

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4480713号
(P4480713)

(45) 発行日 平成22年6月16日(2010.6.16)

(24) 登録日 平成22年3月26日(2010.3.26)

(51) Int.Cl.

HO4N 7/26 (2006.01)

F 1

HO4N 7/13

Z

請求項の数 22 (全 15 頁)

(21) 出願番号	特願2006-500826 (P2006-500826)
(86) (22) 出願日	平成16年1月7日(2004.1.7)
(65) 公表番号	特表2006-516375 (P2006-516375A)
(43) 公表日	平成18年6月29日(2006.6.29)
(86) 國際出願番号	PCT/US2004/000316
(87) 國際公開番号	W02004/064414
(87) 國際公開日	平成16年7月29日(2004.7.29)
審査請求日	平成19年1月4日(2007.1.4)
(31) 優先権主張番号	60/439,062
(32) 優先日	平成15年1月8日(2003.1.8)
(33) 優先権主張国	米国(US)
(31) 優先権主張番号	10/614,929
(32) 優先日	平成15年7月7日(2003.7.7)
(33) 優先権主張国	米国(US)

(73) 特許権者	503260918 アップル インコーポレイテッド アメリカ合衆国 95014 カリフォルニア州 クパチーノ インフィニット ループ 1
(74) 代理人	100064621 弁理士 山川 政樹
(74) 代理人	100098394 弁理士 山川 茂樹
(72) 発明者	ドゥミトラス、アドリアーナ アメリカ合衆国、94087 カリフォルニア州、サンマテオ、アパートメント 2403、ダブリュー、エル 250 キャミーノ リアル

早期審査対象出願

最終頁に続く

(54) 【発明の名称】 改良型符号化モード選択の方法および装置

(57) 【特許請求の範囲】

【請求項 1】

ビデオエンコーダでモード選択を実行する方法であって、
 一連の符号化モードから複数の符号化モードを用いて符号化するステップと、
 前記複数の符号化モードから各符号化モードごとに歪値を計算するステップと、
 前記複数の符号化モードから各符号化モードごとにビット・レート値を計算するステップと、

前記歪値、前記ビット・レート値、およびラグランジュ乗数を用いて前記複数の符号化モードから、各符号化モードごとにラグランジアン値を計算するステップと、

符号化モードのサブセットを定めるために最小ラグランジアン値の閾値の範囲内にある複数の前記ラグランジアン値をクラスタ化するステップと、

前記符号化モードのサブセットから 1 つの符号化モードを選択するステップとから構成され、

前記符号化モードのサブセットの各符号化モードは、前記クラスタ化された複数のラグランジアン値からの 1 つのラグランジアン値と関連しており、

前記符号化モードのサブセットがモード 0 符号化を含んでいる時に、モード 0 符号化は選択され、そして、このモード 0 符号化は、動きベクトル情報を送信しない符号化モードであることを特徴とする方法。

【請求項 2】

前記ビット・レート値を計算するステップは、動きベクトル式および変換係数式を

10

20

符号化するのに必要な総ビット数を計算するステップを有することを特徴する請求項 1 に記載の方法。

【請求項 3】

前記クラスタは、複数の計算されたラグランジアン値の前記最小ラグランジアン値を含むことを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記方法は、前記モード 0 符号化に関連した前記ラグランジアン値が前記クラスタ内にない場合に、前記最小ラグランジアン値を生成する符号化モードを選択するステップをさらに有することを特徴とする請求項 1 に記載の方法。

【請求項 5】

前記モード 0 は直接モードであることを特徴とする請求項 1 に記載の方法。

【請求項 6】

前記モード 0 はスキップ・モードであることを特徴とする請求項 1 に記載の方法。

【請求項 7】

少なくとも 1 つのプロセッサによってモード選択を行うためのコンピュータ・プログラムが実行される時に、前記少なくとも 1 つのプロセッサに、

一連の符号化モードから複数の符号化モードを用いて符号化を行う動作と、

前記複数の符号化モードから各符号化モードごとに歪値を計算する動作と、

前記複数の符号化モードから各符号化モードごとにビット・レート値を計算する動作と、

前記歪値、前記ビット・レート値およびラグランジュ乗数を用いて、前記複数の符号化モードから各符号化モードごとにラグラジアン値を計算する動作と、

符号化モードのサブセットを定めるために最小ラグランジアン値の閾値の範囲内にある複数の前記ラグランジアン値をクラスタ化する動作と、

前記符号化モードのサブセットから 1 つの符号化モードを選択する動作とを実行させるための命令を含み、

前記符号化モードのサブセットは、クラスタ化されたラグラジアン値と関連しており、

前記符号化モードのサブセットがモード 0 符号化を含んでいる時に、モード 0 符号化は選択され、そして、このモード 0 符号化は、動きベクトル情報を送信しない符号化モードである、前記コンピュータ・プログラムを格納するコンピュータ可読媒体。

【請求項 8】

前記ビット・レート値を計算する命令が実行される時に、前記少なくとも 1 つのプロセッサに、動きベクトル式および変換係数式を符号化するのに必要な総ビット数を計算する命令を実行させることを特徴とする請求項 7 に記載のコンピュータ可読媒体。

【請求項 9】

前記クラスタは、複数の計算されたラグランジアン値の前記最小ラグランジアン値を含むことを特徴とする請求項 7 に記載のコンピュータ可読媒体。

【請求項 10】

前記モード 0 符号化に関連したラグランジアン値が、前記クラスタ内にない場合に、前記少なくとも 1 つのプロセッサに、前記最小ラグランジアン値を生成する符号化モードを選択する動作を実行させる命令をさらに含む請求項 7 に記載のコンピュータ可読媒体。

【請求項 11】

前記モード 0 は直接モードであることを特徴とする請求項 7 に記載のコンピュータ可読媒体。

【請求項 12】

前記モード 0 はスキップ・モードであることを特徴とする請求項 7 に記載のコンピュータ可読媒体。

【請求項 13】

モード 0 符号化と非モード 0 符号化を含む複数の符号化モードから 1 つの符号化モード

10

20

30

40

50

を選択する方法であって、この方法は、

前記複数の符号化モードから各特定の符号化モードごとに、歪みとビット・レート間の関係を示すラグランジアン値を計算するステップと、

最小ラグランジアン値の閾値の範囲内にある一連のラグランジアン値に基づいて、前記複数の符号化モードから符号化モードのサブセットを特定するステップと、

前記符号化モードのサブセットから、前記モード0符号化と関連した前記ラグランジアン値が前記最小ラグランジアン値でない場合であっても、前記符号化モードのサブセットが前記モード0符号化を含む時、このモード0符号化を選択するステップと、を含み、

前記モード0符号化は、動きベクトル情報を送信しない符号化モードである、前記方法。

10

【請求項14】

前記モード0符号化は、直接モード符号化である請求項13に記載の方法。

【請求項15】

前記モード0符号化は、スキップ・モード符号化である請求項13に記載の方法。

【請求項16】

前記モード0符号化が前記符号化モードのサブセットにない場合、前記最小ラグランジアン値に関連した1つの符号化モードを前記符号化モードのサブセットから選択するステップをさらに含む請求項13に記載の方法。

【請求項17】

前記ラグランジアン値を計算するステップは、歪値、ビット・レート値およびラグランジュ乗数に基づいて、前記ラグランジアン値を計算することを含む請求項13に記載の方法。

20

【請求項18】

少なくとも1つのプロセッサによってモード0符号化と非モード0符号化を含む複数の符号化モードから1つの符号化モードを選択するためのコンピュータ・プログラムが実行される時に、前記少なくとも1つのプロセッサに、

前記複数の符号化モードから各特定の符号化モードごとに、歪みとビット・レート間の関係を示すラグランジアン値を計算する動作と、

最小ラグランジアン値の閾値の範囲内にある一連のラグランジアン値に基づいて、前記複数の符号化モードから符号化モードのサブセットを特定する動作と、

30

前記符号化モードのサブセットから、前記モード0符号化と関連した前記ラグランジアン値が前記最小ラグランジアン値でない場合であっても、前記符号化モードのサブセットが前記モード0符号化を含む時、このモード0符号化を選択する動作とを実行させるための命令を含む前記コンピュータ・プログラムを格納するコンピュータ可読媒体。

【請求項19】

前記モード0符号化は、直接モード符号化である請求項18に記載のコンピュータ可読媒体。

【請求項20】

前記モード0符号化は、スキップ・モード符号化である請求項18に記載のコンピュータ可読媒体。

40

【請求項21】

前記モード0符号化が前記符号化モードのサブセットにない場合、前記少なくとも1つのプロセッサに、前記最小ラグランジアン値に関連した1つの符号化モードを前記符号化モードのサブセットから選択する動作を実行させる命令をさらに含む請求項18に記載のコンピュータ可読媒体。

【請求項22】

前記ラグランジアン値を計算する命令が実行される時に、前記少なくとも1つのプロセッサに、歪値、ビット・レート値およびラグランジュ乗数に基づいてラグランジアン値を計算する動作を実行させる命令を含む請求項18に記載のコンピュータ可読媒体。

【発明の詳細な説明】

50

【技術分野】**【0001】**

(関連出願)

本特許出願は、2003年1月8日出願の「Method and Apparatus for Improved Coding Mode Selection」(改良型符号化モード選択の方法および装置)と題された米国仮特許出願第60/439062号の利益を主張する。

(技術分野)

本発明は、マルチメディア圧縮システムの分野に関する。特に、本発明は、符号化モード選択を改良する方法およびシステムを開示する。

10

【背景技術】**【0002】**

デジタル・ベースの電子媒体フォーマットはついに大幅にアナログ電子媒体フォーマットに取って代わる段階に達した。デジタル・コンパクト・ディスク(CD)はかなり以前にアナログのビニール製レコードに取って代わっている。アナログ磁気カセット・テープはますます稀になっている。ミニディスクおよびMP3(MPEGオーディオ・レイヤ3)などの第2世代および第3世代デジタル・オーディオ・システムはコンパクト・ディスクの第1世代デジタル・オーディオ・フォーマットから市場シェアを奪いつつある。

ビデオ媒体はオーディオ媒体と比較してデジタル記憶装置および伝送フォーマットへの移行が遅れている。このデジタルへの移行の遅さの理由は、主として画像をデジタル形式で正確に表現するのに必要な大量のデジタル情報が原因である。画像を正確に表現するのに必要な大量のデジタル情報は、極めて大容量のデジタル記憶システムおよび高帯域伝送システムを必要とする。

20

しかし、画像はデジタル記憶装置および伝送フォーマットに急速に移行しつつある。より高速のコンピュータ・プロセッサ、高密度記憶システム、および新しい効率的な圧縮および符号化アルゴリズムによって、デジタル・ビデオ伝送および記憶装置は消費者価格にまで引き下げられた。デジタル・ビデオ・システムであるDVD(デジタル多用途ディスク)は、数年にわたって最もよく売れている消費者向け電子製品の1つである。DVDは、その高品質の画像、極めて高品質の音響、利便性および他の特徴によって、好んで選ばれる事前録画されたビデオ再生システムとしてビデオカセット・レコーダ(VCR)に急速に取って代わっている。古くなったアナログNTSC(全米テレビジョン方式委員会)のビデオ伝送システムは現在、デジタルATSC(高度テレビジョン方式委員会)のビデオ伝送システムに移行しつつある。

30

【0003】

コンピュータ・システムは数年間、さまざまな異なるデジタル・ビデオ符号化フォーマットを使用してきた。コンピュータ・システムによって使用してきた最良のデジタル・ビデオ圧縮および符号化システムには、頭字語MPEGで一般に知られる動画映像専門家グループ(Motion Pictures Expert Group)が支援するデジタル・ビデオ・システムが含まれる。MPEGの3つの最も知られ、使用されているデジタル・ビデオ・フォーマットは、MPEG-1、MPEG-2、およびMPEG-4である。ビデオCDおよび消費者グレードのデジタル・ビデオ編集システムは初期のMPEG-1フォーマットを使用する。デジタル多用途ディスク(DVD)およびデイシッシュ・ネットワーク・ブランドの直接放送衛星(DBS)テレビジョン放送システムはMPEG-2デジタル・ビデオ圧縮および符号化システムを使用する。MPEG-4符号化システムは、最新のコンピュータ・ベースのデジタル・ビデオ・エンコーダおよびそれに関連するデジタル・ビデオ・プレーヤに急速に採用されている。

40

【発明の開示】**【0004】**

符号化モード選択を改良する方法およびシステムが開示される。本開示では、H.26

50

4 (MPEG - 4 / Part 10) の枠内のB画像の直接モード拡張およびP画像のスキップ・モード拡張の新規の方法が開示される。

直接モードおよびスキップ・モード拡張は、既存の圧縮方式にいくつかの変更を加えることで達成される。特に、本発明のシステムは歪値内のアウトライアを除去するステップと、符号化モード選択の速度歪最適化のラグランジュ乗数のより小さい値を指定するステップと、符号化モード選択の前にラグランジアン値をクラスタ化するステップとを導入する。一実施形態では、Huberコスト関数を使用して異なる符号化モードでの歪を計算し、アウトライアを除去する。本発明の一実施形態では、システムはラグランジュ乗数を参考のH.264 (MPEG - 4 / Part 10) の実施態様よりも量子化値Qの関数として緩やかに変化するように変更する。ラグランジアンのクラスタ化はビット・レート低減のためのモード0符号化モードを支持するために使用される。10

高品質ビデオ・シーケンスを用いた実験結果は、ビット・レートの低減がピーク信号対雑音比 (PSNR) のわずかな損失の犠牲を払って本発明の方法を用いて得られることを示している。2つの異なる実験を行うことで、ピーク信号対雑音比が変化したにも関わらず主観的な画像損失がないということが検証された。

(非規範的) MPEG - 4 / Part 10 エンコーダで現在採用されている既存の速度歪の最適化方法に関連して、本発明の方法は簡単で有用な拡張機能を表す。より重要なことには、量子化パラメータの値をさらに増やすなどの他の解決策が適用できない時には、復号化された画像には許容不能のアーティファクトが混入するので、本発明の方法は復号化シーケンスに観察できる歪を導入することなくビット・レート低減を達成する。20

本発明の他の目的、特徴および利点は、添付の図面および下記の詳細な説明を読めば明らかになるだろう。

【0005】

(図面の簡単な説明)

本発明の目的、特徴および利点は、以下の詳細な説明を読めば当業者には明らかとなるであろう。

図1は、変数のHuberコスト関数を示すグラフである。

図2Aは、対象範囲の量子化パラメータ (Q) 値の関数である元の、および修正ラグランジュ乗数 m_{ode} の変化を示す図である。

図2Bは、対象範囲の量子化パラメータ (Q) 値の関数であるBフレームの元の、および修正ラグランジュ乗数 m_{ode} の変化を示す図である。30

図2Cは、対象範囲の量子化パラメータ (Q) 値の関数である元の、および修正ラグランジュ乗数 m_{otion} の変化を示す図である。

図3は、符号化モードの選択方法を示す流れ図である。

【発明を実施するための最良の形態】

【0006】

符号化モード選択を改良する方法およびシステムが開示される。以下の説明では、説明のために、本発明を十分に理解するための特定の用語を使用する。しかし、本発明を実施するために具体的な詳細は必要ないことは当業者には明らかであろう。

序論

MPEG - 4 / Part 10、Joint Video Team (JVT)、Advanced Video Coding (AVC) としても知られる先進のH.264 ビデオ符号化標準およびH.26Lは、既存のMPEG標準に準拠する最先端のビデオ符号化方式より高度な動画圧縮を提供する目標を掲げた動画映像専門家グループ (Motion Picture Experts Group) (MPEG) および国際電気通信連合 (ITU) によって共同開発してきた。2003年の国際標準になると思われるH.264の対象分野は、テレビ会議、ディジタル記憶媒体、テレビジョン放送、インターネット・ストリーム伝送および通信を含む（が、それに限定されない）。

他のビデオ符号化標準と同様に、（その本文または付属書で）H.264 標準は速度歪 (RD) 決定の枠組みを採用する。特に、H.264 標準は符号化モード選択および動き

10

20

30

40

50

予測の速度歪最適化を採用する。この開示では、第1の焦点はH.264標準の枠内の符号化モード選択に向けられる。

大半のディジタル・ビデオ符号化方式では、ビデオ・シーケンスの各ビデオ・フレームは画素ブロックと呼ばれる画素のサブセットに分割される。H.264標準では、画素ブロックはさまざまなサイズを有する(16×16画素に等しいサイズの画素ブロックは従来マクロブロックとして知られる)。符号化モード選択の問題は、非公式には「ビデオ・フレーム内の各画素ブロックを符号化するすべての可能な符号化方法(または符号化モード)のうち最良のものを選択する手順」と定義できる。符号化モード選択の問題はいくつかの異なった方法でビデオ・エンコーダによって解決できる。符号化モード選択の問題を解決する1つの可能な方法は速度歪最適化の使用である。

10

【0007】

H.264ビデオ符号化標準の枠内で各画素ブロックを符号化するために選択できる多数の異なる符号化モードがある。モード0はBフレーム内の「直接モード」として、またPフレーム内の「スキップ・モード」として知られる。別の符号化モードは、B画像またはP画像内の16×16、16×8および8×16画素、8×8、8×4、4×8、4×4画素に等しいサイズの画素ブロックを使用する。

直接モード(B画像のモード0)では、デコーダに動き情報は送信されない。その代わり、予測システムを用いて動き情報が生成される。したがって、直接モードは、隣接する空間または時間情報を用いる優れた動きベクトル予測を可能にするシーケンスのビット・レートの節約を提供できる。しかし、実験結果は、H.264の直接モード選択はビデオ・シーケンスによっては予測されただけの数の選択画素ブロックを生成しないことを示している。

20

本開示は、H.264標準の枠内の双方向予測画像(B画像またはBフレームとして知られる)内の直接モード(モード0)の選択を拡張する方法を提案する。Pフレームに適用された場合、本発明の符号化方法はスキップ・モード(またはモード0)選択の拡張を達成する。直接モードとスキップ・モードの拡張は、ラグランジアン値のクラスタ化と、アウトライアの除去と、符号化モード選択のための速度歪最適化の際のラグランジュ乗数のより小さい値の指定とによって達成される。

高品質サンプル・ビデオ・シーケンスを用いた実験結果は、本発明の圧縮ビットストリームのビット・レートが参考のH.264コーデックを用いて得た圧縮ビットストリームより速度が低減していることを示す。このビット・レート低減はビットストリームのピーク信号対雑音比(PSNR)のわずかな損失に関連する。しかし、2つの実験は、ピーク信号対雑音比の変化にはいかなる主観的な画像損失も対応しないことを示している。より重要なことには、復号化された画像に許容できないアーティファクトが混入するので量子化パラメータの値をさらに増やすなどの他の可能な解決策は適用できないが、本発明の方法は復号化ビデオ・シーケンス内に観察できる歪を混入させることなくさらなるビット・レート低減を有意に達成する。さらに、本発明はH.264の枠組みを利用するという事実にも関わらず、本発明の符号化方法は速度歪最適化を使用する任意のビデオ符号化システムに適用可能である。

30

本明細書の残りの構成は以下の通りである。まず、ビデオ圧縮の概要の部で、H.264標準の速度歪最適化の枠組みに関連する基本概念を述べる。次いで、提案する直接モード拡張方法の部で本発明が提案する符号化方法を詳述する。最後に、実験結果の部と結論の部でそれぞれ実験結果と結論を述べる。

40

【0008】

ビデオ圧縮の概要

本明細書で前述したように、各ビデオ・フレームはH.264標準の画素ブロックのセットに分割される。これらの画素ブロックは動き補償予測符号化を用いて符号化できる。予測画素ブロックは、その符号化のために前の画像からの情報を用いないイントラ(I)画素ブロック(I画素ブロック)、1つの前の画像の情報を用いる片方向予測(P)画素ブロック(P画素ブロック)、または1つの前の画像と1つの後の画像との情報を用いる

50

双方向予測（B）画素ブロック（B画素ブロック）である。

P画像内のP画素ブロックごとに、1つの動きベクトルが計算される。（なお、各ビデオ画像内で画素ブロックは多くの方法で符号化できる。例えば、画素ブロックはより小さいサブブロックに分割して、各サブブロックごとに動きベクトルを計算し送信できる。サブブロックの形状は変わることがあり、正方形でなくてもよい。）コンピュータの動きベクトルを用いて、上述の前の画像内の画素を平行移動して予測画素ブロックを形成できる。次に、ビデオ画像内の実際の画素ブロックと予測画素ブロックとの差が送信のために符号化される。（この差を用いて予測画素ブロックと実際の画素ブロックの小さな差を修正する。）

各動きベクトルは、また予測符号化によって送信できる。すなわち、動きベクトルの予測はすでに送信済みの近くの動きベクトルを用いて形成でき、次いで実際の動きベクトルと予測動きベクトルとの差が送信のために符号化される。

各B画素ブロックで、2つの動きベクトルが通常計算され、1つの動きベクトルは上記前の画像、1つの動きベクトルは後の画像に使用される。（ただし、P画像またはB画像内で、一部の画素ブロックは動き補償がないほうが符号化がうまくいく。そのような画素はイントラ画素ブロックとして符号化できる。B画像内で、一部の画素ブロックは前方または後方片方向動き補償があったほうが符号化がうまくいく。そのような画素は、予測に前の画像を使うか後の画像を使うかによって前方予測または後方予測として符号化できる。）2つのB画素ブロック動きベクトルから、2つの予測画素ブロックが計算される。次いで2つの予測画素ブロックは組み合わされて最終予測画素ブロックを形成する。上記のように、ビデオ画像内の実際の画素ブロックと予測ブロックとの差が送信のために符号化される。

【0009】

P画素ブロックの場合、B画素ブロックの各動きベクトルは予測符号化によって送信できる。すなわち、予測動きベクトルは送信済みの近くの動きベクトルを用いて形成できる。次いで、実際の動きベクトルと予測動きベクトルとの差が送信のために符号化される。

しかし、B画素ブロックでは、記憶された画像の同位置の、または近くの画素ブロック内の動きベクトルから動きベクトルを補間する余地がある。（動きベクトル予測が現在の画素ブロックの同位置のブロックの動きベクトルを用いて構成される時には、直接モード・タイプは時間的な直接モードとして知られる。動きベクトル予測が現在の画素ブロックの空間的に隣接する画素ブロックを用いて構成される時には、直接モード・タイプは空間的な直接モードとして知られる。）次に、補間された値は予測動きベクトルと、送信のために符号化された実際の動きベクトルと予測動きベクトルとの差として使用できる。そのような補間はエンコーダとデコーダの両方で実行される。（ただし、エンコーダは常にデコーダを有するので、エンコーダは再構成されたビデオ画像の表示状態を正確に認識できる。）

場合によっては、補間された動きベクトルはいかなる修正差もなしに十分に使用できる。その場合、全く動きベクトル・データを送信する必要はない。これはH.264（およびH.263）標準では直接モードと呼ばれている。直接モード選択は、記録用カメラが固定した背景を低速でパンする時に特に有効である。実際、動きベクトル補間はそのまま十分に使用できる。すなわち、これらのB画素ブロック動きベクトルについて差分情報を送信する必要がない。スキップ・モード（P画像のモード0）では、動きベクトル予測は 16×16 直接モードと同じに構成され、動きベクトル・ビットの送信は実行されない。

送信に先立って、画素ブロックまたはサブブロックの予測誤差（差）が、通常、変換され、量子化され、エントロピー符号化されてビット数が減らされる。予測誤差は、元の所望の画素ブロックと直接モードによる符号化後の復号化された予測画素ブロックとの平均平方誤差として計算され、直接モードで符号化される。しかし、予測誤差はスキップ・モードで符号化および送信されない。変換に使用されるサブブロックのサイズと形状とは動き補償に使用されるサブブロックのサイズと形状と同じではない。例えば、 8×8 の画素または 4×4 の画素は一般に変換に使用されるが、 16×16 の画素、 16×8 の画素、

10

20

30

40

50

8×16 の画素、またはそれより小さいサイズは動き補償に一般に使用される。動き補償および変換サブブロックのサイズおよび形状は画素ブロックによって変化する。

各画素ブロックを符号化する最良の符号化モードの選択は、圧縮ビットストリームのビット・レート R と復号化されたビデオ・シーケンスの内の歪 D に極めて直接影響を与える H.264 標準の決定の 1 つである。符号化モードの選択の目標は、ビット・レート制約 $R(p) \leq R_{max}$ (p は調整可能な符号化パラメータのベクトル、 R_{max} は最大許容ビット・レート) を受ける歪 $D(p)$ を最小限にする符号化モード M^* を選択することである。制約つき最適化問題は、下式で与えられるラグランジュの方程式 $J(p, \lambda)$ を用いて制約なし最適化問題に変換できる。

【0010】

10

【数1】

$$J(\bar{p}, \lambda) = D(\bar{p}) + \lambda R(\bar{p}) \quad (1)$$

ただし、 λ は速度歪トレードオフを制御するラグランジュ乗数である。符号化モード決定問題は $J(p, \lambda)$ の最小化となる。これは下式で表される。

【0011】

20

【数2】

$$\min_{all \bar{p}} \{D(\bar{p}) + \lambda R(\bar{p})\} \quad (2)$$

上記ラグランジュの方程式は、許容可能な符号化モードごとに以下のステップを実行することで評価できる。

【0012】

(a) 歪 D を元の画素ブロックと特定の符号化モードを用いた符号化および復号化後の再構成された画素ブロックとの間の誤差の L_2 規範として計算する。

30

(b) ビット・レート R を、動きベクトルおよび変換係数を符号化するのに必要な総ビット数として計算する。

(c) 式(1)を用いてラグランジアン J を計算する。

最後に、すべての符号化モードについてラグランジアン J の値を計算した後に得られた最小ラグランジアン J は、式(2)で表す最小化問題を解決する符号化モード M^* を示す。

H.264 ビデオ圧縮標準では、符号化モードの決定は 8×8 以下の画素ブロックを用いて実行され、次いでそれより大きいサイズについて符号化モードが決定される。さらに、最適化プロセスの複雑さを低減しようとして、最小化の決定は固定量子化値 Q で実行され、ラグランジュ乗数はしばしば（例えば） $0.85 \times Q / 2$ または $0.85 \times 2^{Q/3}$ に等しい値に選択される。ただし、 Q は量子化パラメータである。複数の B 画像の場合、はるかに大きい値が選択されることが多い。もちろん、この複雑さの低減によって速度歪平面内のラグランジアン J の最小値の検索も制約される。

40

【0013】

提案する直接モード拡張方法

本発明のシステムは、B フレーム内の直接モード選択および P フレーム内のスキップ・モード選択の拡張の方法を提案する。本発明のシステムは、コスト値のクラスタ化、アウトライアの低減、およびラグランジュ乗数の指定を使用する。一実施形態では、このシステムは 4 つのステップを用いる方法を実行する。以下、図 3 を参照しながらこれらのことについて詳述する。

50

第 1 に、現在の画素ブロックは各々の可能な符号化モード M について符号化および復号化され、歪 D_M はステップ 310 および 320 で述べたように計算される。一実施形態では、歪 D_M は元の画素ブロック内の画素と符号化された画素ブロック内の画素との誤差の Huber 関数値の総計として計算される。Huber 関数は、図 1 に示され、下式で与えられる。

【0014】

【化 1】

$$D_M(x) = \begin{cases} \frac{1}{2}x^2 & , |x| \leq \beta \\ \beta|x| - \frac{1}{2}\beta^2 & , |x| > \beta \end{cases}$$

10

ただし、x は画素ブロックの 1 つの画素の誤差、β はパラメータである。見て分かるように、より小さい誤差値について、Huber 関数の値は平方誤差によって与えられる値に等しい。より大きい誤差値について、Huber 関数の値は同じ誤差値の平方誤差の値より小さい。

【0015】

20

第 2 に、ステップ 330 で述べたように、各符号化モードのビット・レート R が計算される。一実施形態では、このシステムは、ビット・レート R を、画素ブロックの動きベクトルおよび変換係数を符号化するのに必要な総ビット数として計算する。

第 3 に、本発明のシステムは、ステップ 340 で述べたように、式(1)を用いて符号化モードのラグランジアンを計算する。一実施形態では、システムは、ラグランジュ乗数の値が、量子化パラメータの関数として、H.264 標準バージョン 4.1 の非規範部で提案されている元のラグランジアンよりも緩やかに変化するようにこれを選択する。量子化値 Q の関数であるラグランジアンの提案された変化を図 2A、図 2B、および図 2C に示す。ラグランジュ乗数を参考実施態様のよりも緩やかに変化させることで、本発明のシステムはラグランジュの方程式(1)のビット・レート構成要素 R よりも歪構成要素 D により重点を置く。このラグランジュ乗数への移行の結果、ビット・レート R がわずかに増加しても出力ラグランジアン値 J への影響は少なくなる。(また、ビット・レート R が以下の節で説明するラグランジアン・クラスタに与える影響は低減する。)

30

第 4 に、すべての J_M について J_{M*} を J の最小値とする(式(1)を用いて)(ただし、M は可能な符号化モードの 1 つ)。J_{M*} を生成する符号化モードと同じ符号化モード(M*)を選択する代わりに、システムは計算されたラグランジアン J_M の値を以下のようにクラスタ化する。S を、計算されたラグランジアン値が以下の条件を満たす符号化モード k の集合とする。

【0016】

【数 3】

40

$$S = \left\{ k \mid \frac{J^*}{J_k} \geq \varepsilon \right\} \quad (3)$$

ただし、イプシロン(「ε」)は選択された誤差値、J* はすべてのモードで最小の J である。符号化モード 0 が集合 S の要素の場合、システムは画素ブロックを符号化するために使用する符号化モードとして符号化モード 0 を選択する。そうでない場合、システムは

50

J_{M^*} （最小の J 値を生成する符号化モード M^* ）に対応する符号化モード M^* を選択する。

【0017】

上記ステップは参考（非規範的）H.264エンコーダと比べて新規の構成要素を利用する。特に、本発明は Huberコスト関数を用いて歪、修正ラグランジュ乗数、およびラグランジアン値のクラスタ化を計算する。

Huberコスト関数は堅牢な M 推定器（estimator）に属する。これらの関数の重要な特性はアウトライアの影響を低減する能力である。より詳細には、画素ブロック内にアウトライアが存在する場合、Huberコスト関数がそれらを平均平方誤差関数の場合（二次式）よりも少なく重み付けし（一次式）、その画素ブロックについて選択された符号化モードをおそらく隣接マクロブロックの符号化モードと同じにする。10

修正ラグランジュ乗数 は量子化パラメータ Q の関数としてより緩やかに変化するため、ビット・レート構成要素 R よりもラグランジアン J の歪構成要素により重点を置く。（本明細書では、「ラムダ」またはラムダ記号「λ」は符号化モード決定プロセスで使用されるラグランジュ乗数を示す。動きベクトル選択プロセスで使用される乗数は異なる。）

最後に、前述したラグランジアン値のクラスタ化は符号化モード 0 を選択する。したがって、本発明のシステムによって、それぞれ B 画素ブロックおよび P 画素ブロックについて直接モードまたはスキップ・モードを用いてより多くの画素ブロックを符号化することができます。20

【0018】

実験の結果

実験で使用したビデオ試験セットは映画シーケンス「エジプトの発見」、「風と共に去りぬ」、および「イングリッシュ・ペーシェント」から抜粋した 9 つのカラー・ビデオ・クリップからなる。これらのビデオ・シーケンスの具体的な特徴を表 1 に示す。

【0019】

【表 1】

Table 1: 試験シーケンス
(略語 ch および Og はそれぞれ章とオポージング・グラ
ンスを表す。)

シーケンス No.	ビデオ・シーケンス名	フレーム・サイズ	フレームNo.	タイプ
1	エジプトの発見 , ch. 1	704 x 464	58	パン
2	風と共に去りぬ , ch. 11	720 x 480	44	Og
3	エジプトの発見 , ch. 1	704 x 464	630	パン
4	エジプトの発見 , ch. 2	704 x 464	148	ズーム
5	エジプトの発見 , ch. 3	704 x 464	196	ズーム
6	エジプトの発見 , ch. 6	704 x 464	298	パン
7	イングリッシュ・ペーシェント , ch. 2	720 x 352	97	テクスチャ
8	イングリッシュ・ペーシェント , ch. 6	720 x 352	196	Og
9	イングリッシュ・ペーシェント , ch.8	720 x 352	151	Og

【0020】

10

20

30

40

50

ビデオ・フレームはそれぞれ YUV フォーマットで表現され、ビデオ・フレーム速度はすべてのビデオ・シーケンスで 23.976 フレーム / 秒 (fps) に等しい。本発明で提案する方法の有効性は圧縮ビデオ・シーケンスのビット・レート R と復号化されたビデオ・シーケンスの画像品質とを用いて評価されている。後者はビデオ・シーケンスの目視点検とピーク信号対雑音比 (PSNR) 値によって評価されている。

提案する直接モード拡張方法の節に記載した本発明の符号化方法の新規の構成要素はその速度および歪への影響に関して互いに補う。本発明の方法は、ビット・レートの全面的な低減とピーク信号対雑音比のわずかな低減とを生成する。本発明のシステムは以下の節に述べる 2 つの実験を用いて評価されている。

【0021】

10

全シーケンスについての固定量子化パラメータ

第 1 の実験は全てのビデオ・シーケンスで等しく、それぞれ I フレーム、P フレーム、B フレームで Q、Q + 1、Q + 3 に等しい量子化パラメータ Q を選択する。表 2 に示すように、ビット・レートの低減は本発明の符号化方法を使用すると 9 % であるが、ピーク信号対雑音比 (PSNR) の損失は約 0.12 dB である。参考の方法を用いて符号化した場合とは対照的に、本発明の符号化方法を用いて符号化したビデオ・シーケンスには歪は見られない。

【0022】

20

【表 2】

Table 2: 全てのシーケンスについて同じ量子化パラメータ Q を用いた参考の方法および提案する方法による映画シーケンスのビット・レート (BR) [キロビット / 秒] およびピーク信号対雑音比 (PSNR) [dB]

シーケンス No.	参考の方法		提案される方法	
	ビット・レート [kbits/sec]	PSNR [dB]	ビット・レート [kbits/sec]	PSNR [dB]
1	162.04	38.89	155.43 (-4.08%)	38.75 (-0.13 dB)
2	287.71	39.82	283.35 (-1.51 %)	39.71 (-0.11 dB)
3	659.14	37.32	650.92 (-1.24 %)	37.20 (-0.12 dB)
4	1029.02	35.84	1012.17 (-1.63%)	35.76 (-0.07 dB)
5	390.46	36.77	354.25 (-9.27%)	39.59 (-0.18 dB)
6	144.82	39.11	139.02 (-4.00 %)	39.02 (-0.09 dB)
7	257.06	37.30	255.08 (-0.76 %)	37.12 (-0.18 dB)
8	102.75	40.17	99.81 (-2.85%)	40.03 (-0.13 dB)
9	222.29	39.62	218.48 (-1.71%)	39.50 (-0.12 dB)
	最大 BR	変化 :	-9.27%	
	最小 BR	変化 :	-0.76%	
	平均 BR	変化 :	-3.00%	
	最大 PSNR	利得:		0 dB
	最大 PSNR	損失:		-0.183 dB
	平均 PSNR	変化 :		-0.128 dB

30

【0023】

40

シーケンスごとの最大量子化パラメータ

本発明の符号化方法の有用性をさらに評価するために、第 2 の実験が設計され実行された。ビット・レート R とピーク信号対雑音比の値が共に減少する時の一般的な説明は、ビデオ・シーケンスの事前フィルタリング、量子化値 Q の値の増加など多数の方法が同じ結果を生むという説明である。この実験の目標は、これらの解決策がビデオ品質を大幅に損ねることなく適用できない時に、本発明の方法はビット・レートを低減できるということを示すことである。

第 1 に、各試験ビデオ・シーケンスで、ビット・レートは歪が観察できる状態になる Q

50

$m_{max} + 1$ になるまで量子化パラメータの値を増加させることで参考の方法を用いて最大限低減される。次に、本システムは、 Q_{max} （歪がまだ観察できる状態でない最大値）および参考の方法を用いてビデオ・シーケンスを符号化および復号化し、表 3 に示すビット・レートとピーク信号対雑音比（PSNR）値とを生成する。各シーケンスで、 Q_{max} の値は異なり、それぞれ I フレーム、P フレーム、B フレームでも異なる。画像損失なしにビット・レートの低減をこうして最大限に達成したことで、本発明の符号化方法は同じ値 Q_{max} のシーケンスの符号化に適用される。

【 0 0 2 4 】

【表 3】

Table 3: 最大量子化パラメータを用いた参考の方法および提案する方法による映画シーケンスのビット・レート(BR) [キロビット／秒]およびピーク信号対雑音比(PSNR) [dB]

シーケンス No.	参考の方法		提案される方法	
	ビット・レート [kbits/sec]	PSNR [dB]	ビット・レート [kbits/sec]	PSNR [dB]
1	512.59	41.39	479.52 (-6.45%)	41.15 (-0.24 dB)
2	316.70	40.10	298.86 (-5.63%)	39.89 (-0.21 dB)
5	238.78	35.74	210.40 (-11.33%)	35.18 (-0.56 dB)
6	169.28	39.46	146.75 (-13.30%)	39.10 (-0.36 dB)
7	300.56	37.78	290.67 (-3.28%)	37.50 (-0.28 dB)
9	276.91	40.45	270.56 (-2.30%)	40.31 (-0.14 dB)
	最大 BR	変化 :	-13.30%	
	最小 BR	変化 :	-2.30%	
	平均 BR	変化 :	-7.04%	
	最大 PSNR	利得 :		0 dB
	最大 PSNR	損失 :		-0.56 dB
	平均 PSNR	変化 :		-0.29 dB

【 0 0 2 5 】

表 3 に示すように、本発明の方法は、ビット・レートを最大 13.3% と大幅に低減でき、かつ、ピーク信号対雑音比（PSNR）の損失は約 0.29 dB である。フルフレーム速度でのシーケンスの目視点検（任意の B フレーム関連のアーティファクトを評価するための）によって、このビット・レートの低減によって復号化されたビデオ・シーケンス内に可視アーティファクトが混入しないことが分かる。本発明の方法を使用する時には、量子化パラメータの値を Q_{max} より増やすことができ、画像損失なしにビット・レートを低減することができる。

【 0 0 2 6 】

結論

本発明は、H.264 (MPEG-4 / Part 10) ビデオ圧縮標準の枠内の B 画像内の直接モード拡張および P 画像内のスキップ・モード拡張の方法を提案した。本発明のシステムは Huber コスト関数を用いて歪、修正ラグランジュ乗数、およびラグランジアン値のクラスタ化を計算し、画素ブロックを符号化するための符号化モードを選択する。試験によって、本発明の方法を用いて、主観的な画像品質の低下を招くことなくピーク信号対雑音比（PSNR）のわずかな損失でビット・レートを大幅に低減することができることが分かった。これらの特徴によって、本発明の方法は、量子化パラメータの値をこれ以上増やすといったその他の解決策が適用できない場合に拡張機能として符号化モード決定の速度歪最適化の枠組みを採用する任意のビデオ符号化システムでのビット・レートの低減に特に有用である。

以上、ディジタル画像拡張を実行する方法および装置について説明した。本発明の範囲を逸脱することなしに、当業者は本発明の要素の材料および構成を変更し修正することが

10

20

30

40

50

できる。

【図面の簡単な説明】

【0027】

【図1】変数のHuberコスト関数を示すグラフである。

【図2A】対象範囲の量子化パラメータ(Q)値の関数である元の、および修正ラグランジュ乗数 λ_{mode} の変化を示す図である。

【図2B】対象範囲の量子化パラメータ(Q)値の関数であるBフレームの元の、および修正ラグランジュ乗数 λ_{mode} の変化を示す図である。

【図2C】対象範囲の量子化パラメータ(Q)値の関数である元の、および修正ラグランジュ乗数 λ_{motion} の変化を示す図である。

【図3】符号化モードの選択方法を示す流れ図である。

【符号の説明】

【0028】

310, 320, 330, 340, 350 ステップ

10

【図1】

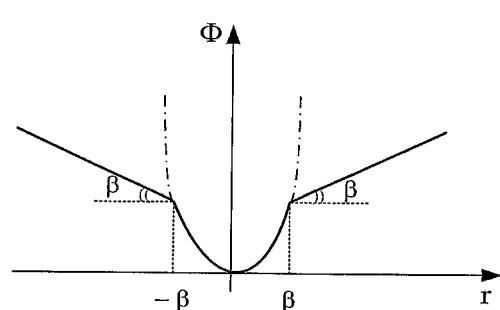
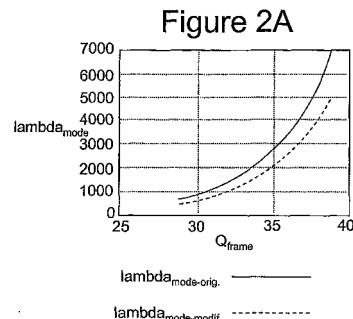
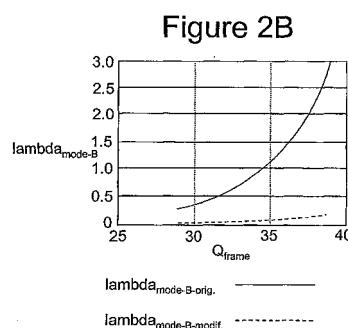


Fig. 1 – Huber 関数

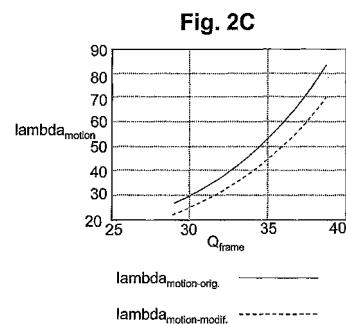
【図2A】



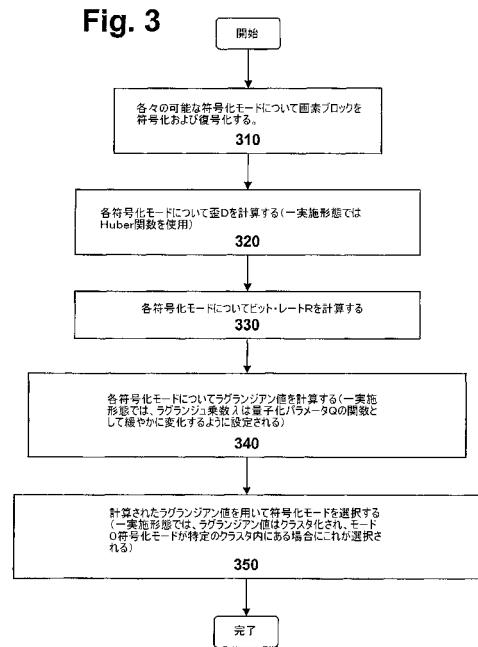
【図2B】



【図2C】



【図3】



フロントページの続き

(72)発明者 ハスケル、バリン、ジー .

アメリカ合衆国、94040 カリフォルニア州、マウンテン ヴュー、フェアブルック 119
0

(72)発明者 プリ、アチュール

アメリカ合衆国、95014 カリフォルニア州、クーペルティーノ、#4203 ブルンエッジ
アヴェニュー 19500

審査官 坂東 大五郎

(56)参考文献 特開2003-299102(JP, A)

Thomas Wiegand, Bernd Girod, Lagrange multiplier selection in hybrid video coder control, Proceedings of International Conference on Image Processing, 2001, ICIP 2001, 2001年10月 7日, vol. 3, p. 542-545

Mei-Yin Shen, JongWon Kim and C.-C. Jay Kuo, Fast compression artifact reduction technique based on nonlinear filtering, Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, 1999. ISCAS '99., 1999年 5月30日, vol. 4, p. 179-182
高木幸一、滝嶋康弘、中島康之, JVT符号化におけるレート制御に関する一検討, 電子情報通信学会技術研究報告, 2002年11月 8日, Vol. 102, No. 445, p. 123
- 126

(58)調査した分野(Int.Cl., DB名)

H04N 7/24-7/68