

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4447689号
(P4447689)

(45) 発行日 平成22年4月7日(2010.4.7)

(24) 登録日 平成22年1月29日(2010.1.29)

(51) Int.Cl.

G06T 7/20 (2006.01)

F 1

G06T 7/20

C

請求項の数 2 (全 19 頁)

(21) 出願番号 特願平11-127728
 (22) 出願日 平成11年5月7日(1999.5.7)
 (65) 公開番号 特開2000-30057(P2000-30057A)
 (43) 公開日 平成12年1月28日(2000.1.28)
 審査請求日 平成18年5月8日(2006.5.8)
 (31) 優先権主張番号 PP3407
 (32) 優先日 平成10年5月7日(1998.5.7)
 (33) 優先権主張国 オーストラリア(AU)

(73) 特許権者 000001007
 キヤノン株式会社
 東京都大田区下丸子3丁目30番2号
 (74) 代理人 100076428
 弁理士 大塚 康徳
 (74) 代理人 100112508
 弁理士 高柳 司郎
 (74) 代理人 100115071
 弁理士 大塚 康弘
 (74) 代理人 100116894
 弁理士 木村 秀二

最終頁に続く

(54) 【発明の名称】自動映像解釈システム

(57) 【特許請求の範囲】

【請求項 1】

奥行きを判定するための情報を含む複数種類の状況情報を有するデジタルビデオ信号を解釈する装置であって、

前記デジタルビデオ信号を、各々が前記複数種類の状況情報の対応する一部分を有する1つ又は複数のビデオセグメントに分割する手段と、

前記複数種類の状況情報の前記対応する部分を状況アナライザにより解析した出力に従って領域アナライザにおいて使用されるアプリケーションドメインのラベルの先駆的確率を変更し、該領域アナライザによりそれぞれのビデオセグメントに含まれる1つ又は複数の領域に關わるラベル付き三次元隣接度グラフを形成する解析手段と
10
を具備する装置。

【請求項 2】

奥行きを判定するための情報を含む複数種類の状況情報を有するデジタルビデオ信号を解釈するためのコンピュータプログラムが記録されているコンピュータ読み取り可能媒体であって、

前記デジタルビデオ信号を、各々が前記複数種類の状況情報の対応する一部分を有する1つ又は複数のビデオセグメントに分割する手段と、

前記複数種類の状況情報の前記対応する部分を状況アナライザにより解析した出力に従って領域アナライザにおいて使用されるアプリケーションドメインのラベルの先駆的確率を変更し、該領域アナライザによりそれぞれのビデオセグメントに含まれる1つ又は複数
20

の領域に関わるラベル付き三次元隣接度グラフを形成する解析手段としてコンピュータを機能させるコンピュータプログラムが記録されているコンピュータ読み取り可能媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明はデジタルビデオ信号の統計解析に関し、特に、意味ラベルに関する自動内容解釈のためのデジタルビデオ信号の統計解析に関する。ラベルは、その後、内容に基づく検索やビデオアブストラクトの生成などのタスクの基礎として利用できる。

【0002】

10

【従来の技術】

一般的には、デジタルビデオは、視覚シーンの時間的展開を表現する信号であると想定される。通常、この信号は関連する音声情報と共に符号化される(例えば、MPEG-2オーディオビジュアル符号化フォーマット)。場合によっては、ビデオ信号及び音声信号と共に、シーン又はシーンの捕捉に関する情報が符号化されるときもあるだろう。デジタルビデオは、典型的には、静止デジタル画像又はフレームのシーケンスにより表現され、各デジタル画像は複数のカラーチャネル(例えば、R、G、B)に対する画素輝度の集合から構成される。この表現の大部分は、視覚シーンが感知されるグリッド方式による。

【0003】

20

視覚信号の内容を音声信号の中に見出すことができ、又、音声信号の内容を視覚信号の中に見出すことができるという意味で、視覚信号と、関連する音声信号とは互いに相関関係にある場合が多い。この相関は、MPEG-4などの、符号化の単位が1つのシーンの中の空間的位置及び時間的位置を限定されたオーディオビジュアルオブジェクトである最近のデジタルオーディオビジュアル符号化フォーマットにおいては、明白に認められる。このようなオーディオビジュアル情報の表現は、デジタル素材の利用に、より適応するものであるが、自然のシーンの視覚成分は依然としてグリッドに基づく感知技法を使用して捕捉されるのが普通である(すなわち、デジタル画像はキャプチャ装置により規定されるフレームレートで感知される)。従って、通常、デジタルビデオ解釈のプロセスは依然としてデジタル画像解釈のプロセスを基礎としており、関連する音声信号とは別個に考慮されるのが普通である。

30

【0004】

デジタル画像信号の解釈は、画像中の意味のあるオブジェクト又は領域の識別を通して画像の内容を理解し且つそれらの空間的配列を解析するプロセスである。従来、画像解釈のタスクは人間による解析を必要としていた。これにはコストも、時間もかかり、そのため、自動画像解釈システムを構築すべく、多くの研究が重ねられてきた。

【0005】

40

既存の多くの画像解釈システムは低レベル処理と高レベル処理とを含む。通常、低レベル処理は、画像を画素輝度のアレイから辺と領域のような空間的に関連する画像プリミティブの集合に変換する。その後、プリミティブ(例えば、平均画素輝度)から様々な特徴を抽出することができる。高レベル処理では、画像ドメインの知識と特徴測定を利用して、プリミティブにオブジェクトラベル又は領域ラベル、すなわち、解釈を割り当て、それにより、「画像の中に何が存在しているか」に関する記述を構築する。

【0006】

初期の段階の画像解釈の試みは、個別のプリミティブをその特徴測定値に従って限られた数のオブジェクトクラスに分類することに基づいていた。この方式の成功には、低レベル処理で頻繁に発生する誤りを含む結果又は不完全な結果や、画像中の雑音に起因する特徴測定値の誤差によって限界があった。最近の技法の多くは高レベル処理に空間的制約を取り入れている。すなわち、不明瞭な領域又はオブジェクトは、隣接する領域又はオブジェクトの認識が成功した結果として認識されることが多い。

【0007】

50

更に、時を経て、画像に関する領域ラベルの空間的依存性をマルコフ確率場(MRF)などの統計的方法を使用してモデリングする技法が現れた。MRFモデルの主な利点は、空間的に関連する確率変数の相互作用の一般的で自然なモデルを提供し、場の(大域的に)最適の実現形態を見出すために使用できる最適化アルゴリズムが相対的にフレキシブルであるというところにある。通常、MRFは、一般には領域隣接度グラフ(RAG)と呼ばれるセグメントに分割された複数の領域のグラフにおいて定義される。分割された領域は、利用できる数多くの領域に基づく画像セグメンテーション方法のいずれか1つによって生成することができる。MRFモデルは、画像からの測定値(低レベル特徴)に対するラベルの依存性と共に意味ラベルの空間的依存性に関する知識を取り入れる強力なメカニズムを構成する。

10

【0008】

デジタル音声信号解釈は、語/句、すなわち、キーとなる音声の識別を通して音声信号の内容を理解し、それらの時間的配列を解析するプロセスである。通常、デジタル音声解析の研究は、結果として考えられる技術に対して、例えば、コンピュータ及び他の電子装置のための自然言語インターフェースのような多数の用途が可能であるために、音声認識に集中していた。

【0009】

隠れマルコフモデルは、本来、デジタル音声信号の逐次且つ統計的な特性を取り入れる能力を備えているため、連続音声認識に広く使用されている。このモデルは、音声の単位(音素、又は場合によっては語)が状態の集合を介して時間のシーケンスとして表現される時変プロセスのモデリングのための確率の枠組みを構成する。状態間の遷移の確率を推定するには、音声の単位ごとの標本音声信号の集合の解析を行う必要がある(すなわち、トレーニングセット)。認識プロセスが話者とは無関係のものでなければならない場合には、トレーニングセットはある範囲の話者からの標本音声信号を含んでいなければならない。

20

【0010】

【発明の概要】

本発明の1つの面によれば、奥行きを判定するための情報を含む複数種類の状況情報を有するデジタルビデオ信号を解釈する装置であって、前記デジタルビデオ信号を、各々が前記複数種類の状況情報の対応する一部分を有する1つ又は複数のビデオセグメントに分割する手段と、前記複数種類の状況情報の前記対応する部分を状況アナライザにより解析した出力に従って領域アナライザにおいて使用されるアプリケーションドメインのラベルの先駆的確率を変更し、該領域アナライザによりそれぞれのビデオセグメントに含まれる1つ又は複数の領域に関わるラベル付き三次元隣接度グラフを形成する解析手段とを具備する装置が提供される。

30

【0012】

本発明の更に別の面によれば、奥行きを判定するための情報を含む複数種類の状況情報を有するデジタルビデオ信号を解釈するためのコンピュータプログラムが記録されているコンピュータ読み取り可能媒体であって、前記デジタルビデオ信号を、各々が前記複数種類の状況情報の対応する一部分を有する1つ又は複数のビデオセグメントに分割する手段と、前記複数種類の状況情報の前記対応する部分を状況アナライザにより解析した出力に従って領域アナライザにおいて使用されるアプリケーションドメインのラベルの先駆的確率を変更し、該領域アナライザによりそれぞれのビデオセグメントに含まれる1つ又は複数の領域に関わるラベル付き三次元隣接度グラフを形成する解析手段としてコンピュータを機能させるコンピュータプログラムが記録されているコンピュータ読み取り可能媒体が提供される。

40

【0013】

【発明の実施の形態】

以下、添付の図面を参照して本発明の実施例を説明する。

【0014】

50

1. 概要

本発明は、空間ドメイン及び時間ドメインにおいてデジタルビデオ信号の(重要な)意味内容を捕捉する、デジタルビデオ信号の省略型高レベル記述を自動的に生成する方法、装置及びシステムに関する。そのような記述は、後に、内容に基づく検索、ビデオシーケンスのブラウジング又はデジタルビデオのアブストラクトの作成を含む数多くの目的に使用できる。

【0015】

デジタルビデオ信号は、ビデオキャプチャ装置で記録される視覚信号であると解される。この信号は、通常、二次元センサアレイ(例えば、CCDアレイ)から指定のサンプリング速度で生成されるが、必ずそうであるとは限らないであろう。各サンプルは1つの(ビデオ)フレームにより表現される。この信号の空間的及び時間的内容の解析は、ある範囲にわたる状況情報から恩恵を受ける。場合によっては、この状況情報はデジタルビデオ信号の中に含まれており(例えば、動き)、また、他の関連する供給源(例えば、関連する音声信号、記録されたカメラのパラメータ、一般に使用される視覚スペクトルのセンサとは別の他のセンサ)から情報を利用できる場合もある。利用可能な状況情報の範囲は静止画像解析プロセスで利用可能な情報と比べてはるかに広く、時間の進展という付加的な特性を有する。

10

【0016】

ビデオ信号におけるデジタル画像信号の時間の進展を利用して、デジタルビデオ信号の画像解釈プロセスの結果を改善することができる。例えば、動き情報を使用して、デジタルで記録されたシーンの中で動いている人間のようなオブジェクトの検出を支援することができる。また、動きを利用して、画像フレームの中の複数の領域をシーンの背景の一部であるとして選択的にグループ分けすることも可能である。デジタルビデオ信号を解釈又は理解するプロセスは、関連する音声信号の中にある音声要素(音声及び非音声)の識別によっても恩恵を被るであろう。例えば、音声信号で識別された言葉によって、解釈プロセスを支援することも可能であろう。また、野生動物のドキュメンタリーと関連する音声信号は様々な動物が発する音を含み、これがビデオ信号の内容を識別する上で助けになると考えられる。

20

【0017】

本発明の実施例は、MRFなどの確率モデルの使用をデジタルビデオ信号の解釈に拡張している。これは、様々に異なる供給源(例えば、ビデオフレーム、音声信号など)からの、測定値及び知識の形態をとる情報を单一の最適化手続きで統合するために、ビデオシーケンスを通して確率モデルを繰り返し使用することを含む。

30

【0018】

本発明の実施例においては、デジタルビデオ信号の全体にわたって、選択された解析イベントで高レベル記述を生成する。高レベル記述は、その解析イベントで現れる様々な領域への様々な意味ラベルの割り当てに基づいている。各解析イベントにおいて、その解析イベントの中心に置かれたビデオフレームは自動的に複数の均質領域に空間的に分割される。それらの領域と、それらの空間的隣接度特性は領域隣接度グラフ(RAG)により表現される。そこで、確率モデルをRAGに適用する。モデルはフレームの領域からの特徴測定値と、フレームの周囲の注目領域(ROI)からの状況情報と、RAGの領域と関連する可能性が考えられる様々な意味ラベルに関する先駆的知識とを含む。それらの意味ラベル(例えば、「人」、「空」、「水」、「木の葉」など)は、典型的には適切なアプリケーションドメイン(例えば、屋外のシーン、結婚式、都市のシーンなど)について構築されているリストから取り出される。

40

【0019】

各解析イベントにおいて、状況情報を使用して、選択された適切なアプリケーションドメインの意味ラベル(以下、ラベルという)の先駆的確率に偏りを与える。所定の解析イベントで実行される解析も、先行する解析イベントに依存する。この依存性は、通常、時間ドメインにおいて2つの解析イベントが互いに近接している場合にはより大きくなる。例え

50

ば、1つのビデオセグメントの中では、必ずそうであるとは限らないものの、先行する最近の解析イベントで領域について選択されたラベルはデジタルビデオの現在セクションの記述の中では選択されなかったラベルに比べてより高い確率を有するといえる。

【0020】

デジタルビデオ解釈システムは、アプリケーションドメインが1つであっても、あるいは複数であっても動作することができる。複数のアプリケーションドメインを使用している場合には、状況情報を使用して、最も確率の高いアプリケーションドメインを判定することができる。アプリケーションドメインは狭くても(すなわち、ラベルが数個)、広くても(すなわち、多数のラベルを使用できる)、どちらでも良い。狭いアプリケーションドメインは、通常、非常に特定的で、信頼性の高い領域のラベル付けが求められる場合に使用されると考えられる。例えば、セキュリティに適用するときには、人間や自動車と関連する領域を識別する能力は望ましいであろうが、それらのオブジェクトの識別は高い信頼性をもって要求されるであろう。

【0021】

以下の実施例の詳細な説明においては、説明を更に徹底して理解させるために、ビデオ符号化技法、センサの種類などの特定の詳細な事項を数多く挙げる。しかし、そのような特定の詳細を挙げなくても本発明を実施しうることは、当業者には明白であろう。また、場合によっては、本発明をわかりにくくしないように、ビデオフォーマット、音声フォーマットなどの周知の特徴の詳細な説明を省略した。

【0022】

2. 好ましい実施例のデジタルビデオ解釈システム

図1は、本発明の好ましい実施例による確率デジタルビデオ解釈システム160を示す。デジタルビデオ解釈システム160はビデオセグメンタ120と、ビデオセグメントアナライザ140とを含み、デジタルビデオ源100から発生されたデジタルビデオ源出力110を処理する。デジタルビデオ源はデジタルビデオカメラであるのが好ましい。ビデオセグメンタ120はデジタルビデオ源100と、ビデオセグメントアナライザ140との間に結合している。

【0023】

デジタルビデオ解釈システム160は、デジタルビデオ源100と関連して、任意に内部又は外部で実現されれば良い。デジタルビデオ解釈システム160がデジタルビデオ源100(例えば、デジタルカメラ)の内部に配置されている場合、解釈システム160は、付加的なカメラ情報を典型的にオーディオビジュアル信号を構成するビデオ信号及び音声信号を明示して格納する必要なく、そのようなカメラ情報を容易に利用することができる。例えば、デジタルビデオカメラにおけるカメラの動きを示す情報を使用して、デジタルビデオ信号110Aの動き解析を補助することもできるであろう。更に、撮影者の視線位置に基づいて、シーン中の重要な被写体がどこに位置しているかに関する情報を得ることや、焦点情報(又は他の距離情報)を使用して、図3Bに示すようにRAGに対する奥行き軸を生成するための奥行き情報を生成することも可能であろう。

【0024】

デジタルビデオ解釈システム160に対する入力は、デジタルビデオカメラなどの装置を使用して捕捉されるデジタルビデオ源出力110である。通常、デジタルビデオ源出力110はデジタルビデオ信号110Aと、デジタル音声信号110Bとから構成されている。キャプチャ装置に応じて、記録されるシーンに関する追加情報110Cも利用できるであろう。この追加情報110Cとしては、カメラパラメータ(焦点情報、露出の詳細、撮影者の視線位置など)や、その他のセンサ情報(例えば、赤外線感知)が考えられるであろう。

【0025】

デジタルビデオ解釈システム160においては、ビデオセグメンタ120は、デジタルビデオ信号をその出力側で提供される複数の時間ビデオセグメント又はスロット130に分割する。ビデオセグメンタ120により生成されたビデオセグメント130は、ビデオセ

10

20

30

40

50

グメントアナライザ 140 への入力として提供される。ビデオセグメントアナライザ 140 は、ビデオセグメントごとに 130 ラベル付き R A G のシーケンスを生成する。

【 0 0 2 6 】

図 1 のビデオセグメントアナライザ 140 は、1 つ又は複数の適切なアプリケーションドメインを使用して、R A G のシーケンスを生成し、次に、そのシーケンスの領域に最適のラベルを付けようとする。これにより得られるラベル付き R A G のシーケンス 150 はデジタルビデオ信号 110 A の内容を表現し、以下の説明においては、これをメタデータという。

【 0 0 2 7 】

図 9 に描かれたコンピュータ 900 により示されるように、本発明の実施例をデジタルビデオ源に対して外部で実現することができる。あるいは、図 10 に描かれているように、デジタルビデオ源 1000 の内部でデジタルビデオ解釈システムを実現しても良い。

10

【 0 0 2 8 】

図 9 を参照すると、汎用コンピュータは遠隔デジタルビデオ源 1000 に結合している。ビデオ解釈システムは、コンピュータにロードされ、コンピュータにより実行することができるコンピュータ読み取り可能媒体に記録されるソフトウェアとして実現されている。コンピュータ 900 はコンピュータモジュール 902 と、ビデオ表示モニタ 904 と、入力装置 920, 922 とを具備する。コンピュータモジュール 902 自体は少なくとも 1 つの中央処理装置 912 と、典型的にはランダムアクセスメモリ (R A M) 及び読み取り専用メモリ (R O M) を含むメモリユニット 916 と、ビデオインターフェース 906 を含む入出力 (I / O) インタフェース 906, 908, 914 とを具備する。I / O インタフェース 908 はデジタルビデオ源 1000 をコンピュータモジュール 902 及びマウス 922 などの指示装置と結合させることができる。記憶装置 910 は、フロッピーディスク、ハードディスクドライブ、C D - R O M ドライブ、磁気テープドライブ又は当業者には知られている同様の不揮発性記憶装置のうち 1 つ又は 2 つ以上を含んでいても良い。コンピュータ 902 の構成要素 906 から 916 は、通常、相互結合バス 918 を介して、当業者に知られているコンピュータシステム 900 の通常の動作モードが得られるように通信する。本発明の実施例を実施できるコンピュータシステムの例としては、I B M P C / A T 及びその互換機、マッキントッシュコンピュータ、サンスパークステーション、又は当業者に良く知られている多数のコンピュータシステムのうちいずれかなどがある。デジタルビデオ源 1000 は、ビデオ信号を記憶装置(例えば、メモリ、磁気記録媒体など)に記録でき且つ追加情報、例えば、ビデオ信号と関連する赤外線データを記録できるデジタルカメラであるのが好ましい。デジタルビデオ信号及び関連する追加(状況)情報はコンピュータ 900 にダウンロードされても良く、コンピュータ 900 は本発明の実施例に従って解釈及びラベル付けのプロセスを実行する。

20

【 0 0 2 9 】

あるいは、本発明の実施例をデジタルビデオカメラであるのが好ましいデジタルビデオ源 1000 の内部で実施しても差し支えない。デジタルビデオ源 1000 は、画像を捕捉するビデオキャプチャ装置 1002(例えば、電荷結合素子を含む)を具備し、ビデオキャプチャ装置 1002 は、その焦点データ及びその他の設定データを提供するためのセンサ及び / 又は機構を有する。デジタルビデオ源 1000 は、音声情報、周囲及び / 又は環境データ、位置決めデータ(例えば、G P S 情報)などを捕捉するセンサ 1004 を含んでいても良い。これらのセンサ 1004 及びビデオキャプチャ装置 1002 は、本発明の実施例を実施しうるデジタルビデオ源 1000 の中央処理装置に接続している。処理装置 1006 は、メモリ 1008 と、通信ポート 1010 と、ユーザインターフェースユニット 1012 とに結合している。ユーザインターフェースユニット 1012 により、ビデオ源 1000 の撮影者はデジタルビデオ源 1000 の動作モードにおける種々の設定値を指定することができる。例えば、デジタルビデオ源の撮影者は解釈システムと共に使用すべき複数の異なるアプリケーションドメイン(例えば、屋外のシーン、都市のシーン、結婚式のシーンなど)を選択できる。アプリケーションドメインを電子的に、又は使用可能な無線リンク

30

40

50

を介してキャプチャ装置にダウンロードすることもできるであろう。メモリ 1008 はランダムアクセスメモリ、読み取り専用メモリ及び／又は不揮発性記憶装置を具備しても良い。処理装置を動作させるためのデータと処理命令は、共に、メモリ 1008 に格納されても良い。通信ポート 1010 は、デジタルビデオ源 1000 と図 9 のコンピュータ 900 のような外部装置との通信を成立させる。通信ポート 1010 は、メモリ 1008 及び処理装置 1006 との間でデータ及び命令の送受信を行うことができる。

【0030】

3. 好ましい実施例のビデオセグメンタ

ビデオセグメンタ 120 は、デジタルビデオ信号を複数の時間ビデオセグメント又はスロット 130 に分割する。フレーム中の画素の動きに関する情報(暗黙にデジタルビデオ信号 110A の中にある)、及び、デジタル音声信号 110B 又はその他の情報 110C で利用できる何らかの他の支援情報の両方又はいずれかを使用して、ビデオセグメンタ 120 による分割を支援することができる。例えば、撮影者が記録を開始した時点及び停止した時点に関する情報を利用できるとすれば、この情報に基づいてビデオ分割を実行できるであろう。ビデオセグメンタにおいて周知のビデオ分割技法を実現しても、本発明の趣旨から逸脱することにはならない。

【0031】

4. 好ましい実施例のビデオセグメントアナライザ

ビデオセグメントアナライザ 140 は、ビデオセグメント 130 ごとにラベル付き RAG のシーケンス 150 を生成する。RAG は三次元であるのが好ましく、距離情報は図 1 に示すデジタルビデオ源 100 から得られる。各 RAG は、互いに素な領域の集合と、それらの領域を結合する辺の集合とから構成される。同じ X-Y 平面に位置する領域は同じ平面にあると想定する。これに対し、異なる Z 平面に位置する領域は、描かれるシーンの中で異なる奥行きにある領域に対応すると想定する。一般的には、RAG における奥行き軸(例えば、Z 軸)の使用は、特定の領域が 1 つ又は複数の他の領域とは異なる奥行きに位置していることを示すための情報を利用できるか否かにかかっている。例えば、特定の領域の奥行きを判定するために焦点情報又は奥行き情報を利用できるようなデジタルビデオ解釈システム 160 では、奥行き軸を利用することができる。しかし、ビデオセグメントアナライザ 140 は、ほぼ全ての互いに素な領域を同じ平面にあると扱って、奥行き情報の支援なしに、ラベル付き RAG のシーケンス 150 を生成することができる。

【0032】

図 2 は、好ましい実施例によるビデオセグメントアナライザ 140 を示す。ブロック 200 では、解析のために、図 1 のビデオセグメンタ 120 からのビデオセグメント 130 の最初のフレームを選択する。フレームイベントアナライザ 202 は、以下に図 6 を参照して説明するように、選択されたフレームと、そのフレームに関する注目時間領域(ROI)とを受け取り、ラベル付き RAG を生成する。次に、生成された RAG はブロック 204 で格納され、決定ブロック 206 では、ビデオセグメント 130 の終わりに達したか否かを判定する。ビデオの終わりに達していれば、すなわち、決定ブロック 206 で真(イエス)が戻れば、ブロック 208 でビデオセグメントの処理は終了する。これに対し、決定ブロック 206 が偽(ノー)を戻した場合には、ビデオセグメント 130 で次に解析すべきフレームを検索し、処理はフレームイベントアナライザ 202 に戻る。ビデオセグメント 130 の各フレームはリアルタイムで選択、解析されるのが好ましい。しかし、実際には、解析すべきフレームの選択はデジタルビデオ解釈システムを適用する用途によって決まる。例えば、デジタルビデオ解釈システムを具備する装置によっては、各フレームを解析するときにリアルタイムの実行が不可能な場合もあり、そのような場合には、解析に際してビデオセグメントの所定のフレームのみを選択することになる。

【0033】

図 3B は、図 3A に示す空間的に分割されたフレーム 300 の三次元 RAG 310 の一例を示す。空間的に分割されたフレーム 300 は R1 から R9 とラベル付けされた 9 つの領域を含む。領域 R1 は空を含む。領域 R2、R3 及び R9 は陸地を含み、領域 R8 は道路

10

20

30

40

50

を含む。領域 R 4 は家に似た構造であり、領域 R 5 及び R 6 はその家の突き出た構造物である。図 3 A において奥行きを指示するために、線を幾分太くして領域の境界線を示している。特に、それぞれの境界線の太さは Z 軸に沿った奥行きの前後関係を示す。R A G 3 1 0 は分割フレーム 3 0 0 中の領域 R 1 から R 9 を結合する辺を示す。領域 R 1 、 R 2 、 R 3 、 R 7 、 R 8 及び R 9 は、全て、R A G 3 1 0 においてほぼ同じ奥行き(実線の辺により指示する)に位置しているが、異なる X - Y 位置にある。領域 R 1 は一方では領域 R 2 、 R 8 、 R 9 に順次結合しており、他方では、領域 R 3 及び R 7 に結合している。又、領域 R 4 は領域 R 2 、 R 3 、 R 7 及び R 8 と辺でつながっているが、破線の辺で指示するように、異なる奥行きにある。最後に、領域 R 5 及び R 6 は領域 R 4 と辺を共有するが、点線の辺により指示される異なる、平行な奥行きにある。このように、破線と点線は異なる Z 平面と交わる。

【 0 0 3 4 】

5 . フレームイベントアナライザ

図 4 及び図 5 を参照して、図 2 のフレームイベントアナライザ 2 0 2 の機能性を更に詳細に説明する。図 4 に示すフレームイベントアナライザのステップは 1 つのアプリケーションドメイン(例えば、屋外のシーン)を使用する。そのようなアプリケーションドメインは空、水、木の葉、草、道路、人間などのフレーム領域にラベル付けするための知識と機能性を含んでいることができるであろう。

【 0 0 3 5 】

図 4 において、情報源(例えば、デジタルビデオ信号、デジタル音声信号など)ごとの現在フレームと R O I 4 0 0 がその R O I 4 0 0 を使用する状況アナライザ 4 1 0 に提供される。状況アナライザ 4 1 0 に加えて、図 2 のフレームイベントアナライザ 2 0 2 はフレームセグメンタ 4 5 0 と、アプリケーションにおけるラベルの先駆的確率を調整する調整ユニット 4 3 0 と、領域アナライザ 4 7 0 とを具備する。

【 0 0 3 6 】

R O I において利用できる状況情報 4 0 0 は状況アナライザ 4 1 0 により解析される。状況情報を供給する源は複数存在するので、状況アナライザ 4 1 0 は、通常、2 つ以上の状況解析ユニットを含む。

【 0 0 3 7 】

図 7 は、調整ユニット 4 3 0 及びアプリケーションドメイン 4 4 0 に関連して、状況アナライザ 4 1 0 を更に詳細に示す。図 7 に示す状況アナライザ 4 1 0 はフレームイベントに関わる状況情報 4 0 0 を受信する。この情報は音声 R O I 、動き解析 R O I 及び / 又は赤外線スペクトル R O I を含むのが好ましい。状況アナライザ 4 1 0 自体は、音声解析ユニット 7 1 0 と、動き解析ユニット 7 2 0 と、赤外線解析ユニット 7 3 0 とを含んでいても良い。状況アナライザ 4 1 0 により生成された出力は、調整ユニット 4 3 0 により、図 4 の領域アナライザ 4 7 0 により使用されるアプリケーションドメイン 4 4 0 におけるラベルの先駆的確率を変更するために使用される。音声解析ユニット 7 1 0 は、音声信号 R O I に入っているデジタル音声信号中のキーワード又はキーフレーズを認識し、次に、それらのキーワード / フレーズがそのフレームでは他のラベルと比べて特定のラベルが発生しやすいことを示唆しているか否かを検査することにより、この結果を得ても良い。他の状況アナライザユニット(例えば、7 2 0 , 7 3 0)は、ラベルの先駆的確率を直接に変更しても良い。

【 0 0 3 8 】

1 つのアプリケーションドメイン 4 4 0 を伴うフレームイベントアナライザ 2 1 0 を有する好ましい実施例においては、アプリケーションドメイン 4 4 0 のラベルごとに、キーワード / キーフレーズごとの先駆的確率重み付け係数と共に格納されているキーワード / キーフレーズ 4 2 0 のリストにもとづいて調整ユニット 4 3 0 によりラベルの先駆的確率を調整しても良い。確率重み付け係数が大きくなるほど、そのラベルにより記述される領域がフレーム中に存在する尤度は高くなる。キーワード 4 2 0 に加えて、又はその代わりに、他の状況解析結果を調整ユニット 4 3 0 に提供しても良い。

10

20

30

40

50

【0039】

フレームセグメンタ450は、領域ベースセグメンテーション方法を使用して、フレームを複数の均質領域に分割する。通常、セグメンテーション方法は異なる情報源(例えば、ビデオ110A及び音声110B)のROIから抽出した状況情報を使用して、セグメンテーションプロセスを支援する。例えば、モーションベクトルによって、背景から動くオブジェクトを識別するのを助けることができる。焦点情報を利用できるのであれば、この情報を使用して距離を推定でき、従って、フレーム中の異なるオブジェクト又は領域平面を識別することができる。フレームセグメンタ450が実行するセグメンテーションプロセスの結果が図4に示すようなRAG460であり、RAG460は領域アナライザ470に入力として提供される。このRAGは三次元であるのが好ましい。領域アナライザ470に対するもう一方の入力は、状況情報に従ってラベルの先駆的確率が調整されている場合もあるアプリケーションドメイン440である。

【0040】

確率モデルに基づいた領域アナライザ470は、RAG中の領域に適切なアプリケーションドメイン440を使用して最適のラベルを付ける。その結果として得られるラベル付きRAGは、内容に基づいた検索などのより高いレベルのプロセスに使用できるフレームの内容の記述、すなわち、メタデータを表現する。領域アナライザはMRF(確率)モデルを使用して、ラベル付きRAG480を生成するのが好ましい。MRFモデルについては、以下に詳細に説明する。

【0041】

図5には、フレームイベントアナライザが複数のアプリケーションドメインを有することを除いて、図4とほぼ同様に説明できるフレームイベントアナライザが示されている。複数のアプリケーションドメインを有するフレームイベントアナライザ(すなわち、図5に示すようなフレームイベントアナライザ)の場合、各アプリケーションドメインはキーワード/キーフレーズを含むと考えられ、選択ユニット530の役割として、解析で使用すべきアプリケーションドメインの選択を含めることができる。すなわち、選択ユニット530は最も確率の高いアプリケーションドメインを選択し、選択したドメインにおけるラベルの先駆的確率を調整するのが好ましい。

【0042】

図6を参照すると、ビデオセグメントに関する時間系列が示されている。ビデオセグメントの現在フレーム601は、利用可能な状況情報603、音声情報604(信号)及びビデオ情報605(信号)から抽出した1つ又は複数の注目領域(ROI)602を参照して解析される。

【0043】

ROIの時間境界は、状況情報(図6を参照)の型に応じて変わることもある。例えば、カメラパラメータのような状況情報は長い時間周期、おそらくはビデオセグメント全体にわたって広がっている可能性がある。これに対し、ビデオ信号のROIははるかに短く、おそらくは現在解析中であるフレームの前後の数フレームにすぎないと考えられる。図6に示すように、ROIの中心は現在フレームにあるとは限らない。例えば、ROIが先行フレームを含むこともありうる。

【0044】

数学的には、RAGは互いに素な領域の集合R及びそれらの領域を結合する辺の集合E、すなわち、 $G = \{R, E\}$ を含むグラフGであると定義される。ビデオフレーム解釈では、Gの中の領域に最適のラベルを付けようとする。アプリケーションドメインが状況情報の解析によって偏りの与えられた先駆的確率 $P_{r_L} = \{P_{r_{L1}}, P_{r_{L2}}, P_{r_{L3}}, \dots, P_{r_{Lp}}\}$ を伴うp個のラベルの集合 $L = \{L_1, L_2, L_3, \dots, L_p\}$ から構成されているとすれば、解釈プロセスはグラフGにおいて最も確率の高いラベルの集合を推定するプロセスであるとみなすことができる。

【0045】

グラフGがN個の互いに素な領域から構成されるならば、 $X = \{X_1, X_2, X_3, \dots, X_N\}$

$N\}$ をRAGにおける一連の確率変数とする。すなわち、 X は確率場であり、 X_i は R_i と関連する確率変数である。 X_i の実現 x_i はラベルの集合 L の一メンバである。Gにおける隣接系 Γ は次のように示される。

【0046】

【数1】

$$\Gamma = \{n(R_i); 1 \leq i \leq N\} \quad (1)$$

【0047】

式中、 $n(R_i)$ は R_i の隣接領域を含む R の部分集合である。領域 R_i の隣接系 Γ はその領域と、 R_i と共に通する境界線を有する他の全ての領域であるのが好ましい。

10

【0048】

更に、 Ω はあり得る全てのラベル付け構成の集合であり、 ω は Ω の中の1つの構成を示す。

【0049】

【数2】

$$\Omega = \{\omega = \{x_i, x_2, x_3, \dots, x_N\}; x_i \in L, 1 \leq i \leq N\} \quad (2)$$

【0050】

そこで、 X は、

$$P(X = \omega) > 0$$

20

なる X のあらゆる実現に対して、

【0051】

【数3】

$$P(X_i = x_i | X_j = x_j, R_j \neq R_i) = P(X_i = x_i | X_j = x_j, R_j \in n(R_i)). \quad (3)$$

【0052】

であれば、隣接系 Γ に関してMRFである。

【0053】

MRFの重要な特徴は、その同時確率密度係数 $P(X = \omega)$ がギブズの分布を有することである。すなわち、

30

【0054】

【数4】

$$P(X = \omega) = Z^{-1} \exp[-U(\omega) / T], \quad (4)$$

【0055】

式中、 T は温度であり、 $U(\omega)$ はギブズのエネルギー関数である。分配関数 Z は次のように求められる。

【0056】

【数5】

$$Z = \sum_{\omega} \exp[-U(\omega) / T]. \quad (5)$$

40

【0057】

エネルギー関数は「クリーク」の概念を使用して表現できる。グラフ G と関連するクリーク c は、1つの領域か、又は、その全てが互いに隣接するいくつかの領域かのいずれかを含むような R の部分集合である。図3Bに示すRAGの中の領域ごとのクリークを図8に示す。例えば、領域 R_1 は関連するクリーク $\{R_1\}$ 、 $\{R_1, R_2\}$ 及び $\{R_1, R_3\}$ を有する。

50

グラフ G に関するクリークの集合を C と呼ぶ。クリーク関数 V_c は、 $(i \in C)$ である x_i の値(すなわち、ラベル)によって $V_c(\cdot)$ が決まるという特性を有する関数である。一群のクリーク関数はポテンシャルと呼ばれるので、 G に関するクリーク関数を加算することにより $U(\cdot)$ を得ることができる。

【0058】

【数6】

$$U(\omega) = \sum_{c \in C} V_c(\omega). \quad (6)$$

【0059】

フレーム及び先駆的知識から得られる領域に基づく特徴測定値をクリーク関数 V_c に取り入れる。領域特徴測定値の集合が与えられたときの特定の領域ラベル L_i の尤度をトレンジングセット(例えば、神経ネットワーク)の使用を含めた様々な方法を使用して推定できる。あるいは、経験的知識に基づいて尤度を推定しても良い。同様に、測定値に基づく場合も、基づかない場合もあるという制約の形で、先駆的知識をクリーク関数 V_c に取り入れることも可能である。例えば、制約はラベル L_i とラベル L_j とが隣接できない(すなわち、隣接する確率が零である)という形態であっても良い。あるいは、 L_i と L_j が隣接している場合、その境界はいくつかの特性(例えば、フラクタル次元)を有すると思われ、制約の値は測定値に従属しても良いであろう。

【0060】

式(4)から(6)は、構成に関するギブズの $U(\cdot)$ エネルギーを最小にすることは、その確率密度関数を最大にすることと等価であることを示す。本発明の好ましい実施例は、フレームから得られる測定値 M と、ラベルに関する先駆的知識 K と、アプリケーションドメインにおけるラベルの先駆的確率 P_r とが与えられた場合の最適の領域ラベル構成を見出そうとする。ラベルの先駆的確率は、状況情報の解析によって偏りが与えられる。(フレームの) RAG 全体にわたりラベルを最適化するという問題は、いずれかのサイト i でラベルを繰り返し最適化することにより解決できる。領域 i におけるラベルの M 、 K 及び P_r への依存性を設計上のクリーク関数 $V_c(\cdot)$ に取り入れる。従って、サイト i における x_i を x_i^* とするととき、条件付き確率密度関数を次のように書き表すことができる。

【0061】

【数7】

$$P(X_i = x_i | X, M, K, Pr) = Z_i^{-1} \exp \left[-\frac{1}{T} \sum_{c \in C_i} V_c(\omega^*) \right],$$

$$Z_i = \sum_{x \in L} \exp \left[-\frac{1}{T} \sum_{c \in C_i} V_c(\omega^x) \right], \quad (7)$$

【0062】

式中、 C_i は X_i を含むクリークから構成される C の部分集合であり、 x^* はサイト i では x であり、その他の場所では x と一致する構成を示す。ラベルの先駆的確率を使用して、サイトの初期ラベルに偏りを与えることも可能である。例えば、先行する解析イベントのラベルを利用して、後の解析イベントのグラフを初期設定することができるであろう。

【0063】

先に述べた通り、クリーク関数はフレームからの特徴測定値 M 、ラベルに関する先駆的知識 K 及びラベルの先駆的確率 P_r に基づくことができる。例えば、アプリケーションドメインが屋外のシーンであるときのラベル「空(そら)」を考えてみると、RAG における領域(サイト) i を含むクリークの集合(すなわち、 C_i)は、通常、その領域 i のみから構成される単項クリークと、それぞれが領域 i を含む領域群を包含するクリークの集合とから構成されると考えられるであろう。この場合、各領域はその領域群の領域に対して互いに隣接している。

【0064】

単項クリーク関数は、領域 i に関する特徴の集合を測定し、次に、それらの特徴測定値を

10

20

30

40

50

、手動操作で分割された画像からの空(そら)領域の例を使用して先にトレーニングされていた神経ネットワークへの入力として使用することにより、計算できるであろう。1つの領域に関して測定可能であると考えられる特徴の例は平均R、B、B値のいずれか1つ乃至3つと、平均輝度と、領域内の輝度の分散と、周波数ドメインで取り出される測定値を含む場合もあるテクスチャの特徴と、完全連続などの領域形状特徴とを含む。通常、神経ネットワークは、手動操作で分割された空(そら)領域の特徴測定値に類似する特徴測定値を有する領域に対しては低い値(例えば、零)を生成し、手動操作で分割された領域の特徴測定値に全く類似していない特徴測定値を有する領域に対しては高い値(例えば、1.0)を生成するようにトレーニングされるであろう。

【0065】

10

また、特徴測定値を2つ以上の領域を含むクリーク関数で使用することもできる。例えば、2つの領域の間の共通する境界のねじれを、一对の領域を含むクリーク関数で使用することが可能であろう。例えば、「空」と「水」との間の共通する境界は、通常、それほど大きなねじれを伴わないであろうが、「木の葉」と「空」との間の共通する境界は非常に大きなねじれを伴うであろう。

【0066】

20

制約の形態で、先駆的知識をクリーク関数に取り入れることができる。例えば、「空」ラベルと、「草」ラベルとを含むクリーク関数は、「草」ラベルが適用されている領域が「空」ラベルが適用されている領域の上方にある場合は、高いエネルギー値(例えば、1.0)を戻すであろう。言い換えれば、「空」領域が通常はフレーム内で「草」領域の上方に位置しているという先駆的知識を利用しているのである。

【0067】

「空」である領域*i*の先駆的確率 $P_{r_{sky}}$ をクリーク関数に取り入れることも可能である。これを実行する方法の1つは、既存の単項クリーク関数と、

【0068】

【数8】

$$1 - \alpha \left(\frac{Pr_{Sky}}{\arg \max_{L \in \mathbf{L}} \mathbf{Pr}_L} \right), \quad (8)$$

30

【0069】

のような乗算係数とを乗算することであろう。式中、 α は先駆的確率のクリーク関数全体への寄与に重み付けする(0, 1)の範囲の何らかのパラメータである。また、先駆的確率を2つ以上の領域を含むクリーク関数に取り入れることも可能であろう。この場合、クリーク関数に対する乗算係数は、通常、クリーク関数の各々のラベルの先駆的確率を含むであろう。

40

【0070】

式7は、あるサイトにおいて最も確率の高いラベルを選択することは、ラベルの先駆的確率により重み付けされた、そのサイトのギブズのエネルギー関数 $U(\cdot)$ を最小にすることと等価であることを実証している。フレームの最適な領域ラベル構成は、グラフGのN個のサイトの各々を繰り返し見て、各サイトのラベルを更新することにより得られる。領域ラベルを更新する方法はいくつかある。ある領域に対して、ラベルの一様な分布から、又はMRFの条件付き確率分布のいずれかから新たなラベルを選択することができる(すなわち、ギブズのサンプル、Geman and Geman, IEEE Trans. Pattern Analysis and Machin

50

e Intelligence, 6, 721から741ページ(1984年)を参照)。より迅速な対応が望まれるのであれば、反復条件付きモード(J. BesagがJ. R. Statistical Soc. B, 48の259から302ページ(1986年)の中で説明している)方法を使用しても良い。この第2の方法の場合、RAGのサイトを繰り返し見直し、各サイトにおいて、領域のラベルを最大の条件付き確率分布を有するラベルとなるように更新する。更新するサイトを巡回する反復手続きは、擬似焼きなましスキーム(simulated annealing scheme)(温度を徐々に下げる)の中で実現することができる。更新する方法は、本発明のこの実施例に関しては重大ではない。重大であるのは、ギブズのエネルギー $U(\cdot)$ の計算に先駆的確率を取り入れたことである。

6. 状況アナライザ

10

図4の状況アナライザ410はそれぞれの情報源(例えば、ビデオ信号110A及び音声信号110B)について現在フレームとROI400を取り出し、アプリケーションドメイン440におけるラベルの先駆的確率をどのようにバイアスすべきかに関して情報を調整ユニット430に提供する。図5に示すような状況アナライザ410の機能については、図2のフレームイベントアナライザ202に関連して既に説明した。以下、音声信号ROIにおける様々なキーワード/キーフレーズの有無に基づいてアプリケーションドメイン440におけるラベルの先駆的確率を調整する方法を更に詳細に説明する。その他の状況情報に対して類似の方法を使用することができる。

【0071】

各ラベルを1つ又は複数の根拠単位と関連付けることができ、1つの根拠単位はキーワード又はキーフレーズと、0から1までの重み係数とを含む。例えば、ラベル「水」の根拠単位はキーワード「ビーチ」と、重み付け係数0.8とから成ると考えても良いであろう。重み付け係数の値は、音声ROIにおけるキーワードの存在がRAG中の少なくとも1つの領域について「水」は適切なラベルであることを示す尤度を意味する。

20

【0072】

根拠を収集する前に、全てのラベルの先駆的確率の和は1.0になるはずである。言い換えれば、

【0073】

【数9】

30

$$\sum_{l=1}^L Pr_l = 1.0 \quad (9)$$

【0074】

状況情報のROIから根拠を収集するとき、根拠単位を具体化する。所定のラベル1に関する異なる具体化根拠単位の重み係数を加算して、そのラベルの総根拠 E_l を生成することができる。

【0075】

そこで、アプリケーションドメイン440におけるラベルの Pr_l 値を、

【0076】

【数10】

40

$$Pr_l = (1 + E_l)x, \quad (10)$$

【0077】

を使用して計算することができる。なお、式中、 x の値は、

【0078】

【数11】

50

$$\sum_{l=1}^L (1+E_l)x = 1.0. \quad (11)$$

【 0 0 7 9 】

を解くことにより求められる。

【 0 0 8 0 】

この結果として得られる P_{r_1} 値はクリーク関数により直接使用することができる(例えば、式 8 を参照)。

【 0 0 8 1 】

7. 本発明の他の実施例

図 11 は、本発明の他の実施例によるビデオセグメントアナライザ 140 を示す。この場合、ビデオセグメントアナライザ 140 はオブジェクトに基づくデジタルビデオ符号化システムと統合されている。ブロック 250 では、図 1 のビデオセグメント 120 により生成されたビデオセグメント 130 の第 1 のフレームをビデオセグメントアナライザ 140 にロードする。フレームイベントアナライザ 252 はロードされたフレームを受け取り、図 2A のフレームイベントアナライザ 202 について説明したように、関連 ROI からの状況情報を使用してフレームを解析し、ラベル付き RAG を得る。次に、フレームイベントアナライザ 252 はラベル付き RAG を領域エンコーダ 254 へ出力し、領域エンコーダ 254 は RAG を符号化する。領域エンコーダ 254 は、RAG の領域の隣接度及び奥行き情報、意味ラベルを含めて、RAG の領域をビットストリームに符号化する。ブロック 256 では、ビデオセグメントの終わりに到達したか否かを判定するために、検査を実行する。検査ブロック 256 が真(イエス)を戻せば、ビデオセグメントの処理はブロック 258 で終了する。検査又は決定ブロック 256 が偽(ノー)を戻した場合には、ブロック 260 でビデオセグメントの次のフレームをロードする。

【 0 0 8 2 】

動き検出器 262 はビデオセグメントにおける動きをフレームごとに検出する。動き検出器は、領域ごとに、先行フレームから検出される何らかの動きを検査する。動きモデル(例えば、領域のアフィン変換)により個々の領域の動きを記述できれば、ブロック 266 でモデルパラメータをビットストリームに符号化する。検出された動きを動きモデルにより記述できない場合には、フレームをフレームイベントアナライザ 252 により解析し、新たな RAG を生成し、領域エンコーダ 254 により符号化する。

【 0 0 8 3 】

図 11 に示すビデオセグメントアナライザ 140 では、意味ラベルを符号化デジタルビデオ信号と統合するのが好ましい。ビデオセグメントアナライザがデジタルビデオ符号化システムと統合されていれば、分解能に左右されずに領域を別個に符号化しても良い。これにより、任意の所望の分解能でデジタルビデオ信号の単純な再構成が可能になる。デジタルビデオ信号を符号化する方法は、当業者に良く知られているそのようないくつかの技法のうちいずれかを使用して実行されれば良い。ビデオセグメントアナライザ 140 を必ずしもデジタルビデオ符号化システムと統合しなければならないとは限らないことは明らかである。先に述べた通り、統合するのではなく、ビデオセグメントアナライザ 140 はメタデータを生成するだけであっても良い。そのような実施例においては、1つのセグメントの全てのビデオフレームを処理する必要はないであろう。言い換えれば、1つのセグメントの中で選択されたフレームだけを解析すれば良いのである。そのようなフレームの選択の大部分は実現の態様によって決まるため、どのようにしてフレームを選択するかを指定することは、本発明の実施例の目的ではない。例えば、ビデオ解釈システムはリアルタイムに近い方式で動作する必要があるだろう。

【 0 0 8 4 】

本発明の更に別の実施例は、ビデオフレームのセグメンテーションのプロセスと領域のラ

10

20

30

40

50

ベル付けのプロセスとを1つの最小化プロセスに組み合わせたものである。

【図面の簡単な説明】

【図1】好ましい実施例によるデジタルビデオ解釈システムのブロック線図である。

【図2】好ましい実施例による図1のビデオセグメントアナライザを示す図である。

【図3 A】

【図3 B】本発明の実施例に従った代表的なセグメント分割画像と、対応する領域隣接度グラフ(RAG)とをそれぞれ示す図である。

【図4】1つのアプリケーションドメインを有する図2のフレームイベントアナライザを示す図である。

【図5】複数のアプリケーションドメインを有する図2の別のフレームイベントアナライザを示す図である。 10

【図6】特定の解析イベントに関する注目時間領域(ROI)の選択を示す図である。

【図7】図4又は図5のフレームイベントアナライザで使用するための好ましい状況アナライザを示す図である。

【図8】図3BのRAGと関連するクリークを示す図である。

【図9】本発明の実施例を実施しうるデジタルビデオ源と共に使用するための代表的なコンピュータのブロック線図である。

【図10】本発明の実施例を実施しうる代表的なデジタルビデオ源のブロック線図である。 15

【図11】オプションとしてデジタルビデオ符号化システムに統合されている、別の実施例による図1のビデオセグメントアナライザを示す図である。 20

【図1】

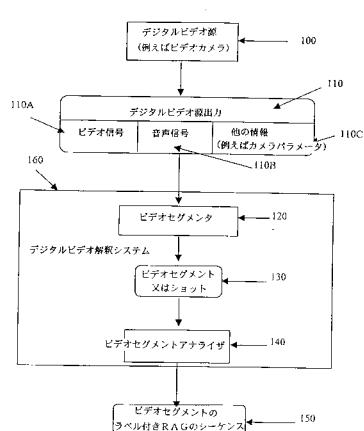


Fig. 1

【図2】

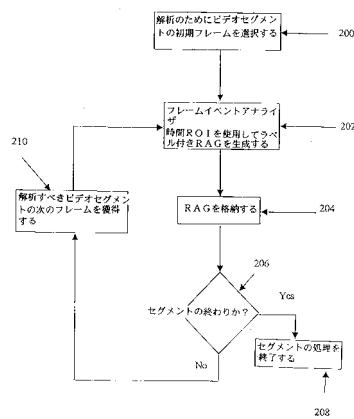


Fig. 2

【図3 A】

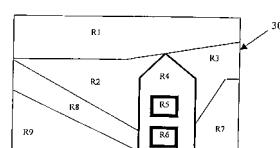


Fig. 3A

【図3B】

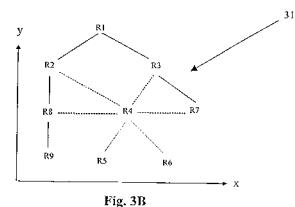


Fig. 31

【 図 4 】

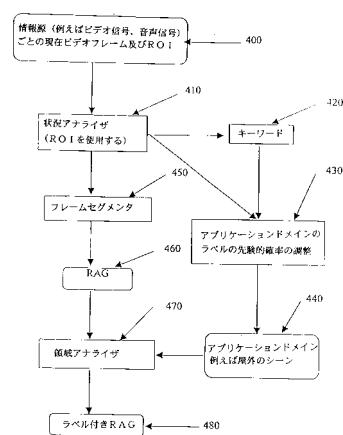


Fig. 4

【 図 5 】

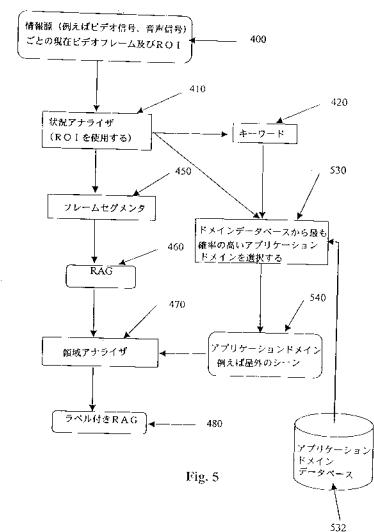
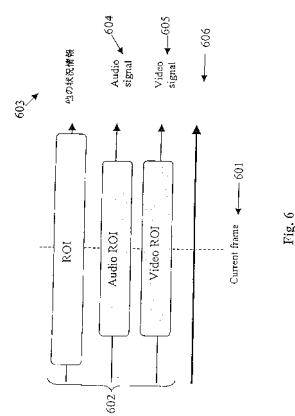


Fig. 2

【図6】



6

【図7】

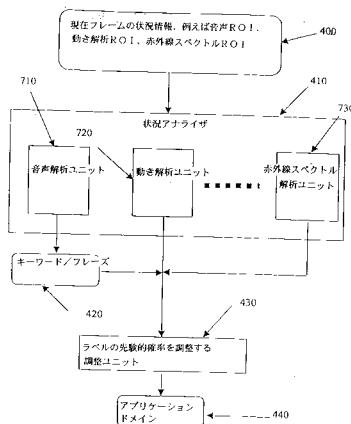


Fig. 7

【図8】

Node or Region	Associated Cliques
R1	{R1}, {R1, R2}, {R1, R3}
R2	{R2}, {R2, R4}, {R2, R3}, {R2, R4, R3}
R3	{R3}, {R3, R4}, {R3, R7}, {R3, R4, R7}
R4	{R4}, {R4, R5}, {R4, R6}, {R4, R7}, {R4, R8}
R5	{R5}
R6	{R6}
R7	{R7}
R8	{R8}, {R8, R9}
R9	{R9}

Fig. 8

【図9】

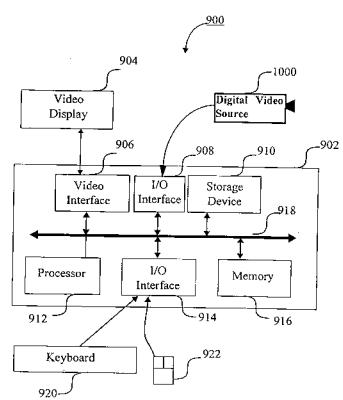


FIG. 9

【図10】

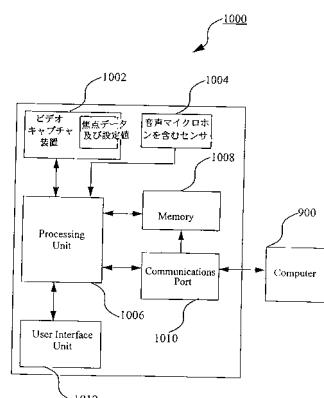


FIG. 10

【図11】

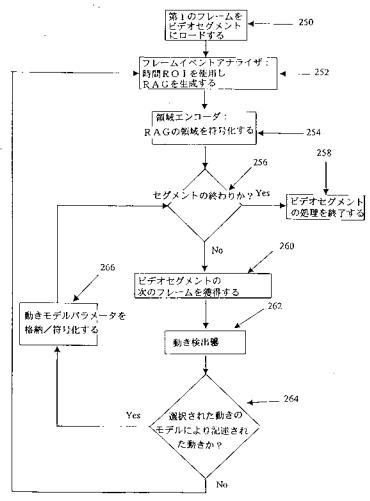


Fig. 11

フロントページの続き

(72)発明者 アリソン ジョアン レノン

オーストラリア国 2113 ニュー サウス ウエールズ州, ノース ライド, トーマス ホルト ドライブ 1 キヤノン インフォメーション システムズ リサーチ オーストラリア プロプライエタリー リミテッド内

(72)発明者 デルフィン アン ダオ リー

オーストラリア国 2113 ニュー サウス ウエールズ州, ノース ライド, トーマス ホルト ドライブ 1 キヤノン インフォメーション システムズ リサーチ オーストラリア プロプライエタリー リミテッド内

審査官 松尾 俊介

(56)参考文献 特開平08-235198 (JP, A)

特開平09-249083 (JP, A)

特開平07-193748 (JP, A)

特開平09-044639 (JP, A)

特開平07-038842 (JP, A)

Kim, 外 1名, An Integration Scheme for Image Segmentation and Labeling Based on Markov Random Field Model, Pattern Analysis and Machine Intelligence, IEEE Transaction on, 米国, IEEE, 1996年 1月, Vol. 18, p. 69-73, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=476014

花村, 外 2名, マルチメディア標準化のためのビデオ・ドキュメント・アーキテクチャの構想, テレビジョン学会技術報告, 日本, 社団法人テレビジョン学会, 1990年 9月13日, Vol. 1.14 No. 45, p. 21~28

(58)調査した分野(Int.Cl., DB名)

G06T 7/20