(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6236154号
(P6236154)

(45)発行日　平成29年11月22日(2017.11.22)　(24)登録日　平成29年11月2日(2017.11.2)

(51)Int.Cl.		F I		
G06F 17/30	(2006.01)	G06F	17/30	220B
H04N 5/76	(2006.01)	G06F	17/30	170D
G06T 7/20	(2017.01)	H04N	5/76	
		G06T	7/20	

請求項の数 14　(全 12 頁)

(21)出願番号　特願2016-527359 (P2016-527359)
(86)(22)出願日　平成26年10月10日 (2014.10.10)
(65)公表番号　特表2017-504085 (P2017-504085A)
(43)公表日　平成29年2月2日 (2017.2.2)
(86)国際出願番号　PCT/EP2014/071774
(87)国際公開番号　WO2015/062848
(87)国際公開日　平成27年5月7日 (2015.5.7)
審査請求日　平成28年6月22日 (2016.6.22)
(31)優先権主張番号　13306502.9
(32)優先日　平成25年10月31日 (2013.10.31)
(33)優先権主張国　欧州特許庁 (EP)

(73)特許権者　391030332
アルカテル－ルーセント
フランス国、９２１００・ブローニュ－ビヤンクール、ルート・ドゥ・ラ・レーヌ・１４８／１５２
(74)代理人　110001173
特許業務法人川口國際特許事務所
(72)発明者　マリイ，エマニュエル
フランス国、９１６２０・ノゼイ、ルット・ドゥ・ビルジュスト、サントル・ドゥ・ビラルソー、アルカテル・ルーセント・ベル・ラブズ・フランス気付

最終頁に続く

(54)【発明の名称】動画コンテンツに現れる物体を表す動画タグクラウドを生成する方法

(57)【特許請求の範囲】

【請求項１】

動画コンテンツに現れる物体を表す動画タグクラウドを生成する方法であって、

前記動画コンテンツの動画フレームを抽出し、前記動画フレームを領域へと個別にセグメント化する手段により、前記動画コンテンツの動画フレームを抽出し、前記動画フレームを領域へと個別にセグメント化するステップ（Ｂ）と、

各抽出フレームに関して、前記フレームのセグメント化された領域間の空間的関係をモデル化するトポロジグラフを構築する手段により、各抽出フレームに関して、前記フレームのセグメント化された領域間の空間的関係をモデル化するトポロジグラフを構築するステップ（Ｃ）と、

構築したトポロジグラフ集合から、空間的および時間的制約に従って、頻出パターンを抽出する手段により、構築したトポロジグラフ集合から、空間的および時間的制約に従って、頻出パターンを抽出するステップ（Ｄ）であり、各パターンが少なくとも１つのセグメント化された領域を含む、抽出するステップ（Ｄ）と、

軌跡制約を用いて同じ物体の部分を表す頻出パターンを再グループ化し、前記動画コンテンツの頻出物体を検出する手段により、軌跡制約を用いて同じ物体の部分を表す頻出パターンを再グループ化し、前記動画コンテンツの頻出物体を検出するステップ（Ｅ）と、

各検出された頻出物体に関して、少なくとも前記物体のパターンの抽出に用いられる空間的および時間的制約ならびに前記パターンの再グループ化に用いられる軌跡制約に従って、前記物体に適用する重み付け係数を決定する手段により、各検出された頻出物体に関

して、少なくとも前記物体のパターンの抽出に用いられる空間的および時間的制約ならびに前記パターンの再グループ化に用いられる軌跡制約に従って、前記物体に適用する重み付け係数を決定するステップ（Ｆ）と、

前記頻出物体のそれぞれの視覚表示を含む動画タグクラウドをそれらの重み付け係数に従って生成する手段により、前記頻出物体のそれぞれの視覚表示を含む動画タグクラウドをそれらの重み付け係数に従って生成するステップ（Ｈ）と、

を提供する、方法。

【請求項２】

対応する重み付け係数とともにデータリポジトリ（１５）にさらに格納された検出された頻出物体を抽出してセグメント化する手段により、対応する重み付け係数とともにデータリポジトリ（１５）にさらに格納された検出された頻出物体を抽出してセグメント化するステップ（Ｇ）であり、前記格納された物体および前記重み付け係数から動画タグクラウドが生成される、抽出してセグメント化するステップ（Ｇ）を提供することを特徴とする、請求項１に記載の方法。

【請求項３】

頻出パターンが、動画フレームにおける前記パターンの時間的および空間的発生に従って抽出されることを特徴とする、請求項１または２に記載の方法。

【請求項４】

パターンの時間的発生を、動画フレームにおける前記パターンの２回の発生間の平均時間距離に従って評価する手段により、パターンの時間的発生が、動画フレームにおける前記パターンの２回の発生間の平均時間距離に従って評価されることを特徴とする、請求項３に記載の方法。

【請求項５】

パターンの空間的発生を、同じ動画フレームにおける前記パターンの２回の発生間の平均空間距離に従って評価する手段により、パターンの空間的発生が、同じ動画フレームにおける前記パターンの２回の発生間の平均空間距離に従って評価され、前記空間距離が、以下の式：

$$\max_{s \in V} d(o_1(s), o_2(s))$$

に従って計算され、ここで、Ｖが前記パターンの領域集合、o_1、o_2が同じ動画フレームにおける前記パターンの２回の発生、$d(o_1(s), o_2(s))$が前記パターンの領域ｓの発生間のユークリッド距離であることを特徴とする、請求項３または４に記載の方法。

【請求項６】

同じ物体の部分を表す頻出パターンが、動画フレームにおける前記パターンの軌跡間の相違度合いに従って再グループ化され、前記相違度合いが、以下の式：

【数１】

$$\frac{\sum_{t=1}^{n} x_t}{n}$$

に従って計算され、ここで、x_tが動画フレームｔにおける２つのパターンの図心間のユークリッド距離であり、パターンの図心が、動画フレームｔにおける前記パターンのすべての空間的発生の重心に対応することを特徴とする、請求項３から５のいずれか一項に記載の方法。

【請求項７】

動画コンテンツに現れる物体を表す動画タグクラウドを生成する請求項１から６のいずれか一項に記載の方法を実行するように構成されたコンピュータプログラム。

【請求項８】

動画コンテンツに現れる物体を表す動画タグクラウドを生成する請求項７に記載のコンピュータプログラムを実行するように構成されたアプリケーション装置（１）であって、

前記生成を管理するエンジンモジュール（２）と、

前記動画コンテンツの動画フレームを抽出する手段および前記動画フレームを領域へと個別にセグメント化する手段を備えた抽出モジュール（９）と、
各抽出フレームに関して、前記フレームのセグメント化された領域間の空間的関係をモデル化するトポロジグラフを構築する手段を備えたグラフモジュール（１０）と、
構築したトポロジグラフ集合から、空間的および時間的制約に従って、頻出パターンを抽出する手段であり、各パターンが少なくとも１つのセグメント化された領域を含む、抽出する手段を備えたデータマイニングモジュール（１１）と、
軌跡制約を用いて同じ物体の部分を表す頻出パターンを再グループ化し、前記動画コンテンツの頻出物体を検出する手段を備えたクラスタリングモジュール（１２）と、
各検出された頻出物体に関して、少なくとも前記物体のパターンの抽出に用いられる空間的および時間的制約ならびに前記パターンの再グループ化に用いられる軌跡制約に従って、前記物体に適用する重み付け係数を決定する手段を備えた重み付けモジュール（１３）と、
前記頻出物体のそれぞれの視覚表示を含む動画タグクラウドをそれらの重み付け係数に従って生成する手段を備えた表示モジュール（１６）と、
を備えた、アプリケーション装置（１）。
【請求項９】
検出された頻出物体をそれぞれ抽出してセグメント化する手段を備えたセグメンテーション抽出モジュール（１４）を備え、前記アプリケーション装置が、対応する重み付け係数とともに前記セグメント化された物体を格納するデータリポジトリ（１５）をさらに備え、表示モジュール（１６）が、前記格納された物体および前記重み付け係数から動画タグクラウドを生成することを特徴とする、請求項８に記載のアプリケーション装置（１）。
【請求項１０】
データマイニングモジュール（１１）の抽出手段が、動画フレームにおける前記パターンの時間的および空間的発生に従って頻出パターンを抽出するように構成されたことを特徴とする、請求項８または９に記載のアプリケーション装置（１）。
【請求項１１】
データマイニングモジュール（１１）が、動画フレームにおける前記パターンの２回の発生間の平均時間距離に従ってパターンの時間的発生を評価する手段を備えたことを特徴とする、請求項１０に記載のアプリケーション装置（１）。
【請求項１２】
データマイニングモジュール（１１）が、同じ動画フレームにおける前記パターンの２回の発生間の平均空間距離に従ってパターンの空間的発生を評価する手段を備え、前記空間距離が、以下の式：

$$\max_{s \in V} d(o_1(s), o_2(s))$$

に従って計算され、ここで、Ｖが前記パターンの領域集合、o_1、o_2が同じ動画フレームにおける前記パターンの２回の発生、$d(o_1(s), o_2(s))$が前記パターンの領域ｓの発生間のユークリッド距離であることを特徴とする、請求項１０または１１に記載のアプリケーション装置（１）。
【請求項１３】
クラスタリングモジュール（１２）の再グループ化手段が、動画フレームにおける前記パターンの軌跡間の相違度合いに従って、同じ物体の部分を表す頻出パターンを再グループ化するように構成され、前記相違度合いが、以下の式：
【数２】

$$\frac{\sum_{t=1}^{n} x_t}{n}$$

に従って計算され、ここで、x_tが動画フレームｔにおける２つのパターンの図心間のユークリッド距離であり、パターンの図心が、動画フレームｔにおける前記パターンのすべ

ての空間的発生の重心に対応することを特徴とする、請求項１０から１２のいずれか一項に記載のアプリケーション装置（１）。
【請求項１４】
動画コンテンツから動画タグクラウドを生成する前記アプリケーション装置のユーザおよび／またはインターフェース（３、４、５）による使用を可能にする少なくとも１つのアプリケーションプログラミングインターフェース（ＡＰＩ）（６、７、８）を備えたことを特徴とする、請求項８から１３のいずれか一項に記載のアプリケーション装置（１）。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、動画コンテンツに現れる物体を表す動画タグクラウドを生成する方法に関する。
【背景技術】
【０００２】
一般的なテキストタグクラウドは、インターネットのユーザによく知られており、重み付けキーワードまたはメタデータの一群の視覚表示を本質とする。また、「ワードクラウド」または「重み付けリスト」としても知られており、通常、ウェブサイト上のキーワードメタデータの表現またはフリーフォームテキストの視覚化に用いられる。タグは通例、その重要性がフォントのサイズおよび／または色によって強調された単語である。
【０００３】
一般的に、このようなタグクラウドは、テキスト情報を解析するツールによって提供され、前記ツールでは、入力メタデータ、キーワード、およびテキストを取り込むとともに、たとえば意味解析によってすべてを処理することにより、グローバルタグクラウド上に表示される有意な視覚表示（タグ）を構築する。
【０００４】
たとえば、米国特許第８３５９１９１号は、タグが、異なる言語カテゴリに分離され、および／または共通ドメインに従ってクラスタリングされて表現される、タグクラウドを生成する方法を提供している。
【０００５】
ワードタグクラウドに関する既存の文書の大部分は、テキスト、音声、および／または動画等のマルチメディアコンテンツからワードを抽出した後、前記ワードに専用のアルゴリズムを適用して、前記ワードに適用する適切な重み付けを評価するとともに、前記重み付けに対応した適当なタグを作成する方法を記載している。
【０００６】
動画および画像コンテンツのタグクラウドに関しては、動画または画像コンテンツと関連付けられたテキスト注釈に基づくワードタグクラウドの構築を本質とする基本的な方法が存在する。たとえば、マルチメディアファイル共有ウェブサイトＦｌｉｃｋｒは、そのユーザにより共有される写真および／または動画コンテンツに関連したキーワードに基づいて、上記のようなタグクラウドを提供する。また、画像タグクラウドの構築を本質とし、タグが完全な重み付け画像の視覚表示である、より精緻な方法が存在する。
【０００７】
しかし、上述のような画像タグクラウドは、画像に付随するテキスト注釈の意味解析にも依拠しているが、画像自体の解析には依拠していない。さらに、無料ソフトウェアＷｉｎｋによる提案のように、テキストの意味解析に依拠しない画像タグクラウドは、意味解析なしで構築された単純な表示モデルである。
【０００８】
論文「Ｓｕｉｖｉ　Ｔｒｉｄｉｍｅｎｔｉｏｎｎｅｌ　ｅｎ　Ｓｔｅｒｅｏｖｉｓｉｏｎ」（Ｓ．ＣＯＮＳＥＩＬ、Ｓ．ＢＯＵＲＥＮＮＡＮＥ、Ｌ．ＭＡＲＴＩＮ、ＧＲＥＴＳＩ　２００５年）によれば、動画コンテンツの関心のある物体は、前記コンテンツが移動

しないカメラで取得されたものである場合、バックグラウンド減算手法によって容易に検出可能であることが明らかである。実際、この論文において、著者らは、システムの開始時に取り込まれた基準画像を構成するバックグラウンドを減算することによって、画像中の手を検出している。

【０００９】

しかし、この論文の検出ソリューションでは、物体間の関係を確立不可能であり、動画コンテンツが移動するカメラで取得されたものである場合、このバックグラウンド減算技術は、取得動画コンテンツ中の物体に関して、有用な情報を一切提供しない。

【００１０】

移動するカメラで取得された動画コンテンツを処理する場合は、２つの異なる手法が一般的に用いられるが、第１の手法は、動画コンテンツ中の関心のある物体に対するタグ付けをユーザに要請した後、圧縮アルゴリズムまたは追跡学習検出（ＴＬＤ）アルゴリズム等の運動と見かけモデルを用いることを本質としている。しかし、この技術は、非常に正確な追跡情報を提供するものの、先行してユーザ入力が必要であることから、完全に自動化されたシステムには使用できない。

【００１１】

第２の手法では、取得動画コンテンツに関する事前知識を用いて、この問題を単純化している。このような手法は一般的に、関心のある物体のモデルを事前に学習することを本質としており、前記モデルは、取得動画コンテンツの各フレームにおける類似物体の検出に利用可能である。これら技術の非常に印象的な一例は、論文「Ｍａｘｉｍｕｍ Ｗｒｉｇｈｔ Ｃｌｉｑｕｅｓ ｗｉｔｈ Ｍｕｔｅｘ Ｃｏｎｓｔｒａｉｎｔｓ ｆｏｒ Ｏｂｊｅｃｔ Ｓｅｇｍｅｎｔａｔｉｏｎ」（Ｔ．ＭＡ、Ｌ．Ｊ．ＬＡＴＥＣＫＩ、ＣＶＰＲ ２０１２年）に提示されており、多様な物体カテゴリに対して、事前トレーニングされた一般的な物体モデルをアプリケーションが使用する。しかし、これらの技術は、ユーザ入力なく複数の物体を同時に検出して追跡可能であっても、依然としてトレーニングステップに依存しており、任意の種類の物体では動作しない。

【００１２】

米国特許第５８６７５８４号は、動画像シーケンスを通じて物体を自動追跡可能なシステムを記載しているが、前記システムでは、物体を含むウィンドウの指定ひいてはユーザ相互作用および／または追跡する物体の事前知識が必要となる。

【００１３】

また、米国特許第８３５１６４９号は、トレーニングフェーズを用いた物体追跡技術を記載している。

【００１４】

総じて、上述の方法は一般的に、動画コンテンツの物体の検出、追跡、抽出および前記コンテンツにおける前記物体間の関係の確立するために、事前知識を有するアルゴリズムすなわち学習フェーズおよび／または事前の相互作用によって具体的に精緻化されたアルゴリズムを使用するので満足できない。さらに、これらの方法のいくつかは、移動するカメラの制約に適応しておらず、これも不便である。

【先行技術文献】

【特許文献】

【００１５】

【特許文献１】米国特許第８３５９１９１号明細書
【特許文献２】米国特許第５８６７５８４号明細書
【特許文献３】米国特許第８３５１６４９号明細書

【非特許文献】

【００１６】

【非特許文献１】「Ｓｕｉｖｉ Ｔｒｉｄｉｍｅｎｔｉｏｎｎｅｌ ｅｎ Ｓｔｅｒｅｏ ｖｉｓｉｏｎ」（Ｓ．ＣＯＮＳＥＩＬ、Ｓ．ＢＯＵＲＥＮＮＡＮＥ、Ｌ．ＭＡＲＴＩＮ、ＧＲＥＴＳＩ ２００５年）

【非特許文献２】「Ｍａｘｉｍｕｍ Ｗｒｉｇｈｔ Ｃｌｉｑｕｅｓ ｗｉｔｈ Ｍｕｔｅｘ Ｃｏｎｓｔｒａｉｎｔｓ ｆｏｒ Ｏｂｊｅｃｔ Ｓｅｇｍｅｎｔａｔｉｏｎ」（Ｔ．ＭＡ、Ｌ．Ｊ．ＬＡＴＥＣＫＩ、ＣＶＰＲ ２０１２年）
【非特許文献３】「Ｒｅｇｉｏｎｓ Ａｄｊａｃｅｎｃｙ Ｇｒａｐｈ Ａｐｐｌｉｅｄ ｔｏ Ｃｏｌｏｒ Ｉｍａｇｅ Ｓｅｇｍｅｎｔａｔｉｏｎ」（Ａ．ＴＲＥＭＥＡＵ、Ｐ．ＣＯＬＡＮＴＯＮＩ、ＩＥＥＥ、Ｔｒａｎｓａｃｔｉｏｎｓ ｏｎ Ｉｍａｇｅ Ｐｒｏｃｅｓｓｉｎｇ、２０００年）
【発明の概要】
【発明が解決しようとする課題】
【００１７】
本発明は、動画コンテンツに現れる有意な物体の抽出、前記物体間の相対的な相互作用の決定および集約、ならびに前記物体の表示を含む改良されたマルチメディアタグクラウドの生成を可能にする解決手段であって、前記有意な物体が、前記物体の事前知識なく、前記動画コンテンツのさまざまな記録条件を考慮することによって自動的に検出される、解決手段を提案することにより先行技術を改良することを目的とする。
【課題を解決するための手段】
【００１８】
上記目的のため、第１の態様によれば、本発明は、動画コンテンツに現れる物体を表す動画タグクラウドを生成する方法であって：
－　前記動画コンテンツの動画フレームを抽出し、前記動画フレームを領域へと個別にセグメント化するステップと、
－　各抽出フレームに関して、前記フレームのセグメント化された領域間の空間的関係をモデル化するトポロジグラフを構築するステップと、
－　構築したトポロジグラフ集合から、空間的および時間的制約に従って、頻出パターンを抽出するステップであり、各パターンが少なくとも１つのセグメント化された領域を含む、抽出するステップと、
－　軌跡制約を用いて同じ物体の部分を表す頻出パターンを再グループ化し、前記動画コンテンツの頻出物体を検出するステップと、
－　各検出された頻出物体に関して、少なくとも前記物体のパターンの抽出に用いられる空間的および時間的制約ならびに前記パターンの再グループ化に用いられる軌跡制約に従って、前記物体に適用する重み付け係数を決定するステップと、
－　前記頻出物体のそれぞれの視覚表示を含む動画タグクラウドをそれらの重み付け係数に従って生成するステップと、
を提供する、方法に関する。
【００１９】
第２の態様によれば、本発明は、上記のような方法を実行するように構成されたコンピュータプログラムに関する。
【００２０】
第３の態様によれば、本発明は、動画コンテンツに現れる物体を表す動画タグクラウドを生成する上記のようなコンピュータプログラムを実行するように構成されたアプリケーション装置であって：
－　前記生成を管理するエンジンモジュールと、
－　前記動画コンテンツの動画フレームを抽出する手段および前記動画フレームを領域へと個別にセグメント化する手段を備えた抽出モジュールと、
－　各抽出フレームに関して、前記フレームのセグメント化された領域間の空間的関係をモデル化するトポロジグラフを構築する手段を備えたグラフモジュールと、
－　構築したトポロジグラフ集合から、空間的および時間的制約に従って、頻出パターンを抽出する手段であり、各パターンが少なくとも１つのセグメント化された領域を含む、抽出する手段を備えたデータマイニングモジュールと、
－　軌跡制約を用いて同じ物体の部分を表す頻出パターンを再グループ化し、前記動画

コンテンツの頻出物体を検出する手段を備えたクラスタリングモジュールと、
－　各検出された頻出物体に関して、少なくとも前記物体のパターンの抽出に用いられる空間的および時間的制約ならびに前記パターンの再グループ化に用いられる軌跡制約に従って、前記物体に適用する重み付け係数を決定する手段を備えた重み付けモジュールと、
－　前記頻出物体のそれぞれの視覚表示を含む動画タグクラウドをそれらの重み付け係数に従って生成する手段を備えた表示モジュールと、
を備えた、アプリケーション装置に関する。

【００２１】

本発明の他の態様および利点については、添付の図面を参照する以下の説明において明らかとなるであろう。

【図面の簡単な説明】

【００２２】

【図１】本発明によるプロセスのステップを模式的に表した図である。

【図２】外部プラットフォームと相互作用して動画タグクラウドを生成する本発明によるアプリケーション装置を模式的に表した図である。

【発明を実施するための形態】

【００２３】

以下、上記図面に関連して、動画コンテンツに現れる物体を表す動画タグクラウドを生成するプロセス、たとえばこのようなプロセスを実行する手段を備えたアプリケーション装置１等を説明する。

【００２４】

特に、このプロセスは、適応コンピュータプログラムによって実行可能であり、アプリケーション装置１としては、前記コンピュータプログラムまたは前記プログラムを含むコンピュータ可読記憶媒体も可能である。

【００２５】

アプリケーション装置１は、このような生成を管理する中央エンジンモジュール２を備える。

【００２６】

図１および図２に関連して、このプロセスは、ユーザおよび／またはインターフェースによる動画コンテンツの提供によって、前記動画コンテンツから動画タグクラウドを生成する事前ステップＡを含む。動画コンテンツは、特に、Ｙｏｕｔｕｂｅ、Ｄａｉｌｙｍｏｔｉｏｎ、もしくはＡｌｃａｔｅｌ－Ｌｕｃｅｎｔ社のＯｐｅｎｔｏｕｃｈ　Ｖｉｄｅｏ　Ｓｔｏｒｅプラットフォーム等の動画プラットフォーム３または前記アプリケーションのユーザのローカル端末上のハードドライブもしくは端末を介して前記ユーザが接続されたローカルネットワーク等のローカルリポジトリ４から提供可能である。

【００２７】

動画コンテンツは、その他任意の種類のアプリケーションのウェブサービスプラットフォーム５からも提供可能である。たとえば、アプリケーション装置１は、ゲートウェイを介してＩＭＳネットワークのコアに接続されたＷｅｂＲＴＣ（ウェブリアルタイム通信）クライアント、ＩＭＳモバイルクラウド、および異種カメラに対して画像クラウドを生成するために、ＩＭＳ（インターネットプロトコルマルチメディアサブシステム）製品とインターフェース接続可能である。

【００２８】

このようなインターフェースとの相互作用のため、アプリケーション装置１は、ユーザおよび／またはインターフェースが前記アプリケーション装置を用いて動画コンテンツから動画タグクラウドを生成可能な少なくとも１つのアプリケーションプログラミングインターフェース（ＡＰＩ）を備える。図２に関連して、アプリケーション装置１は、動画プラットフォーム３が前記アプリケーションおよび／またはその動画解析機能を使用することを可能とする第１のＡＰＩ６と、ローカルリポジトリ４からユーザにより直接アップロ

ードされた動画コンテンツを用いて、前記ユーザが前記アプリケーション装置を直接使用することを可能とする第２のＡＰＩ７と、その他任意の種類のアプリケーションを用いて、他のウェブプラットフォーム５が前記アプリケーションとのインターフェース接続または前記アプリケーションを使用することを可能とする第３のＡＰＩ８とを備える。

【００２９】

このプロセスは、提供された動画コンテンツの動画フレームを抽出し、前記動画フレームを領域へと個別にセグメント化するステップＢをさらに提供する。このため、アプリケーション装置１は、エンジンモジュール２が相互作用する抽出モジュール９であって、動画フレームを抽出する手段および領域へと個別にセグメント化する手段を備えた、抽出モジュール９を備える。

【００３０】

特に、抽出モジュール９は、このような抽出およびセグメンテーションのための専用アルゴリズムを実装する手段を備え得る。このアルゴリズムとしては、特に、Ｐ．Ｆ．ＦＥＬＺＥＮＳＺＷＡＬＢおよびＤ．Ｐ．ＨＵＴＴＥＮＬＯＣＨＥＲの両氏によって開発された色セグメンテーションアルゴリズムまたはその他任意の種類の既知のセグメンテーションアルゴリズムをわずかに改良したものが可能である。

【００３１】

動画フレームの抽出およびセグメント化の後、このプロセスは、各抽出フレームに関して、前記フレームのセグメント化された領域間の空間的関係をモデル化するトポロジグラフを構築する後続ステップＣを提供する。このため、アプリケーション装置１は、エンジンモジュール２により提供された各抽出フレームに関して、上記のようなトポロジグラフを構築する手段を備えたグラフモジュール１０を備える。

【００３２】

特に、トポロジグラフとしては、セグメント化された領域がノードで表されるとともに隣接領域の対がエッジで表され、フレームの下層ゾーンの色を表すラベルが各ノードに割り当てられた領域隣接グラフ（ＲＡＧ）が可能である。このようなトポロジグラフは、論文「Ｒｅｇｉｏｎｓ Ａｄｊａｃｅｎｃｙ Ｇｒａｐｈ Ａｐｐｌｉｅｄ ｔｏ Ｃｏｌｏｒ Ｉｍａｇｅ Ｓｅｇｍｅｎｔａｔｉｏｎ」（Ａ．ＴＲＥＭＥＡＵ、Ｐ．ＣＯＬＡＮＴＯＮＩ、ＩＥＥＥ、Ｔｒａｎｓａｃｔｉｏｎｓ ｏｎ Ｉｍａｇｅ Ｐｒｏｃｅｓｓｉｎｇ、２０００年）において、さらに詳しく提示されている。

【００３３】

このプロセスは、構築したトポロジグラフ集合から、空間的および時間的制約に従って、頻出パターンを抽出するステップＤであり、各パターンが少なくとも１つのセグメント化された領域を含む、抽出するステップＤをさらに提供する。このため、アプリケーション装置１は、グラフモジュール１０により提供されたパターン集合からのこのような抽出を行う手段を備えたデータマイニングモジュール１１であって、エンジンモジュール２がモジュール１０、１１と相互作用すると抽出を行う上記手段を備えた、データマイニングモジュール１１を備える。データマイニングモジュール１１は、特に、ナレッジディスカバリインデータベース（ＫＤＤ）モデルに従って、頻出パターンを抽出するように構成可能である。

【００３４】

データマイニングモジュール１１の動作は、動画コンテンツの最も関心の高い物体が前記動画コンテンツすなわち、特に、前記コンテンツの多数の動画フレーム中に頻繁に現れるはずであるという事実に依拠している。特に、データマイニングモジュール１１は、たとえば頻出パターンの抽出用に構成された平面グラフマイニングアルゴリズムを実装することにより、動画フレームにおける前記パターンの時間的および空間的発生に従って頻出パターンを抽出する手段を備える。実際、パターンの空間的および時間的発生を考慮に入れることは、前記パターンの頻度のみを考慮に入れることよりも正確であり、後者では、前記パターンを含むグラフの数のみが関係しており、前記パターンが同じグラフに２回以上現れる場合を考慮していない。

【００３５】
さらに、データマイニングモジュール１１は、空間的および時間的発生に基づいて、空間的および時間的に離れ過ぎたパターンの発生がより近い発生よりも同じ物体を表す可能性が低いことを考慮し、前記パターンのその他任意の発生から、空間的および時間的に離れた発生を廃棄することができる。
【００３６】
特に、データマイニングモジュール１１は、動画フレームにおける前記パターンの２回の発生間の平均時間距離に従って、パターンの時間的発生を評価する手段を備える。
【００３７】
同様に、データマイニングモジュール１１は、同じ動画フレームにおける前記パターンの２回の発生間の平均空間距離に従って、パターンの空間的発生を評価する手段を備える。平均空間距離は、特に、以下の式：
$$\max_{s \in V} d(o_1(s), o_2(s))$$
に従って計算可能であり、ここで、Ｖは前記パターンの領域集合、o_1、o_2は同じ動画フレームにおける前記パターンの２回の発生、$d(o_1(s), o_2(s))$は前記パターンの領域ｓの発生間のユークリッド距離である。
【００３８】
また、データマイニングモジュール１１は、評価したパターンの空間的および時間的発生から発生グラフを構築する手段であって、パターンの各発生がノードによって表されるとともに、同じパターンのノードが、空間的および時間的に十分近い場合はエッジによって接続される、手段を備え得る。したがって、このような発生グラフにおいては、パターンが接続ノードのチェーンによって表され、前記パターンが、頻出パターンと考えられるとともに、前記パターンが少なくとも１回発生する異なるフレームの数に対応した前記チェーンの長さが頻度閾値よりも大きい場合、上記のように抽出される。
【００３９】
このプロセスは、軌跡制約を用いて同じ物体の部分を表す頻出パターンを再グループ化し、動画コンテンツの頻出物体を検出するステップＥをさらに提供する。このため、アプリケーション装置１は、上記のような頻出パターンを再グループ化し、前記頻出物体のより完全な追跡を実現する手段を備えたクラスタリングモジュール１２を備える。
【００４０】
特に、クラスタリングモジュール１２の再グループ化手段は、動画フレームにおける前記パターンの軌跡間の相違度合いに従って、同じ物体の部分を表す頻出パターンを再グループ化するように構成可能である。この相違度合いは、特に、以下の式：
【数１】
$$\frac{\sum_{t=1}^{n} x_t}{n}$$
に従って計算可能であり、ここで、x_tは動画フレームｔにおける２つのパターンの図心（ｃｅｎｔｒｏｉｄ）間のユークリッド距離であり、パターンの図心は、動画フレームｔにおける前記パターンのすべての空間的発生の重心に対応する。
【００４１】
特に、データマイニングモジュール１１により提供された発生グラフにおいては、頻出パターンの発生を接続可能であるため、これらの間にいくつかのフレームが存在していても、前記パターンが広がる各フレームにおいては、前記パターンが必ずしも上記のように発生しない。したがって、クラスタリングモジュール１２は、２つのパターンの図心間の距離について、前記パターンがともに広がる各フレームで演算可能となるように、欠落した図心を補間するように構成されている。
【００４２】
各頻出パターン対間の相違度合いの演算の後、クラスタリングモジュール１２の再グループ化手段は、階層的凝集クラスタリングアルゴリズムを用いて、頻出パターン間の階層

を生成するようにしてもよく、また、前記階層を解析して、より高頻出の物体を表す頻出パターンのクラスタを求めることにより、前記頻出物体を検出するとともに、動画コンテンツの他の物体との相互作用を集約するようにしてもよい。

【００４３】

このプロセスは、各検出された頻出物体に関して、少なくとも前記物体のパターンの抽出に用いられる空間的および時間的制約ならびに前記パターンの再グループ化に用いられる軌跡制約に従って、前記物体に適用する重み付け係数を決定するステップＦをさらに提供する。このため、アプリケーション装置１は、エンジンモジュール２と相互作用すると、各検出された頻出物体に関するこのような重み付け係数を決定する手段を備えた重み付けモジュール１３を備える。

【００４４】

特に、重み付けモジュール１３のこの手段は、クラスタリングモジュール１２により提供された相違度合いおよび階層解析等、データマイニングモジュール１１により評価された時間的および空間的発生によって、重み付け係数を処理するように構成されている。一般的に、重み付けモジュール１３のこの手段は、頻度、サイズ、時間的および空間的発生、構成パターンおよび／もしくは発生間のユークリッド距離、動画コンテンツ中の存在時間、前記動画コンテンツにおける他の物体、特に他の頻出物体との関係、色、またはその他任意の関連した入力に従って、物体の重み付け係数を決定するようにしてもよい。

【００４５】

さらに、アプリケーション装置１は、たとえば前記アプリケーション装置のグラフィカルユーザインターフェース（ＧＵＩ）上の専用機能によって、重み付け係数を決定する特定の規則を確立または変更する手段をユーザに提供可能である。

【００４６】

また、このプロセスは、検出された頻出物体を抽出してセグメント化するステップＧを含んでいてもよい。このため、アプリケーション装置１は、エンジンモジュール２と相互作用すると、データマイニングモジュール１１およびクラスタリングモジュール１２の入力により、上記のような検出物体のセグメンテーションおよび抽出をそれぞれ行う手段を備えたセグメンテーション抽出モジュール１４を備える。

【００４７】

特に、セグメンテーション抽出モジュール１４は、特に、グラフカットアルゴリズム、グラブカットアルゴリズム、および／または画像／スペクトル艶消しアルゴリズム等の既知のセグメンテーションアルゴリズムによって、物体およびそれぞれの位置を識別する手段ならびに前記物体を抽出する手段を備える。

【００４８】

その後、セグメント化および抽出した頻出物体は、それぞれの対応する重み付け係数と併せて、データリポジトリ１５に格納されるようになっていてもよい。このため、アプリケーション装置１は、このようなデータリポジトリ１５を備え、エンジンモジュール２と相互作用すると、モジュール１４からのセグメント化および抽出物体が、重み付けモジュール１３からのそれぞれ対応する重み付け係数と併せて格納される。

【００４９】

このプロセスは、前記頻出物体のそれぞれの視覚表示を含む動画タグクラウドをそれらの重み付け係数に従って生成するステップＨをさらに提供する。このため、アプリケーション装置１は、このような視覚表示を含む動画タグクラウドを生成する手段を備えた表示モジュール１６を備える。特に、表示モジュール１６は、エンジンモジュール２と相互作用すると、データリポジトリ１５に格納された物体およびそれぞれの対応する重み付け係数から、動画タグクラウドを生成する。

【００５０】

特に、物体の視覚表示のサイズ、位置、および動きは、その対応する重み付け係数に応じて変更可能であり、前記係数が、動画コンテンツ中の前記物体の重要性によって決まり、前記重要性が、たとえば前記物体の頻度および／または前記コンテンツの前記物体と他

の物体との間の関係によって推定される。

【００５１】

たとえば、アナウンサがカメラの前で話したり動いたりしている動画コンテンツを起点として、アプリケーション装置１は、放送チャンネルおよび番組の各ロゴ等、前記動画コンテンツの最も重要な物体として前記アナウンサの顔および両手が識別され、すべてが大きなサイズの視覚表示によって表された動画タグクラウドを生成するようにしてもよい。これに対して、アナウンサの胴体およびネクタイは、重要ではあるものの、補助的な物体として識別されていてもよく、より小さな視覚表示によって表される。

【００５２】

以上の説明および図面は、本発明の原理を示したに過ぎない。したがって、当然のことながら、当業者であれば、本明細書においては明示的に記述または図示していないものの、本発明の原理を具現化するとともに、その主旨および範囲に含まれる種々構成を考案可能であろう。さらに、本明細書に挙げたすべての例は主として、読者が本発明の原理を理解することを支援する教示上の目的および（１人または複数の）発明者による概念によって当技術分野が進展することのみを特に意図しており、具体的に列挙したこのような例および条件に限定されないものとして解釈されるものとする。さらに、本発明の原理、態様、および実施形態の他、その具体例を挙げた本明細書のすべての記述は、本発明の均等物を網羅することを意図している。

【図１】

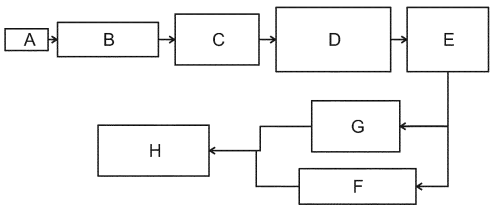


Fig. 1

【図２】

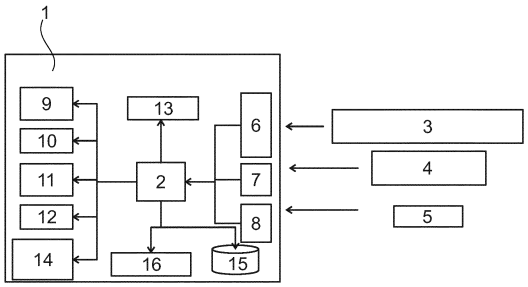


Fig. 2

フロントページの続き

(72)発明者 ディオ，ファビアン
フランス国、９１６２０・ノゼイ、ルット・ドゥ・ビルジュスト、サントル・ドゥ・ビラルソー、アルカテル・ルーセント・ベル・ラブズ・フランス気付
(72)発明者 オウトタガーツ，アブデルカーデル
フランス国、９１６２０・ノゼイ、ルット・ドゥ・ビルジュスト、サントル・ドゥ・ビラルソー、アルカテル・ルーセント・ベル・ラブズ・フランス気付
(72)発明者 オブレド，コリンヌ
フランス国、９１６２０・ノゼイ、ルット・ドゥ・ビルジュスト、サントル・ドゥ・ビラルソー、アルカテル・ルーセント・ベル・ラブズ・フランス気付
(72)発明者 セクダン，シルヴァン
フランス国、９１０７０・ボンドゥフル、スクワール・ジョルジュ・ギヌメ・３４

審査官 齊藤 貴孝

(56)参考文献 米国特許出願公開第２００７／０１６２８７３（ＵＳ，Ａ１）
特開２００９－２０１０４１（ＪＰ，Ａ）
徳平征恒 外１名，映像の時空間分割とネットワーク表現による動物体抽出システムの検討，電子情報通信学会技術研究報告 ＭＶＥ２０００－５３～６４ マルチメディア・仮想環境基礎，社団法人電子情報通信学会，２０００年 7月 7日，第100巻 第184号，第61～66頁
辻 智和、外３名，頻出グラフマイニングを利用した動画像解析，電子情報通信学会論文誌，日本，社団法人電子情報通信学会，２０１０年 2月 1日，第Ｊ９３－Ｄ巻，第２号，ｐ．８６－９９
ANJULAN A; CANAGARAJAH N, A UNIFIED FRAMEWORK FOR OBJECT RETRIEVAL AND MINING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 米国, IEEE SERVICE CENTER, ２００９年 1月 1日, VOL:19, NR:1, PAGE(S):63 - 76, ＵＲＬ, http://dx.doi.org/10.1109/TCSVT.2008.2005801

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １７／３０
Ｇ０６Ｔ ７／２０
Ｈ０４Ｎ ５／７６