

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7687633号
(P7687633)

(45)発行日 令和7年6月3日(2025.6.3)

(24)登録日 令和7年5月26日(2025.5.26)

(51)国際特許分類	F I
H 0 4 N 19/126 (2014.01)	H 0 4 N 19/126
H 0 4 N 19/136 (2014.01)	H 0 4 N 19/136
H 0 4 N 19/152 (2014.01)	H 0 4 N 19/152
H 0 4 N 19/176 (2014.01)	H 0 4 N 19/176

請求項の数 20 (全20頁)

(21)出願番号	特願2022-581327(P2022-581327)	(73)特許権者	314015767
(86)(22)出願日	令和2年6月30日(2020.6.30)		マイクロソフト テクノロジー ライセン
(65)公表番号	特表2023-535290(P2023-535290		シング,エルエルシー
	A)		アメリカ合衆国 ワシントン州 9 8 0 5
(43)公表日	令和5年8月17日(2023.8.17)		2 レッドモンド ワン マイクロソフト
(86)国際出願番号	PCT/CN2020/099390		ウェイ
(87)国際公開番号	WO2022/000298	(74)代理人	100107766
(87)国際公開日	令和4年1月6日(2022.1.6)		弁理士 伊東 忠重
審査請求日	令和5年5月22日(2023.5.22)	(74)代理人	100070150
			弁理士 伊東 忠彦
		(74)代理人	100135079
			弁理士 宮崎 修
		(72)発明者	リー, ジアハオ
			中国 1 0 0 0 8 0 ベイジン ハイディ
			エン ディストリクト ダンリン ストリ
			最終頁に続く

(54)【発明の名称】 レート制御に基づく強化学習

(57)【特許請求の範囲】

【請求項1】

コンピュータで実行される方法であって：

ビデオ・エンコーダの符号化状態を決定するステップであって、前記符号化状態は前記ビデオ・エンコーダにより第1のビデオ・ユニットを符号化することに関連付けられている、ステップ；

強化学習モデルにより、前記ビデオ・エンコーダの前記符号化状態に基づいて、前記ビデオ・エンコーダのレート制御に関する符号化パラメータを決定するステップ；及び

前記第1のビデオ・ユニットとは異なる第2のビデオ・ユニットを、前記符号化パラメータに基づいて符号化するステップ；

を含み、前記強化学習モデルは、前記第2のビデオ・ユニットの符号化に基づいて、前記符号化パラメータに関する報酬を決定するステップを行うことにより訓練されており、前記報酬は、バッファのオーバーシュートにはペナルティを科し、且つ前記符号化パラメータがより高い視覚的品質をもたらす場合には増加するように、基本報酬をスケールリング因子でスケールリングすることにより取得され、前記スケールリング因子は、前記第2のビデオ・ユニットを符号化することに関連付けられる帯域幅の、送信チャネルの最大帯域幅に対する比率に基づいている、方法。

【請求項2】

請求項1に記載の方法において、前記符号化パラメータを決定するステップは：

前記強化学習モデルにより、前記ビデオ・エンコーダの前記符号化状態に基づいて行動

を決定するステップ；及び

前記行動を前記符号化パラメータにマッピングするステップ；
を含む、方法。

【請求項 3】

請求項 1 に記載の方法において、前記第 1 のビデオ・ユニットを符号化することに関連付けられている前記符号化状態は：

少なくとも前記第 1 のビデオ・ユニットを符号化することに関する成果を表現する状態；
前記ビデオ・エンコーダにより符号化されたビデオ・ユニットを送信前にバッファリングするように構成されたバッファの状態；及び

前記符号化されたビデオ・ユニットを送信するためのネットワークの状況に関連付けられた状態；

を含む、方法。

10

【請求項 4】

請求項 3 に記載の方法において、少なくとも前記第 1 のビデオ・ユニットを符号化することに関する成果は、符号化された第 1 のビデオ・ユニットの前記符号化パラメータと前記符号化された第 1 のビデオ・ユニットのサイズとを含み、前記バッファの状態は前記バッファの使い方を含み、前記ネットワークの状況に関連付けられた状態は目標ビット・パー・ピクセルを含み、方法。

【請求項 5】

請求項 4 に記載の方法において、前記バッファの使い方は：

前記バッファの最大スペースに対する占有スペースの比率；及び

前記ビデオ・ユニットにおいて測定された前記バッファの残存スペース；

のうちの少なくとも 1 つを含む、方法。

20

【請求項 6】

請求項 1 に記載の方法において、前記強化学習モデルにおいて、前記バッファのオーバーシュートが生じない場合には、より小さな値の符号化パラメータは、より高い視覚的品質に対応する基本報酬をもたらす一方、前記バッファのオーバーシュートが生じる場合に
対応する程度に小さな値の符号化パラメータは、負の値の基本報酬をもたらす、方法。

【請求項 7】

請求項 6 に記載の方法において、前記報酬を決定するステップは：

前記バッファのオーバーシュートが生じない場合には前記基本報酬が前記符号化パラメータに負の係数で比例するような方法で、前記基本報酬を決定するステップ；

を含む、方法。

30

【請求項 8】

コンピュータで実行される方法であって：

ビデオ・エンコーダの符号化状態を決定するステップであって、前記符号化状態は前記ビデオ・エンコーダにより第 1 のビデオ・ユニットを符号化することに関連付けられている、ステップ；

強化学習モデルにより、前記ビデオ・エンコーダの前記符号化状態に基づいて、前記ビデオ・エンコーダのレート制御に関する符号化パラメータを決定するステップ；及び

前記第 1 のビデオ・ユニットとは異なる第 2 のビデオ・ユニットを、前記符号化パラメータに基づいて符号化するステップ；

を含み、前記強化学習モデルは、前記第 2 のビデオ・ユニットの符号化に基づいて、前記符号化パラメータに関する報酬を決定するステップを行うことにより訓練されており、前記報酬は、バッファのオーバーシュートにはペナルティを科し、且つ前記符号化パラメータがより高い視覚的品質をもたらす場合には増加するように設定されており、前記強化学習モデルは、

前記符号化パラメータに関連付けられた行動を、前記ビデオ・エンコーダの前記符号化状態に基づいて決定するステップ；

前記第 2 のビデオ・ユニットを符号化することに関する前記符号化状態の評価価値を決

40

50

定するステップ；

前記報酬及び前記評価価値に基づいて、価値損失を決定するステップ；

前記行動に基づいてポリシー損失を決定するステップ；及び

前記価値損失及び前記ポリシー損失に基づいて、前記強化学習モデルを更新するステップ；

により訓練されている、方法。

【請求項 9】

請求項1に記載の方法において、前記強化学習モデルはエージェントのニューラル・ネットワークを含み、前記ニューラル・ネットワークは：

前記符号化状態から特徴を抽出するように構成された少なくとも1つの入力全結合層；

抽出された特徴を受けると結合された少なくとも1つのリカレント・ニューラル・ネットワーク；及び

前記エージェントの行動を決定するように構成された少なくとも1つの出力全結合層；を含む、方法。

【請求項 10】

請求項9に記載の方法において、前記ニューラル・ネットワークはアクター・クリティック・アーキテクチャに基づいて訓練されており、前記アクターは前記符号化状態に基づいて前記行動を生成するように構成されており、前記クリティックは前記符号化状態に関する評価価値を生成するように構成されており；及び

前記アクター及び前記クリティックは、前記少なくとも1つの入力全結合層を含む前記ニューラル・ネットワークと前記少なくとも1つのリカレント・ニューラル・ネットワークとによる共通部分を共有している、方法。

【請求項 11】

請求項1に記載の方法において、前記符号化パラメータは、量子化パラメータ及びラムダ・パラメータのうちの少なくとも1つを含む、方法。

【請求項 12】

請求項1に記載の方法において、前記ビデオ・エンコーダは、リアル・タイム通信のスクリーン・コンテンツを符号化するように構成されている、方法。

【請求項 13】

プロセッサ；及び

前記プロセッサによる実行のために命令を記憶しているメモリ；

を含むデバイスであって、前記命令は、前記プロセッサにより実行されると、前記デバイスに動作を実行させ、前記動作は：

ビデオ・エンコーダの符号化状態を決定するステップであって、前記符号化状態は前記ビデオ・エンコーダにより第1のビデオ・ユニットを符号化することに関連付けられている、ステップ；

強化学習モデルにより、前記ビデオ・エンコーダの前記符号化状態に基づいて、前記ビデオ・エンコーダのレート制御に関する符号化パラメータを決定するステップ；及び

前記第1のビデオ・ユニットとは異なる第2のビデオ・ユニットを、前記符号化パラメータに基づいて符号化するステップ；

を含み、前記強化学習モデルは、前記第2のビデオ・ユニットの符号化に基づいて、前記符号化パラメータに関する報酬を決定するステップを行うことにより訓練されており、前記報酬は、バッファのオーバーシュートにはペナルティを科し、且つ前記符号化パラメータがより高い視覚的品質をもたらす場合には増加するように、基本報酬をスケールリング因子でスケールリングすることにより取得され、前記スケールリング因子は、前記第2のビデオ・ユニットを符号化することに関連付けられる帯域幅の、送信チャネルの最大帯域幅に対する比率に基づいている、デバイス。

【請求項 14】

請求項13に記載のデバイスにおいて、前記第1のビデオ・ユニットを符号化することに関連付けられている前記符号化状態は：

10

20

30

40

50

少なくとも前記第1のビデオ・ユニットを符号化することに関する成果を表現する状態；
前記ビデオ・エンコーダにより符号化されたビデオ・ユニットを送信前にバッファリングするように構成されたバッファの状態；及び

前記符号化されたビデオ・ユニットを送信するためのネットワークの状況に関連付けられた状態；

を含む、デバイス。

【請求項15】

請求項14に記載のデバイスにおいて、少なくとも前記第1のビデオ・ユニットを符号化することに関する成果は、符号化された第1のビデオ・ユニットの前記符号化パラメータと前記符号化された第1のビデオ・ユニットのサイズとを含み、前記バッファの状態は前記バッファの使い方を含み、前記ネットワークの状況に関連付けられた状態は目標ビット・パー・ピクセルを含む、デバイス。

10

【請求項16】

請求項13に記載のデバイスにおいて、前記強化学習モデルにおいて、前記バッファのオーバーシュートが生じない場合には、より小さな値の符号化パラメータは、より高い視覚的品質に対応する基本報酬をもたらす一方、前記バッファのオーバーシュートが生じる場合に対応する程度に小さな値の符号化パラメータは、負の値の基本報酬をもたらす、デバイス。

【請求項17】

請求項16に記載のデバイスにおいて、前記報酬を決定するステップは：
前記バッファのオーバーシュートが生じない場合には前記基本報酬が前記符号化パラメータに負の係数で比例するような方法で、前記基本報酬を決定するステップ；
を含む、デバイス。

20

【請求項18】

請求項13に記載のデバイスにおいて、前記強化学習モデルはエージェントのニューラル・ネットワークを含み、前記ニューラル・ネットワークは：

前記符号化状態から特徴を抽出するように構成された少なくとも1つの入力全結合層；
抽出された特徴を受けるように結合された少なくとも1つのリカレント・ニューラル・ネットワーク；及び

前記エージェントの行動を決定するように構成された少なくとも1つの出力全結合層；
を含む、デバイス。

30

【請求項19】

請求項13に記載のデバイスにおいて、前記ビデオ・エンコーダは、リアル・タイム通信のスクリーン・コンテンツを符号化するように構成されている、デバイス。

【請求項20】

プログラム命令を組み込んだコンピュータ・プログラムであって、前記プログラム命令は、プロセッサに動作を実行させるためにプロセッサにより実行されることが可能であり、前記動作は：

ビデオ・エンコーダの符号化状態を決定するステップであって、前記符号化状態は前記ビデオ・エンコーダにより第1のビデオ・ユニットを符号化することに関連付けられている、ステップ；

40

強化学習モデルにより、前記ビデオ・エンコーダの前記符号化状態に基づいて、前記ビデオ・エンコーダのレート制御に関する符号化パラメータを決定するステップ；及び

前記第1のビデオ・ユニットとは異なる第2のビデオ・ユニットを、前記符号化パラメータに基づいて符号化するステップ；

を含み、前記強化学習モデルは、前記第2のビデオ・ユニットの符号化に基づいて、前記符号化パラメータに関する報酬を決定するステップを行うことにより訓練されており、前記報酬は、バッファのオーバーシュートにはペナルティを科し、且つ前記符号化パラメータがより高い視覚的品質をもたらす場合には増加するように、基本報酬をスケールリング因子でスケールリングすることにより取得され、前記スケールリング因子は、前記第2のビデオ

50

オ・ユニットを符号化することに関連付けられる帯域幅の、送信チャネルの最大帯域幅に対する比率に基づいている、コンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

[0001] リアル・タイム通信 (real time communication, RTC) において、一般的な要求は様々なユーザーとの画面共有である。例えば、参加者は、彼又は彼女のデスクトップ画面を、マルチ・ユーザー・ビデオ会議で他の参加者に提示することを必要とするかもしれない。この状況では、技術的なゴールは、より良い体感品質 (quality of experience, QOE) を提供することであり、これは、視覚的な品質、ドロップ・レート、伝送遅延などのような様々な要因によってしばしば決定される。レート制御は、目標ビットレートを達成するためのビデオ・エンコーダの符号化パラメータを決定することによってそのゴールを達成するために重要な役割を果たす。

10

【0002】

[0002] 既存のレート制御方法は、主に、自然なシーン (natural scenes) を伴うビデオ用に設計されている。しかしながら、多くがスムーズなコンテンツの動きを伴っている自然なビデオとは異なり、スクリーン・コンテンツは、通常、複雑な突然の変化や静止したシーンと組み合わせられている。この独特な動きの特徴に起因して、自然なビデオ用に設計された既存のレート制御方法は、スクリーン・コンテンツに対してはうまく機能することができない。

20

【背景技術】

【0003】

[0003] 本件で説明される対象事項の実装によれば、強化学習 (reinforcement learning) に基づくレート制御のための解決策が提供される。この解決策では、ビデオ・エンコーダの符号化状態が決定され、符号化状態はビデオ・エンコーダによる第1のビデオ・ユニットの符号化に関連付けられている。ビデオ・エンコーダのレート制御に関する符号化パラメータは、強化学習モデルにより、ビデオ・エンコーダの符号化状態に基づいて決定される。第1のビデオ・ユニットとは異なる第2のビデオ・ユニットが、符号化パラメータに基づいて符号化される。強化学習モデルは、1つ以上のビデオ・ユニットの符号化状態を受信して、別のビデオ・ユニットで使用する符号化パラメータを決定するように構成されている。符号化状態は、限られた状態の次元を有し、演算オーバーヘッドが低減されたリアル・タイム通信のためのより良いQOEを達成することが可能である。

30

【0004】

[0004] この要約は、以下の詳細な説明で更に記述される概念の選択肢を簡略化された形式で導入するために提供される。この要約は、クレームされた対象事項の主要な特徴又は本質的な特徴を特定するようには意図されておらず、また、クレームされた対象事項の範囲を限定するために使用されることになるようにも意図されていない。

【図面の簡単な説明】

【0005】

[0005] 本件において添付図面に記載されている対象事項の幾つかの実装のより詳細な説明を通じて、上記及びその他の目的、並びに本件で説明される対象事項の特徴及び利点は、より明白になるであろう。

40

【図1】[0006] 図1は、本件で説明される対象事項の種々の実装を実施することが可能な演算デバイスのブロック図を示す。

【図2】[0007] 図2は、本件で説明される対象事項の実装による強化学習モジュールのブロック図を示す。

【図3】[0008] 図3は、本件で説明される対象事項の実装による強化学習モジュールで使用するためのエージェントの例を示す。

【図4】[0009] 図4は、本件で説明される対象事項の実装による強化学習ベースのレート制御方法のフローチャートを示す。 [0010] 図面を通じて、同一又は類似の参照番号

50

は同一又は類似の要素を表す。

【発明を実施するための形態】

【0006】

[0011] 本件で説明される対象事項は、以下、幾つかの例示的な実装を参照しながら説明される。これらの実装は、当業者が本件で説明される対象事項をより良く理解して実施することを可能にする目的だけのために議論されており、対象事項の範囲に関する如何なる限定も示唆するものではない、ということが理解されるべきである。

【0007】

[0012] 本件で使用されているように、用語「～を含む」及びその変形は、「～を含むが、それに限定されない」を意味するオープンな用語として読まれるべきである。用語「～に基づいて」は、「～に少なくとも部分的に基づいて」として読まれるべきである。用語「ある実装」及び「実装」は、「少なくとも1つの実装」として読まれるべきである。用語「別の実装」は、「少なくとも1つの他の実装」として読まれるべきである。用語「第1の」、「第2の」等は、異なる又は同じ対象を指す可能性がある。他の定義は、明示的であれ暗黙的であれ、以下に含まれる可能性がある。

【0008】

[0013] 図1は、本件で説明される対象事項の種々の実装が実施されることが可能な演算デバイス100のブロック図を示す。図1に示される演算デバイス100は、本件で説明される対象事項の実装の機能及び範囲に如何なる方法によっても如何なる制限も示唆することなく、単に説明のためのものであることが理解されるであろう。図1に示すように、演算デバイス100は、汎用演算デバイス100を含む。演算デバイス100の構成要素は、1つ以上のプロセッサ又は処理ユニット110、メモリ120、記憶デバイス130、1つ以上の通信ユニット140、1つ以上の入力デバイス150、及び1つ以上の出力デバイス160を含んでもよいが、これらに限定されない。

【0009】

[0014] 幾つかの実装において、演算デバイス100は、演算能力を有する任意のユーザー端末又はサーバー端末として実装されてもよい。サーバー端末は、サーバー、大規模な演算デバイス等であって、サービス・プロバイダによって提供されるものであってもよい。ユーザー端末は、例えば、携帯電話、ステーション、ユニット、デバイス、マルチメディア・コンピュータ、マルチメディア・タブレット、インターネット・ノード、通信機、デスクトップ・コンピュータ、ラップトップ・コンピュータ、ノートブック・コンピュータ、ネットブック・コンピュータ、タブレット・コンピュータ、パーソナル通信システム・システム（PCS）デバイス、パーソナル・ナビゲーション・デバイス、パーソナル・デジタル・アシスタント（PDA）、オーディオ/ビデオ・プレーヤ、デジタル・カメラ/ビデオ・カメラ、測位デバイス、テレビ受信機、無線放送受信機、電子書籍デバイス、ゲーミング・デバイス、又はこれらの任意の組み合わせであって、これらのデバイスのアクセサリ及び周辺機器を含むもの、等を含む任意の種類の移動端末、固定端末、又は携帯端末であってよいし、又はそれらの組み合わせであってよい。演算デバイス100は、ユーザーに対する任意のタイプのインターフェース（例えば、「ウェアラブル」回路など）をサポートすることが可能であることが想定されている。

【0010】

[0015] 処理ユニット110は、物理的又は仮想的なプロセッサであってもよく、メモリ120に記憶されたプログラムに基づいて種々のプロセスを実行することができる。マルチ・プロセッサ・システムでは、複数の処理ユニットが、コンピュータ実行可能命令を並列に実行し、演算デバイス100の並列処理能力を向上させる。処理ユニット110は、中央処理ユニット（CPU）、マイクロプロセッサ、コントローラ又はマイクロコントローラとも呼ばれる。

【0011】

[0016] 演算デバイス100は、典型的には、種々のコンピュータ記憶媒体を含む。そのような媒体は、揮発性及び不揮発性媒体、又は着脱可能な及び着脱可能でない媒体を含む

10

20

30

40

50

がこれらに限定されない演算デバイス100によってアクセス可能な任意の媒体であるとする
ことが可能である。メモリ120は、揮発性メモリ（例えば、レジスタ、キャッシュ、ラン
ダム・アクセス・メモリ（RAM））、不揮発性メモリ（例えば、リード・オンリー・メ
モリ（ROM）、電氣的に消去可能なプログラマブル・リード・オンリー・メモリ（EEPROM）、又はフラッシュ・メモリ）、又はそれらの任意の組み合わせであるとする
ことが可能である。記憶デバイス130は、任意の着脱可能な又は着脱可能でない媒体であってよく、メモリ、フラッシュ・メモリ・ドライブ、磁気ディスク、又はその他の媒体のような
機械読み取り可能な媒体を含んでもよく、これらは、情報及び/又はデータを記憶するために使用することが可能であり、且つ演算デバイス装置100内でアクセスされることが可能
である。

10

【0012】

[0017] 演算デバイス100は、更に、追加の着脱可能な/着脱可能でない、揮発性/不揮
発性メモリ媒体を含んでもよい。図1には示されていないが、着脱可能で不揮発性の磁気
ディスクから読み取り及び/又はそこへ書き込みを行うための磁気ディスク・ドライブ、及
び着脱可能な不揮発性の光ディスクから読み取り及び/又はそこへ書き込みを行うための光
ディスク・ドライブを提供することが可能である。このような場合、各ドライブは、1つ
以上のデータ媒体インターフェースを介してバス（図示せず）に接続されてもよい。

【0013】

[0018] 通信ユニット140は、通信媒体を介して別の演算デバイスと通信する。更に演
算デバイス100内の構成要素の機能は、通信接続を介して通信することが可能な単一の演
算クラスタ又は複数の演算マシンによって実現することが可能である。従って、演算デバ
イス100は、1つ以上の他のサーバー、ネットワーク化されたパーソナル・コンピュータ
（PC）、又は更に一般的なネットワーク・ノードとの論理接続を使用して、ネットワーク
化された環境で動作することが可能である。

20

【0014】

[0019] 入力デバイス150は、マウス、キーボード、トラッキング・ボール、音声入力
デバイス等の種々の入力装置のうちの1つ以上であってもよい。出力デバイス160は、デ
ィスプレイ、ラウドスピーカ、プリンタ等の種々の出力デバイスのうちの1つ以上であっ
てもよい。通信ユニット140によって、演算デバイス装置100は、記憶デバイス及び表示
デバイスのような1つ以上の外部デバイス（図示せず）と、ユーザーが演算デバイス100
と対話することを可能にする1つ以上のデバイスと、又は、何らかのデバイスであって、
必要に応じて、演算デバイス100が1つ以上の他の演算デバイスと通信することを可能に
するもの（例えば、ネットワーク・カード、モデム等）と、更に通信することが可能であ
る。このような通信は、入力/出力（I/O）インターフェース（図示せず）を介して行うこ
とが可能である。

30

【0015】

[0020] 一部の実装では、単一のデバイスに統合される代替例として、演算デバイス10
0の一部又は全部の構成要素が、クラウド演算アーキテクチャに配置されてもよい。クラ
ウド演算アーキテクチャでは、構成要素は、遠隔的に用意されていてもよく、本件で説明
される対象事項において説明される機能を実現するために協働してもよい。一部の实装で
は、クラウド演算は、演算、ソフトウェア、データ・アクセス、及びストレージ・サービ
スを提供するが、これらのサービスを提供するシステム又はハードウェアの物理的な位置
又は構成に気付いていることはエンド・ユーザーには要求されない。様々な実装において
、クラウド演算は、適切なプロトコルを使用して、ワイド・エリア・ネットワーク（例え
ば、インターネットなど）を介してサービスを提供する。例えば、クラウド演算プロバイ
ダは、ウェブ・ブラウザ又は他の任意の演算コンポーネントを介してアクセスすることが
可能なワイド・エリア・ネットワーク上でアプリケーションを提供する。クラウド演算ア
ーキテクチャのソフトウェア又はコンポーネント、及び対応するデータは、遠隔地のサー
バーに格納されていてもよい。クラウド演算環境における演算リソースは、リモート・デー
タ・センター内の場所に併合又は分配されていてもよい。クラウド演算インフラストラク

40

50

チャは、ユーザーにとって単一のアクセス・ポイントとして動作するが、共有データ・センターを介してサービスを提供する可能性がある。従って、クラウド演算アーキテクチャは、本件で説明される構成要素及び機能を、遠隔地のサービス・プロバイダから提供するために使用されてもよい。あるいは、これらは、従来のサーバーから提供されてもよいし、直接的に又は他の方法でクライアント・デバイスにインストールされてもよい。

【0016】

[0021] 演算デバイス100は、本件で説明される対象事項の実装に強化学習ベースのレート制御を実施するために使用されてもよい。メモリ120は、1つ以上のプログラム命令を有する1つ以上の強化学習モジュール122を含んでもよい。これらのモジュールは、本件で説明される様々な実装の機能を実行するために、処理ユニット110によってアクセス可能であり実行可能である。例えば、入力デバイス150は、ビデオ会議アプリケーションを可能にするために、演算デバイス100の環境のビデオ又は一連のフレームを、強化学習モジュール122に提供することが可能であり、一方、処理ユニット110及び/又はメモリ120は、スクリーン・コンテンツの少なくとも一部を、強化学習モジュール122に提供して、スクリーン・コンテンツ共有アプリケーションを実行可能にしてもよい。マルチメディア・コンテンツは、良好なQOEでレート制御を達成するために、強化学習モジュール122によって符号化されることが可能である。

10

【0017】

[0022] 図2をここで参照すると、本件で説明される実装による強化学習モジュール200のブロック図が示されている。強化学習モジュール200は、例えば、強化学習モジュール122として、演算デバイス100内に実装されてもよい。強化学習モジュール200はエンコーダ204を含み、これは、演算デバイスの100の他の構成要素、例えば、処理ユニット110、メモリ120、ストレージ130、入力デバイス150及び/又はこれに類するものからのマルチメディア・コンテンツを符号化するように構成されている。例えば、入力デバイス150は、ビデオの1つ以上のフレームを強化学習モジュール200に提供することが可能である一方、処理ユニット110及び/又はメモリ120は、スクリーン・コンテンツの少なくとも一部を強化学習モジュール200に提供することが可能である。例えば、エンコーダ204は、ビデオ・エンコーダ、特に、演算デバイス100からのスクリーン・コンテンツのために最適化されたビデオ・エンコーダであってもよい。

20

【0018】

[0023] 量子化パラメータ (quantization parameter, QP) 又はラムダのようなレート制御に関連する符号化パラメータは、ビデオ・ユニット、例えばフレーム、ブロック又はフレーム内のマクロブロックの圧縮の粒度を制御する。大きな値は、より高い量子化、より多くの圧縮、及びより低い品質が存在することになることを意味する。より低い値はその逆を意味する。従って、エンコーダの符号化パラメータ、例えば、量子化パラメータやラムダを調整するためにレート制御を実行することによって、良好なQOEを達成することが可能である。ここで、量子化パラメータやラムダに対する参照が行われているが、量子化パラメータやラムダは、例示の目的で提供されており、レート制御に関連付けられる適切な他の任意の符号化パラメータが調整されたり制御されたりする可能性がある、ということに留意されたい。

30

40

【0019】

[0024] 図2に示されるように、強化学習モジュール200は、エンコーダ204の符号化パラメータを制御する決定を行うように構成されたエージェント202を含むことが可能である。幾つかの実装では、エージェント202は、ニューラル・ネットワーク、例えば、リカレント・ニューラル・ネットワークによって実現される強化学習モデルを採用してもよい。

【0020】

[0025] 次いで、符号化されたビットストリームが送信バッファに出力される。エンコーダ204はまた、ビットストリーム送信プロセスを実現するために、そのような送信バッファ (不図示) を含んでもよい。符号化された後、最新の符号化されたビデオ・ユニット

50

のビットストリームは、送信バッファに記憶又は追加される。送信の際に、送信バッファに記憶されているビットストリームは、或る帯域幅で1つ以上のチャンネルを介して1つ以上の受信機へ送信され、送信されたビットストリームは、その帯域幅での送信とともにバッファから除去されることになる。送信バッファの状態は、送信バッファに出入りする進入及び進出ビットストリームに起因して、絶えず変化するプロセスにある。

【0021】

[0026] 各時間ステップ t において、エージェント202は、エンコーダ204の符号化状態 s_t を観察する。時間ステップ t における符号化状態 s_t は、時間ステップ $t-1$ における少なくともビデオ・ユニットの符号化に基づいて決定されてもよい。この入力情報に基づいて、エージェント202は、推定を行って行動 (action) a_t を出力する。行動 a_t は、エンコーダ204が時間ステップ t においてビデオ・ユニットをどれだけ細かく圧縮すべきかを示す。行動 a_t は、レート制御のためのエンコーダ204の符号化パラメータ、例えば量子化パラメータ (QP) であってもよいし、あるいはエンコーダ204の符号化パラメータにマッピングされてもよい。符号化パラメータを取得した後、エンコーダ204は、ビデオ・ユニット (例えば、スクリーン・コンテンツ・フレーム) を符号化し始めることが可能である。時間ステップ t におけるビデオ・ユニットの符号化は、時間ステップ $t+1$ におけるエージェント202の符号化状態 s_{t+1} を更新するために使用されることになる。強化学習モジュール200は、リアル・タイムのスクリーン・コンテンツ共有以外の適切な他の任意のマルチメディア・アプリケーションに適用される可能性がある、ということが理解されるべきである。

【0022】

[0027] 従来の手作業によるルールではなく、エンコーダの符号化状態に基づいて、エージェントの行動によって符号化パラメータを制御することによって、本件で説明される対象事項の実装による強化学習ベースの解決策は、無視できる程度のドロップ・レート変化を伴って、より良い視覚的品質を達成することができる。エンコーダの符号化状態は、制限された状態空間を有し、従って、符号化パラメータの決定が、削減された演算オーバーヘッド及び改善された効率とともに行われることを可能にする。特に、シーンの突然の変化がスクリーン・コンテンツで生じた場合に、十分に訓練された強化学習モデルは、より良いQOEを達成するために、符号化パラメータを非常に速やかに更新することができ、このことは、リアル・タイム通信におけるスクリーン・コンテンツ共有にとって特に有益である。強化学習ベースのアーキテクチャは、如何なるコーデックにも限定されず、種々の様々なコーデック、例えばH.264、HEVC、及びAV1とともに協働することができる。

【0023】

[0028] 一部の实装では、強化学習モジュール200のエージェント202が正確かつ信頼性のある決定を行うことを支援するために、エージェント202への入力としての時間ステップ t における符号化状態は、様々な観点から符号化状態を表すための多数の要素を含むことが可能である。例えば、ビデオ・ユニットはフレームであってもよく、符号化状態 s_t は、少なくとも時間ステップ $t-1$ におけるフレームを符号化することに関する成果 (outcome) を表現する状態；時間ステップ t における送信バッファの状態；及び符号化されたフレームを送信するための、時間ステップ t におけるネットワークの状況に関連付けられた状態を含んでもよい。

【0024】

[0029] 例えば、少なくとも時間ステップ $t-1$ におけるフレームを符号化することに関する成果は、時間ステップ $t-1$ より前のフレーム、例えば時間ステップ $t-2$ におけるフレーム、を符号化することに関する成果を更にも含む。一例では、その成果は、時間ステップ $t-1$ における符号化フレームの符号化パラメータ (例えば、QP又はラムダ)、及び時間ステップ $t-1$ における符号化フレームのサイズを含んでもよい。フレームが欠落した場合、時間ステップ $t-1$ において符号化されるフレームの符号化パラメータは、ゼロのような所定の値に設定されてもよい。一例では、時間ステップ $t-1$ におけるフレーム・サイズは、フレームのフレーム・サイズ比率によって表現されてもよく、これは、平均的な目標フレーム・サイズに対するフレーム・サイズの比率によって定義される。換言すれば、時間ス

10

20

30

40

50

テップ $t-1$ におけるフレーム・サイズは、平均目標フレーム・サイズによって正規化されてもよい。例えば、フレーム・サイズは、フレームのビットストリーム・サイズによって表現されてもよく、平均目標フレーム・サイズは、フレーム中の目標ビット数の平均を表してもよく、また、目標ビットレートをフレーム・レートで除算することによって計算されてもよい。目標ビットレートは、送信されるべき目標ビット数を表し、フレーム・レートはフレームを送信する頻度又はレートを表す。目標ビットレートとフレーム・レートの両方を、ビデオ・エンコーダから決定することが可能である。

【0025】

【0030】 一例では、送信バッファの状態は、バッファの使い方、例えば、バッファの最大スペースに対する占有スペースの比率；フレーム内で測定されたバッファの残存スペース；又はその組み合わせを含んでもよい。フレーム内で測定されたバッファの残存スペースは、バッファの残存スペースを平均目標フレーム・サイズで除算することによって計算されてもよい。この値は、フレーム・レートの影響が考慮されている、別の態様からのバッファの使い方を述べている。

10

【0026】

【0031】 一例では、ネットワークの状況に関連する状態は、目標ビット・パー・ピクセル (bits per pixel, BPP) を含む。このパラメータは、ピクセルによって使用されるビット数によって定義され、また、目標ビットレートを、単位時間当たりのフレーム内のピクセル数で除算することによって計算されてもよい。フレーム内のピクセル数及び目標ビットレートは、例えば、ビデオ・エンコーダから決定されてもよい。

20

【0027】

【0032】 一部の実装では、上述の符号化状態はフレームに関しており、強化学習モジュール200はフレーム・ベースで決定を行う。他の実装では、強化学習モジュール200は、圧縮又は符号化のために、適切な他の任意のビデオ・ユニットに適用され又は適合させられてもよい。例えば、強化学習モジュールは、ブロック・レベルで、例えばマクロブロック (H.264)、符号化ツリー・ユニット (HEVC)、スーパーブロック (AV1) などで決定を行ってもよい。従って、エージェント202への入力として使用される符号化状態 s_t は、時間ステップ $t-1$ における少なくとも1つのブロックを符号化することに関する成果を表す状態、時間ステップ t における送信バッファの状態、及び、符号化されたブロックを送信するための、時間ステップ t におけるネットワークの状況に関連付けられた状態を含んでもよい。

30

【0028】

【0033】 例えば、少なくとも1つのブロックを符号化することに関する成果は、1つ以上の近隣のブロック (neighbor block) を符号化することに関する成果を含んでもよい。近隣のブロックは、処理されるブロックの空間的に左、右、上、及び/又は下にあるブロックを含んでもよい。空間的に近隣のブロックの符号化は、時間ステップ $t-1$ 又はその他の先行する時間ステップで実行されていてもよい。空間的に近隣のブロックの符号化の成果は、ストレージに記憶されてもよく、空間的に近隣のブロックの符号化の成果は、ストレージから取り出されてもよい。追加的又は代替的に、近隣のブロックは、時間的に近隣のブロックであるとも言及される、先行するフレームにおける1つ以上の対応するブロックを含んでもよい。時間的に近隣のブロックの符号化の成果は、ストレージに記憶され、そこから取り出されることが可能である。

40

【0029】

【0034】 一例において、成果は、符号化された少なくとも1つのブロックの符号化パラメータ、例えばQP又はラムダ、及び符号化された少なくとも1つのブロックのサイズを含んでもよい。例えば、符号化されたブロックのサイズは、符号化されたブロックのサイズの、平均目標ブロック・サイズに対する比率によって定義されるブロック・サイズ比率によって表現されてもよい。言い換えれば、ブロック・サイズは、平均目標ブロック・サイズによって正規化することが可能である。例えば、ブロック・サイズは、ブロックを符号化するためのビットストリーム・サイズによって表現されてもよく、平均目標ブロック・

50

サイズは、ブロックにおける目標ビット数の平均を表してもよく、また、目標ビットレートを、単位時間あたりに送信されるブロック数で除算することによって計算されてもよい。

【0030】

[0035] 一例では、送信バッファの状態は、バッファの使い方、例えば、バッファの最大スペースに対する占有スペースの比率、ブロック内で測定されるバッファの残存スペース、又はその組み合わせを含んでもよい。ブロック内で測定されるバッファの残存スペースは、バッファの残存スペースを、平均目標ブロック・サイズで除算することによって計算されてもよい。

【0031】

[0036] 一例では、ネットワーク状態に関連する状態は、目標ビット・パー・ピクセル (BPP) を含む。このパラメータは、ピクセルによって使用されるビット数によって定義され、フレームに関する実装と同様に計算することができる。

10

【0032】

[0037] 符号化状態は、量子化パラメータ又はラムダのような符号化パラメータに関して説明されている。符号化状態は、エンコーダによって使用されるレート制御に関連付けられた適切な他の任意の符号化パラメータに適用されてもよい、ということに留意されたい。

【0033】

[0038] 再び図2に関し、エージェント202により出力される行動 a_t は、エンコーダ204の符号化品質を制御することが可能である。例えば、エージェント202によって決定される行動 a_t は、正規化されて、0ないし1の範囲内にあってもよい。一部の实装では、行動は、エンコーダが理解することが可能なQPにマッピングされることが可能である。例えば、マッピングは、次式により実装されてもよい：

20

【数1】

$$QP_{cur} = QP_{min} + (QP_{max} - QP_{min}) * action \quad (1)$$

[0039] ここで、 QP_{max} 及び QP_{min} はそれぞれ最大及び最小のQPを表し、 QP_{cur} はエンコーダ204による符号化のために使用されることになるQPを表す。このマッピング関数は線形関数として例示されているが、適切な他の任意の関数を代わりに使用することが可能である、ということが理解されるべきである。より小さなQP値は、より繊細な方法で圧縮を実行し且つより高い再構築品質を獲得することを、エンコーダに行わせる。しかしながら、その犠牲は、より大きな符号化されたビットストリームを生成することである。過剰に大きなビットストリームは、バッファを簡単にオーバーシュートさせてしまい、それに応じてフレームが欠落してしまう可能性がある（例えば、フレーム・レベル・レート制御の場合である）。一方、より大きなQP値は、より粗い符号化を採用するが、より小さな符号化ビットストリームが生成されることになる。

30

【0034】

[0040] 幾つかの更なる実装では、符号化パラメータはラムダとして実装されてもよい。エージェント202により出力される行動 a_t は、エンコーダが理解することが可能なラムダにマッピングされることが可能である。例えば、マッピングは次式により実装されてもよい：

40

【0035】

【数2】

$$\lambda_{cur} = e^{\ln(\lambda_{min}) + (\ln(\lambda_{max}) - \ln(\lambda_{min})) * action} \quad (2)$$

50

[0041] ここで、 λ_{max} 及び λ_{min} はそれぞれ最大及び最小のラムダを表し、 λ_{cur} はエンコーダ204により使用されることになるラムダを表す。このマッピング関数は、ラムダの対数ドメインで直線的に振る舞う。上述のマッピング関数に加えて又はその代わりに、適切な他の任意の関数が代わりにマッピングに使用される可能性がある。より低いラムダ値は、より繊細な方法で符号化を制御し、より高い再構築品質を獲得する。しかしながら、より大きな符号化されたビットストリームを生成する可能性があり、バッファは簡単にオーバーシュートさせられる可能性がある一方、より高いラムダ値は、より粗い符号化を採用するが、より小さな符号化ビットストリームが生成されることになる。

【0036】

[0042] 引き続き図2を参照すると、強化学習モジュール200の訓練において、エージェント202により行われる行動がどの程度良いかを評価する必要がある。この目的のために、エンコーダ204が行動 a_t で各ビデオ・ユニットを符号化することを終えた後に、報酬 r_t が提供される。エージェント202が或る数量の訓練サンプルを獲得した場合、エージェント202は、報酬 r_t に基づいてそのポリシー（又は方策）を更新することができる。エージェント202は、蓄積された報酬を最大化することが可能な方向に向かって収束するように訓練されることが可能である。より良いQOEを得るために、QOEを反映する1つ以上の要因が報酬に組み込まれることが可能である。例えば、報酬 r_t は、バッファ・オーバーシュートにはペナルティを科し、符号化パラメータがより高い視覚的品質をもたらす場合には増加するように設定される。例えば、視覚的品質は、量子化パラメータ又はラムダが減少するにつれて増加する。

【0037】

[0043] 一例では、報酬 r_t は次のようにして計算されてもよい：

【0038】

【数3】

$$r_{base} = \begin{cases} a * (QP_{max} - QP_{cur}), & Non_BufferOvershoot \\ b, & BufferOvershoot \end{cases}$$

$$r_{base} = \begin{cases} a * (\ln(\lambda_{max}) - \ln(\lambda_{cur})), & Non_BufferOvershoot \\ b, & BufferOvershoot \end{cases} \quad (3)$$

$$r_{final} = \left(1 + \frac{Bandwidth_{cur}}{Bandwidth_{max}}\right) * r_{base} \quad (4)$$

[0044] ここで、 a は定数因子であり、 b は負数であり、 r_{base} は基本報酬であり、 $Bandwidth_{cur}$ は、時間ステップ t においてビットストリームを送信するチャンネルの帯域幅を表し、 $Bandwidth_{max}$ は最大帯域幅を表し、 r_{final} は最終的な報酬を表す。

【0039】

[0045] 基本報酬 r_{base} は数式(3)によって計算される。例えば、より高い視覚的品質は、より高いQOEを、特に画面コンテンツ共有のシナリオにおいてもたらす可能性がある。従って、より高い視覚的品質を達成するために、より小さなQP又はラムダを使用することが望ましく、また、数式(3)に示されるように、現在の量子化パラメータ QP_{cur} が減少するにつれて報酬は増加する。しかしながら、非常に小さなQP値は、大きなビットストリーム・サイズをもたらす可能性があり、それは容易にバッファ・オーバーシュートを招き、その結果、フレーム・レベルのレート制御に関してフレームを欠落させてしまう可能性がある。従って、バッファ・オーバーシュートの場合に、報酬は負数（即ち、 b ）として設定される。負数をペナルティとして設定することは、バッファのオーバーシュートを回避するように、エージェント202を訓練するために使用される。

10

20

30

40

50

【 0 0 4 0 】

[0046] r_{base} を計算した後、例えば数式(4)に示されるように、最終的な報酬 r_{final} は、スケーリング r_{base} により求めることが可能である。スケーリング因子は、時間ステップ t における帯域幅の、最大帯域幅に対する比率に関連している。時点 t における帯域幅が大きい場合、報酬 r_t はより大きな値にスケールされ、また、バッファ・オーバーシュートが起こる場合にはペナルティもより大きくなる。広帯域幅の条件下ではより良い視覚的品質を追求することはより積極的である一方、バッファ・オーバーシュートの発生についてはより深刻である。本件で説明される実装の精神から逸脱することなく、報酬を計算するために、適切な他の任意の機能が代わりに使用される可能性がある、ということに留意されたい。

10

【 0 0 4 1 】

[0047] 一部の实装では、プロキシマル・ポリシー最適化(Proximal Policy Optimization, PPO)アルゴリズムが、報酬 r_t に基づいてエージェント202を訓練するために採用されてもよい。PPOは、アクター・クリティック・アーキテクチャ(actor-critic architecture)に基づいて実装され、これは、俳優に関するアクター・ネットワークと批評家に関するクリティック・ネットワークとを含む。アクターはエージェント202として動作する。アクター・ネットワークへの入力は符号化状態であり、アクター・ネットワークの出力は行動である。アクター・ネットワークは、ポリシー $\pi(a_t | s_t)$ を推定するように構成され、ここで、 π はポリシー・パラメータ(例えば、アクター・ネットワークにおけるウェイト)を表し、 a_t 、 s_t はそれぞれ時間ステップ t における行動と符号化状態を表す。批評家のクリティック・ネットワークは、符号化状態 s_t がどの程度良いかを評価するように設定されており、訓練プロセスの間に動作するだけである。

20

【 0 0 4 2 】

[0048] PPOアルゴリズムでは、以下のように、ポリシー損失 L_{policy} がアクターを更新するために使用される可能性があり、価値損失 L_{value} がクリティックを更新するために使用される可能性がある:

【 0 0 4 3 】

【数4】

$$L_{value} = [\sum_{i=t}^{\infty} \gamma^{i-t} r_i - V_{\theta}(s_t)]^2 \quad (5)$$

30

$$L_{policy} = \min\left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)} A_t, \text{clip}\left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)}, 1 - \epsilon, 1 + \epsilon\right) A_t\right) \quad (6)$$

[0049] ここで、価値損失は、

【 0 0 4 4 】

【数5】

$$\sum_{i=t}^{\infty} \gamma^{i-t} r_i - V_{\theta}(s_t)$$

40

の二乗として計算され、 $\gamma^{i-t} r_i$ は割引報酬(discounted reward)であり(γ は割引を表す)、 $V_{\theta}(s_t)$ は入力符号化状態 s_t に対してクリティックにより生成された評価価値であり、 V_{θ} は価値関数を表す。強化学習において、価値関数は、エージェントの状態がどの程度良いかを表す。 A_t は、時間ステップ t におけるアドバンテージ関数の推定を表し、これは、

【 0 0 4 5 】

【数6】

50

$$\sum_{i=t}^{\infty} \gamma^{i-t} r_i - V_{\theta}(s_t)$$

として計算され、即ち、与えられた状態-行動のペアと、エージェントの状態の価値関数との差分差として計算される。 θ は確率的ポリシー・パラメータを表し、 $V_{\theta, \text{old}}$ は更新前のポリシー・パラメータを表す。Clip()はクリップ関数を表し、 β はハイパーパラメータを表す。適切な任意の変形を損失関数に適用することが可能であることに留意されたい。

10

【0046】

[0050] 強化学習モジュール200で使用される符号化状態は、エージェント202に関する軽量ネットワーク・アーキテクチャと、エージェント202を訓練するための軽量ネットワーク・アーキテクチャとを使用可能にする。例えば、エージェント202を実装するニューラル・ネットワークは、符号化状態 s_t から特徴を抽出するように構成された1つ以上の入力全結合層(input fully connected layers)を含んでもよい。抽出された特徴は、時間的な特徴又は特徴からの相関を抽出するために、1つ以上のリカレント・ニューラル・ネットワークに提供されてもよい。次いで、特徴は、例えば行動 a_t を生成する決定を行うために、1つ以上の出力全結合層(output fully connected layers)に提供されてもよい。リカレント・ニューラル・ネットワークは、例えば、ゲート付きリカレント・ユニット(gated recurrent unit, GRU)又は長期短期メモリ(long-short term memory, LSTM)であってもよい。ニューラル・ネットワークは、軽量ではあるが効率的なアーキテクチャを有し、リアル・タイム・アプリケーションの要求、特にスクリーン・コンテンツ符号化(screen content coding, SCC)の要求を満たすことができる。

20

【0047】

[0051] 図3は、本件で説明される対象事項の実装に従ってエージェント202を訓練するためのニューラル・ネットワーク300の一例を示す。ニューラル・ネットワーク300は、アクター・ネットワーク302とクリティック・ネットワーク304を含む。アクター・ネットワーク302とクリティック・ネットワーク304は、最適化されるべきパラメータを低減するために、共通するネットワーク・モジュールを共有してもよい。この例では、入力は2つの全結合(FC)層を通過し、特徴ベクトルに変換される。図3では、リーキー正規化線形ユニット(leaky Rectified Linear Unit, RELU)が示されているが、適切な他の任意の活性化関数がネットワークで使用されてもよい、ということが理解されるべきである。

30

【0048】

[0052] レート制御は時系列問題であることを考慮して、2つのゲート付きリカレント・ユニット(GRU)が導入され、履歴情報と組み合わせて特徴を更に抽出している。適切な他の任意のリカレント・ニューラル・ネットワークが同様に使用可能であることが理解されるべきである。GRUより後に、アクター及びクリティック・ネットワークは、各自のネットワーク・モジュールを持ち始める。アクター及びクリティックの両方が、それぞれFC層を用いて特徴ベクトルの次元を低減する。最終的に、両ネットワークは1つのFC層を用いて各自それぞれの出力を生成し、アクター・ネットワークではシグモイド層(sigmoid layer)を使用して行動の範囲を[0,1]に正規化している。適切な他の任意の活性化関数がシグモイド関数の代わりに使用されてもよいことが理解されるべきである。

40

【0049】

[0053] ニューラル・ネットワーク300は、リアル・タイム・アプリケーションの要求を満たすために、軽量であるが効率的なアーキテクチャを有する。スクリーン・コンテンツ符号化(SCC)に関し、強化学習ベースの解決策は、従来のルール・ベースのレート制御法と比較して、無視できる程度のドロップ・レートの変化を伴って、より良い視覚的品質を達成することができる。特に、この方法は、スクリーン・コンテンツにおいて急激な

50

シーン変化が発生した後に、極めて速やかに品質改善をもたらすことが可能である。強化学習ネットワーク・ベースのアーキテクチャは、如何なるコーデックにも限定されず、種々の異なるコーデック、例えば、H.264、HEVC、及びAV1と協働することが可能である。

【0050】

[0054] 図4は、本件で説明される対象事項の実装による強化学習ベースのレート制御方法400のフローチャートを示す。方法400は、演算デバイス100によって、例えば演算デバイス100内の強化学習モジュール122によって実現されてもよい。また、方法400は、任意の他のデバイス、デバイスのクラスタ、又は演算デバイス100に類似する分散並列システムによって実装されてもよい。説明の目的で、方法400は図1を参照しながら説明される。

10

【0051】

[0055] ブロック402において、演算デバイス100は、ビデオ・エンコーダの符号化状態を決定する。符号化状態は、ビデオ・エンコーダによって第1のビデオ・ユニットを符号化することに関連付けられてもよい。ビデオ・エンコーダは、リアル・タイム通信のためのスクリーン・コンテンツを符号化するように構成されていてもよい。例えば、ビデオ・エンコーダは、図2に示されるように、強化学習モジュール200内にあるようなエンコーダ204であってもよい。第1のビデオ・ユニットを符号化することに関連付けられている符号化状態は：少なくとも第1のビデオ・ユニットを符号化することに関する成果を表現する状態；ビデオ・エンコーダにより符号化されたビデオ・ユニットを送信前にバッファリングするように構成されたバッファの状態；及び符号化されたビデオ・ユニットを送信するためのネットワークの状況に関連付けられた状態を含む。ビデオ・ユニットは、フレーム、ブロック、又はフレーム中のマクロブロックを含んでもよい。一部の実装において、第1のビデオ・ユニットを符号化することに関する成果は、符号化された第1のビデオ・ユニットの符号化パラメータと符号化された第1のビデオ・ユニットのサイズとを含み、バッファの状態はバッファの使い方を含み、ネットワークの状況に関連付けられた状態は目標ビット・パー・ピクセルを含む。一部の実装では、バッファの使い方は：バッファの最大スペースに対する占有スペースの比率；及びビデオ・ユニットにおいて測定されたバッファの残存スペース；のうちの少なくとも1つを含む。

20

【0052】

[0056] ブロック404において、演算デバイス100は、強化学習モデルによって、ビデオ・エンコーダの符号化状態に基づいて、ビデオ・エンコーダのレート制御に関連する符号化パラメータを決定する。符号化パラメータは、量子化パラメータ又はラムダであってもよい。一部の実装では、符号化パラメータは、ビデオ・エンコーダの符号化状態に基づいて、エージェントにより出力された行動に基づいて決定される。エージェントは、強化学習モデルを実装するニューラル・ネットワークを含んでもよく、エージェントにより出力される行動は、符号化パラメータにマッピングされる。

30

【0053】

[0057] ブロック406において、演算デバイス100は、符号化パラメータに基づいて、第1のビデオ・ユニットとは異なる第2のビデオ・ユニットを符号化する。第1のビデオ・ユニットは第1のフレームであってもよく、第2のビデオ・ユニットは第1のフレームに続く第2のフレームであることが可能である。あるいは、第1のビデオ・ユニットは第1のブロックであってもよく、第2のビデオ・ユニットは近辺の第2のブロック、例えば、空間的に近辺のブロック又は時間的に近辺のブロックであってもよい。

40

【0054】

[0058] 一部の実装では、強化学習モデルは、第2のビデオ・ユニットの符号化に基づいて、符号化パラメータに関する報酬を決定するステップにより訓練されており、ここで、報酬は、バッファのオーバーシュートにはペナルティを科し、且つ符号化パラメータがより高い視覚的品質をもたらす場合には増加するように設定されている。

【0055】

[0059] 一部の実装では、報酬を決定するステップは：バッファのオーバーシュートが

50

生じる場合には基本報酬が負の値を有し、バッファのオーバーシュートが生じない場合には基本報酬が符号化パラメータに負の係数で比例するような方法で、基本報酬を決定するステップ；及び基本報酬をスケーリング因子でスケーリングして報酬を得るステップを含み、スケーリング因子は、第2のビデオ・ユニットを符号化することに関連付けられる帯域幅の、送信チャネルの最大帯域幅に対する比率に基づいている。例えば、報酬は数式(3)及び(4)に基づいて計算されてもよい。

【0056】

[0060] 一部の実装では、強化学習モデルは、符号化パラメータに関連付けられた行動を、ビデオ・エンコーダの符号化状態に基づいて決定するステップ；第2のビデオ・ユニットを符号化することに関する符号化状態の評価値を決定するステップ；報酬及び評価値に基づいて、価値損失を決定するステップ；行動に基づいてポリシー損失を決定するステップ；及び価値損失及びポリシー損失に基づいて、強化学習モデルを更新するステップにより更に訓練されている。

10

【0057】

[0061] 一部の実装では、エージェントはニューラル・ネットワークを含み、ニューラル・ネットワークは：符号化状態から特徴を抽出するように構成された少なくとも1つの入力全結合層；抽出された特徴を受けよう結合された少なくとも1つのリカレント・ニューラル・ネットワーク；及びエージェントの行動を決定するように構成された少なくとも1つの出力全結合層を含む。

【0058】

[0062] 一部の実装では、ニューラル・ネットワークはアクター・クリティック・アーキテクチャに基づいて訓練されており、アクターは符号化状態に基づいて行動を生成するように構成されており、クリティックは符号化状態に関する評価値を生成するように構成されており；アクター及びクリティックは、少なくとも1つの入力全結合層を含む前記ニューラル・ネットワークと少なくとも1つのリカレント・ニューラル・ネットワークとによる共通部分を共有している。

20

【0059】

[0063] 本件で説明される対象事項の幾つかの例示的な実装を以下に列挙する。

【0060】

[0064] 第1の態様では、本件で説明される対象事項は、コンピュータで実行される方法を提供する。方法は、ビデオ・エンコーダの符号化状態を決定するステップであって、符号化状態はビデオ・エンコーダにより第1のビデオ・ユニットを符号化することに関連付けられている、ステップ；強化学習モデルにより、ビデオ・エンコーダの符号化状態に基づいて、ビデオ・エンコーダのレート制御に関する符号化パラメータを決定するステップ；及び第1のビデオ・ユニットとは異なる第2のビデオ・ユニットを、符号化パラメータに基づいて符号化するステップを含む。

30

【0061】

[0065] 一部の実装では、符号化パラメータを決定するステップは：強化学習モデルにより、ビデオ・エンコーダの符号化状態に基づいて行動を決定するステップ；及び行動を符号化パラメータにマッピングするステップを含む。

40

【0062】

[0066] 一部の実装では、第2のビデオ・ユニットは第1のビデオ・ユニットの後に続くものであり、第1のビデオ・ユニットを符号化することに関連付けられている符号化状態は：少なくとも第1のビデオ・ユニットを符号化することに関する成果を表現する状態；ビデオ・エンコーダにより符号化されたビデオ・ユニットを送信前にバッファリングするように構成されたバッファの状態；及び符号化されたビデオ・ユニットを送信するためのネットワークの状況に関連付けられた状態を含む。

【0063】

[0067] 一部の実装では、第1のビデオ・ユニットを符号化することに関する成果は、符号化された第1のビデオ・ユニットの符号化パラメータと符号化された第1のビデオ・ユ

50

ニットのサイズとを含み、バッファの状態はバッファの使い方を含み、ネットワークの状況に関連付けられた状態は目標ビット・パー・ピクセルを含む。

【0064】

[0068] 一部の実装では、バッファの使い方は：バッファの最大スペースに対する占有スペースの比率；及びビデオ・ユニットにおいて測定されたバッファの残存スペース；のうちの少なくとも1つを含む。

【0065】

[0069] 一部の実装では、強化学習モデルは、第2のビデオ・ユニットの符号化に基づいて、符号化パラメータに関する報酬を決定するステップにより訓練されており、報酬は、バッファのオーバーシュートにはペナルティを科し、且つ符号化パラメータがより高い視覚的品質をもたらす場合には増加するように設定されている。

10

【0066】

[0070] 一部の実装では、報酬を決定するステップは：バッファのオーバーシュートが生じる場合には基本報酬が負の値を有し、バッファのオーバーシュートが生じない場合には基本報酬が符号化パラメータに負の係数で比例するような方法で、基本報酬を決定するステップ；及び基本報酬をスケーリング因子でスケーリングして報酬を得るステップであって、スケーリング因子は、第2のビデオ・ユニットを符号化することに関連付けられる帯域幅の、送信チャネルの最大帯域幅に対する比率に基づいている、ステップを含む。

【0067】

[0071] 一部の実装では、強化学習モデルは、符号化パラメータに関連付けられた行動を、ビデオ・エンコーダの符号化状態に基づいて決定するステップ；第2のビデオ・ユニットを符号化することに関する符号化状態の評価価値を決定するステップ；報酬及び評価価値に基づいて、価値損失を決定するステップ；行動に基づいてポリシー損失を決定するステップ；及び価値損失及びポリシー損失に基づいて、強化学習モデルを更新するステップにより更に訓練されている。

20

【0068】

[0072] 一部の実装では、強化学習モデルはエージェントのニューラル・ネットワークを含み、ニューラル・ネットワークは：符号化状態から特徴を抽出するように構成された少なくとも1つの入力全結合層；抽出された特徴を受けるように結合された少なくとも1つのリカレント・ニューラル・ネットワーク；及びエージェントの行動を決定するように構成された少なくとも1つの出力全結合層を含む。

30

【0069】

[0073] 一部の実装では、ニューラル・ネットワークはアクター・クリティック・アーキテクチャに基づいて訓練されており、アクターは符号化状態に基づいて行動を生成するように構成されており、クリティックは符号化状態に関する評価価値を生成するように構成されており；及びアクター及びクリティックは、少なくとも1つの入力全結合層を含むニューラル・ネットワークと少なくとも1つのリカレント・ニューラル・ネットワークとによる共通部分を共有している。

[0074] 一部の実装では、符号化パラメータは、量子化パラメータ及びラムダ・パラメータのうちの少なくとも1つを含む。

40

【0070】

[0075] 一部の実装では、ビデオ・エンコーダは、リアル・タイム通信のスクリーン・コンテンツを符号化するように構成されている。

【0071】

[0076] 第2の態様では、本件で説明される対象事項は、電子デバイスを提供する。電子デバイスは、処理ユニットと処理ユニットに結合され命令を記憶したメモリとを備え、命令は、処理ユニットによって実行されると、上記方法の任意のステップを電子デバイスに実行させる。

【0072】

[0077] 第3の態様では、本件で説明される対象事項は、コンピュータ記憶媒体に実体

50

的に記憶され且つ機械実行可能な命令を含むコンピュータ・プログラム製品を提供し、命令は、デバイスによって実行されると、第1の態様における態様による方法をデバイスに実行させる。コンピュータ記憶媒体は非一時的なコンピュータ記憶媒体であってもよい。

【0073】

[0078] 第4の態様では、本件で説明される対象事項は、機械実行可能な命令を記憶した非一時的なコンピュータ記憶媒体を提供し、命令は、デバイスによって実行されると、第1の態様の態様による方法をデバイスに実行させる。

【0074】

[0079] 本件で説明される機能は、少なくとも部分的に、1つ以上のハードウェア論理コンポーネントによって実行されることが可能である。例えば、限定されるものではないが、使用することが可能なハードウェア論理コンポーネントの例示的なタイプは、フィールド・プログラマブル・ゲート・アレイ (FPGA)、特定用途向け集積回路 (ASIC)、特定用途向け標準製品 (ASSP)、システム・オン・チップ・システム (SOC)、複合プログラマブル・ロジック・デバイス (CPLD) 等を含む。

【0075】

[0080] 本件で説明される対象事項の方法を実行するためのプログラム・コードは、1つ以上のプログラミング言語の任意の組み合わせで書かれてもよい。プログラム・コードは、汎用コンピュータ、専用コンピュータ、又はその他のプログラマブル・データ処理装置のプロセッサ又はコントローラに提供されてもよく、その結果、プログラム・コードは、プロセッサ又はコントローラによって実行されると、フローチャート及び/又はブロック図に示される機能/動作が実行されることを引き起こす。プログラム・コードは、全体的又は部分的にマシン上で実行されてもよいし、部分的にマシン上で、部分的にリモート・マシン上で、又は完全にリモート・マシン又はサーバー上でスタンド・アロン・ソフトウェア・パッケージとして実行されてもよい。

【0076】

[0081] 本件で説明される対象事項の文脈では、機械読み取り可能な媒体は、命令実行システム、装置、又はデバイスによって、又はそれらに関連して使用するためのプログラムを含む又は記憶することが可能な任意の有形媒体である可能性がある。機械読み取り可能な媒体は、機械読み取り可能な信号媒体又は機械読み取り可能な記憶媒体であってもよい。機械読み取り可能な媒体は、電子的、磁氣的、光学的、電磁的、赤外線的、もしくは半導体的なシステム、装置、デバイス、又は前述したものの任意の適切な組み合わせを含むが、これらに限定されない。機械読み取り可能な記憶媒体のより具体的な例は、1つ以上のワイヤを含む電気接続、ポータブル・コンピュータ・ディスクレット、ハード・ディスク、ランダム・アクセス・メモリ (RAM)、リード・オンリー・メモリ (ROM)、消去可能プログラマブル・リード・オンリー・メモリ (EPROM又はフラッシュ・メモリ)、光ファイバ、ポータブル・コンパクト・ディスク・リード・オンリー・メモリ (CD-ROM)、光記憶デバイス、磁気記憶デバイス、又は前述したものの適切な任意の組み合わせを含む。

【0077】

[0082] 更に、動作は特定の順序で説明されているが、これは、このような動作が図示された特定の順序で又は連続順で実行されること、又は所望の結果を達成するために、図示された全ての動作が実行されること、を要求するものとして理解されるべきではない。特定の状況下では、マルチタスク及び並列処理が有利である可能性がある。同様に、幾つかの特定の実装の詳細が上記の説明に含まれているが、これらは、本件で説明される対象事項の範囲に対する制限としてではなく、むしろ特定の实装に特異的である可能性のある特徴の説明として解釈されるべきである。別々の実装の文脈で説明される特定の特征是、単一の実装の中で組み合わせて実装されてもよい。むしろ、単一の実装で説明された種々の特征是、複数の実装で別々に、又は任意の適切なサブ・コンビネーションで実装されてもよい。

【0078】

10

20

30

40

50

[0083] 対象事項は、構造的特徴及び/又は方法論的動作に特有の言葉で説明されているが、添付のクレームで特定される対象事項は、必ずしも上記の特定の特徴や動作に限定されない、ということが理解されるべきである。むしろ、上述の特定の特徴や動作は、クレームを実施する際の例示的な形態として開示されている。

【図面】

【図 1】

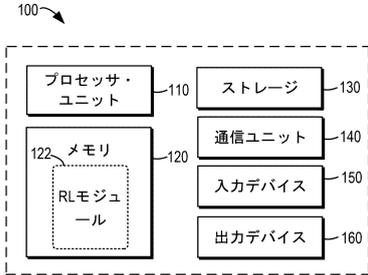


FIG. 1

【図 2】

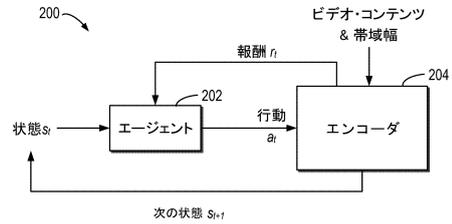


FIG. 2

【図 3】

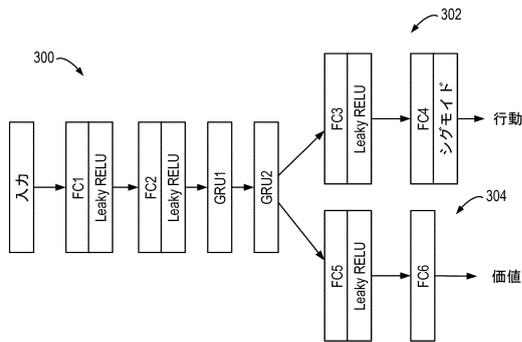


FIG. 3

【図 4】

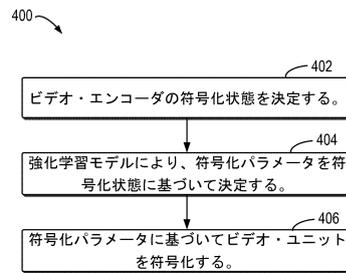


FIG. 4

10

20

30

40

50

フロントページの続き

- ート ナンバー 5 ティー 2 - 1 4 1 1 2
- (72)発明者 リー, ピン
中国 1 0 0 0 8 0 ベイジン ハイディエン ディストリクト ダンリン ストリート ナンバー 5
ティ 2 - 1 4 1 1 5
- (72)発明者 ルー, ヤン
中国 1 0 0 0 8 9 ベイジン ハイディエン ディストリクト ワンリウ グアン ダ シー ユアン
1 - 6 - 0 6 8 8
- (72)発明者 ホルコム, ダブリュ. トム
アメリカ合衆国 9 8 0 7 4 ワシントン州 サマミッシュ ノースイースト 4 4 ス プレイス 1 9
6 5 6
- (72)発明者 ルー, メイシュアン
台湾 1 0 6 タイペイ ダーアン ディストリクト ルーズベルト・ロード セクション 3 レーン 2
8 3 アリー 1 4 ナンバー 1 4 4 エフ - 3
- (72)発明者 メゼンツェフ, アンドレイ
アメリカ合衆国 9 8 0 5 2 ワシントン州 レッドモンド ノースイースト 8 3 ウェイ 1 5 8 9 0
- (72)発明者 リー, ミンチェ
アメリカ合衆国 9 8 0 0 5 ワシントン州 ベルビュー ノースイースト 1 3 4 ス アヴェニュー
3 8 5 2
- 審査官 片岡 利延
- (56)参考文献 特開 2 0 0 1 - 2 3 1 0 3 9 (J P , A)
特表 2 0 0 6 - 5 2 4 4 6 1 (J P , A)
米国特許出願公開第 2 0 0 6 / 0 1 9 2 8 5 0 (U S , A 1)
Mingliang Zhou(他 4 名), "Rate Control Method Based on Deep Reinforcement Learning for
Dynamic Video Sequences in HEVC", IEEE Transactions on Multimedia , Volume: 23 , IEE
E , 2020年05月06日 , P1106-P1121
- (58)調査した分野 (Int.Cl. , D B 名)
H 0 4 N 1 9 / 0 0 - 1 9 / 9 8