分的には補間された参照フレームの1つまたは複数のピクセルを推定された動きに基づいてワーピングすることによって、ワーピングされたフレームを生成するための手段と、ワーピングされたフレームおよび予測残差に基づいて、入力フレームを表す再構成されたフレームを生成するための手段とを備え、再構成されたフレームは双方向予測フレームを含む。
【０８０７】
態様48:態様47の装置であって、態様41～44のうちのいずれかによる動作を実行するための手段をさらに備える。
【符号の説明】
【０８０８】
100 画像処理システム
102 中央処理ユニット(CPU)
104 グラフィックス処理ユニット(GPU)
106 デジタル信号プロセッサ(DSP)
108 ニューラル処理ユニット(NPU)
110 接続性ブロック
112 マルチメディアプロセッサ
114 センサ
116 画像信号プロセッサ(ISP)

118 メモリブロック
120 ストレージ
200 深層畳み込みネットワーク(DCN)
202 全結合ニューラルネットワーク
204 局所結合ニューラルネットワーク
206 畳み込みニューラルネットワーク
218 特徴マップの第1のセット
220 特徴マップの第2のセット
222 出力
224 第1の特徴ベクトル
226 画像
228 第2の特徴ベクトル
230 画像キャプチャデバイス
232 畳み込み層
350 深層畳み込みネットワーク
352 入力データ
354 畳み込みブロック
356 畳み込み層(CONV)
358 正規化層(LNorm)
360 最大プーリング層(MAX POOL)
362 全結合層
364 ロジスティック回帰(LR)層
366 分類スコア
400 Pフレームコーディングシステム
401 入力フレーム
402 動き圧縮エンジン
403 参照フレーム
404 動き補償エンジン
406 残差圧縮エンジン
500 Bフレームコーディングシステム
501 入力フレーム
502 動き圧縮エンジン
504 動き補償エンジン
506 残差圧縮エンジン
510 Bフレームコーディングシステム
511 入力フレーム
512 フレーム補間エンジン
516 残差圧縮エンジン
600 Bフレームコーディングシステム
611 入力フレーム
613 補間された参照フレーム
615 再構成された入力フレーム
622 フレーム補間エンジン
624 Pフレームコーディングシステム
700 Pフレームコーディングシステム
702 動き予測システム
703 デコーダネットワーク
704 ワーピングエンジン
705 エンコーダネットワーク
706 残差予測システム

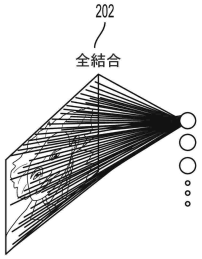
707 エンコーダネットワーク
708 減算演算
709 デコーダネットワーク
710 加算演算
720 Bフレームコーディングシステム
722 フレーム補間エンジン
725 フレーム補間エンジン
730 オプティカルフロー推定層
732 オプティカルフロー予測層
734 2Dワーピング層
736 改善層
738 3Dワーピング層
740 双方向動き情報
742 時間情報
744 2Dオプティカルフローマップ
746 ワーピングされたフレーム
748 3Dオプティカルフローマップ
802 ブロックベースの動き推定技法
804 オプティカルフロー動き推定技法
806 動的畳み込み動き推定技法
1700 コンピューティングデバイスアーキテクチャ
1705 接続部
1710 処理ユニット
1712 キャッシュ
1715 コンピューティングデバイスメモリ
1720 読取り専用メモリ(ROM)
1725 ランダムアクセスメモリ(RAM)
1730 記憶デバイス
1732 サービス1
1734 サービス2
1735 出力デバイス
1736 サービス3
1740 通信インターフェース
1745 入力デバイス

【図面】

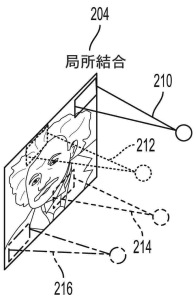
【図１】



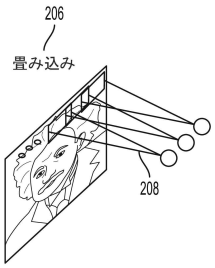
【図２Ａ】



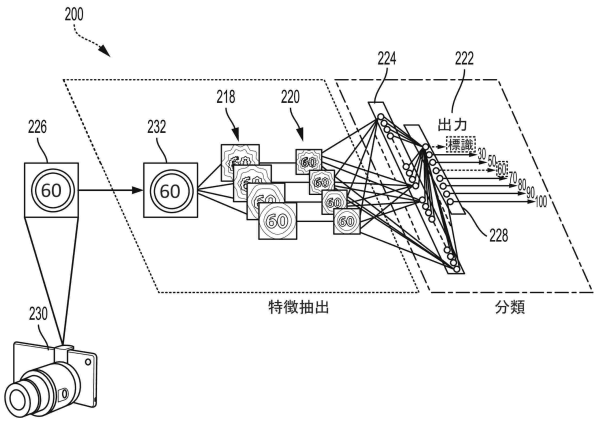
【図２Ｂ】



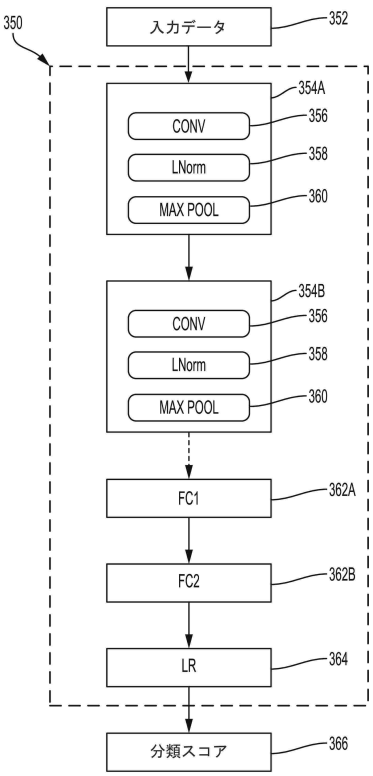
【図２Ｃ】



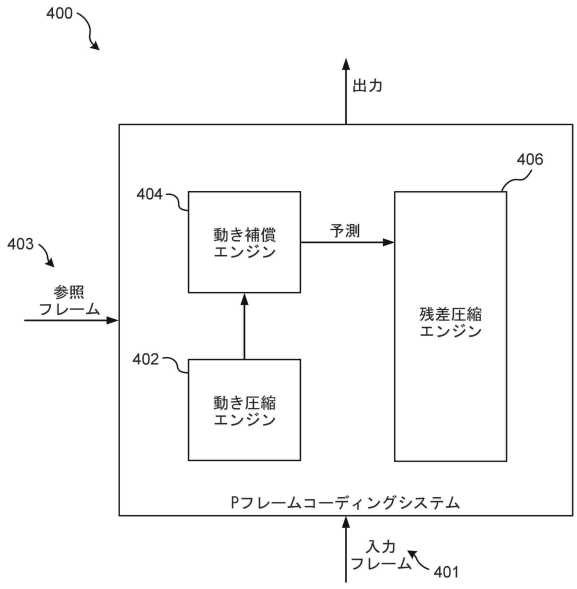
【図２Ｄ】


200
224
222
218
220
出力
226
232
60
30
50
70
80
90
100
228
230
特徴抽出
分類

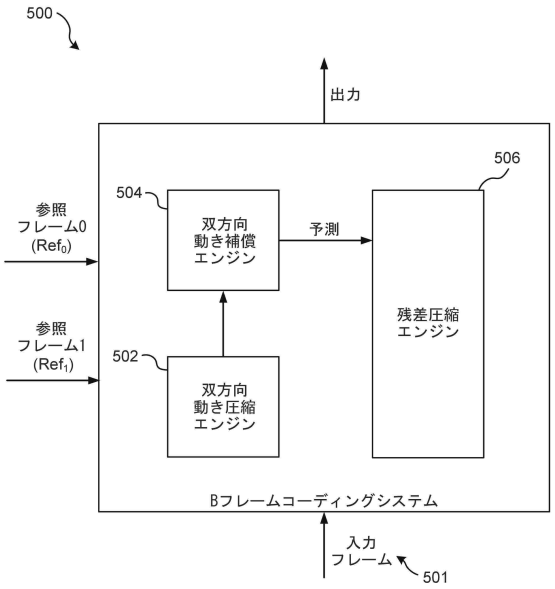
【図３】


350
入力データ
352
354A
CONV
356
LNorm
358
MAX POOL
360
354B
CONV
356
LNorm
358
MAX POOL
360
FC1
362A
FC2
362B
LR
364
分類スコア
366

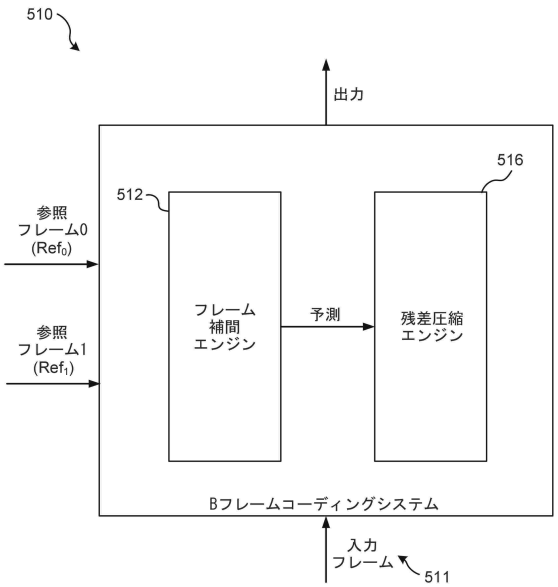
【図４】


400
出力
406
404
動き補償
エンジン
予測
403
参照
フレーム
残差圧縮
エンジン
402
動き圧縮
エンジン
Pフレームコーディングシステム
入力
フレーム
401

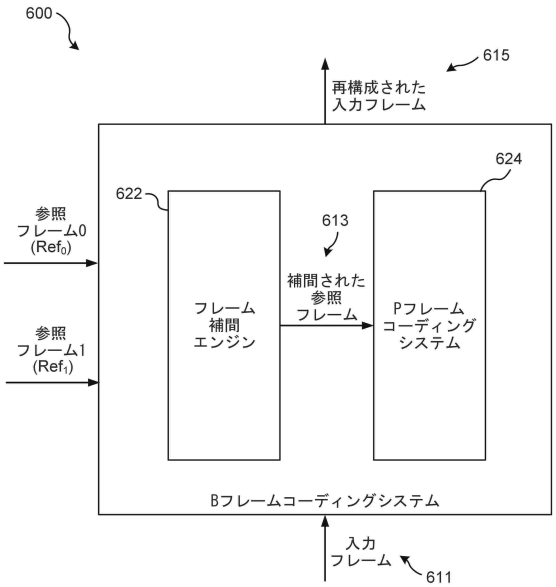
【図５Ａ】


500
出力
506
504
参照
フレーム0
(Ref0)
双方向
動き補償
エンジン
予測
残差圧縮
エンジン
参照
フレーム1
(Ref1)
502
双方向
動き圧縮
エンジン
Bフレームコーディングシステム
入力
フレーム
501

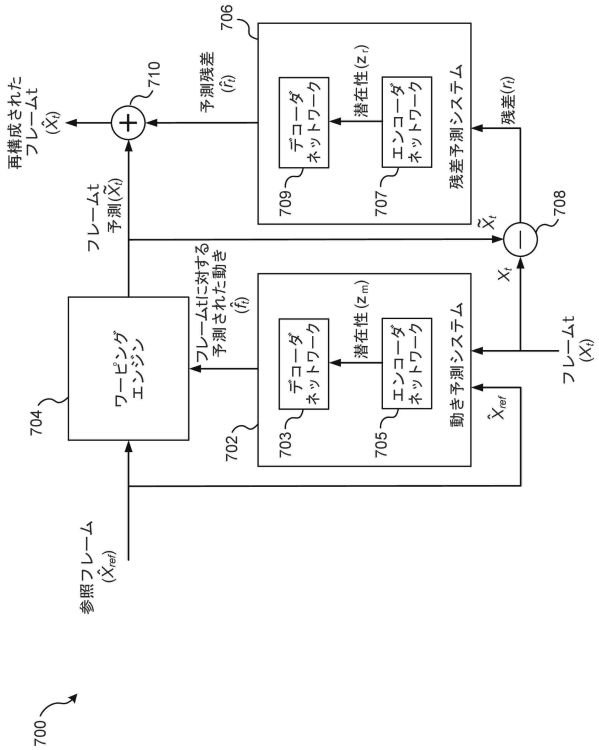
【図５Ｂ】



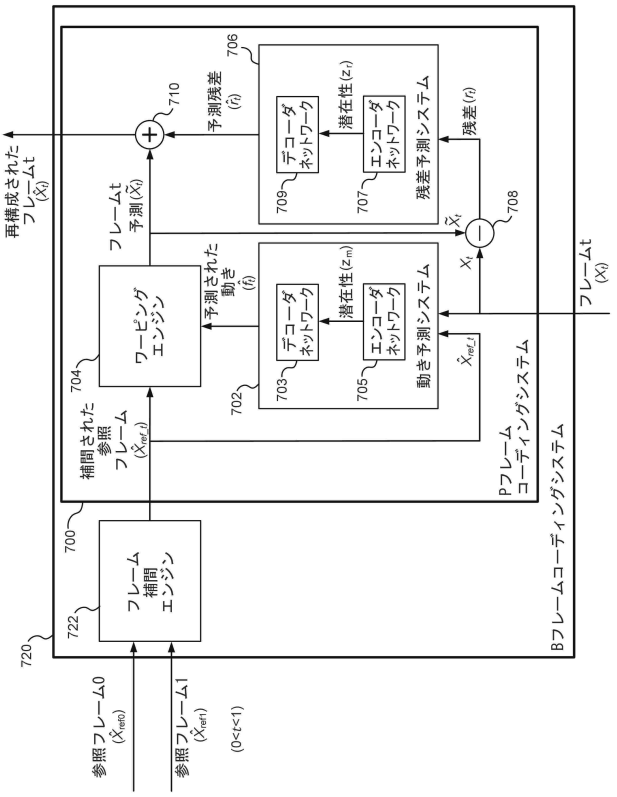
【図６】



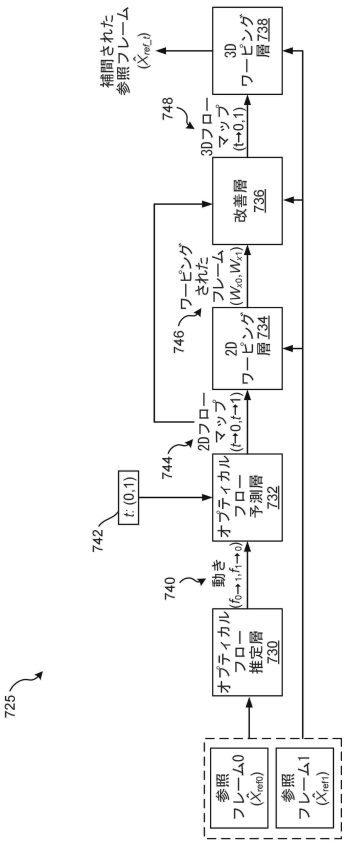
【図７Ａ】



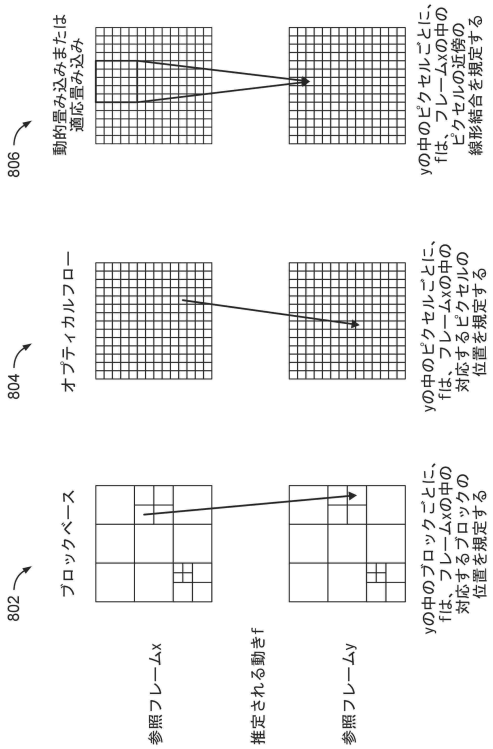
【図７Ｂ】



【図７Ｃ】



【図８】



【図９】

	GoP 1					GoP 2				...
タイプ	I	...	B	...	I	...	B	...	I	...
順序	1番目		2番目		1番目		2番目		1番目	...
参照										...

【図１０】

	GoP 1					GoP 2				...
タイプ	I	...	B	...	P	...	B	...	P	...
順序	1番目		3番目		2番目		2番目		1番目	...
参照										...

【図１１】

	GoP								
タイプ	参照	B	B	B	...	...	...	...	参照
順序	1	2	3	4					1
参照									

【図１２】

	GoP								
タイプ	参照	B	B	B	B	B	B	B	参照
順序	1	4	3	4	2	4	3	4	1
参照									

【図１３】

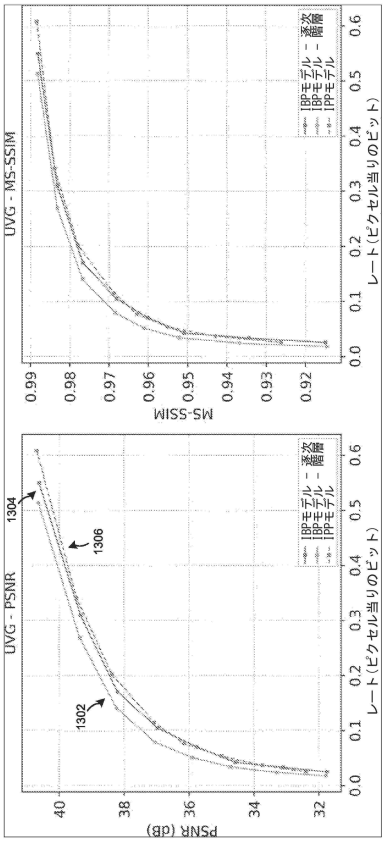
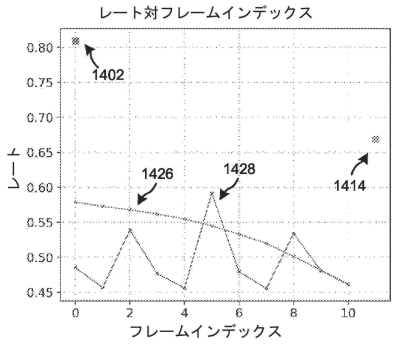


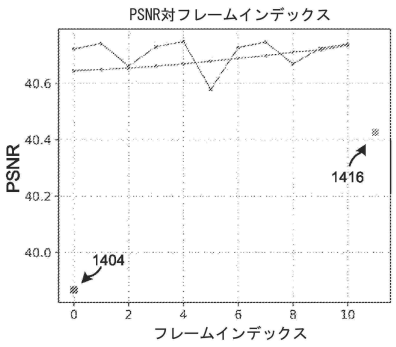
図13B

図13A

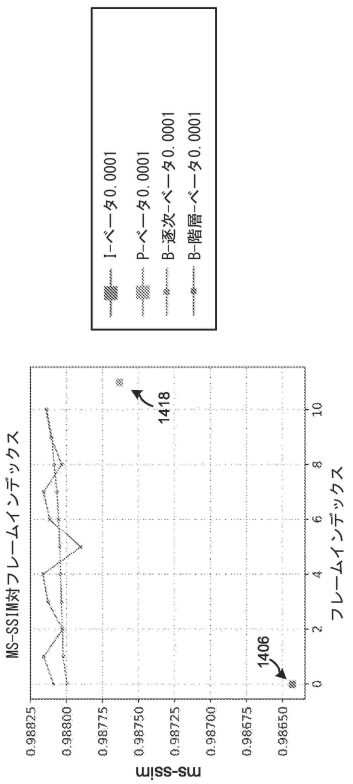
【図１４Ａ】



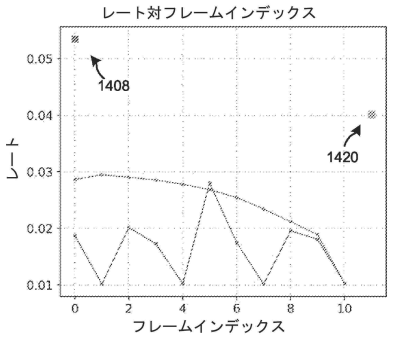
【図１４Ｂ】



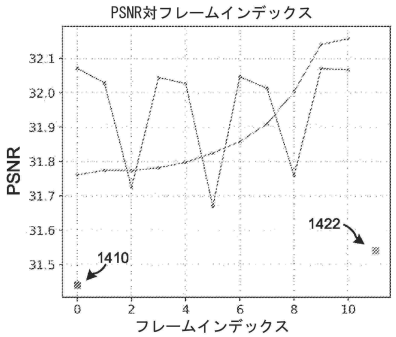
【図１４Ｃ】



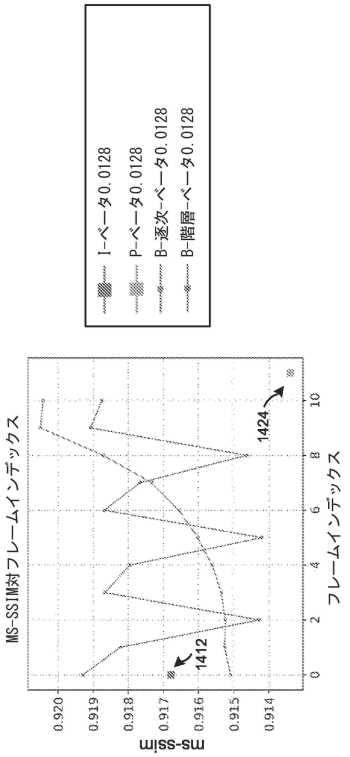
【図１４Ｄ】



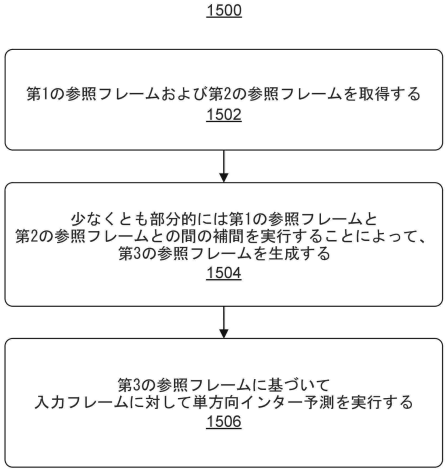
【図１４Ｅ】



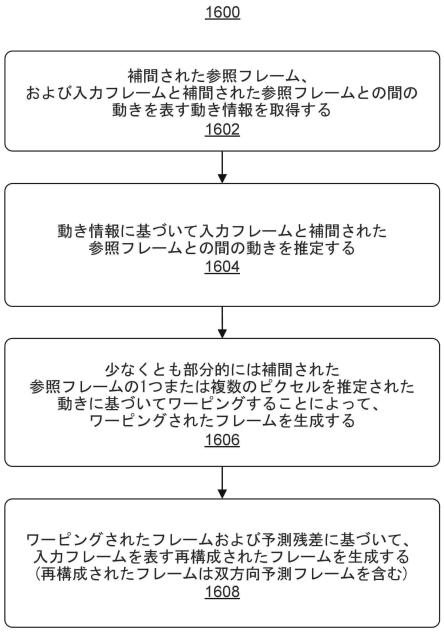
【図１４Ｆ】



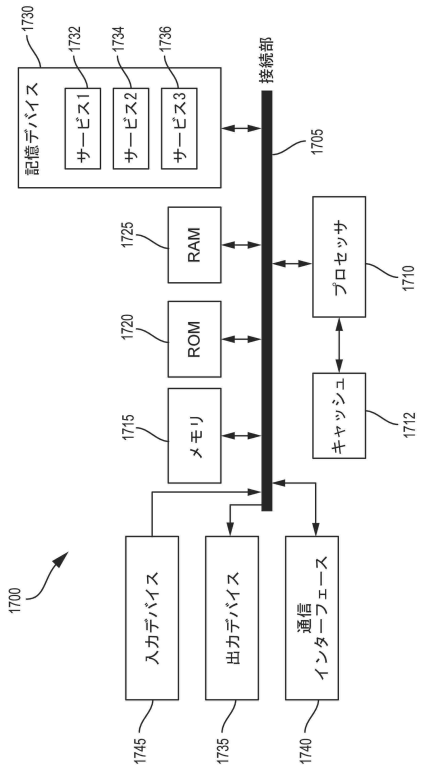
【図１５】



【図１６】



【図１７】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/014143

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N19/117 H04N19/139 H04N19/172 H04N19/577 H04N19/82
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI BOHAN ET AL: "Optical Flow Based Co-Located Reference Frame for Video Compression", IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE, USA, vol. 29, 12 August 2020 (2020-08-12), pages 8303-8315, XP011804954, ISSN: 1057-7149, DOI: 10.1109/TIP.2020.3014723 [retrieved on 2020-08-14] sections III, IV -----	1-52
X	US 2019/306526 A1 (CHO SEUNG-HYUN [KR] ET AL) 3 October 2019 (2019-10-03) paragraphs [0755] - [0758], [0855] - [0871] -----	1-52
	-/--	

[x] Further documents are listed in the continuation of Box C.

[x] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 May 2022	13/05/2022

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Montoneri, Fabio

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/014143

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2019/040134 A1 (GOOGLE LLC [US]) 28 February 2019 (2019-02-28) paragraphs [0005], [0102] - [0118], [0151] - [0154] -----	1-52
X	PARK WOONSUNG ET AL: "Deep Predictive Video Compression Using Mode-Selective Uni- and Bi-Directional Predictions Based on Multi-Frame Hypothesis", IEEE ACCESS, IEEE, USA, vol. 9, 21 December 2020 (2020-12-21), pages 72-85, XP011829522, DOI: 10.1109/ACCESS.2020.3046040 [retrieved on 2021-01-04] section III; figure 1 -----	1-52
A	LU GUO ET AL: "An End-to-End Learning Framework for Video Compression", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE COMPUTER SOCIETY, USA, vol. 43, no. 10, 20 April 2020 (2020-04-20), pages 3292-3308, XP011875084, ISSN: 0162-8828, DOI: 10.1109/TPAMI.2020.2988453 [retrieved on 2021-09-01] sections 3, 4 -----	1-52

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2022/014143

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2019306526 A1	03-10-2019	NONE	
WO 2019040134 A1	28-02-2019	CN 110741640 A	31-01-2020
		EP 3673655 A1	01-07-2020
		JP 6905093 B2	21-07-2021
		JP 2020522200 A	27-07-2020
		KR 20200002036 A	07-01-2020
		KR 20210109049 A	03-09-2021
		WO 2019040134 A1	28-02-2019

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW
１２１・サン・ディエゴ・モアハウス・ドライヴ・５７７５

Ｆターム（参考） 5C159 LA02 MA04 MA05 MA15 MA21 MC11 ME01 NN19 PP04 RC11
TA17 TB03 TB08