

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5634075号
(P5634075)

(45) 発行日 平成26年12月3日(2014.12.3)

(24) 登録日 平成26年10月24日(2014.10.24)

(51) Int.Cl.

F 1

G06T	7/20	(2006.01)
HO4N	5/76	(2006.01)
HO4N	5/91	(2006.01)

GO 6 T	7/20	C
HO 4 N	5/76	B
HO 4 N	5/91	P

請求項の数 15 外国語出願 (全 23 頁)

(21) 出願番号	特願2010-14279 (P2010-14279)
(22) 出願日	平成22年1月26日 (2010.1.26)
(65) 公開番号	特開2010-191954 (P2010-191954A)
(43) 公開日	平成22年9月2日 (2010.9.2)
審査請求日	平成25年1月17日 (2013.1.17)
(31) 優先権主張番号	0901262.6
(32) 優先日	平成21年1月26日 (2009.1.26)
(33) 優先権主張国	英国 (GB)

(73) 特許権者 501253316
ミツビシ・エレクトリック・アールアンド
ディー・センター・ヨーロッパ・ビーヴィ
MITSUBISHI ELECTRIC
R&D CENTRE EUROPE
B. V.
イギリス国、サリー・ジーユー2・5ワイ
ディ、ギルドフォード、ザ・サリー・リサ
ーチ・パーク、フレデリック・サンガー・
ロード 20
20 Frederick Sanger
Road, The Surrey R
esearch Park, Guild
ford, Surrey GU2 5Y
D, Great Britain
最終頁に続く

(54) 【発明の名称】画像のシーケンスを処理する方法および装置、画像データを処理する装置、ならびにコンピュータプログラム製品

(57) 【特許請求の範囲】

【請求項 1】

画像シーケンスの画像データを処理して、前記シーケンスを表現する値を生成する方法
であって、

前記方法は、前記シーケンス内の各画像について複数の画素近傍集合のそれぞれに対する前記画像データを処理して複数の記述子要素を生成するステップと、

前記複数の記述子要素から複数のワードを生成するステップと、

いずれのワードが前記シーケンス内に現れるか、その頻度を示す2値化ヒストグラムデータを生成するステップとを含み、

各前記ワードは前記複数の記述子要素の一部である複数の記述子要素からなる順序集合が連結されて生成され、

前記ワードのすべてにおける総ビット数が前記記述子要素のすべてにおける総ビット数を下回ることを特徴とする方法。

【請求項 2】

画像シーケンスの画像データを処理して、前記シーケンスを表現する値を生成する方法
であって、

前記方法は、前記シーケンス内の各画像について複数の画素近傍集合のそれぞれに対する前記画像データを処理して複数の記述子要素を生成するステップと、

前記複数の記述子要素から、複数のワードを生成するステップと、

いずれのワードが前記シーケンス内に現れるか、その頻度を示す2値化ヒストグラムデータを生成するステップとを含み、

10

20

ータを生成するステップとを含み、

各前記ワードは、前記複数の記述子要素の一部である複数の記述子要素からなる順序集合が連結されて、前記記述子要素の固有の組合せとして生成されることを特徴とする方法。

【請求項 3】

前記方法は、各前記記述子要素を量子化するステップをさらに含み、

各前記ワードは、前記量子化された記述子要素から生成される、請求項 1 又は 2 に記載の方法。

【請求項 4】

各前記記述子要素を量子化する前記ステップは、各前記記述子要素について 1 ビットの 10
2 値結果を生成するために各前記記述子要素を量子化するステップを含む、請求項 3 に記載の方法。

【請求項 5】

前記記述子要素は、複数の異なる画素解像度で前記画像に対して生成され、

異なる画素解像度で生成される前記記述子要素からの 1 つ又は複数のビットを組み合わせることによって、少なくとも 1 つのワードが生成される、請求項 1 ~ 4 のいずれか 1 項に記載の方法。

【請求項 6】

前記画像のシーケンスはビデオシーケンスの断片を含む、請求項 1 又は 2 に記載の方法

。

20

【請求項 7】

前記ヒストグラムデータ、前記ワード、及び前記記述子要素はビットストリーム内に記憶される、請求項 1 ~ 5 のいずれか 1 項に記載の方法。

【請求項 8】

画像の第 1 のシーケンス及び画像の第 2 のシーケンスを処理して、前記第 1 のシーケンスと前記第 2 のシーケンスとを比較する方法であって、

前記方法は、ステップ(a)およびステップ(b)を含み、

前記ステップ(a)は、前記第 1 のシーケンス内の複数の前記画像のそれぞれ、及び前記第 2 のシーケンス内の複数の前記画像のそれぞれを処理するステップであり、

前記ステップ(a)は、前記画像内の複数の画素近傍集合のそれぞれに対する前記画像データを処理して複数の記述子要素を生成するサブステップと、前記記述子要素から複数のワードを生成するサブステップと、いずれのワードが前記シーケンス内に現れるか、その頻度を示す2 値化ヒストグラムデータを生成するサブステップとを含み、

各前記ワードは、前記複数の記述子要素の一部である複数の記述子要素からなる順序集合が連結されて、前記記述子要素の固有の組合せとして生成され、

前記第 2 のシーケンスに対する前記ワードのそれぞれは、前記第 1 のシーケンスに対する前記ワードのそれぞれと同じ組合せの記述子要素から生成され、

前記ステップ(b)は、前記第 1 のシーケンス内の前記複数の画像に対して生成された前記2 値化ヒストグラムデータと、前記第 2 のシーケンス内の前記複数の画像に対して生成された前記2 値化ヒストグラムデータとを比較することによって、前記第 1 のシーケンスと前記第 2 のシーケンスとを比較する処理を実施するステップである、方法。 40

【請求項 9】

前記2 値化ヒストグラムデータについての比較結果はマージされて全体の比較結果を生成する、請求項 8 に記載の方法。

【請求項 10】

各前記ワードに対して、双方のシーケンス内に発生する前記ワード値を求めるステップと、

各前記ワードに対して、双方のシーケンス内における前記ワード値の時間的な発生順序を求めるステップと、

前記時間的な発生順序を比較するステップと、

50

をさらに含む、請求項8又は9に記載の方法。

【請求項 1 1】

前記方法は、一致する画像対を特定するステップと、少なくとも1つの特定された画像対間の類似度を求めるために、さらなる対応性試験を実施するステップとを含み、

前記一致する画像対は、前記第1のシーケンスからの1つの画像と前記第2のシーケンスからの1つの画像とを含む、請求項8～10のいずれか1項に記載の方法。

【請求項 1 2】

画像の第1のシーケンス及び画像の第2のシーケンスを処理して、前記第1のシーケンスと前記第2のシーケンスとを比較する装置であって、

前記装置は、記述子要素生成手段と、ワード生成手段と、ヒストグラムデータ生成手段と、シーケンス比較手段とを備え、10

前記記述子要素生成手段は、前記第1のシーケンス内の複数の前記画像のそれぞれ、及び前記第2のシーケンス内の複数の前記画像のそれぞれを、前記画像内の複数の画素近傍集合のそれぞれに対する前記画像データを処理して、複数の記述子要素を生成することによって処理するように構成され、

前記ワード生成手段は、各前記画像の前記記述子要素から、複数のワードを生成するように構成され、

各前記ワードは、前記複数の記述子要素の一部である複数の記述子要素からなる順序集合が連結されて、前記記述子要素の固有の組合せとして生成され、20

前記ワード生成手段は、前記第2のシーケンスに対する前記ワードのそれぞれを、前記第1のシーケンスに対する前記ワードのそれぞれと同じ組合せの記述子要素から生成するように構成され、

前記ヒストグラムデータ生成手段は、いずれのワードが前記シーケンス内に現れるか、その頻度を示す2値化ヒストグラムデータを生成するよう構成され、

前記シーケンス比較手段は、前記第1のシーケンス内の前記複数の画像に対して生成される前記2値化ヒストグラムデータと、前記第2のシーケンス内の前記複数の画像に対して生成される前記2値化ヒストグラムデータとを比較することによって、前記第1のシーケンスと前記第2のシーケンスとを比較する処理を実施するように構成される、装置。

【請求項 1 3】

画像シーケンスの画像データを処理して、前記シーケンスを表現する値を生成する装置であって、30

前記装置は、記述子要素生成手段と、ワード生成手段と、ヒストグラムデータ生成手段とを備え、

前記記述子要素生成手段は、前記画像内の複数の画素近傍集合のそれぞれに対する前記画像データを処理することにより、複数の記述子要素を生成するように構成され、

前記ワード生成手段は、前記記述子要素から複数のワードを生成するように構成され、

前記ヒストグラムデータ生成手段は、いずれのワードが前記画像のシーケンス内に現れるか、その頻度を示す2値化ヒストグラムデータを生成するように構成され、

各前記ワードは前記複数の記述子要素の一部である複数の記述子要素からなる順序集合が連結されて生成され、40

前記ワードのすべてにおける総ビット数が前記記述子要素のすべてにおける総ビット数を下回ることを特徴とする装置。

【請求項 1 4】

画像のシーケンスの画像データを処理して、前記シーケンスを表現する値を生成する装置であって、

前記装置は、記述子要素生成手段と、ワード生成手段と、ヒストグラムデータ生成手段とを備え、

前記記述子要素生成手段は、前記画像内の複数の画素近傍集合のそれぞれに対する前記画像データを処理することにより、複数の記述子要素を生成するように構成され、

前記ワード生成手段は、前記記述子要素から、複数のワードを生成するように構成され50

、
前記ヒストグラムデータ生成手段は、いずれのワードが前記画像のシーケンス内に現れるか、その頻度を示す2値化ヒストグラムデータを生成するように構成され、

各前記ワードは、前記複数の記述子要素の一部である複数の記述子要素からなる順序集合が連結されて、前記記述子要素の固有の組合せとして生成されることを特徴とする装置。
。

【請求項15】

処理装置に、請求項1～11のいずれか1項に記載の方法を実施させるコンピュータプログラム。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、一般的な編集操作（カメラ取込み、アナログ／デジタル変換、再圧縮等）に対してロバストである、高速且つ信頼できるビデオ識別のための方法、装置、及びコンピュータプログラム製品に関する。

【背景技術】

【0002】

専門家及び消費者によって記憶されているビデオの数が急速に増大している。専門家に関しては、ビデオコンテンツの作成、記憶、及び編集のコストの問題及び技術的な問題は減少してきている。消費者に関しては、セットトップボックス、パーソナルビデオレコーダ、ビデオカメラ、及びコンピュータの選択の幅が広がっていること、並びに、これらのコストが減少していることによって、ビデオコンテンツの増大に拍車がかかっている。過去数年にわたって、インターネット上で利用可能な合法及び非合法のコンテンツが共に爆発的に増大している。このコンテンツをインデックス付けし、検索し、監視する能力がますます重要な問題になってきている。M P E G - 7 標準規格は、コンテンツベースの検索及び取出しの領域における初期の仕事であった。この標準規格の元のバージョンにおいて欠けていた領域が、複製類似ビデオ（near-duplicate video）検出である。

20

【0003】

複製類似ビデオ検出は、問合せビデオシーケンスが与えられると、データベース内のすべての複製を発見するものと定義することができる。複製（複製類似）の概念及び解釈はさまざまである。しかしながら、本発明に関しては、複製とは、オリジナルに対する一般的なビデオ編集／処理操作によって作成されたシーケンスであるとみなす。このような操作の例は、より多くの例の中でも、色変更、圧縮、トランスコーディング、フォーマット変更、フレームレート変更、アナログV C R 再取込み、及びカメラ再取込みを含む。本発明は、複製部分が問合せシーケンスの一部のみを形成する場合がある問題にも対処する。

30

【0004】

この領域における従来の仕事（非特許文献1）は、複数のカットを撮影し、境界を使用してビデオシーケンスのシグネチャを形成していた。これは、ビデオの非常にコンパクトな表現を提供するが、短いシーケンスに対しては非常に不満足にしか機能せず、使用されるショット検出アルゴリズムの影響を非常に受けやすい（非特許文献2）。

40

【0005】

ビデオ内の複製（複製類似）検出に対する、特徴点を用いる一般的な当該技術水準の手法は、非特許文献3に与えられており、i) キーフレームを検出し、ii) 該フレーム内のキーポイントを検出し、iii) 該ポイントの周囲の領域から特徴を検出し、iv) 特徴を使用してシーケンスを照合し、v) シーケンス内の物体の空間的結束性（spatial cohesion）に関する試験を適用する、というように概説することができる。この概説した手法にはいくつかの欠点がある。まず、キーフレームを使用するということは、この方法が短いクリップに対してはあまり良好に機能しない可能性があることを意味している。特徴の抽出（iii）は計算コストが高い方法であり、その結果ストレージ要求が大きくなる。ステップ（iv）に関して、データのクラスタリングから学習される視覚的語彙が使用

50

される。これは特定のデータセットに対する過剰適合をもたらす可能性があり、これは一般化の失敗を伴う。非特許文献4のような関連方法は、使用されるハッシュテーブルに対する高メモリ要件と引き換えに高速の検索を提供する。これはいくつかのシナリオに対しては適切であり得るものなの、一般的にメモリ資源が非常に限られている家庭用電化製品の環境においては適切でない。

【先行技術文献】

【非特許文献】

【0006】

【非特許文献1】T. Hoad及びJ. Zobel著「Video similarity detection for digital rights management」(Proceedings of Australasian Computer Science Conference, pages 237-245, Adelaide, Australia, 2003) 10

【非特許文献2】T. Hoad及びJ. Zobel, Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, pages 262-269, Berkeley, US, 2003

【非特許文献3】J. Sivic、A. Zisserman著「Efficient Visual Search for Objects in Videos」(Proceedings of the IEEE, April 2008, 96 (4), pages 548-566)

【非特許文献4】Ondrej Chum、James Philbin、Michael Isard、及びAndrew Zisserman著「Scalable near identical image and shot detection」(Proceedings of the 6th ACM international conference on Image and video retrieval, pages 549-556, Amsterdam, The Netherlands, 2007) 20

【発明の概要】

【発明が解決しようとする課題】

【0007】

本発明は、少なくとも、このような従来技術の方法の制限のうちの1つ又は複数に対処することを目的とする。

【課題を解決するための手段】

【0008】

本発明の特定の態様を添付の特許請求の範囲に記載している。他の態様は下記の実施の形態に記載しており、当業者であれば本明細書を読むことによって理解されよう。

【0009】

要約すると、本発明は、画像の第1のシーケンス及び画像の第2のシーケンスを処理して、該第1のシーケンスと該第2のシーケンスとを比較する装置であって、

該装置は、

第1のシーケンス内の複数の画像のそれぞれ、及び第2のシーケンス内の複数の画像のそれぞれを、画像内の複数の画素近傍集合のそれぞれに対する画像データを処理することによって、該画素近傍集合のそれぞれに対して少なくとも1つの記述子要素を生成することによって処理するように構成される記述子要素生成手段であって、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、

各画像の記述子要素から、複数のワード(word)であって、各ワードが記述子要素ビットの固有の組合せを含むような、複数のワードを形成するように構成されるワード生成手段であって、該ワード生成手段は、第2のシーケンスに対するワードのそれぞれを第1のシーケンスに対するワードのそれぞれと同じ組合せの記述子要素ビットから生成するように構成される、ワード生成手段と、 40

第1のシーケンス内の複数の画像に対して生成されるワードと、第2のシーケンス内の複数の画像に対して生成されるワードとを比較することによって、第1のシーケンスと第2のシーケンスとを比較する処理を実施するように構成されるシーケンス比較手段とを備える、装置を提供する。

【0010】

シーケンス比較手段は、第1のシーケンス及び第2のシーケンス内に異なるワード値が発生する頻度を比較する処理を含む、第1のシーケンスと第2のシーケンスとを比較する 50

処理を実施するように構成することができる。

【0011】

シーケンス比較手段は、

第1のシーケンスに対する発生データを生成するために、第1のシーケンス内の各ワードに対する各可能なワード値の発生回数を求めること、

第2のシーケンスに対する発生データを生成するために、第2のシーケンス内の各ワードに対する各可能なワード値の発生回数を求めること、及び、

第1のシーケンスに対する発生データと、第2のシーケンスに対する発生データとを比較することを含む、第1のシーケンスと第2のシーケンスとを比較する処理を実施するよう構成することができる。

10

【0012】

シーケンス比較手段は、第1のシーケンスに対する発生データと、第2のシーケンスに対する発生データとを比較して、各ワードについての比較結果を生成すると共に、該それぞれの比較結果をマージして全体の比較結果を生成するよう構成することができる。

【0013】

本装置は、

各ワードに対して、双方のシーケンス内に発生するワード値を求める手段と、

各ワードに対して、双方のシーケンスにおけるワード値の時間的な発生順序を求める手段と、

該時間的な発生順序を比較する手段と
をさらに備えることができる。

20

【0014】

本装置は、

一致する画像対であって、第1のシーケンスからの1つの画像と第2のシーケンスからの1つの画像とを含む画像対を特定する手段と、

少なくとも1つの特定された画像対間の類似度を求めるためにさらなる対応性試験を実施する手段と

をさらに備えることができる。

【0015】

本発明はまた、画像を規定する画像データを処理して、該画像を表す少なくとも1つの値を生成する装置であって、

30

該装置は、

画像内の複数の画素近傍集合のそれぞれに対する画像データを処理するよう構成される記述子要素生成手段であって、該画素近傍集合のそれぞれに対して少なくとも1つの記述子要素を生成し、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、

記述子要素から、少なくとも1つのワードであって、各ワードが、1つ又は複数の記述子要素からの1つ又は複数のビットを含むと共に、該ワードのすべてにおける総ビット数が記述子要素のすべてにおける総ビット数を下回るような、少なくとも1つのワードを形成するよう構成されるワード生成手段と

40

を備える、装置を提供する。

【0016】

本発明はまた、画像を規定する画像データを処理して、該画像を表す複数の値を生成する装置であって、

該装置は、

画像内の複数の画素近傍集合のそれぞれに対する画像データを処理するよう構成される記述子要素生成手段であって、該画素近傍集合のそれぞれに対して少なくとも1つの記述子要素を生成し、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、

記述子要素から、複数のワードであって、各ワードが、記述子要素ビットの固有の組合

50

せを含むような、複数のワードを形成するように構成されるワード生成手段とを備える、装置を提供する。

【0017】

本装置は、各記述子要素を量子化する量子化手段をさらに備えることができ、ワード生成手段は、量子化された記述子要素から各ワードを形成するように構成することができる。

【0018】

量子化器は、各記述子要素を量子化して、それぞれの1ビットの2値結果を形成するように構成することができる。

【0019】

記述子要素生成手段は、複数の異なる画素解像度で画像に対する記述子要素を生成するように構成することができ、ワード生成手段は、複数の異なる画素解像度で生成された記述子要素からの1つ又は複数のビットを組み合わせることによって、少なくとも1つのワードを生成するように構成することができる。

【0020】

本発明はまた、画像を規定する画像データを処理して、該画像を表す少なくとも1つの値を生成する装置であって、

該装置は、

画像内の複数の画素近傍集合のそれに対する画像データを処理するように構成される記述子要素生成手段であって、該画素近傍集合のそれに対して少なくとも1つの記述子要素を生成し、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、

各記述子要素を量子化するように構成される量子化手段と、

量子化された記述子要素から画像を表す少なくとも1つの値を形成するように構成される表現値形成手段と

を備える、装置を提供する。

【0021】

量子化器は、各記述子要素を量子化して、それぞれの1ビットの2値結果を形成するように構成することができる。

【0022】

本発明はまた、画像のシーケンスを規定する画像データを処理して、シーケンスの表現を生成及び記憶する装置であって、

該装置は、

シーケンス内の各画像に対して、画像内の複数の画素近傍集合のそれに対する画像データを処理するように構成される記述子要素生成手段であって、該画素近傍集合のそれに対して少なくとも1つの記述子要素を生成し、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、

シーケンス内の各画像に対して、画像の記述子要素から、少なくとも1つのワードであって、各ワードが、1つ又は複数の記述子要素からの1つ又は複数のビットを含むと共に、画像の該ワードのすべてにおける総ビット数が、該画像の記述子要素のすべてにおける総ビット数を下回るような、少なくとも1つのワードを形成するように構成されるワード生成手段と、

いずれのワードが画像のシーケンス内に現れるか、及び現れる回数を規定するヒストグラムデータを生成するように構成されるヒストグラムデータ生成手段と、

シーケンス内の画像に対するヒストグラムデータ、ワード、及び記述子要素を記憶装置に書き込むように構成されるデータ書き込み手段と
を備える、装置を提供する。

【0023】

本装置は、各記述子要素を量子化するように構成される記述子要素量子化手段をさらに備えることができ、ワード生成手段は、量子化された記述子要素から各ワードを形成する

10

20

30

40

50

ように構成することができる。加えて、本装置は、ヒストグラムデータを量子化するよう構成されるヒストグラムデータ量子化手段をさらに備えることができ、データ書込み手段は、量子化されたヒストグラムデータ、ワード、及び量子化された記述子要素を記憶装置に書き込むように構成することができる。

【0024】

本装置は、ビデオシーケンスの断片を含む画像シーケンスを処理するように動作可能とすることができる。

【0025】

データ書込み手段は、ヒストグラムデータ、ワード、及び記述子要素をビットストリーム内に記憶するように構成することができる。

10

【0026】

本発明はまた、画像のシーケンスを規定する画像データを処理して、シーケンスの表現を生成及び記憶する装置であって、

該装置は、

シーケンス内の各画像に対して、画像内の複数の画素近傍集合のそれに対する画像データを処理するように構成される記述子要素生成手段であって、該画素近傍集合のそれに対して少なくとも1つの記述子要素を生成し、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、

シーケンス内の各画像に対して、画像の記述子要素から、複数のワードであって、各ワードが、記述子要素ビットの固有の組合せを含む、複数のワードを生成するように構成されるワード生成手段と、

20

いずれのワードが画像のシーケンス内に現れるか、及び現れる回数を規定するヒストグラムデータを生成するように構成されるヒストグラムデータ生成手段と、

シーケンス内の画像に対するヒストグラムデータ、ワード、及び記述子要素を記憶装置に書き込むように構成される、データ書込み手段とを備える、装置を提供する。

【0027】

本装置は、各記述子要素を量子化するように構成される記述子要素量子化手段をさらに備えることができ、ワード生成手段は、量子化された記述子要素から各ワードを形成するように構成することができる。加えて、本装置は、ヒストグラムデータを量子化するように構成されるヒストグラムデータ量子化手段をさらに備えることができ、データ書込み手段は、量子化されたヒストグラムデータ、ワード、及び量子化された記述子要素を記憶装置に書き込むように構成することができる。

30

【0028】

本装置は、ビデオシーケンスの断片を含む画像のシーケンスを処理するように動作可能とすることができる。

【0029】

データ書込み手段は、ヒストグラムデータ、ワード、及び記述子要素をビットストリーム内に記憶するように構成することができる。

【0030】

本発明はまた、画像のシーケンスを規定する画像データを処理して、シーケンスの表現を生成及び記憶する装置であって、

40

該装置は、

シーケンス内の各画像に対して、画像内の複数の画素近傍集合のそれに対する画像データを処理するように構成される記述子要素生成手段であって、該画素近傍集合のそれに対して少なくとも1つの記述子要素を生成し、各記述子要素は1つ又は複数のビットを含む、記述子要素生成手段と、各記述子要素を量子化するように構成される記述子要素量子化手段と、

シーケンス内の各画像に対して、量子化された記述子要素から画像を表す少なくとも1つの値を形成するように構成される表現値形成手段と、画像を表すいずれの値が画像のシーケンス内に現れるか、及び現れる回数を規定するヒストグラムデータを生成するように

50

構成されるヒストグラムデータ生成手段と、

ヒストグラムデータを量子化するように構成されるヒストグラムデータ量子化手段と、

量子化されたヒストグラムデータ、画像を表す値、及び量子化された記述子要素を記憶装置に書き込むように構成される、データ書込み手段と
を備える、装置を提供する。

【0031】

本装置は、ビデオシーケンスの断片を含む画像のシーケンスを処理するように動作可能と
とすることができる。

【0032】

データ書込み手段は、ヒストグラムデータ、画像を表す値、及び記述子要素をビットス
トリー内に記憶するように構成することができる。 10

【0033】

本発明の一実施の形態は、
ビデオシーケンスの非常にコンパクトな表現を提供し、
トレーニングデータに依存せず、
非常に高速の検索及び照合を提供し、
一般的な編集／処理操作に対してロバストであり、且つ
照合フレーム位置の正確な位置特定を提供する、
ビデオ識別のための新規の方法及び装置を提供する。

【0034】

ここで、本発明の実施形態を、添付の図面を参照して例示としてのみ説明する。 20

【図面の簡単な説明】

【0035】

【図1】一実施形態における、記述子要素を生成する処理を示す図である。

【図2】一実施形態における、記述子要素からワードを生成する処理を示す図である。

【図3】一実施形態における2値化を示す図である。

【図4】一実施形態における、時間シフト及びフレームレート変化の決定を示す図である。
。

【図5】一実施形態の処理動作を実施する処理装置の一例を示す図である。

【発明を実施するための形態】

【0036】

これより、本発明の一実施形態における処理装置によって実施される方法を説明する。
本方法は、いくつかの処理動作を含む。本明細書の末尾において説明するように、これらの
処理動作は、ハードウェア、ファームウェア、コンピュータプログラム命令に従って動作する
処理ユニット、又はそれらの組合せを使用する処理装置によって実施することができる。

【0037】

複数のビデオフレーム $f_i(c, x, y)$ から成るシーケンスを考える。ここで、 i は
フレームインデックスであり、たとえば $i \in [0, T - 1]$ であり、ここで、 T はシーケンス内
のフレームの総数である。 c は 1 (エル) 個のカラーチャネルにおけるインデックスであり、たとえば $c = 3$ では $c \in \{Y, U, V\}$ である。 x 及び y は空間座標
であり、たとえば $x \in [0, M - 1]$ 、 $y \in [0, N - 1]$ であり、ここで M は水平フレーム解像度であり、 N は垂直フレーム解像度である。 40

【0038】

本発明の好ましい一実施形態では、各フレームをその画素強度(pixel intensity)及び
差分に関して記述する。このような記述子は、本発明の同時継続出願である欧州特許出願
第 1640913 号及び欧州特許出願第 1640914 号に記載されており、これらは参考
によりその全体が本明細書に援用される。本発明の好ましい一実施形態では、空間的に
再サンプリングされた $m \times m$ 画素解像度のフレームを考える。ここで、 m は 2 の累乗である。
好ましくは、 m は小さな値、たとえば $m = 16$ 又は $m = 32$ であるが、これは限定で 50

はない。ここで、各フレームは重複しない 2×2 画素近傍集合 (pixel neighbourhoods) に分割され、各近傍集合において記述子要素が以下のように計算される。

【0039】

【数1】

$$\overset{m}{d}_i(c, p, q) = (f_i(c, p, q) + f_i(c, p+1, q) + f_i(c, p+1, q+1) + f_i(c, p, q+1))/4 \quad (1)$$

$$\overset{m}{d}_i(c, p+1, q) = (f_i(c, p, q) - f_i(c, p+1, q))/2 \quad (2)$$

$$\overset{m}{d}_i(c, p+1, q+1) = (f_i(c, p+1, q) - f_i(c, p+1, q+1))/2 \quad (3)$$

10

$$\overset{m}{d}_i(c, p, q+1) = (f_i(c, p+1, q+1) - f_i(c, p, q+1))/2 \quad (4)$$

【0040】

これは、図1において、 $m = 8$ 及び $c = Y$ について一般性を損なうことなく示されている。式(1)による記述子要素は平均強度であり、フレーム全体に対してとられ、そのフレームを $(m/2) \times (m/2)$ 画素解像度で再サンプリングしたものを与え、これが式(1)～(4)に従って再処理される。このプロセスは、フレーム全体に対する平均強度が計算されるまで継続する。したがって、フレーム f_i に対する完全な記述子は、異なる複数の画素解像度（すなわち、 $m \times m$, $m/2 \times m/2$, ..., 4×4 , 2×2 ）における、そのフレームに対する複数の記述子要素を含み、したがって、

20

【0041】

【数2】

$$d_i = \left\{ d_{i,2}, d_{i,4}, \dots, d_{i,m/2}, d_{i,m} \right\}$$

【0042】

によって与えられる。本発明の好ましい一実施形態では、フレームは自身の輝度チャネル Y によって表され、記述子 d_i はこの輝度情報から抽出される。

【0043】

30

任意選択で、記述子要素は所望のビット数に量子化される。該所望のビット数は異なるものであってもよく、要素ごとに、またはカラーチャネルごとに異なってもよい。本発明の好ましい一実施形態では、記述子要素は以下のように2値化される。

【0044】

【数3】

$$\tilde{d}_{i,Avg} = \begin{cases} 1 & d_{i,Avg} \geq r/2 \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

40

【0045】

... (5)

ここで、 r はダイナミックレンジであり、たとえば8ビット画素値について $r = 256$ である。

【0046】

【数4】

$$\tilde{d}_i = \begin{cases} 1 & d_i \geq 0 \text{ のとき} \\ DIFF & \\ 0 & \text{その他} \end{cases}$$

【0047】

…(6)

ここで、

【0048】

【数5】

$$\frac{\tilde{d}_i}{AVG}$$

【0049】

は式(1)による平均を示し、

【0050】

【数6】

$$\frac{\tilde{d}_i}{DIFF}$$

【0051】

は式(2)～(4)による差分を示す。(5)及び(6)の利点は、それらを、要素の整数部分の2値表現のMSB(最上位ビット)を維持することと等価となるように実施することができるることである。

【0052】

代替的な実施形態では、位置センシティブハッシング(locality sensitive hashing)のような他の量子化／2値化技法を採用することもできる。この方法は、本明細書においては検討しないが、Samet H.著「Foundations of Multidimensional and Metric Data Structures」(Morgan Kaufmann, 2006)に記載されている。

【0053】

次いで、フレームに対する完全な記述子が使用されて、少なくとも1つのコンパクトなワード(word)が、記述子要素のうちの1つ又は複数の、1つ又は複数のビットに基づいて形成される。これは図2(a)に示されている。「ワード」という用語は、本明細書においては任意のビット数を有し得るワードを指すのに使用され、いくつかのコンピュータ分野の刊行物にあるような特定のビット数に具体的に限定されるわけではないことに留意されたい。より具体的には、本発明の好ましい実施形態では、2値化された

【0054】

【数7】

$$\tilde{d}_i$$

【0055】

に含まれる要素のいくつかからなる、小さな順序集合が連結されて、1つのワードを生成する。いずれの要素が選択されるかに対する限定は存在せず、ワードは、異なる複数のカラーチャンネルから、異なるレベルの(たとえば 16×16 又は 4×4 の解像度で抽出される)粗さ(coarseness)からの要素を、式(1)による平均であっても式(2)～(4)による差分であっても混合することができる。たとえば、本発明の1つの実施形態では、要素はすべての利用可能な要素の中からランダムに選択することができる。本発明の別の実施形態では、要素は、たとえばビデオフレームの特定の空間領域のみを覆う或る幾何パ

10

20

30

40

50

ターンに従って選択することができる。本発明のさらに別の実施形態では、要素はそれらのエントロピーに従って順序付けることができ、エントロピーが最も高い要素が選択されてワードを形成する。したがって、ワード形成のプロセスは 次元空間から 次元空間への射影であり、好ましくは << である。2つのビデオフレームに関して、2つの対応するワード（すなわち、順序を有する選択された要素からなるビットパターンのうち対応するもの）間の距離は、フレーム記述子全体の距離の近似である。

【0056】

ワードを構成する、順序付けされている要素のすべての可能な値のすべての可能な組合せは、そのワードに対する語彙(vocabulary)を与える。すなわち、ワードは複数の記述子要素からなる固有の組合せであり、或るワードに対する語彙はそのワードがとることができ異なる複数の値から成る集合である。たとえば、ワードが2つの記述子要素を含み、各記述子要素が1ビット長であり、各ビットが値0又は1を自由にとることができる場合、そのワードに対する語彙は00、01、10、及び11である。

【0057】

本発明の好ましい一実施形態では、図2(b)に示すように、各フレームから複数のワードが抽出され、各ワードは2値化された

【0058】

【数8】

$$\tilde{d}_i$$

10

20

【0059】

に含まれる要素のいくつかからなる、小さな順序集合から与えられる。好ましくは、各ワードの要素は任意の適切な方法（ランダム射影、幾何学的制約、エントロピー順序付け等）に従って選択される。各ワードは語彙を有し、各ワードを生成するビットの組合せは固有であるが、部分的に重複する可能性を排除するものではない。たとえば、いくつかの要素は2つ以上のワードに現れる。さらに、各ワードは任意の他のワードと同じ長さを有してもよいし、有しなくてもよい。すなわち、から₁、₂へ等の次元削減を行うことができる。

【0060】

30

本発明の代替的な実施形態では、その元の解像度及びダイナミックレンジにおける記述子 d_i の任意のバージョン（量子化されても2値化されても）からワードを形成することができる。このような実施形態では、1つの要素が1つのワードに対して2ビット以上寄与することができ、また異なるワードに対して異なるビット数寄与することができ、1つのワードはさまざまな要素からのさまざまな数のビットを含むことができる。

【0061】

ビデオシーケンスは上述のように、そのフレームのうちの1つ又は複数によって記述される。次いで、1つ又は複数の語彙のそれぞれにおいて異なるワードが発生する頻度を計算する。より具体的には、本発明の好ましい実施形態では、シーケンス内の複数のフレームが処理され、各フレーム f_i から 次元の完全な記述子及び複数のワード w_{ik} ($k \in [0, Q-1]$) が抽出され、各ワードがQ個の語彙のうちの1つに対応する。次いで、各語彙 k について、フレームシーケンス内に見られるワードのヒストグラム h_k がプロットされる。簡単に言えば、このようなヒストグラムは、いずれのワードがフレームシーケンス内に現れるか、及び現れる頻度を示す。より具体的には、各ワードについて、一方の軸に、そのワードがとり得る各値（すなわち、ワードの語彙内の各値）に対するそれぞれのビンを有し、他方の軸に、各値の発生回数を記録するヒストグラムが生成される。

【0062】

2つのフレームシーケンス f^1_i 及び f^2_j について、各語彙 k に対するヒストグラム h^1_k と h^2_k とを比較してその結果をマージして決定に至ることによって、それらの類似度を評価することができる。この比較は、ヒストグラムインターセクションのような適切な距離

40

50

測度を使用して達成することができる。本発明の好ましい一実施形態では、ヒストグラムは2値化される。ここで、このような2値化ヒストグラムは多く存在するワードを示す。2値化は、複数の適切な方法（単純な閾値化、適応的な閾値化等）のうちの任意の方法によって達成することができる。2値化ヒストグラム

【0063】

【数9】

$$\tilde{h}_k^1$$

【0064】

10

及び

【0065】

【数10】

$$\tilde{h}_k^2$$

【0066】

について、それらの距離は、好ましくは以下の式によって与えられるJaccard距離測度によって測定される。

【0067】

20

【数11】

$$D_k \left(\tilde{h}_k^1, \tilde{h}_k^2 \right) = 1 - \frac{\tilde{h}_k^1 \cap \tilde{h}_k^2}{\tilde{h}_k^1 \cup \tilde{h}_k^2}$$

【0068】

… (7)

簡単に言えば、式(7)は、所与の語彙におけるフレームシーケンス f^1_j 及び f^2_j の距離を、それらが共通に有する各ワードと、それらが全体で含む全てのワードとの関数として測定する。

30

【0069】

Q 個の語彙について、 Q 個のJaccard距離

【0070】

【数12】

$$D_0, D_1, \dots, D_{Q-1}$$

【0071】

がある。本発明の好ましい一実施形態では、以下のように、これらの距離が融合されて、複合距離

40

【0072】

【数13】

$$D$$

【0073】

が与えられる。

【0074】

【数14】

$$D_J = \sum_{k=0}^{Q-1} D_k$$

【0075】

... (8)

したがって、

【0076】

【数15】

10

$$D_J$$

【0077】

の処理（好ましくは閾値化）によって、フレームシーケンスが類似しているか否かについての決定に至ることができる。この事例において、

【0078】

【数16】

$$D_J$$

20

【0079】

が或る閾値未満である場合、シーケンスは一致すると宣言され、そうでない場合は一致しないと宣言される。

【0080】

明らかに、トリム平均、メジアン、最大値等のような、式(8)に対する多くの代替物が存在する。さらに、代替的な実施形態は、距離融合ではなく、決定融合（各

【0081】

【数17】

$$D_k$$

30

【0082】

を適切な閾値を用いて個々に閾値化して、次いで個々の決定をたとえば多数決によって融合すること）によって、フレームシーケンスの類似度に対する決定に至ることができる。本発明の代替的な一実施形態では、距離融合と決定融合と組み合わせること、すなわち、複合の

【0083】

【数18】

$$D_J$$

40

【0084】

を計算し、続いて閾値化することによって決定に至ることも可能であるが、一致の決定を、少なくともある設定された数の

【0085】

【数19】

$$D_k$$

50

【0086】

距離によってサポートする必要もある。

【0087】

任意選択で、各語彙に関してビデオシーケンス内の一致するワードの発生の時間的順序を考慮することによって、検出された一致が検証される。好ましい一実施形態では、各語彙 k に対して、双方のビデオシーケンス f^1_i 及び f^2_j において発生するワードの集合が見つけられる。好ましくは、これらは、双方のヒストグラム h^1_k 及び h^2_k において非ゼロである要素（2値ヒストグラム

【0088】

【数20】

10

$$\tilde{h}_k^1$$

【0089】

及び

【0090】

【数21】

$$\tilde{h}_k^2$$

20

【0091】

の共通部分）を特定することによって見つけられる。各ヒストグラムの各要素が語彙からの

1つのワードに対応することを想起すると、共通部分は、 z 個の要素を有する、双方のフレームに共通であるワードの集合 $W^{12}_k = \{ w_{k, [0, z-1]} \} = \{ w_{\cdot, \dots, \cdot} \}$ を生成する。好ましくは、この集合は、語彙に従って、たとえば数値順又はアルファベット順に順序付けされる。次いで、各ビデオシーケンスについて、順序集合 W^{12}_k におけるワードそれについての最初の発生を記録したフレーム番号の集合 $\{ k, [0, z-1] \} = \{ f_{\cdot, \dots, \cdot} \}$ が生成される。

【0092】

30

フレーム番号の2つの集合 1_k 及び 2_k （それぞれ f^1_i 及び f^2_j に対応する）が比較されて、それらの間の距離が求められる。好ましい実施形態では、 1_k 及び 2_k は2値化され、それらの間の距離が正規化ハミング距離によって与えられる。好ましい一実施形態では、2値化は、集合

【0093】

【数22】

$$\tilde{\omega}_{\lambda} = \begin{cases} 0 & \omega_{\lambda} < \omega_{\lambda+1} \text{ のとき} \\ 1 & \text{その他} \end{cases}$$

40

【0094】

…（9）

に含まれる互いに隣接する要素の間の差を評価することによって、図3に示されるように実施される。

【0095】

2値化集合は

【0096】

【数23】

$$\check{\Omega}_k = \left\{ \tilde{\omega}_{\lambda k}, \lambda \in [0, z-1] \right\}$$

【0097】

と表され、正規化ハミング距離は

【0098】

【数24】

$$\overline{D}_H^{\Omega}$$

10

【0099】

と表される。簡単に言えば、このような2値化集合は、順序集合 W^{12}_k 内の各ワードがビデオシーケンスにおいて最初に現れるのが、ワードの順序集合 W^{12}_k 内におけるその次のワードの前であるか又は後であるかを記録する。

【0100】

本発明の代替的な実施形態では、 1_k 及び 2_k は2値化することなく、適切な距離測度(たとえば、一方のシーケンス内のワードの発生状況が他方のシーケンスと異なるようなフレームすべての数を与えるL1)を使用して比較することができる。

【0101】

Q 個の語彙について、 Q 個の正規化ハミング距離

20

【0102】

【数25】

$$\overline{D}_H^{\Omega_0}, \overline{D}_H^{\Omega_1}, \dots, \overline{D}_H^{\Omega_{Q-1}}$$

【0103】

がある。本発明の好ましい一実施形態では、以下のように、これらの距離が融合されて、複合距離

【0104】

30

【数26】

$$\overline{D}_H^{\Omega}$$

【0105】

が、

【0106】

【数27】

$$\overline{D}_H^{\Omega} = \sum_{k=0}^{Q-1} \overline{D}_H^{\Omega_k}$$

40

【0107】

…(10)

として与えられる。

したがって、

【0108】

【数28】

$$\overline{D}_H^\Omega$$

【0109】

の処理（好ましくは閾値化）によって、フレームシーケンスが類似しているか否かについての決定に至ることができる。この事例において、

【0110】

【数29】

$$\overline{D}_H^\Omega$$

10

【0111】

が或る閾値未満である場合、シーケンスは一致すると宣言され、そうでない場合は一致しないと宣言される。

【0112】

明らかに、トリム平均、メジアン、最大値等のような、式(10)に対する多くの代替物が存在する。さらに、代替的な実施形態は、距離融合ではなく、決定融合（すなわち、各

【0113】

【数30】

$$\overline{D}_H^{\Omega_k}$$

20

【0114】

を適切な閾値を用いて個々に閾値化して、次いで個々の決定をたとえば多数決によって融合することによって、フレームシーケンスの類似度に対する決定に至ることができる。本発明の代替的な一実施形態では、距離融合と決定融合と組み合わせること、すなわち、複合の

30

【0115】

【数31】

$$\overline{D}_H^\Omega$$

【0116】

を計算し、続いて閾値化することによって決定に至ることも可能であるが、一致の決定を、少なくともある設定された数の

【0117】

【数32】

$$\overline{D}_H^{\Omega_k}$$

40

【0118】

距離によってサポートする必要もある。

【0119】

任意選択で、さらなる改良及び位置特定段階が適用されて、ビデオシーケンスが一致するか否かが非常に高いレベルの精度まで決定され、2つのシーケンスにおける対応するフ

50

レームの位置が求められる。双方のシーケンスからの一致するワードの集合が使用されて、可能性のあるフレームの対応性が求められる。次いで、シーケンス間の時間的变化を推定することができる。好ましい実施形態では、図4に示すように、シーケンス間の時間シフト及びフレームレート差が求められる。次いで、対応する可能性のある複数のフレームからの記述子を使用して、対応の妥当性を判断し且つ／又は時間変化パラメータを改良することができる。

【0120】

より具体的には、各語彙 k について、共通部分 $W^{12}_k = \{ w_k, [0, z-1] \} = \{ w_1, \dots, w_z \}$ 内のワードのそれぞれは、ビデオシーケンスのそれぞれからの1つ又は複数のフレームに対応し、ここで「1つ又は複数」とは、各シーケンスについて異なる数であり得る。たとえば、あるシーケンス内の1つのみのフレームと、第2のフレーム内の多数のフレームとによって1つのワードを生成することができる。一対のビデオシーケンスについて、また各語彙 k について、また W^{12}_k 内の z 個のワードのそれぞれについて、シーケンス f^1_i 及び f^2_j 内においてそのワードに対応するフレーム番号対を含む集合を構築する。したがって、 W^{12}_k 内の z 個のワードのうちの1つが、 f^1_i 及び f^2_j 内のフレームの单一の対によって生成される場合、そのワードに対する集合は一対のみのフレーム番号を含む。別のワードが、 f^1_i 及び f^2_j 内の複数対のフレームによって生成される場合、そのワードに対する集合はこれと同じ複数の対のフレーム番号を含む。

【0121】

次いで、 W^{12}_k 内の各ワードについて、また各語彙 k について、そのワードに対応するフレーム番号対は、単一の2次元ヒストグラム $G^{12}_{i,j}$ (ただし $i \in [0, T_1 - 1], j \in [0, T_2 - 1]$) 内にプロットされ、ここで T_1 はシーケンス f^1_i 内のフレームの数であり、 T_2 はシーケンス f^2_j 内のフレームの数である。簡単に言えば、 $G^{12}_{i,j}$ において、ビン値 0 は、検討中のフレーム対、すなわち (i, j) が Q 個の語彙のいずれにおいても一致するワードを生成しなかったことを表し、 $k < Q$ のビン値は、検討中のフレーム対が Q 個の語彙のうちの k 個のみにおいて一致するワードを生成したことを表し、ビン値 Q は、検討中のフレーム対が Q 個すべての語彙において一致するワードを生成したことを表す。

【0122】

f^1_i 及び f^2_j 内のフレームは、1つ又は複数の語彙において同じワードを有する場合には対応しているとみなされ、これは対応性の弱試験であるとみなすことができる、すなわち、これらのフレームは弱く対応している。弱く対応しているフレーム対のそれぞれについて、それらをより詳細な記述子に従って比較することによって、対応性のより強い試験が実行される。好ましくは、上述のように f^1_i 及び f^2_j から抽出される2値記述子

【0123】

【数33】

$$\tilde{d}_i^1$$

【0124】

及び

【0125】

【数34】

$$\tilde{d}_j^2$$

【0126】

が使用される。このような2値記述子間の距離は好ましくは、ハミング距離

【0127】

【数35】

$$\check{D}_{\check{H}}^{\check{d}}$$

【0128】

として計算される。

【0129】

本発明の代替的な実施形態では、その元の解像度及びダイナミックレンジにおける、式(1)～(4)に記述されている記述子dの任意のバージョン(量子化されても2値化されても)、又はこの記述子の適切な部分集合を使用することができ、ハミング距離

【0130】

【数36】

$$\check{D}_{\check{H}}^{\check{d}}$$

【0131】

は適切な距離測度、たとえばL1に置き換えられる。

【0132】

弱く対応するフレーム対が多数存在する場合、処理時間を有益に低減するために、限られた数のフレーム対のみを強く対応するフレーム対であると確認することができる。好ましくは、これらの対は共通のワードが最も多い順に試験される。すなわち、Q個の共通のワードを有する対が最初に試験され、次いで、Q-1個の共通のワードを有する対が試験され、以下同様である。共通のワードの数によるこのタイプの順序付けは、2Dヒストグラム G^{12}_{ij} によって容易に提供される。弱く対応するフレーム対は、フレーム記述子間の距離が特定の基準を満たす場合に(すなわち、好ましくは所定の閾値を下回る場合に)、強く対応するフレーム対として受容される。これは、残りの弱く対応する対がなくなるか、又は強く対応する対が特定数発見されるまで繰り返される。上記から、Q個よりも少ない一致ワードを有するフレーム対、又はただ1つの一致ワードしか有しないフレーム対であっても、強く対応するフレーム対であると確認される場合があり、一方で、Q個の語彙内でQ個の一致ワードを有するフレーム対が、それらの詳細な記述子によれば類似性に乏しいために、強く対応するフレーム対であると確認されない場合があることになる。このように、一般的に、より多数の弱く対応するフレーム対は、より少数の強く対応するフレーム対に低減される。

【0133】

次いで、強く対応する対が使用されて、フレームシーケンス間の時間パラメータ変化が推定される。好ましい一実施形態では、時間パラメータは時間シフト及びフレームレート変化である。これらは線形特性であり、したがって2つの強く対応する対を使用して推定することができる。好ましくは、ハフ変換(Hough transform)が使用されて時間パラメータが推定される。ハフ空間内のピークは最も可能性のある時間パラメータに対応する。ハフ変換及びラドン変換(Radon transform)はここでは検討しないが、van Ginkel, M.、Hen driks, C. L.、van Vliet, L. J.著「A short introduction to the Radon and Hough transforms and how they relate to each other」(Number QI-2004-01 in the Quantitative Imaging Group Technical Report Series, Delft University of Technology)に専門的に記載されている。

【0134】

ハフ変換は一般的に、パラメータの粗い推定値を得るために使用される。これらのパラメータを使用して、2つのビデオシーケンス間のフレーム対応性が確立される。好ましい一

10

20

30

40

50

実施形態では、フレーム対応性をより正確に求めることができる。単一の記述子は一定レベルの精度を提供するが、複数の記述子を組み合わせることによってより高いレベルの精度が達成される。好ましい一実施形態では、記述子間のハミング距離が使用され、このハミング距離は全記述子にわたって加算される。好ましくは、2つのシーケンスからの複数の対応するフレームが選択されて、該フレームからの記述子の組合せの間の距離を試験することによって、該シーケンス間の距離が求められる。好ましい一実施形態では、一群の（たとえば7つの）連続するフレームが第1のシーケンスの開始部分から選択され、第2のシーケンスからの対応するフレームと比較される。代替的に、各フレームはシーケンスの複数の部分（中央、終了部分等）からのものであってもよく、且つノ又はフレームは不連続であってもよい。好ましい一実施形態では、対応する対のうち一方のフレームを固定しておき、他方の対応するフレームの周辺領域（locality）においてより良好に対応するフレームを求めて検索することによって、フレームの対応性が改良される。

【0135】

一実施形態では、効率的な記述子記憶方式が使用されて、ビデオシーケンスのコンパクトな表現が形成される。好ましい実施形態では、シーケンスは、たとえば1秒の断片に分割される。各断片について、Q個の2値ヒストグラム

【0136】

【数37】

\tilde{h}_k

【0137】

がビットストリーム内に記憶される。2値ヒストグラムに続いて、断片内のフレームのそれぞれに対する記述子

【0138】

【数38】

\tilde{d}_i

【0139】

が記憶される。シーケンス内のすべての断片が、このようにしてビットストリーム内に連続して記憶される。任意選択で、ビットストリームの先頭に、シーケンスのフレームレート、フレームの数、断片の数等のような情報を記憶することができる。

【0140】

代替的な一実施形態では、シーケンスは、ビットストリーム内で複数のフレーム表現によって表現することができる。ここで、各フレームはQ個のワードと1つの2値記述子

【0141】

【数39】

\tilde{d}_i

【0142】

とによって表される。シーケンス内のフレームのすべてをこのように記憶することができる。別の代替形態では、最初に全てのフレーム内のすべてのワードを記憶し、続いてすべての記述子を記憶することが望ましい場合がある。明らかに、ヒストグラム、ワード、及び記述子の量子化されたバージョン又は元のバージョンを記憶することが可能である。

【0143】

本発明の代替的な一実施形態では、ビデオフレームから抽出された記述子を、適切なフレーム記述子（たとえば色ヒストグラム又はエッジ強度ヒストグラム）に変更することができ、これらはその元の形態、量子化された形態、又は2値化された形態で記憶すること

10

20

30

40

50

ができ、これらからワードを形成することができる。本発明の異なる実施形態では、さまざまな決定段階を省略することができる。たとえば、各語彙に関するビデオシーケンス内の一一致ワードの発生の時間的な順序付けに基づく処理及び決定段階は、その全体を省略することができる。これは、処理を、対応するフレームの改良、位置特定、及び決定に直接進めることによってなされる。本発明の代替的な実施形態では、各ビデオシーケンス内で発見される一致ワードによる決定段階を省略することができ、処理を、ビデオシーケンス内の一致ワードの発生の時間的な順序付けに基づく処理及び決定段階に直接進めることができる。

【 0 1 4 4 】

上述の処理動作を実施するデータ処理装置 1 を図 5 に示す。装置 1 は、たとえば、パーソナルデスクトップコンピュータ又はポータブルコンピュータとすることができる。 10

【 0 1 4 5 】

装置 1 は、データ処理装置の従来の要素を備える。これらの要素は当業者に既知であるため、詳細な説明は必要ない。手短に、図 5 の装置 1 は、コンピュータプログラム製品(記憶媒体 5 又は信号 7 等)からコンピュータプログラム命令及び処理されるビデオデータを受信する入力データインタフェース 3 を備える。処理システムはたとえば、C P U 9、ランダムアクセスメモリ 11、及び読み出し専用メモリ 13(これらはバス 15 によって接続される)によって提供される。C P U 9 は動作全体を制御する。R A M 11 は、C P U 9 によって、プログラムを実行すると共に R O M 4(プログラム及び他のデータを記憶する)を制御するのに使用される作業メモリである。装置 1 の処理装置は、本明細書において上述したように画像を規定する画像データを処理する方法を実施するように構成される。この処理の結果は出力インターフェース 17 によって出力される。 20

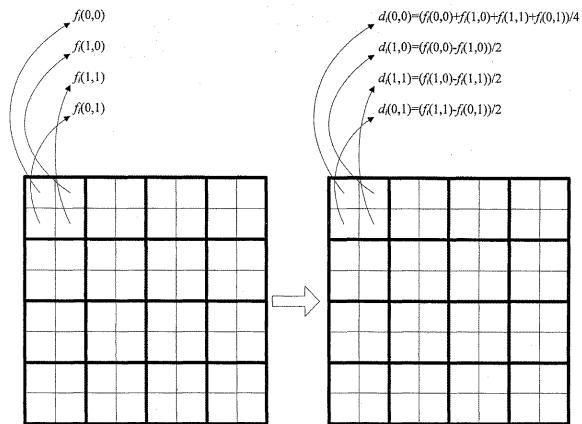
【 0 1 4 6 】

上述の処理装置 1 はコンピュータプログラム命令に従って処理を実施するが、代替的な処理装置を、ハードウェア、ソフトウェア、又はハードウェア及びソフトウェアの任意の適切な組合せとして、任意の適切な又は望ましい方法で実施することができる。本発明を、コンピュータプログラム(プログラム可能処理装置内にロードされ、該装置上で実行されると、上述の画像データ処理方法のうちの 1 つを実行するもの)として具現化することができ、また、コンピュータプログラム製品(たとえばこのようなコンピュータプログラムを記憶するデータキャリア)として具現化することもできることにさらに留意されたい。 30

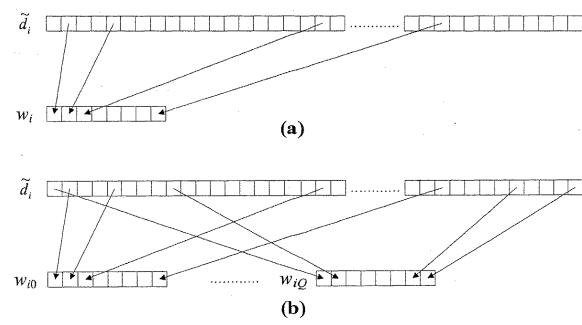
【 0 1 4 7 】

本発明の実施形態の上記の記載は、例示及び説明を目的として提示されている。網羅的であること、又は本発明を開示されている厳密な形態に限定することは意図されていない。本発明の範囲から逸脱することなく変更、改変、及び変形を行うことができる。

【図1】



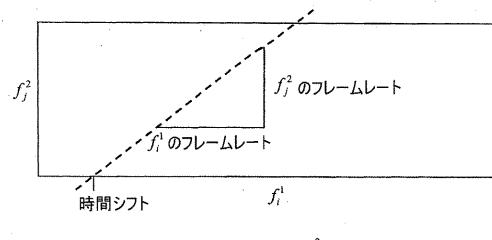
【図2】



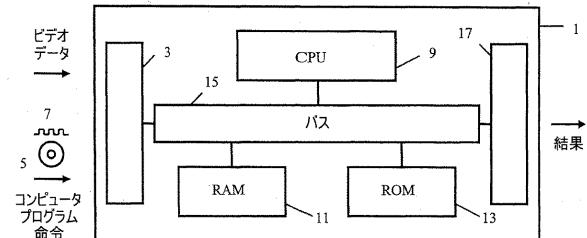
【図3】

ワード	最初に発生するフレーム番号	2値表現
w_α	ω_0	$\tilde{\omega}_0 = \begin{cases} 0 & \omega_0 < \omega_1 \\ 1 & \text{その他} \end{cases}$
$w_{\alpha+1}$	ω_1	$\tilde{\omega}_1 = \begin{cases} 0 & \omega_1 < \omega_2 \\ 1 & \text{その他} \end{cases}$
\vdots	\vdots	\vdots
w_{z-1}	ω_{z-1}	$\tilde{\omega}_{z-1} = \begin{cases} 0 & \omega_{z-1} < \omega_0 \\ 1 & \text{その他} \end{cases}$

【図4】



【図5】



フロントページの続き

(74)代理人 100110423
弁理士 曾我 道治
(74)代理人 100084010
弁理士 古川 秀利
(74)代理人 100094695
弁理士 鈴木 憲七
(74)代理人 100111648
弁理士 梶並 順
(74)代理人 100147500
弁理士 田口 雅啓
(72)発明者 ポール・プラスネット
イギリス国、サリー・ケイティー5・8ジェイダブリュー、サービトン、ジ・アヴェニュー 41
エイ
(72)発明者 スタヴロス・パシャラキス
イギリス国、サリー・ジーユー1・2ピーディー、ギルドフォード、ビーティー・アヴェニュー
3
(72)発明者 ミロスロー・ボバー
イギリス国、サリー・ジーユー1・2エスイー、ギルドフォード、ウィカム・ロード 21

審査官 片岡 利延

(56)参考文献 特開2006-135938(JP,A)
特開2008-282316(JP,A)
特表2008-507718(JP,A)
Kunio Kashino, Takayuki Kurozumi, Hiroshi Murase, A Quick Search Method for Audio and
Video Signals Based on Histogram Pruning, IEEE TRANSACTIONS ON MULTIMEDIA, 2003
年 9月30日, VOL.5, NO.3, pp.348-357, URL, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1223562&userType=inst>

(58)調査した分野(Int.Cl., DB名)

G 0 6 T	7 / 2 0
H 0 4 N	5 / 7 6
H 0 4 N	5 / 9 1