

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第7部門第3区分

【発行日】平成19年3月15日(2007.3.15)

【公表番号】特表2003-527005(P2003-527005A)

【公表日】平成15年9月9日(2003.9.9)

【出願番号】特願2001-566560(P2001-566560)

【国際特許分類】

H04N 7/32 (2006.01)

H04N 7/173 (2006.01)

【F I】

H04N 7/137 Z

H04N 7/173 610Z

【手続補正書】

【提出日】平成19年1月26日(2007.1.26)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【書類名】明細書

【発明の名称】信号処理装置及び方法

【特許請求の範囲】

【請求項1】

コンテンツを処理する信号処理方法において、
端末上の端末情報を記録するステップと、

上記コンテンツ及び該コンテンツを処理するためのコンテンツ情報を記録するステップと、
上記コンテンツ情報及び上記端末情報に基づき、トランスクードヒントを抽出するステップと、

上記トランスクードヒントに基づき、上記コンテンツを変換するステップとを有する信号処理方法。

【請求項2】

上記トランスクードヒントを抽出するステップは、

第1のGOP構造を有する圧縮画像データの第1のビットストリームを受け取るステップと、

上記第1のビットストリームから第1の動き情報を得るステップと、
第1のセグメントのテクスチャ/エッジ情報を得るステップと、
上記第1のビットストリームから特徴点及び関連する動きの情報を得るステップと、
上記第1のビットストリームから重要な情報の領域を得るステップとを有することを特徴とする請求項1記載の信号処理方法。

【請求項3】

上記トランスクードヒントを抽出するステップは、上記第1の動き情報をトランスクードヒントとして保存するステップを更に有することを特徴とする請求項2記載の信号処理方法。

【請求項4】

上記トランスクードヒントを抽出するステップは、動き関連トランスクードヒントをパラメトリック動きモデルのパラメータとして表すステップを更に有することを特徴とする請求項2記載の信号処理方法。

【請求項 5】

上記トランスクードヒントを抽出するステップは、上記パラメトリック動きモデルにより連続する長方形ビデオフレーム内の包括的な動きを記述するステップを更に有することを特徴とする請求項 4 記載の信号処理方法。

【請求項 6】

上記トランスクードヒントを抽出するステップは、上記パラメトリック動きモデルにより、任意形状の定義された領域内の動きを記述するステップを更に有することを特徴とする請求項 4 記載の信号処理方法。

【請求項 7】

上記パラメトリック動きモデルは、MPEG - 4において使用される任意形状の定義された領域内の動きを記述するために使用されることを特徴とする請求項 6 記載の信号処理方法。

【請求項 8】

上記トランスクードヒントを抽出するステップは、動き関連トランスクードヒントを上記圧縮画像データの第 1 のビットストリームに含まれる動きベクトルのアレイとして表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 9】

上記トランスクードヒントを抽出するステップは、動き関連トランスクードヒントを上記圧縮画像データの第 1 のビットストリームに含まれる動きベクトルから導出される動きベクトルのアレイとして表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 10】

上記トランスクードヒントを抽出するステップは、動き関連トランスクードヒントを連続するフレーム内で探索された動きベクトルに関連付けられた特徴点のリストとして表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 11】

上記トランスクードヒントを抽出するステップは、動き関連トランスクードヒントを連続するフレーム内の任意形状の領域間で探索された動きベクトルに関連付けられた特徴点のリストとして表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 12】

上記トランスクードヒントを抽出するステップは、テクスチャ関連トランスクードヒントを DCT 係数のリスト又は DCT 係数から導出された平均、最大、最小、平方偏差、標準偏差のうちの少なくとも 1 つを含む測定値を表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 13】

上記トランスクードヒントを抽出するステップは、エッジ関連トランスクードヒントを DCT 係数のリスト又は DCT 係数から導出された平均、最大、最小、平方偏差、標準偏差のうちの少なくとも 1 つを含む測定値を表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 14】

上記トランスクードヒントを抽出するステップは、特徴点及び該特徴点に関連する動き関連ベクトルをリストとして表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 15】

上記トランスクードヒントを抽出するステップは、符号化複雑性関連トランスクードヒントをあるフレームから次のフレームに移行する際に失われた及び新たに現れた特徴点の数に基づき、連続するフレームで追跡された特徴点のライフタイムリストから導出された複雑性メトリックとして表すステップを更に有することを特徴とする請求項 2 記載の信号処理方法。

【請求項 1 6】

コンテンツを処理する信号処理装置において、
端末上の端末情報を記録する手段と、
上記コンテンツ及び該コンテンツを処理するためのコンテンツ情報を記録する手段と、
上記コンテンツ情報及び上記端末情報に基づき、トランスコードヒントを抽出する手段と、
上記トランスコードヒントに基づき、上記コンテンツを変換する手段とを備える信号処理装置。

【請求項 1 7】

上記トランスコードヒントを抽出する手段は、
第1のG O P構造を有する圧縮画像データの第1のビットストリームを受け取る手段と、
上記第1のビットストリームから第1の動き情報を得る手段と、
第1のセグメントのテクスチャ／エッジ情報を得る手段と、
上記第1のビットストリームから特徴点及び関連する動きの情報を得る手段と、
上記第1のビットストリームから重要な情報の領域を得る手段とを備えることを特徴とする請求項16記載の信号処理装置。

【請求項 1 8】

上記トランスコードヒントを抽出する手段は、上記第1の動き情報をトランスコードヒントとして保存する手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項 1 9】

上記トランスコードヒントを抽出する手段は、動き関連トランスコードヒントをパラメトリック動きモデルのパラメータとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項 2 0】

上記トランスコードヒントを抽出する手段は、上記パラメトリック動きモデルにより連続する長方形ビデオフレーム内の包括的な動きを記述する手段を更に備えることを特徴とする請求項19記載の信号処理装置。

【請求項 2 1】

上記トランスコードヒントを抽出する手段は、上記パラメトリック動きモデルにより任意形状の定義された領域内の動きを記述する手段を更に備えることを特徴とする請求項19記載の信号処理装置。

【請求項 2 2】

上記パラメトリック動きモデルは、M P E G - 4において使用される任意形状の定義された領域内の動きを記述するために使用されることを特徴とする請求項21記載の信号処理装置。

【請求項 2 3】

上記トランスコードヒントを抽出する手段は、動き関連トランスコードヒントを上記圧縮画像データの第1のビットストリームに含まれる動きベクトルのアレイとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項 2 4】

上記トランスコードヒントを抽出する手段は、動き関連トランスコードヒントを上記圧縮画像データの第1のビットストリームに含まれる動きベクトルから導出される動きベクトルのアレイとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項 2 5】

上記トランスコードヒントを抽出する手段は、動き関連トランスコードヒントを連続するフレーム内で探索された動きベクトルに関連付けられた特徴点のリストとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項 2 6】

上記トランスコードヒントを抽出する手段は、動き関連トランスコードヒントを連続するフレーム内の任意形状の領域間で探索された動きベクトルに関連付けられた特徴点のリストとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項27】

上記トランスコードヒントを抽出する手段は、テクスチャ関連トランスコードヒントをDCT係数のリスト又はDCT係数から導出された平均、最大、最小、平方偏差、標準偏差のうちの少なくとも1つを含む測定値を表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項28】

上記トランスコードヒントを抽出する手段は、エッジ関連トランスコードヒントをDCT係数のリスト又はDCT係数から導出された平均、最大、最小、平方偏差、標準偏差のうちの少なくとも1つを含む測定値を表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項29】

上記トランスコードヒントを抽出する手段は、特徴点及び該特徴点に関連する動き関連ベクトルをリストとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【請求項30】

上記トランスコードヒントを抽出する手段は、符号化複雑性関連トランスコードヒントをあるフレームから次のフレームに移行する際に失われた及び新たに現れた特徴点の数に基づき、連続するフレームで追跡された特徴点のライフタイムリストから導出された複雑性メトリックとして表す手段を更に備えることを特徴とする請求項17記載の信号処理装置。

【発明の詳細な説明】

【0001】

【技術分野】

本発明は、マルチメディア記述及び異なる方式（例えば、MPEG）により圧縮されたコンテンツ表現間のトランスコードのためのトランスコードヒントメタデータのコンパクトな表現を抽出し、フレームレート、ビットレート、セッションサイズ、量子化パラメータ、ピクチャ符号化タイプ構造（例えば、グループオブピクチャ：GOP）等の（例えば、MPEG圧縮）ビットストリームのパラメータを操作し、A/Vコンテンツを分類し、マルチメディア情報を取り出す信号処理装置及び方法に関する。

【0002】

【背景技術】

光、無線及び有線ネットワークを介して伝送されるオーディオ／ビデオ（又は、オーディオビジュアル：以下、A/Vという。）コンテンツのデータ量が増加している。これらのネットワークは、それぞれ異なるネットワーク帯域幅の制限を有しているので、A/Vコンテンツを異なるビットレートで表す必要が生じ、この結果、主観的な画質が変化する。また、A/V端末機器の画面のサイズ、演算能力及びメモリ容量の制約により、A/Vコンテンツを圧縮して表す必要が生ずることもある。

【0003】

したがって、例えばムービングピクチャエキスパートグループ（Moving Pictures Experts Group：以下、MPEGという。）により定義されている圧縮形式で保存されたA/Vコンテンツを、例えば異なるA/V端末機器の異なるビットレート、フレームレート、画面サイズ、復号処理の複雑性及びメモリ容量の制約に応じて、変換する必要がある。

【0004】

異なる帯域幅及び異なるA/V端末機器毎に同一のA/Vコンテンツに対応する複数の圧縮表現を保存することを避けるためには、MPEGフォーマットで保存された圧縮A/Vコンテンツを異なるMPEGフォーマットに変換することが好ましい。

【0005】

ビデオトランスコード処理に関する参考文献を以下に示す。

【0006】

1999年、国際特許公開公報WO09838801A1号、オー・エイチ・ワーナ（O.H.Werner）、エヌ・ディー・ウェルス（N.D.Wells）、エム・ジェイ・ニー（M.J.Knee）：「改善された量子化によるデジタル圧縮符号化方法（Digital Compression Encoding with improved quantization）」。この文献には、適応量子化法が開示されている。

【0007】

1999年、米国特許公報5870146号、チュー・キン・ファン（Zhu; Qin-Fan）：「デジタルビデオトランスコード方法及び装置（Device and method for digital video transcoding）」。

【0008】

1999年、国際特許公開公報WO09929113A1号、ニルソン・ミヒヤエルエリイング（Nilsson, Michael, Eriing）、ガンバリ・モハメド（Ghanbari, Mohammed）：「トランスコード方法（Transcoding）」。

【0009】

1998年、米国特許公報5805224号、キースマン（Keesman）、ジェリット・ジェイ（Gerrit J）、バン・オッタールー（Otterloo）、ペトラス・ジェイ（Petrus J）：「ビデオ信号のトランスコード方法及び装置（Method and Device for Transcoding Video Signal）」。

【0010】

1999年、国際特許公開公報WO09943162A1号、ゴリン・スチュアート・ジャイ（Golin, Stuart, Jay）：「ビデオシーケンスをトランスコードするための動きベクトル外挿方法（Motion vector extrapolation for transcoding video sequences）」。

【0011】

1998年、米国特許公報5838664号、ポロムスキ（Polomski）、マーク・ディー（Mark D.）：「デジタルトランスコードによるビデオ会議システム（Video teleconferencing system with digital transcoding）」。

【0012】

1999年、国際特許公開公報WO09957673A2号、バリオル・ニコラス（Baliol, Nicolas）：「データストリームのトランスコード方法（Transcoding of a data stream）」。

【0013】

1998年、米国特許公報5808570号、バクムトスキ－・ミヒヤエル（Bakhmutsky; Michael）：「ペアマッチングハフマントランスコード装置及び方法、並びにこれを用いた2ワードビットストリームセグメント化を行う高性能可変長デコーダ（Device and Method for pair-matching Huffman-Transcoding and high performance variable length decoder with two-word bitstream segmentation which utilizes the same）」。

【0014】

1999年、国際特許公開公報WO09905870A2号、レマグエット・ヤン（Le maguet,Yann）：「ビデオシーケンス間の切換方法及びこれに対応する装置（Method of Switching between Video Sequences and corresponding Device）」。

【0015】

1999年、国際特許公開公報WO09923560A1号、ルドウイング・ラスター（LUDWIG,Lester）、ブラウン・ウィリアム（BROWN,William）、ユル・イン・ジェイ（YUL, Inn,J.）、ブオング・アン・ティー（VUONG,Anh,T.）、バンデルリップ・リチャード（VANDERLIPPE,Richard）、バーネット・ジェラルド（BURNETT,Gerald）、ロウワース・クリス（LAUWERS,Chris）、ルイ・リチャード（LUI,Richard）、アプレバウム・ダニエル（APPLEBAUM,Daniel）：「スケーラブルネットワークマルチメディアシステム及びアプリケーション（Scalable networked multimedia system and application）」。

【 0 0 1 6 】

しかしながら、上述の特許文献のいずれにも、トランスコードヒントメタデータ (transcoding hints metadata) 情報を用いて A / V トランスコード処理を容易に行う方法は開示も示唆もされていない。

【 0 0 1 7 】

ソサエティオブモーションピクチャーアンドテレビジョン (Society of Motion Picture and Television: 以下、SMPTE という。) は、MPEG-2 ビデオレコーディングデータセット (MPEG-2 Video Recording Data Set: 327M-2000) によりテレビジョン用の標準規格を提唱している。この標準規格は、ソースフォーマットの各マクロブロックのメタデータを、256 ビットを用いて再符号化することを規定している。しかしながら、トランスコードヒントメタデータの抽出及び表現には、幾つかの問題点がある。例えば、この標準規格では、トランスコードヒントメタデータ (例えば、GOP の構造、量子化器の設定、動きベクトル等) は、単一のフレーム毎及び A / V ソースコンテンツのマクロブロック毎に抽出される。これにより、詳細でコンテンツに適応したトランスコードヒントメタデータが得られ、主観的な A / V 相対性 (subjective A/V duality) を維持しつつ、トランスコード処理を容易に行うことができる。しかしながら、トランスコードヒントメタデータのデータサイズは非常に大きい。提唱されている標準規格の一実現例においては、MPEG ビデオデータのマクロブロック毎に 256 ビットのトランスコードヒントメタデータが保存される。このような膨大な量のトランスコードヒントメタデータをローカル、例えば家庭用、A / V コンテンツサーバに放送により配信することは、現実的ではない。したがって、提唱されているトランスコードヒントメタデータの標準規格は、放送局等においてのみ適用されている。

【 0 0 1 8 】

トランスコードヒントメタデータを抽出及び表現する他の技術として、特定のビットレートを有する圧縮 A / V ソースコンテンツを他の圧縮フォーマット及びビットレートの圧縮 A / V ソースコンテンツにトランスコードするための一般的なトランスコードヒントメタデータを集める手法がある。しかしながら、この手法は、トランスコードされたコンテンツの特徴的プロパティを考慮していないという問題がある。例えば、ソースコンテンツにおいて、動きが制限され、細部が少ない A / V セグメント (例えば、ニュース番組の司会のシーン) と、動きが速く細部が多い他の A / V セグメント (例えば、スポーツイベントのシーン) とでは、A / V 特性が変化する。上述の手法では、これらのビデオセグメントのいずれをも適切に表現していない誤ったトランスコードヒントメタデータが選択され、したがって、A / V 品質が劣化し、間違ったビットレートの割当が行われてしまうことがある。

【 0 0 1 9 】**【発明が解決しようとする課題】**

本発明は、上述の課題に鑑みてなされたものであり、コンパクトな A / V コンテンツに適応したマルチメディア記述及びトランスコードヒントメタデータ表現を抽出する方法及び装置を提供することである。

【 0 0 2 0 】

さらに、本発明は、長い遅延を生じさせることも、トランスコード処理に要求される演算を非現実的な程に複雑化することもなく、実時間抽出を実現できる信号処理装置及び方法を提供することを目的とする。トランスコード処理に対する第 2 の要求は、主観的な A / V 品質を可能な限り維持するという要求である。様々な圧縮ターゲットフォーマットに対して、これらの要求を満たすトランスコード処理を容易に行うために、トランスコードヒントメタデータを事前に生成し、圧縮 A / V コンテンツと別に、あるいは共に保存してもよい。さらに、本発明は、非常にコンパクトな表現を提供することにより、必要なストレージ容量を低減し、マルチメディア記述及びトランスコードヒントメタデータの配信 (例えば、ローカルオーディオ / ビデオコンテンツサーバへの放送) を容易にすることを目的とする。

【 0 0 0 1 】**【課題を解決するための手段】**

本発明に係る信号処理方法は、コンテンツを処理する信号処理方法において、端末上の端末情報を記録するステップと、コンテンツ及び該コンテンツを処理するためのコンテンツ情報を記録するステップと、コンテンツ情報及び端末情報に基づき、トランスコードヒントを抽出するステップと、トランスコードヒントに基づき、コンテンツを変換するステップとを有する。

【 0 0 0 2 】

また、本発明に係る信号処理装置は、コンテンツを処理する信号処理装置において、端末上の端末情報を記録する手段と、コンテンツ及び該コンテンツを処理するためのコンテンツ情報を記録する手段と、コンテンツ情報及び端末情報に基づき、トランスコードヒントを抽出する手段と、トランスコードヒントに基づき、コンテンツを変換する手段とを備える。

【 0 0 2 1 】

本発明は、1)トランスコード処理においてオーディオ／ビデオ品質を劣化させず、2)演算の複雑性を低減し、最小限の遅延しか有さない実時間アプリケーションを実現できる情報処理装置を提供することを目的とする。本発明の実施の形態においては、トランスコードヒントメタデータをカバーする追加的データ(メタデータ)を圧縮A／Vコンテンツに関連付ける。

【 0 0 2 2 】

本発明の目的及び効果は、以下の詳細な説明及び図面からも明らかである。本発明は、トランスコードヒントメタデータの自動的な抽出及びコンパクトな表現を実現する装置及び方法を提供する。

【 0 0 2 3 】

本発明は、トランスコードメタデータを用いて、ある圧縮フォーマットを有するA／Vコンテンツを他の圧縮フォーマットを有するA／Vコンテンツにトランスコードする技術分野に関する。トランスコードという用語は、特に以下に限定されるものではないが、圧縮フォーマットの変換(例えば、MPEG-2フォーマットからMPEG-4フォーマットへの変換)、フレームレートの変換、ビットレートの変換、セッションサイズの変換、画面サイズの変換、ピクチャ符号化タイプの変換等を含む。

【 0 0 2 4 】

さらに、本発明は、上述のトランスコードヒント状態をビデオデータにおける異なるシーン動きのクラスとして用いる自動的なビデオ分類にも適用できる。

【 0 0 2 5 】

本発明は、以下の詳細な説明において開示され、請求の範囲において示される複数のステップ及び1つ以上のステップの関係、各ステップを実行する要素及び構成部分を備える本発明の特徴を実現する装置の全てを包含する。

【 0 0 2 6 】**【発明を実施するための最良の形態】**

図1は、本発明に基づき、ホームネットワーク環境においてトランスコード処理を行うシステム100の全体的な構成を示す図である。図1に示すように、A／Vコンテンツサーバ102は、A／Vコンテンツストレージ103と、A／Vトランスコードユニット106と、トランスコードヒントメタデータ抽出ユニット104と、A／Vトランスコードヒントメタデータ記憶バッファ105とを備える。A／Vコンテンツストレージ103は、様々なソースから供給される、様々なビットレート及び様々な主観的品質を有する圧縮A／V素材(compressed A/V material:以下、圧縮A／V素材という。)を格納する。A／Vコンテンツストレージ103は、例えば、ポータブルデジタルビデオ(DV)用のビデオカメラ111からのホームビデオデータ、MPEG-4インターネットカメラ112からのビットレートの低い(例えば、10kbit/s)MPEG-4圧縮データ、及び放送サービス101からのメインレベルのMPEG-2メインプロファイルデータ(M

P @ M L)とを格納する。放送サービス 101 からのデータには、既にトランスコードヒントメタデータが関連付けられていることもある。A / V コンテンツサーバ 102 は、ビットレートがかなり高い高解像度の圧縮MPEG ビデオデータを格納することもできる。

【0027】

図 1 に示すように、A / V コンテンツサーバ 102 は、ホームネットワーク 113 に接続されている。ホームネットワーク 113 は、有線ネットワークであっても、無線ネットワークであってもよい。異なる特性を有する A / V 端末機器とは、以下に限定されるものではないが、例えば次のようなものがある。すなわち、無線MPEG - 4 A / V 個人用情報端末機器 (personal digital assistant: 以下、PDA という。) 107、高精細度テレビジョン放送用の高精細度 A / V 端末機器 108、A / V ゲーム機 109 及び国際電気通信連合の技術標準化グループ (Telecommunications Union Technical Standards Group: 以下、ITU-T という。) の規格に基づくテレビ電話 (videophone) 110 等である。A / V 端末機器 107、108、109、110 は、ビットレートの異なる伝送能力 (ケーブル又は無線リンクに起因する) をもってホームネットワーク 113 に接続されている。

【0028】

さらに、例えば、無線MPEG - 4 A / V PDA 107 は、演算能力、メモリ容量、画面サイズ、ビデオフレームレート及びネットワークビットレートが制限されている。したがって、A / V トランスコードユニット 106 は、例えば、A / V コンテンツサーバ 102 に格納されているヨーロッパ規格の 25 フレーム / 秒 (fps)、720 × 480 画素の 5 Mbit / s の MPEG - 2 放送ビデオデータを無線MPEG - 4 A / V PDA 107 の画面に表示可能な 352 × 240 画素の及び無線伝送のための MPEG - 4 の 500 kbit / s の MPEG - 4 データにトランスコードする。A / V トランスコードユニット 106 は、A / V トランスコードヒントメタデータ記憶バッファ 105 からのトランスコードヒントメタデータを用いて、A / V コンテンツの圧縮ソースデータのビットレートを各特定のターゲット A / V 端末機器 107、108、109、110 の能力に応じて、実時間でトランスコードすることができる。トランスコードヒントメタデータは、トランスコードヒントメタデータ抽出ユニット 104 から生成してもよく、放送サービス 101 により配信されたものであってもよい。

【0029】

図 1 に示すように、ソースフォーマットの圧縮ビットストリーム (以下、第 1 のビットストリームという。) 116 は、A / V コンテンツストレージ 103 から A / V トランスコードユニット 106 に供給される。ターゲットフォーマットのビットストリーム (以下、第 2 のビットストリームという。) 115 は、A / V トランスコードユニット 106 においてトランスコードされた後、ホームネットワーク 113 に供給される。例えば、DV フォーマットに圧縮されたコンテンツは、ホームネットワーク 113 からリンク 114 を介して A / V コンテンツストレージ 103 に供給される。

【0030】

図 2 は、本発明に基づくトランスコードヒントメタデータの抽出、トランスコードヒントメタデータの保存、及びトランスコード処理を説明する図である。図 2 に示すように、A / V コンテンツバッファ 201 は、ソースフォーマットの A / V コンテンツを格納する。ソースフォーマット記述バッファ 202 は、ソースフォーマットに関する記述、例えば、ビットレート、圧縮方法、GOP 構造、画面サイズ、インターレース走査であるかプログレッシブ走査であるか、といった情報を格納する。トランスコードターゲット記述バッファ 203 は、ターゲットフォーマットの記述、例えば、ビットレート、圧縮方法、GOP 構造、画面サイズ、インターレース走査であるかプログレッシブ走査であるか、といった情報を格納する。トランスコードヒントメタデータ抽出ユニット 207 は、A / V コンテンツバッファ 201 からソースフォーマットの圧縮 A / V コンテンツを読み出すとともに、ソースフォーマット記述バッファ 202 からソースフォーマットに関する記述及びトランスコードターゲット記述バッファ 203 からターゲットフォーマットに関する記述を

読み出す。トランスコードヒントメタデータ抽出ユニット 207 は、トランスコードヒントメタデータを算出し、これをトランスコードヒントメタデータバッファ 206 に格納する。A/V トランスコードユニット 205 は、A/V コンテンツバッファ 201 からソースフォーマットの第 1 のビットストリーム 204 を読み出し、トランスコードヒントメタデータバッファ 206 に格納されているトランスコードヒントメタデータを用いて、ソースフォーマットをターゲットフォーマットに変換する。A/V トランスコードユニット 205 は、新たなターゲットフォーマットを有する圧縮された第 2 のビットストリーム 208 を A/V ターゲットフォーマットバッファ 209 に供給し、保存する。

【0031】

図 3 及び図 4 は、本発明に基づくトランスコードヒントメタデータの生成を説明する図である。MPEG に基づくビデオ圧縮処理は、連続するフレーム間の変化を符号化する予測符号化法を使用する。あるフレームから次のフレームへの変化が大きいビデオコンテンツは、フレーム間の変化が小さいビデオコンテンツとは異なり、(ビットレートを制限しながら主観的品質を維持するために)異なる再符号化パラメータの設定を必要とする。したがって、事前に再符号化を決定することが重要である。トランスコードヒントメタデータは、主に予測不可能なビジュアルコンテンツの量及び特性に基づいて選択される。新たなビジュアルコンテンツは、先行するフレームからは予測することができず、DCT 係数を用いて、ビットレートに応じた符号化処理が行われる。ここで、本発明においては、先行するフレームから現在のフレームに連続していない新たな特徴点の数に基づき、フレーム毎の新たなコンテンツの量を判定する。

【0032】

図 3 は、ビデオデータのフレーム番号(水平軸、時間軸)に対するフレーム毎の新たな特徴点の数を示すグラフである。セクション 301 は、連続するフレーム間で新たなコンテンツが現れる量が非常に少ないビデオセグメントの部分であり、したがって、これに応じて各トランスコードヒントメタデータ(例えば、大きな GOP サイズ、低いフレームレート、低いビットレート等)を選択することができる。セクション 302 は、フレーム毎の新たな特徴点が比較的多い部分であり、これは、この状態に最適なトランスコードパラメータ(例えば、比較的小さい GOP サイズ、高いビットレート等)を提供する状態記述トランスコードヒントメタデータ(state describing transcoding hints metadata)が選択されることを意味する。セクション 303 は、フレーム毎の特徴点が多く、したがってシーン毎の新たなコンテンツの量が多いトランスコードヒントメタデータ状態を表している。この場合、より小さな M 値(I/P フレーム距離)及び高いビットレートが選択される。

【0033】

図 4 は、3つの離散的トランスコードヒントメタデータ状態(discrete transcoding hints metadata states)からなるトランスコードヒントメタデータ状態の基本的な組織を示す図である。各離散的トランスコード状態は、GOP 構造、量子化パラメータ、ビットレート、画面サイズ等のメタデータを有している。これらのトランスコードヒントメタデータのパラメータは、固定された値を有していてもよく、他のパラメータの関数としてもよい。例えば、GOP 長は、フレーム毎の新たな特徴点の数の離散関数としてもよく、量子化パラメータは、DCT 係数から導出されるエッジ及びテクスチャ(texture)の動きの関数としてもよい。この具体例における3つのトランスコードヒントメタデータ状態のそれぞれは、3つの異なる符号化条件を適用するように選択される。図 4 に示すように、状態「3」403 は、動き量が多く、フレーム毎の新たなコンテンツの量が少ない場合に選択され、このようなコンテンツに対して最適なトランスコードヒントメタデータの状態を表している。状態「2」402 は、全体の動き量は少ないが、コンテンツのエッジの動きが激しく、したがって、大量のビット数を必要とする場合に選択される。状態「1」401 は、例えば、シーンの動きが少ない A/V コンテンツのトランスコード処理に適用されるよう選択される。この他、例えば異なるシーンのクロスフェード効果、シーンの突然の変更、又は2つのシーン間の黒画像等、様々なビデオ編集効果に対して、特別なトラン

スコードヒントメタデータを用意してもよい。ビデオ編集効果の位置は、手動で検出してもよく、半自動的に検出してもよく、完全に自動的に検出してもよい。

【0034】

図5は、本発明に基づき、圧縮ソースコンテンツ及び圧縮されていないソースコンテンツからトランスコードヒントメタデータを抽出する方法を説明する図である。図5に示すように、システム500は、A/Vソースコンテンツバッファ501と、ソースフォーマット記述バッファ502と、ターゲットフォーマット記述バッファ503とを備える。

【0035】

メモリ504は、圧縮領域又は非圧縮領域から抽出された動きベクトル、DCT係数及び特徴点を記憶する。圧縮領域においては、Pマクロブロック及びBマクロブロックからの動きベクトルは、ビットストリームから直接抽出することができる。しかしながら、Iマクロブロック(Itra-macroblocks)には動きベクトルは存在しない。したがって、Pマクロブロック及びBマクロブロックから得られた動きベクトルは、Iマクロブロック用に補間される(この手法については、1999年10月、神戸で開催された画像処理に関するIEEE国際会議ICIP99会報、ロイ・ワン(Roy Wang)、トーマス・ファン(Thomas Huang)著「MPEG領域における高速カメラ動き解析(Fast Camera motion Analysis in MPEG domain)」に開示されている)。Iマクロブロックに対するDCT係数は、ビットストリームから直接抽出することができる。Pマクロブロック及びBマクロブロックについては、制限された数のDCT係数(DCT係数と2つのAC係数)を抽出することができる。この手法は、1996年、通信に関する分野別IEEEジャーナル第8巻(IEEE Journal on Selected Areas in Communications, vol. 8)シー・フー・チャン(Shih-Fu Chang)、デビッド・ジー・メッサーシュミド(David G. Messerschmid)著「MC-DCT圧縮ビデオデータの処理及び合成(Manipulation and Composition of MC-DCT compressed video)」に開示されている。圧縮領域の特徴点抽出及び動き推定の手法の具体例は、例えば1999年12月ピーター・クーン(Peter Kuhn)により出願された国際特許出願「圧縮領域の特徴点登録及び動き推定のための方法及び装置(Method and Apparatus for compressed domain feature point registration and motion estimation)」に開示されており、この出願は、引用することにより、本願に援用されるものとする。ここで、A/Vソースコンテンツは圧縮されていないフォーマットしか利用できない場合もあり、また、MPEG-1、MPEG-2、MPEG-4、ITU-T H.261、ITU-T H.263等に採用されているDCT及び動き補償原理に基づかない圧縮形式しか利用できない場合もある。DVフォーマットでは、DCT係数のみしか利用できない。これらの場合、動きベクトルは、例えば、1999年にクルーウー・アカデミック出版社(Kluwer Academic Publishers)から発行されたピーター・クーン(Peter Kuhn)著「MPEG-4動き推定のためのアルゴリズム、複雑性解析及びVLSIアーキテクチャ(Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation)」に開示されている手法により算出することができる。DCT係数は、ブロックベースのDCT変換により得ることができる。この手法は、1990年、アカデミックプレス(Academic Press)から出版されたケー・アール・ラオ(K.R.Rao)、ピー・イップ(P.Yip)著「離散コサイン変換 - アルゴリズム、利点及び応用(Discrete Cosine Transform - Algorithms, Advantages, Applications)」に開示されている。画素領域(非圧縮領域)における特徴ベクトルの算出は、例えば1981年、人工知能に関する国際共同会議(International Joint Conference on Artificial Intelligence)会報674~679頁記載のブルース・ディー・ルーカス(Bruce D. Lucas)、カナデ・タケオ(Takeo Kanade)著「ステレオビジョンのアプリケーションによる繰返し登録法(An iterative registration technique with an application to stereo vision)」に開示されている。

【0036】

動き解析部505は、メモリ504に記憶されている動きベクトルからパラメトリック動きモデルのパラメータを抽出する。パラメトリック動きモデルは、6個乃至8個のパラメータを有し、パラメトリック動き推定の手法は、例えば1995年、ブレンティスホー

ル(Prentice Hall)発行のエム・テカルプ(M.Tekalp)著「デジタルビデオ処理(Digital Video Processing)」に開示されている。動きデータを用いることにより、トランスクーダにおいて動き推定をする必要がなくなり、遅延を生じることなく速やかな処理を行うことができる。このように、ソースビットストリームからの動きに関する入力データを用いて、出力データ(ターゲットビットストリーム)を導き出すことができる。例えば、画面サイズの変更、インターレース走査からプログレッシブ走査への変換等は、動きデータに大きく依存する。動きデータのパラメータは、GOP構造の符号化方法を決定するためにも使用できる。テクスチャ/エッジ解析部506は、ビットストリームから抽出されたDCT係数に基づいて動作する。この動作原理については、1990年、アカデミックプレス(Academic Press)から出版されたケー・アル・ラオ(K.R.Rao)、ピー・イップ(P.Yip)著「離散コサイン変換 - アルゴリズム、利点及び応用(Discrete Cosine Transform - Algorithms, Advantages, Applications)」又は1993年8月、民生用電子機器に関するIEEEトランザクション第39巻3号(IEEE Transactions on Consumer Electronics, Vol. 39, No. 3)ケー・ダブリュー・チュン(K.W.Chun)、ケー・ダブリュー・リム(K.W.Lim)、エイチ・ディー・チョ(H.D.Cho)、ジェイビー・ラ(J.B.Ra)著「ビデオ符号化のための知覚適応量子化アルゴリズム(An adaptive perceptual quantization algorithm for video encoding)」に開示されている。

【0037】

圧縮領域用の特徴点追跡部507は、1999年12月ピーター・クーン(Peter Kuhn)により出願された国際特許出願「圧縮領域の特徴点登録及び動き推定のための方法及び装置(Method and Apparatus for compressed domain feature point registration and motion estimation)」に開示されており、この出願は引用により本願に援用できるものとする。プロセッサ508は、フレーム毎の新たな特徴点の数を算出する。プロセッサ509は、時間的なビデオセグメント化を行い、プロセッサ510は、各セグメントに対するトランスクードヒント状態を算出する。本発明に基づくこれらの演算処理については、図6～図8を用いて後に詳細に説明する。

【0038】

メモリ511は、動きに関するトランスクードヒントメタデータを格納している。メモリ512は、テクスチャ/エッジに関するトランスクードヒントメタデータを格納している。メモリ513は、特徴点に関するトランスクードヒントメタデータを格納している。これらのメタデータについては、図15を用いて後に詳細に説明する。メモリ514は、ビデオセグメントトランスクードヒント選択メタデータを格納する。このメタデータについては、図16を用いて後に詳細に説明する。以下、トランスクードヒントメタデータの自動抽出、コンパクトな表現及び使用方法について説明する。

【0039】

図6は、本発明に基づくビデオセグメント化及びトランスクードヒント状態の選択処理を説明する図である。ステップ601において、幾つかの変数が初期化される。変数「frame」は、ソースビットストリームの現在のフレーム番号であり、変数「nframes」は、新たなビデオセグメント(又はGOP、グループオブピクチャ)内のフレーム番号である。他の変数は、このルーチンにおいてのみ使用される。ステップ602において、GOP内のフレーム番号がインクリメントされる。ステップ603において、このフレームにおいて新たなセグメント/GOPが開始されるか否かが判定される。この判定については、図7を用いて後に詳細に説明する。新たなセグメント/GOPが開始される「yes」場合、処理はステップ604に進み、この他の場合処理はステップ615に進む。ステップ604においては、変数「last_gop_start」が「new_gop_start」の値により初期化される。ステップ608及びステップ609においては、変数「frame」が1より大きい場合、変数「last_gop_stop」は「frame-1」に設定される。この他の場合、ステップ610において、変数「last_gop_stop」は、1に設定される。次に、図8に詳細を示すステップ611において、動きパラメータ605と、テクスチャ/エッジパラメータ606と、特徴点データ607とに基づいて、トランスクードヒント状態を決定する。ステップ612にお

いて、トランスコードヒントメタデータは、トランスコードヒントメタデータバッファ 206 に出力される。本発明の一実施の形態においては、ステップ 613 において、トランスコードヒントメタデータは、「nframes」(GOP の内のフレーム数)と、全てのパラメータを有するトランスコードヒント状態と、新たな GOP の開始フレーム番号(「new_gop_start」)とを含んでいる。続いて、変数「nframes」は 0 にリセットされ、現在のフレーム番号「frame」を変数「new_gop_start」に設定する。次に、ステップ 615 において、ソースビットストリームの全てのフレームが処理されたか否かが判定される。全てのフレームが処理されてはいない「no」の場合、処理はステップ 614 に進み、フレーム番号がインクリメントされ、ステップ 602 からの処理が繰り返される。この他の場合、処理は終了する。

【0040】

図 7 は、本発明に基づき、新たなビデオセグメント又は GOP の開始フレーム及び終了フレームを判定する処理を説明する図である。ステップ 701 において、図 6 における変数「nframes」が M (I / P フレーム距離) の整数倍であるか否かが判定される。ここで、「no」が選択された場合、ステップ 702 において、現在のフレーム番号が最初のフレームであるか否かが判定される。ここで、「no」が選択された場合、処理はステップ 703 に進み、変数「nframes」が GOP の内の最小フレーム数「gop_min」より大きいか否かが判定される。ステップ 703 における判定の結果が「yes」の場合、ステップ 705 において新たな GOP が開始される。ステップ 703 における判定が結果が「no」である場合、処理はステップ 704 に進み、変数「nframes」が GOP の内の最大フレーム数「gop_max」より大きいか否かが判定される。ステップ 704 における判定の結果が「yes」である場合、ステップ 706 において GOP が閉じられ、この他の場合、処理が終了する。

【0041】

図 8 は、本発明の一実施の形態に基づき、新たな特徴点の数のみを考慮して特定の GOP 又は A / V セグメントに対するトランスコードヒント状態を選択する処理を説明する図である。この実施の形態に示す基本的な方法に基づいて、上述のパラメトリック動き推定からの動きベクトル及び DCT 係数から得られたテクスチャ / エッジパラメータを用いて、同様の判定処理 (decision structures) を実現することができる。なお、ここで説明しているクラス及びアルゴリズムは、A / V 素材の動き、エッジの動き、フレーム毎の新たなコンテンツ等の分類にも使用することができ、これにより、より高度な A / V 分類レベルを実現することができる。このような場合、トランスコードヒント状態は、異なるコンテンツ素材の特定のクラスを表す。図 8 に示すステップ 801 において、変数「frame_no」、「sum」、「new_seg」が初期化される。変数「frame_no」には、パラメータ「last_gop_start」の内容が与えられ、変数「sum」及び「new_seg」は、0 に初期化される。ステップ 802 において、変数「sum」の内容は、現在のフレームの新たな特徴点の数(「frame_no」)だけインクリメントされる。ステップ 803 において、変数「frame_no」が変数「last_gop_stop」より小さいか否かが判定される。この判定の結果が「yes」である場合、ステップ 802 が繰り返され、これ以外の場合、処理はステップ 804 に進む。ステップ 804 において、変数「sum」の値が所定のパラメータ「summax」の 1 / 8 より小さいか否かが判定される。パラメータ「summax」は、フレームからフレームに追跡することができる特徴点の最大数に「last_gop_start」と「last_gop_stop」間のフレーム数を乗算した値を表す定数である。一実施の形態では、この定数の値は 200 である。ステップ 804 における判定の結果が yes である場合、ステップ 806 においてトランスコードヒント状態 1 が選択される。図 8 は、このトランスヒントコード状態 1 のパラメータも示している。この他の場合、ステップ 805 において、変数「sum」の値が所定のパラメータ「summax」の 1 / 4 より小さいか否かが判定される。この判定の結果が「yes」である場合、ステップ 807 において、図 8 の表 1 に示すようなトランスコードヒント状態 2 が選択される。この他の場合、すなわち、ステップ 805 における判定の結果が「no」である場合、ステップ 808 において、トランスコードヒント状態 3 (表 1 に示す

) が選択され、処理は終了する。なお、ステップ 804 及びステップ 805 における判定に使用する閾値は、トランスコードヒント状態の定義及び数に応じて決定される。

【0042】

トランスコードヒントメタデータの説明

擬似 C コードスタイル (pseudo C-code style) を用いて、メタデータを説明する。ここでは、間もなく市場に導入される M P E G - 7 に定義されているように、記述 (Description) を D で、記述スキーム (Description Schemes) を D S で表す。

【0043】

図 9 は、本発明の一実施の形態に基づく汎用 A / V D S 901 内のトランスコードヒントメタデータの構造的な組織 (structural organization) を説明する図である。図 9 に示すように、汎用 A / V D S 901 からセグメント D S 904 及び媒体情報 D S 902 が導出される。さらに、セグメント D S 904 からセグメント分解 (Segment Decomposition) 906 が導出され、このセグメント分解 906 からビデオセグメント D S 907 と動き領域 D S 908 が導出される。さらに、ビデオセグメント D S 907 からは、図 14 を用いて後に詳細に説明するセグメント別トランスコードヒント D S (Segment-based transcoding hints DS) 909 が導出される。ビデオセグメント D S 907 は、図 16 を用いて後に詳細に説明する 1 つ以上のトランスコードヒント状態 D S 911 にアクセスする。動き領域 D S 908 からは、図 14 を用いて後に詳細に説明する動き領域用のセグメント別トランスコードヒント D S 910 が導出され、このセグメント別トランスコードヒント D S 910 は、図 16 を用いて後に詳細に説明する 1 つ以上のトランスコードヒント状態 D S 912 にアクセスする。媒体情報 D S 902 からは、媒体プロファイル D S 903 が導出される。媒体プロファイル D S 903 からは、図 10 を用いて後に詳細に説明する汎用トランスコードヒント D S 905 が導出される。

【0044】

図 10 は、トランスコードヒント D S 1001 の構造的な組織を説明する図である。トランスコードヒント D S 1001 は、図 11 を用いて後に説明するソースフォーマット定義 D S 1002 に 1 つのインスタンスと、図 12 を用いて後に説明するターゲットフォーマット定義 D S 1003 の 1 つ以上のインスタンスとを有する。さらに、トランスコードヒント D S 1001 は、図 13 を用いて後に説明する汎用トランスコードヒント D S 1004 の 1 つの任意インスタンスと、図 15 を用いて後に説明するトランスコード複雑性 D S (Transcoding Encoding Complexity DS) 1005 の 1 つの任意インスタンスとを有する。

【0045】

図 11 は、本発明に基づき、A / V コンテンツ全体又は特定の A / V セグメントに関連付けられたソースフォーマット定義トランスコードヒントメタデータ (例えば、図 10 に示すソースフォーマット定義 D S 1002) を示す図である。図 11 に示すように、このメタデータには、以下のような関連する記述子及び記述スキームが含まれる。

- ・ `bitrate` は、タイプ `int` であり、ソース A / V データストリームのビットレートをビット数 / 秒で記述する。
- ・ `size_of_pictures` は、タイプ `2*int` であり、ソース A / V フォーマットのピクチャの x 方向及び y 方向のサイズを記述する。
- ・ `number_of_frames_per_second` は、タイプ `int` であり、ソースコンテンツをフレーム数 / 秒を記述する。
- ・ `pel_aspect_ratio` は、タイプ `float` であり、画素アスペクト比を記述する。
- ・ `pel_color_depth` は、タイプ `int` であり、色の深さ (color depth) を記述する。
- ・ `usage_of_progressive_interlaced_format` は、サイズが 1 ビット であり、ソースフォーマットがインターレース走査フォーマットであるか、プログレッシブ走査フォーマットであるかを記述する。
- ・ `usage_of_frame_field_pictures` は、サイズが 1 ビット であり、フレームピクチャが使用されているか、フィールドピクチャが使用されているかを記述する。

- compression_methodは、タイプ int であり、ソースフォーマットにおいて使用されている圧縮方法を定義する。この圧縮方法は、MPEG-1、MPEG-2、MPEG-4、DV、H.263、H.261等が含まれるリストから選択してもよい。各圧縮方法に対し、以下のように、更なるパラメータを定義することもできる。

- GOP_structureは、I、P、B状態のランレンジス符号化されたデータフィールドを表す。例えば、MPEG-2ビデオデータにおいて、Iフレームのみしか存在しない場合、圧縮領域における直接変換が可能である。

【0046】

図12は、本発明に基づき、A/Vコンテンツ全体又は特定のA/Vセグメントに関連付けられたターゲットフォーマット定義トランスコードヒントメタデータを示す図である。図12に示すように、このメタデータには、以下のような関連する記述子及び記述スキームが含まれる。

- bitrateは、タイプ int であり、ターゲットA/Vデータストリームのビットレートをビット数／秒で記述する。
- size_of_picturesは、タイプ 2*int であり、ターゲットA/Vフォーマットのピクチャのx方向及びy方向のサイズを記述する。
- number_of_frames_per_secondは、タイプ int であり、ターゲットコンテンツをフレーム数／秒を記述する。
- pel_aspect_ratioは、タイプ float であり、画素アスペクト比を記述する。
- pel_color_depthは、タイプ int であり、色の深さを記述する。
- usage_of_progressive_interlaced_formatは、サイズが 1ビット であり、ソースフォーマットがインターレース走査フォーマットであるか、プログレッシブ走査フォーマットであるかを記述する。
- usage_of_frame_field_picturesは、サイズが 1ビット であり、フレームピクチャが使用されているか、フィールドピクチャが使用されているかを記述する。
- compression_methodはタイプ int であり、ソースフォーマットにおいて使用されている圧縮方法を定義する。この圧縮方法は、MPEG-1、MPEG-2、MPEG-4、DV、H.263、H.261等が含まれるリストから選択してもよい。各圧縮方法に対し、以下のように更なるパラメータを定義することもできる。
- GOP_structureは、I、P、B状態のランレンジス符号化されたデータフィールドを表す。この任意パラメータにより、固定されたGOP構造を強制的に使用することもできる。GOP構造を固定することにより、例えば、特定の位置でIフレームを強制的に使用して、編集を容易にすることができます。

【0047】

図13は、本発明に基づき、A/Vコンテンツ全体又は特定のA/Vセグメントに関連付けられた汎用トランスコードヒントメタデータ（例えば、図11に示す汎用トランスコードヒントDS1004）を示す図である。図13に示すように、このトランスコードヒントメタデータには、以下のような関連する記述子及び記述スキームが含まれる。

- use_region_of_interest_DSは、1ビット 長であり、検討中の記述スキーム（interest description scheme）の領域がトランスコードヒントメタデータとして利用できるか否かを示す。
- region_of_interest_DSが使用される場合、motion_trajectory_Dと共に、shape_D（例えば次のうちの1つであってもよい。boundary_box_D、MB_shape_D又は他のshape_D）を使用して検討中の領域を空間的及び時間的に記述してもよい。MB_shape_Dは、オブジェクト形状を記述するために、マクロブロック（16×16）サイズのブロックを使用してもよい。Motion_trajectory_Dは、既に時間についての概念を含んでいるので、region_of_interest_DSの開始フレーム及び終了フレームを定義することができる。region_of_interest_DSは、各shape_Dのサイズ及び各motion_trajectory_Dを含んでいてもよい。トランスコードコードアプリケーションについては、region_of_interest_DSを用いて、例えば、背景よりも重要な領域内のブロックにより多くのビットを使用するようにしてよい。M

MPEG-4 用の他のトランスコードアプリケーションは、重要な領域を独立した MPEG-4 オブジェクトで記述し、この領域に対しては、例えば背景等の他の MPEG-4 オブジェクトより高いビットレート及び高いフレームレートを使用する。region_of_interest_DS の抽出は、自動的に行ってもよく、ユーザの操作により行ってもよい。

- ・use_editing_effects_transcoding_hints_DS は、1 ビット 長であり、編集効果別トランスコードヒントが情報として利用可能であるか否かを示す。
- ・camera_flash は、カメラフラッシュが生じるフレーム番号を記述する各エントリのリストである。したがって、この記述子の長さは、カメラフラッシュイベントの数に int を乗算したものとなる。多くのビデオ（再）エンコーダ／トランスコーダは、輝度差（luminance difference）に基づく動き推定を行うので、トランスコードアプリケーションにとって、camera_flash 記述子は、非常に有用である。この動き推定については 1999 年にクルーワー・アカデミック出版社 (Kluwer Academic Publishers) から発行されたピーター・クーン (Peter Kuhn) 著「MPEG-4 動き推定のためのアルゴリズム、複雑性解析及び VLSI アーキテクチャ (Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation)」に開示されている輝度に基づく動き推定では、連続する 2 つのフレーム（一方がフラッシュを有し、他方がフラッシュを有さない）の 2 つのマクロブロック間の平均絶対誤差（mean absolute error）は、大きすぎて予測に使用することができず、したがってカメラフラッシュは、高いビットレートを使用するエフレームとして符号化しなくてはならない。したがって、トランスコードヒント記述スキーム（DS）内でカメラフラッシュを指示することにより、例えば輝度補正動き推定方法（luminance corrected motion estimation method）又は他のビットコストがそれ程高くないう ancor frame(s) からカメラフラッシュを有するフレームの予測を実行することができる。
- ・cross_fading は、クロスフェーディングの開始フレーム及び終了フレームを記述するエントリのリストである。したがって、記述子の長さは、クロスフェーディングイベントの数の 2 倍 int である。クロスフェーズ中のビットレート及び量子化処理の制御に有用である。クロスフェーズの期間において、予測の使用は通常制限され、これにより予測誤差符号化用のビットレートは大きくなる。また、クロスフェーディングの期間は、通常、シーンが不鮮明になるため、量子化スケール、ビットレート又は制御パラメータをそれぞれ調整することにより、ビットレートの上昇を制限してもよい。
- ・black_pictures は、黒ピクチャのシーケンスにおける開始フレームと終了フレームを記述するエントリのリストである。特に、家庭用ビデオにおいて、シーン間に黒ピクチャが生じる。経験的に、黒ピクチャの連続により、予測の使用が制限されるため、動き補償 DC エンコーダにおけるビットレートが増加する。したがって、このトランスコードヒント記述子を使用し、黒ピクチャの期間において、量子化スケール、ビットレート又は制御パラメータをそれぞれ調整することにより、ビットレートを制限することができる。
- ・fade_in は、cross_fading に類似する記述子であり、フェードイン効果の開始フレーム及び終了フレームを定義する複数のエントリを記述する。クロスフェーディングとは異なり、フェードインは、黒ピクチャから開始され、したがって、視覚に対するある種のマスク効果を利用して、量子化スケール、ビットレート又は制御パラメータをそれぞれ調整することにより、フェードイン期間におけるビットレートを制限することができる。
- ・fade_out は、シーンの後に黒ピクチャが来る点を除けば fade_in と同様の記述子である。
- ・abrupt_change は、フェーディングを行うことなく突然現れるシーン又はショットの位置を示すタイプ int の単一のフレーム番号のリストを記述する。これらのイベントは、例えば図 3 では、高く急峻なピークにより示される。これらのピークは、新たなカメラショット又はシーンの開始を示している。abrupt_change 編集効果は、フェーディング効果と対称的である。2 つのビデオセグメント間で突然の変更が生じた場合、人間の知覚は、新たな A/V セグメントに適応し、その詳細を認識するために、数ミリ秒を必要とする。こ

の人間の目の認識速度の遅さは、ビデオトランスコード処理にとって好適である。例えば、このため、シーン又はショットの突然の変更の後の最初のフレームのビットレートや量子化スケールパラメータを減少させたり変更したりすることができる。

- ・ `use_motion_transcoding_hints_DS` は、1 ビット 長であり、動き関連トランスコードヒントメタデータの使用を指示する。
- ・ `number_of_regions` は、タイプ `int` であり、後続する動き関連トランスコードヒントメタデータが有効である領域の数を示す。
- ・ `for_every_region` は、1 ビット 長のフィールドで表され、領域が長方形であるか任意の形状であるかを示す。領域が任意の形状である場合、領域記述子（例えば、形状記述子及び動き軌道記述子 (`motion trajectory descriptor`) からなる）を使用する。領域が長方形の場合、長方形領域のサイズを示す記述子を使用する。この領域内の動きフィールドは、各フレーム又はフレームのシーケンスに対する複数のパラメータによって決定されるパラメトリック動きモデルにより記述される。トランスコード処理においては、このソースビデオデータの実際の動きに関する動きデータを用いて、複雑な演算が必要な（再）符号化部の動き推定の探索領域を制限することができ、さらにビデオデータ内の動き量に応じて、高速且つ効率的なインターレース走査 / プログレッシブ走査（フレーム / フィールド）変換処理及び G O P（グループオブピクチャ）構造の決定を行うことができる。動きデータは、ビデオデータのサイズの変換処理にも好適に使用することができる。

【 0 0 4 8 】

図 14 は、セグメント別トランスコードヒントメタデータ（例えば、図 9 に示すセグメント別トランスコードヒント DS 9 0 9、9 1 0）を示す図である。このメタデータを用いて、本発明に基づき、一定の特性を有する A / V セグメントの（再）エンコーダ / トランスコーダの設定を決定することができる。図 14 に示すように、このメタデータには、以下のような関連する記述子及び記述スキームが含まれる。

- ・ `start_frame` は、タイプ `int` であり、A / V セグメントのトランスコードヒントメタデータの最初のフレーム番号を記述する。
- ・ `nframes` は、タイプ `int` であり、A / V セグメントの長さを記述する。
- ・ `I_frame_location` は、A / V セグメント内の I フレームの位置に関する複数の可能性を記述する。
- ・ `select_one_out_of_the_following` は、サイズが 2 ビット であり、以下の 4 つの I フレーム位置記述法のうちの 1 つを選択する。
 - ・ `first_frame` は、サイズが 1 ビット であり、デフォルトの I フレーム位置を示す。このメソッドは、最初のフレームのみが I フレームである A / V セグメントを記述し、後続の予測のためのアンカ (`anchor`) として使用され、この A / V セグメント内の他のフレームは全て P フレーム又は B フレームである。
 - ・ `List_of_frames` は、A / V セグメント内の I フレームのフレーム番号のリストを示す。このメソッドによれば、A / V セグメント内の I フレームを任意に記述することができる。このリスト内の k 個のフレームについて、この記述子のサイズは `k * int` である。
 - ・ `first_frame_and_every_k_frames` は、タイプ `int` であり、セグメント内の第 1 のフレームは I フレームであり、k は A / V セグメント内の I フレームの間隔 (`interval`) を記述する。
 - ・ `no_I_frame` は、サイズが 1 ビット であり、A / V セグメント内に I フレームが使用されていないことを示す。この記述子は、A / V セグメントの符号化処理が先行するセグメント内のアンカ (I フレーム) に基づいている場合に有用である。
 - ・ `quantizer_scale` は、タイプ `int` であり、A / V セグメントに対する初期の量子化スケールを記述する。
 - ・ `target_bitrate` は、タイプ `int` であり、A / V セグメントのビットレートをビット数 / 秒で記述する。
 - ・ `target_min_bitrate` は、タイプ `int` であり、A / V セグメントの最低の毎秒当たりのビットレートを記述する（この記述子は任意の記述子である）。

- ・target_max_bitrateは、タイプ int であり、A / Vセグメントの最高の毎秒当たりのビットレートを記述する（この記述子は任意の記述子である。）
- ・use_transcoding_statesは、サイズが 1ビット であり、A / Vセグメントに対してトランスコードヒント状態を使用するか否かを記述する。
- ・transcoding_state_nrは、タイプ int であり、セグメントに対するトランスコードヒントメタデータ状態を定義する。トランスコードヒントメタデータ状態は、トランスコードヒント状態テーブルにおけるエントリを指示するポインタである。トランスコードヒント状態テーブルは、複数のエントリを有することができ、さらにトランスコードヒントパラメータにより新たなエントリを追加し又は削除してもよい。単一のトランスコードヒント状態のトランスコードヒントメタデータについては、図16を用いて後に説明する。
- ・add_new_transcoding_stateは、サイズが 1ビット であり、関連する情報を有する新たなトランスコード状態をトランスコードヒント状態テーブルに追加する必要があるか否かを記述する。add_new_transcoding_stateが「yes」を示す場合、新たなトランスコードヒント状態のパラメータのリストが与えられる。パラメータリストのサイズは、1つのトランスコードヒント状態のパラメータ数及びトランスコードヒント状態の数により決定される。
- ・remove_transcoding_stateは、サイズが 1ビット のフラグであり、トランスコード状態を削除するか否かを示す。トランスコード状態を削除してもよい場合、削除されるトランスコード状態の状態番号（タイプ int ）が与えられる。
- ・use_encoding_complexity_descriptionは、サイズが 1ビット であり、図15に示すようなより詳細な符号化複雑性記述スキームを使用する必要があるか否かを示す。

【0049】

図15は、符号化複雑性トランスコードヒントメタデータ（coding complexity transcoding hints metadata）を示し、このメタデータは、本発明に基づき、A / Vコンテンツ全体又は特定のA / Vセグメントに関連付けることができる。符号化複雑性トランスコードヒントメタデータは、量子化器及びビットレートの設定のレート制御及び決定に使用することができる。

- ・use_feature_pomtsは、サイズが 1ビット であり、特徴点に基づく複雑性推定データ（complexity estimation data）の使用を示す。
- ・select_feature_point_methodは、サイズが 2ビット であり、特徴点に関する方法を選択する。
- ・number_of_new_feature_points_per_frameは、サイズが nframes*int であり、図3に示すようなフレーム毎の新たな特徴点の数を記述する。このメトリックは、フレーム毎の新たなコンテンツの量を示す。
- ・feature_point_metricsは、1つのセグメント内のフレーム毎の新たな特徴点に基づくメトリックのリストを記述する。このメトリックは、フレーム毎の新たな特徴点の平均、最大、最小、平方偏差、標準偏差を順番に表す int 値のリストとして表される。
- ・use_equation_descriptionは、int ポインタであり、数式に基づいて、フレーム毎の符号化複雑性を記述する。
- ・use_motion_descriptionは、サイズが 1ビット であり、動きに基づく複雑性記述を使用するか否かを示す。
- ・select_motion_methodは、サイズが 4ビット であり、動き記述方法を選択する。
- ・param_k_motionは、サイズが nframes*k*int であり、包括的パラメトリック動きモデルの各单一のフレーム用のk個のパラメータを記述する。
- ・motion_metricsは、動きベクトルの大きさに基づいてセグメント全体のメトリックのリストを記述する。このメトリックは、マクロブロック動きベクトルの平均、最大、最小、平方偏差、標準偏差を順番に表す int 値のリストとして表される。
- ・block_motion_fieldは、(m*m) ブロックサイズの動きフィールドの各ベクトルを記述し、サイズが nframes*int*size_x*size_y/(m*m) を有する。
- ・use_texture_edge_metricsは、テクスチャ又はエッジメトリックが使用された場合に設

定されるサイズが 1 ビット のフラグである。

- ・ `select_texture_edge_metrics` は、 サイズが 4 ビット であり、 以降にどのテクスチャメトリックを使用するかを決定する。
- ・ `DCT_block_energy` は、 1 つのブロックの全ての D C T 係数の総和を表し、 フレーム内の各ブロックについて定義されている。この記述子のサイズは、 `size_y * サイズ_x * nframes * int / 64` である。
- ・ `DCT_block_activity` は、 D C 係数以外の 1 つのブロックの全ての D C T 係数の和を定義する。この記述子は、 フレーム内の各ブロックに対して定義され、 サイズが `size_y * size_x * nframes * int / 64` を有する。
- ・ `DCT_energy_metric` は、 各ブロックの個別の D C T エネルギに基づいて、 セグメント全体のメトリックのリストを記述する。このメトリックは、 全ての個別の D C T エネルギメトリックの平均、 最大、 最小、 平方偏差、 標準偏差を順番に表す `int` 値のリストとして表される。この記述子のサイズは `6 * int` である。この記述子の変形例として、 ビデオセグメントの各单一のフレームに対する D C T エネルギメトリックを記述してもよい。
- ・ `DCT_activity_metric` は、 各ブロックの個別の D C T アクティビティ (`DCT_activities`) に基づいて、 セグメント全体のメトリックのリストを記述する。このメトリックは、 全ての個別の D C T アクティビティメトリックの平均、 最大、 最小、 平方偏差、 標準偏差を順番に表す `int` 値のリストとして表される。この記述子のサイズは `6 * int` である。この記述子の変形例として、 ビデオセグメントの各单一のフレームに対する D C T アクティビティメトリックを記述してもよい。

【 0 0 5 0 】

図 16 は、 本発明に基づき、 A / V コンテンツ全体又は特定の A / V セグメントに関連付けることができるトランスクードヒントメタデータを示す図である。このメタデータに関連する記述子及び記述スキームとしては、 例えば以下のようなものがある。

- ・ `M` は、 タイプ `int` であり、 I フレーム / P フレーム 距離を記述する。
- ・ `bitrate_fraction_for_I` は、 タイプ `float` であり、 A / V セグメントに対して定義される I フレームに使用できるビットレートの割合を記述する。
- ・ `bitrate_fraction_for_P` は、 タイプ `float` であり、 A / V セグメントに対して定義される P フレームに使用できるビットレートの割合を記述する。 B フレーム用のビットレートの割合は、 100 % からこれらの割合を引いた残りである。
- ・ `quantizer_scale_ratio_I_P` は、 タイプ `float` であり、 I フレームと P フレーム間の量子化スケールの関係 (このセグメントに対して定義される) を示す。
- ・ `quantizer_scale_ratio_I_B` は、 タイプ `float` であり、 I フレームと B フレーム間の量子化スケールの関係 (このセグメントに対して定義される) を示す。 なお、 ビットレート記述子 (`bitrate_fraction_for_I < bitrate_fraction_for_P`) 、 `quantizer_scale_ratio` 記述子 (`quantizer_scale_ratio_I_P`、 `quantizer_scale_ratio_I_B`) 、 又は以下のレート制御パラメータのいずれかは、 必ず設けなくてはならない。
- ・ `X_I`、 `X_P`、 `X_B` は、 `frame_vbv_complexities` であり、 それぞれタイプ `int` であり、 フレーム単位の圧縮ターゲットフォーマット (図 12) の場合に定義される。 これらの及び以下の仮想バッファ検証 (Virtual Buffer Verifier : V B V) 複雑性の調整は任意であり、 ソースコンテンツの特性及びターゲットフォーマットの定義に応じてレート制御スキームを変更するために使用できる。
- ・ `X_I_top`、 `X_P_top`、 `X_B_top` は、 トップフィールド用の `field_vbv_complexities` であり、 それぞれタイプ `int` であり、 フィールド単位の圧縮ターゲットフォーマット (図 12) の場合に定義される。
- ・ `X_I_bot`、 `X_P_bot`、 `X_B_bot` は、 ボトムフィールド用の `field_vbv_complexities` であり、 それぞれタイプ `int` であり、 フィールド単位の圧縮ターゲットフォーマット (図 12) の場合に定義される。

【 0 0 5 1 】

以上、 本発明を詳細に説明したが、 上述の詳細事項は、 本発明の主旨及び範囲から逸脱

することなく、様々に構造を変更して本発明の方法を実施できることは明らかである。上述の説明に含まれ、添付の図面に示した全ての事項は、例示的なものにすぎず、本発明を限定するものと解釈してはならない。

【0052】

さらに、請求の範囲は、ここに説明した発明の包括的及び特定の特徴の全てのカバーするものであり、請求の範囲における全ての表現は、この範囲に適合するものと解釈しなくてはならない。

【図面の簡単な説明】

【図1】

様々なA/V端末機器を備えるホームネットワーク環境において本発明に基づくトランスコード処理を行うシステムの構成を示す図である。

【図2】

本発明に基づくトランスコードヒントメタデータの抽出（グループオブピクチャ：GOP）を説明する図である。

【図3】

本発明に基づくフレーム毎の新たな特徴点の数に基づくトランスコード状態の選択を説明する図である。

【図4】

本発明に基づく3つの状態を有するトランスコードヒント状態の具体例を示す図である。

【図5】

本発明に基づき、圧縮ソースコンテンツ及び非圧縮ソースコンテンツからトランスコードヒントメタデータを抽出する手法を説明する図である。

【図6】

本発明に基づくビデオセグメント化及びトランスコードヒント状態の選択の手順を説明する図である。

【図7】

本発明に基づき新たなビデオセグメント（又は新たなGOP）の境界を判定する処理を説明する図である。

【図8】

本発明に基づきトランスコードヒント状態を選択するアルゴリズムを示す図である。

【図9】

トランスコードヒントメタデータの構造的な組織を説明する図である。

【図10】

汎用トランスコードヒントメタデータ記述スキームの構造的な組織を説明する図である。

【図11】

本発明に基づくソースフォーマット定義のためのトランスコードヒントメタデータを示す図である。

【図12】

本発明に基づくターゲットデータ定義のためのトランスコードヒントメタデータを示す図である。

【図13】

本発明に基づく汎用トランスコードヒントメタデータ表現を示す図である。

【図14】

本発明に基づくセグメント別トランスコードヒントメタデータ表現を示す図である。

【図15】

本発明に基づく符号化複雑性トランスコードヒントメタデータを示す図である。

【図16】

本発明に基づくトランスコードヒント状態メタデータを示す図である。

【手続補正2】

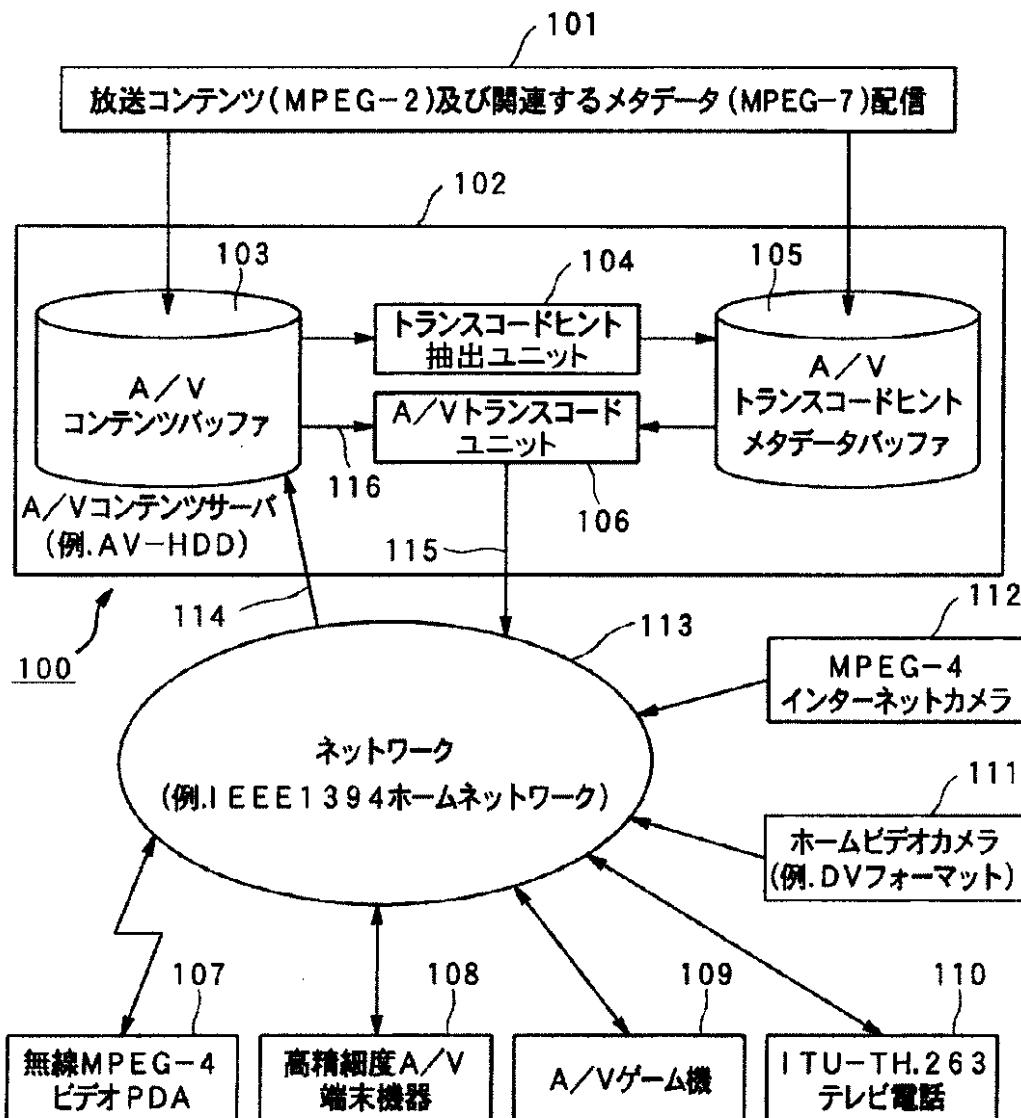
【補正対象書類名】図面

【補正対象項目名】図1

【補正方法】変更

【補正の内容】

【図1】



【手続補正3】

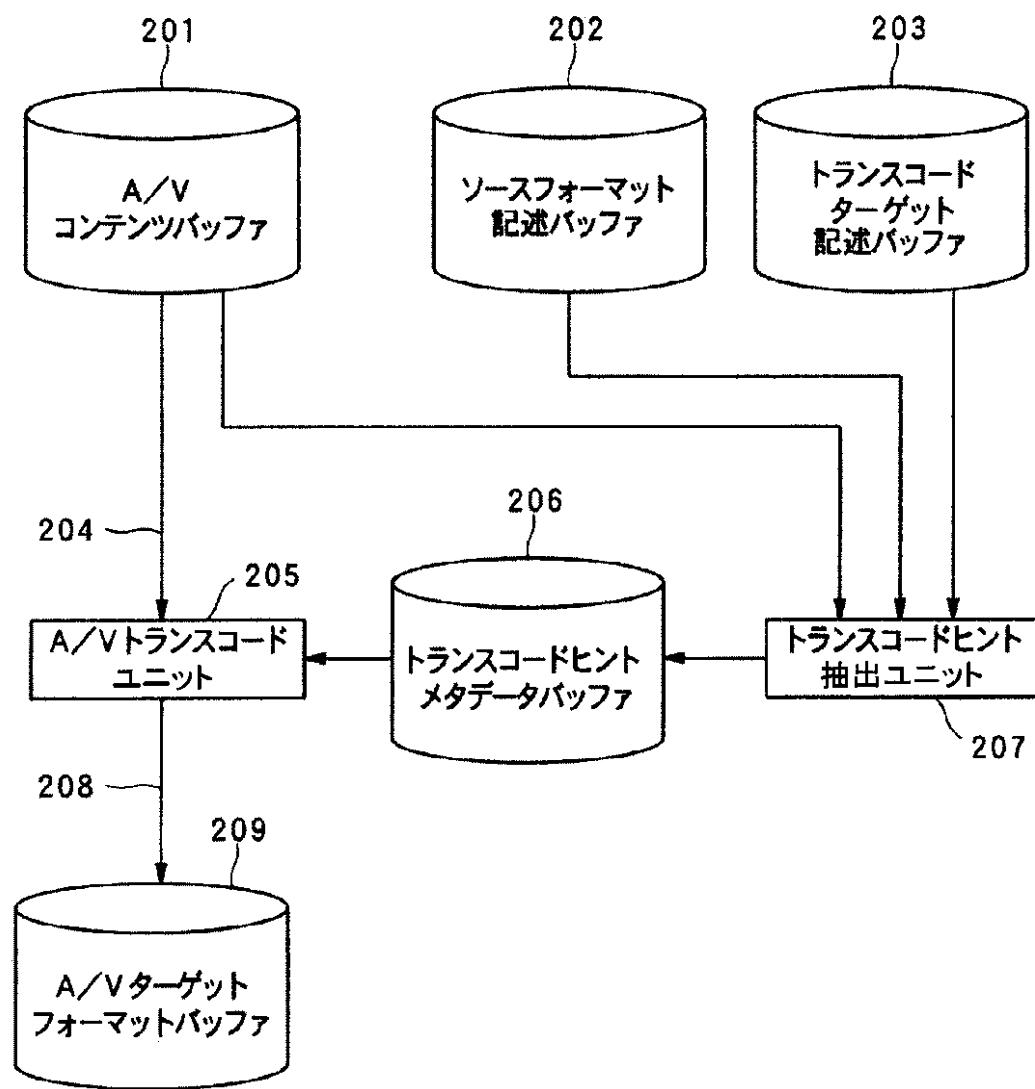
【補正対象書類名】図面

【補正対象項目名】図2

【補正方法】変更

【補正の内容】

【図2】



【手続補正4】

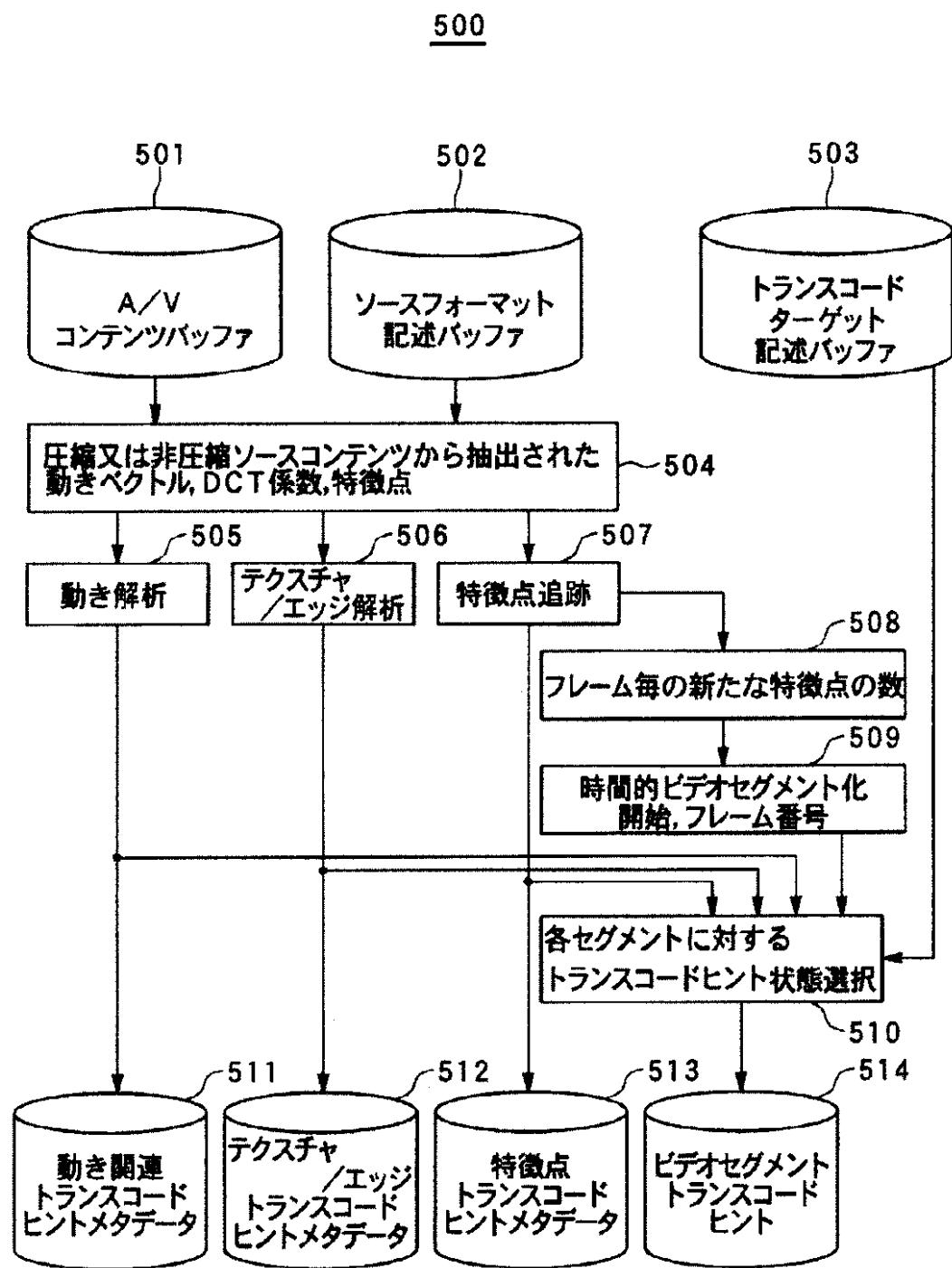
【補正対象書類名】図面

【補正対象項目名】図5

【補正方法】変更

【補正の内容】

【図5】



【手続補正5】

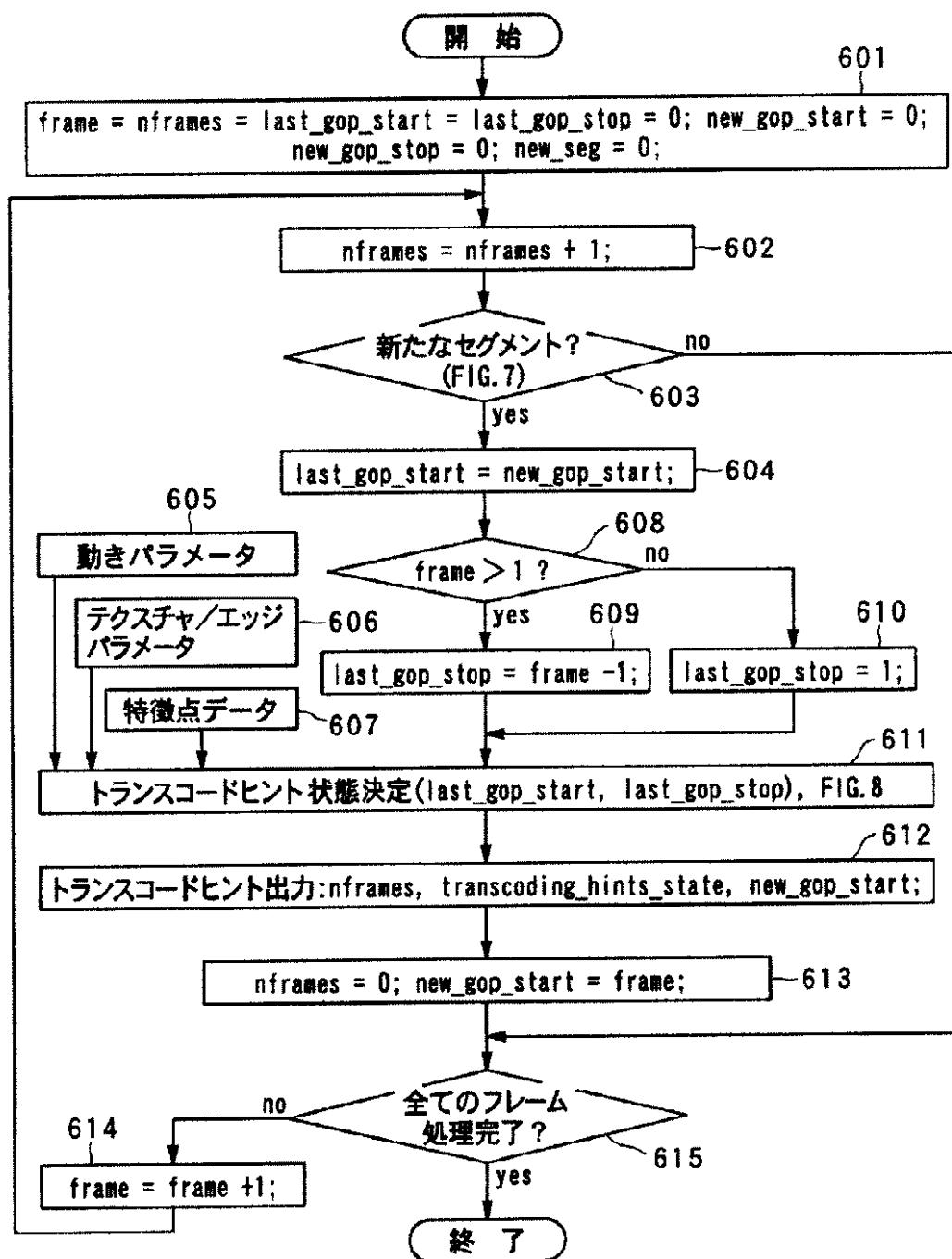
【補正対象書類名】図面

【補正対象項目名】図6

【補正方法】変更

【補正の内容】

【図6】



【手続補正6】

【補正対象書類名】図面

【補正対象項目名】図8

【補正方法】変更

【補正の内容】

【図8】

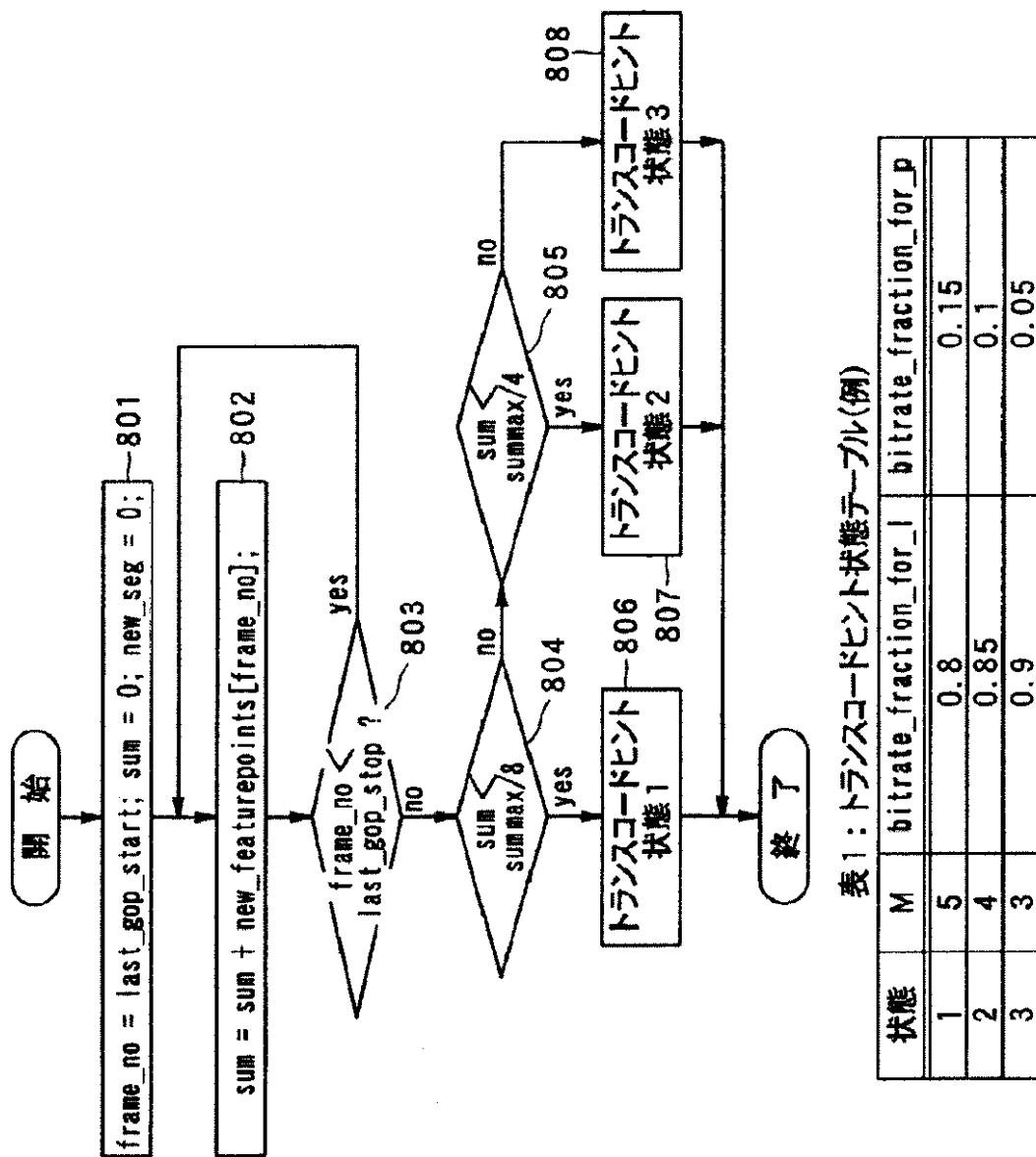


表1：トランスクードヒント状態テーブル(例)

状態	M	bitrate_fraction_for_l	bitrate_fraction_for_p
1	5	0.8	0.15
2	4	0.85	0.1
3	3	0.9	0.05

【手続補正7】

【補正対象書類名】図面

【補正対象項目名】図9

【補正方法】変更

【補正の内容】

【図9】

