

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7123122号

(P7123122)

(45)発行日 令和4年8月22日(2022.8.22)

(24)登録日 令和4年8月12日(2022.8.12)

(51)国際特許分類

F I

H 0 4 N 21/232 (2011.01)

H 0 4 N 21/232

H 0 4 N 21/278 (2011.01)

H 0 4 N 21/278

請求項の数 13 (全23頁)

(21)出願番号	特願2020-503818(P2020-503818)	(73)特許権者	521555742
(86)(22)出願日	平成30年7月5日(2018.7.5)		キンドリル・インク
(65)公表番号	特表2020-528705(P2020-528705 A)		Kyndryl Inc.
(43)公表日	令和2年9月24日(2020.9.24)		アメリカ合衆国ニューヨーク州、ニュー
(86)国際出願番号	PCT/IB2018/054963		ヨーク、ヴァンダービルト・アベニュー
(87)国際公開番号	WO2019/021088		1
(87)国際公開日	平成31年1月31日(2019.1.31)		One Vanderbilt Avenue, 15th Floor, New Y
審査請求日	令和2年12月22日(2020.12.22)		ork, New York 10017, USA
(31)優先権主張番号	15/657,626	(74)代理人	110000420弁理士法人M I P
(32)優先日	平成29年7月24日(2017.7.24)	(72)発明者	ハミルトン セカンド、リック
(33)優先権主張国・地域又は機関	米国(US)		アメリカ合衆国13760 ニューヨー
			ク州エンディコット ノース・ストリー
			ト1701
			最終頁に続く

(54)【発明の名称】 認知的洞察を使用したビデオ・シーンの移動

(57)【特許請求の範囲】

【請求項1】

データベースからシーンを取得するためのコンピュータ実装方法であって、前記方法が、ユーザから、自然言語処理を使用するサーバにより実行された概念タギング、感情解析、情緒解析、および関係抽出による注釈が付されたビデオ・コンテンツを含んだデータベースに格納されたシーンに関する検索要求を受信することと、

1つまたは複数のキーワードを前記検索要求から抽出することと、

前記1つまたは複数のキーワードの各々に関して、1つまたは複数の拡張キーワードを生成することと、

前記1つまたは複数の拡張キーワードおよび前記1つまたは複数のキーワードを、前記注釈付きのビデオ・コンテンツと比較して、それぞれ対象のシーンの注釈を含んだ1つまたは複数の対象のシーンを識別することと、

前記1つまたは複数の対象のシーンの各々の前記対象のシーンの注釈と、前記1つまたは複数の拡張キーワードおよび1つまたは複数のキーワードとの間の相関関係に少なくとも部分的に基づいて、信頼性評価を前記1つまたは複数の対象のシーンの各々に割り当てることと、

前記1つまたは複数の対象のシーンのうちの前記信頼性評価に基づいて決定される、少なくとも1つを前記ユーザに表示することとを含んでいる、コンピュータ実装方法。

【請求項2】

前記ユーザから、前記1つまたは複数の対象のシーンのうちの1つの対象のシーンの選

10

20

択の指示を受信することと、

前記選択に少なくとも部分的に基づいて前記対象のシーンの注釈を更新することとをさらに含んでいる、請求項 1 に記載のコンピュータ実装方法。

【請求項 3】

前記 1 つまたは複数の対象のシーンを前記ユーザに表示することと、

前記ユーザから、前記 1 つまたは複数の対象のシーンのうちの 1 つの対象のシーンの拒否の指示を受信することと、

前記拒否に少なくとも部分的に基づいて前記対象のシーンの注釈を更新することとをさらに含んでいる、請求項 1 に記載のコンピュータ実装方法。

【請求項 4】

前記ユーザを監視して、前記ユーザが前記対象のシーンを見ているときに、前記ユーザの感情的反応を決定することと、

前記ユーザの前記感情的反応を前記対象のシーンの注釈と比較して、前記感情的反応と前記対象のシーンの注釈の間の相関関係を決定することと、

前記感情的反応と前記対象のシーンの注釈の間の前記相関関係に少なくとも部分的に基づいて、前記対象のシーンの注釈を更新することとをさらに含んでいる、請求項 2 に記載のコンピュータ実装方法。

【請求項 5】

前記検索要求がユーザ音声入力であり、1 つまたは複数のキーワードを前記抽出することとが、

前記検索要求に対して自然言語処理を実行して、前記ユーザ音声入力をテキストに変換することと、

前記テキストを 1 つまたは複数の別々の語句に分割することと、

前記 1 つまたは複数の別々の語句を解析して、キーワードを決定することとを含んでいる、請求項 1 に記載のコンピュータ実装方法。

【請求項 6】

前記 1 つまたは複数の対象のシーンを前記ユーザに表示することと、

前記信頼性評価に従って前記 1 つまたは複数の対象のシーンの前記表示を順序付けることとをさらに含んでいる、請求項 1 に記載のコンピュータ実装方法。

【請求項 7】

ビデオ・シーンに注釈を付けるためのコンピュータ実装方法であって、前記方法が、プロセッサによって 1 つまたは複数のビデオを受信することと、前記 1 つまたは複数のビデオの各々をシーンのセットに分割することと、前記シーンのセット内の第 1 のシーンを、自然言語処理を使用するサーバにより実行された概念タギング、感情解析、情緒解析、および関係抽出により解析して、前記第 1 のシーンのラベルのセットを決定することと、

前記第 1 のシーンに、前記ラベルのセットで注釈を付けることと、

前記第 1 のシーンをデータベースに格納することとを含んでいる、コンピュータ実装方法。

【請求項 8】

前記第 1 のシーンの前記ラベルのセットに少なくとも部分的に基づいて拡張キーワードを決定することと、

前記拡張キーワードで、前記第 1 のシーンに注釈を付けることとをさらに含んでいる、請求項 7 に記載のコンピュータ実装方法。

【請求項 9】

前記第 1 のシーンの前記ラベルのセットを決定することが、

前記第 1 のシーンに対して画像認識を実行して、前記第 1 のシーン内の 1 つまたは複数の物体のラベルを生成することとを含んでいる、請求項 7 に記載のコンピュータ実装方法。

【請求項 10】

前記第 1 のシーンの前記ラベルのセットを決定することが、

10

20

30

40

50

前記第 1 のシーンに対して音声の解析を実行して、1 つまたは複数のキーワードを抽出することと、

前記 1 つまたは複数のキーワードに少なくとも部分的に基づいて、前記第 1 のシーンのラベルを生成することと、

前記第 1 のシーンに、前記 1 つまたは複数のキーワードで注釈を付けることとをさらに含んでいる、請求項 7 に記載のコンピュータ実装方法。

【請求項 1 1】

前記第 1 のシーンの前記 1 つまたは複数のキーワードに少なくとも部分的に基づいて拡張キーワードを決定することと、

前記拡張キーワードに、前記第 1 のシーンで注釈を付けることとをさらに含んでいる、請求項 1 0 に記載のコンピュータ実装方法。

【請求項 1 2】

データベースからシーンを取得するためのコンピュータ・システムであって、前記コンピュータ・システムが、メモリに通信によって接続されたプロセッサを含んでおり、前記プロセッサが、

ユーザから、自然言語処理を使用するサーバにより実行された概念タギング、感情解析、情緒解析、および関係抽出による注釈が付されたビデオ・コンテンツを含んだデータベースに格納されたシーンに関する検索要求を受信することと、

1 つまたは複数のキーワードを前記検索要求から抽出することと、

前記 1 つまたは複数のキーワードの各々に関して、1 つまたは複数の拡張キーワードを生成することと、

前記 1 つまたは複数の拡張キーワードおよび前記 1 つまたは複数のキーワードを、前記注釈付きのビデオ・コンテンツと比較して、それぞれ対象のシーンの注釈を含んだ 1 つまたは複数の対象のシーンを識別することと、

前記 1 つまたは複数の対象のシーンの各々の前記対象のシーンの注釈と、前記 1 つまたは複数の拡張キーワードおよび 1 つまたは複数のキーワードとの間の相関関係に少なくとも部分的に基づいて、信頼性評価を前記 1 つまたは複数の対象のシーンの各々に割り当てることとを実行するように構成されている、コンピュータ・システム。

【請求項 1 3】

ビデオ・シーンに注釈を付けるためのコンピュータ・システムであって、前記コンピュータ・システムが、メモリに通信によって接続されたプロセッサを含んでおり、前記プロセッサが、

1 つまたは複数のビデオを受信することと、

前記 1 つまたは複数のビデオの各々をシーンのセットに分割することと、

前記シーンのセット内の第 1 のシーンを、自然言語処理を使用するサーバにより実行された概念タギング、感情解析、情緒解析、および関係抽出により解析して、前記第 1 のシーンのラベルのセットを決定することと、

前記第 1 のシーンに、前記ラベルのセットで注釈を付けることと、

前記第 1 のシーンをデータベースに格納することとを実行するように構成されている、コンピュータ・システム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本開示は、ビデオ・シーンを移動することに関連しており、より詳細には、認知的洞察を使用してビデオ・シーンを移動するための方法およびシステムに関連している。

【背景技術】

【0 0 0 2】

コンピュータの性能が向上し続けるのに伴って、デジタル・ビデオの利用が、ますます一般的になってきている。インターネットの帯域幅の 5 0 % 以上がビデオ・サービスによって消費されているということを示す研究によって証明されているように、ビデオを見る

10

20

30

40

50

ことは、私たちの日常生活で最も人気のある娯楽活動の1つになっている。容易に利用できるビデオ・カメラ、スマートフォン、およびその他のハンドヘルド・デバイスを使って、人は、ますます多くの旅行、重要なお祝い、および家族との時間を記録している。さらに、デジタル・ビデオ技術が、ほとんどの監視システムにおいて使用されている。

【0003】

しかし、それらの大量のビデオ・データを使用および管理することに伴う問題が発生している。人は、ビデオ全体ではなく、映画またはホーム・ビデオの特定のシーンを見たいと思うことがある。同様に、警官または保安要員は、膨大な量のビデオ・データから証拠を見つけ、集めたいと思うことがある。

【0004】

多くの場合、1つまたは複数のビデオ・ファイル内の特定のシーンを見つけ、そこに移動することは、時間がかかり、困難である。既存の技術を使用すると、ユーザは、目的のシーンが見つかるまで、早送りモードを使用するか、またはスライダー・バーを何度もドラッグして、ビデオを見なければならない。さらに、一部のビデオは章ごとに分けられており、ユーザは、シーンがあるかもしれない特定の章に移動することができる。これらの技術では、シーンを見つけるために、ビデオに関するユーザの記憶に基づいて、何度も移動する必要がある。また、ユーザは、前にビデオを見たことがない場合があり、別の視聴者からの推奨に基づいてシーンを検索していることがある。

【0005】

この問題の解決策がいくつか存在するが、それらは、必要性に直接対処しない。例えば、次回、映画が中止された位置から自動的に再開されるように、多くの最新のビデオ・プレーヤーは、映画を再生するために、最後の中断位置を「記憶」する。しかし、この解決策は、過去の記録に基づいており、単一のユーザまたはクライアントのために設計されているため、特定のユーザに関する以前の記録が存在しない場合、または複数のユーザが存在する状況において、適切に動作しない。

【0006】

他のシステムは、顔認識を使用してビデオ・ストリーム内の特定の人を見つけるが、この機能は、概要を使用して特定のシーンを見つけるという問題に対処せず、その個人が複数のシーン内に存在する場合に特定のシーンを見つけることに対処しない。そのような状況においてこの技術を使用することは、ユーザに提供された多数の結果をさらに手動でフィルタリングする必要があるため、問題になる。したがって、単一のビデオまたは大量のビデオ・データのいずれかから必要なシーンを素早く正確に発見できる新しい方法が、必要とされている。

【発明の概要】

【発明が解決しようとする課題】

【0007】

本発明は、認知的洞察を使用してビデオ・シーンを移動するための方法およびシステムを提供することを目的とする。

【課題を解決するための手段】

【0008】

実施形態は、データベースからシーンを取得するためのコンピュータ実装方法を含む。コンピュータ実装方法の非限定的な例は、注釈付きのビデオ・コンテンツを含んでいるデータベースに格納されたシーンに関する検索要求を受信することを含む。1つまたは複数のキーワードが、検索要求から抽出される。キーワードごとに、拡張キーワードが生成される。拡張キーワードおよびキーワードが、注釈付きのビデオ・コンテンツと比較されて、対象のシーンの注釈を含んでいる対象のシーンを識別する。対象のシーンの各々の対象のシーンの注釈と、拡張キーワードおよびキーワードとの間の相関関係に少なくとも部分的に基づいて、信頼性評価が対象のシーンの各々に割り当てられる。対象のシーンのうちの少なくとも1つがユーザに表示され、対象のシーンのうちの少なくとも1つは、信頼性評価に基づいて決定される。

10

20

30

40

50

【 0 0 0 9 】

実施形態は、データベースからシーンを取得するためのコンピュータ・システムを含み、このコンピュータ・システムはプロセッサを含み、このプロセッサは、方法を実行するように構成される。システムの非限定的な例は、注釈付きのビデオ・コンテンツを含んでいるデータベースに格納されたシーンに関する検索要求を受信することを含む。1つまたは複数のキーワードが、検索要求から抽出される。キーワードごとに、拡張キーワードが生成される。拡張キーワードおよびキーワードが、注釈付きのビデオ・コンテンツと比較されて、対象のシーンの注釈を含んでいる対象のシーンを識別する。対象のシーンの各々の対象のシーンの注釈と、拡張キーワードおよびキーワードとの間の相関関係に少なくとも部分的に基づいて、信頼性評価が対象のシーンの各々に割り当てられる。対象のシーンのうちの少なくとも1つがユーザに表示され、対象のシーンのうちの少なくとも1つは、信頼性評価に基づいて決定される。

10

【 0 0 1 0 】

実施形態は、データベースからシーンを取得するためのコンピュータ・プログラム製品も含み、このコンピュータ・プログラム製品は、コンピュータ可読プログラム・コードが具現化されている非一過性コンピュータ可読記憶媒体を含む。コンピュータ可読プログラム・コードは、方法を実行するように構成されたコンピュータ可読プログラム・コードを含む。方法の非限定的な例は、注釈付きのビデオ・コンテンツを含んでいるデータベースに格納されたシーンに関する検索要求を受信することを含む。1つまたは複数のキーワードが、検索要求から抽出される。キーワードごとに、拡張キーワードが生成される。拡張キーワードおよびキーワードが、注釈付きのビデオ・コンテンツと比較されて、対象のシーンの注釈を含んでいる対象のシーンを識別する。対象のシーンの各々の対象のシーンの注釈と、拡張キーワードおよびキーワードとの間の相関関係に少なくとも部分的に基づいて、信頼性評価が対象のシーンの各々に割り当てられる。対象のシーンのうちの少なくとも1つがユーザに表示され、対象のシーンのうちの少なくとも1つは、信頼性評価に基づいて決定される。

20

【 0 0 1 1 】

実施形態は、ビデオ・シーンに注釈を付けるためのコンピュータ実装方法を含む。コンピュータ実装方法の非限定的な例は、プロセッサによって、1つまたは複数のビデオを受信することを含む。1つまたは複数のビデオの各々は、シーンのセットに分割される。シーンのセット内の第1のシーンが解析されて、第1のシーンのラベルのセットを決定する。第1のシーンに、ラベルのセットで注釈が付けられ、それらがデータベースに格納される。

30

【 0 0 1 2 】

実施形態は、ビデオ・シーンに注釈を付けるためのコンピュータ・システムを含み、このコンピュータ・システムはプロセッサを含み、このプロセッサは、方法を実行するように構成される。システムの非限定的な例は、プロセッサによって、1つまたは複数のビデオを受信することを含む。1つまたは複数のビデオの各々は、シーンのセットに分割される。シーンのセット内の第1のシーンが解析されて、第1のシーンのラベルのセットを決定する。第1のシーンに、ラベルのセットで注釈が付けられ、それらがデータベースに格納される。

40

【 0 0 1 3 】

その他の特徴および長所が、本発明の手法によって実現される。本発明のその他の実施形態および態様は、本明細書において詳細に説明され、請求される発明の一部と見なされる。さらに良く理解するために、説明および図面を参照すること。

【 0 0 1 4 】

本明細書に記載された専有権の詳細は、本明細書の最後にある特許請求の範囲において具体的に指摘され、明確に請求される。本発明の各実施形態の前述およびその他の特徴と長所は、添付の図面と併せて行われる以下の詳細な説明から明らかになる。

【図面の簡単な説明】

50

【 0 0 1 5 】

【図 1】本発明の 1 つまたは複数の実施形態に記載されたクラウド・コンピューティング環境を示す図である。

【図 2】本発明の 1 つまたは複数の実施形態に記載された抽象モデル・レイヤを示す図である。

【図 3】本明細書の内容を実践することにおいて使用するためのコンピュータ・システムのブロック図である。

【図 4】1 つまたは複数の実施形態に記載されている、デジタル・ビデオを移動するためのシステムのブロック図である。

【図 5】1 つまたは複数の実施形態に記載されている、デジタル・ビデオを移動するための例示的なシステムの図である。

10

【図 6】1 つまたは複数の実施形態に記載されている、データベースからシーンを取得するための方法のフロー図である。

【図 7】1 つまたは複数の実施形態に記載されている、ビデオ・シーンに注釈を付けるための方法のフロー図である。

【発明を実施するための形態】

【 0 0 1 6 】

本明細書において示される図は、実例である。本発明の思想から逸脱することなく、本明細書に記載された図または動作の多くの変形が存在することが可能である。例えば、動作は異なる順序で実行されることが可能であり、あるいは動作は追加、削除、または変更されることが可能である。また、「結合される」という用語およびその変形は、2 つの要素間に通信経路が存在することを表しており、それらの要素間に要素 / 接続が介在しない要素間の直接的接続を意味していない。これらのすべての変形は、本明細書の一部であると見なされる。

20

【 0 0 1 7 】

添付の図および開示された実施形態に関する以下の詳細な説明では、図に示されたさまざまな要素が、2 桁または 3 桁の参照番号付きで提供されている。わずかな例外を除いて、各参照番号の左端の数字は、その要素が最初に示された図に対応している。

【 0 0 1 8 】

本明細書では、関連する図面を参照して、本発明のさまざまな実施形態が説明される。本発明の範囲を逸脱することなく、本発明の代替の実施形態を考案することができる。以下の説明および図面において、要素間のさまざまな接続および位置関係（例えば、上、下、隣接など）が示される。それらの接続または位置関係あるいはその両方は、特に規定されない限り、直接的または間接的であることができ、本発明はこの点において限定するよう意図されていない。したがって、各実体の結合は、直接的結合または間接的結合を指すことができ、各実体間の位置関係は、直接的位置関係または間接的位置関係であることができる。さらに、本明細書に記載されたさまざまな作業および工程段階は、本明細書に詳細に記載されない追加の段階または機能を含んでいるさらに包括的な手順または工程に組み込まれ得る。

30

【 0 0 1 9 】

以下の定義および略称が、特許請求の範囲および本明細書の解釈に使用される。本明細書において使用されているように、「備える」、「備えている」、「含む」、「含んでいる」、「有する」、「有している」、「含有する」、「含有している」という用語、またはこれらの任意のその他の変形は、非排他的包含をカバーするよう意図されている。例えば、要素のリストを含んでいる組成、混合、工程、方法、製品、または装置は、それらの要素のみに必ずしも限定されず、明示されていないか、またはそのような組成、混合、工程、方法、製品、または装置に固有の、その他の要素を含むことができる。

40

【 0 0 2 0 】

さらに、「例示的」という用語は、本明細書では「例、事例、または実例としての役割を果たす」ことを意味するために使用される。「例示的」として本明細書に記載された実

50

施形態または設計は、必ずしも他の実施形態または設計よりも好ましいか、または有利であると解釈されるべきではない。「少なくとも1つ」および「1つまたは複数」という用語は、1以上の任意の整数（すなわち、1、2、3、4など）を含んでいると理解されてよい。「複数」という用語は、2以上の任意の整数（すなわち、2、3、4、5など）を含んでいると理解されてよい。「接続」という用語は、間接的「接続」および直接的「接続」の両方を含んでよい。

【0021】

「約」、「実質的に」、「近似的に」、およびこれらの変形の実語は、本願書の出願時に使用できる機器に基づいて、特定の量の測定に関連付けられた誤差の程度を含むよう意図されている。例えば、「約」は、特定の値の±8%または5%、あるいは2%の範囲を含むことができる。

10

【0022】

簡潔さの目的で、本発明の態様の作成および使用に関連する従来手法は、本明細書に詳細に記載されることもあれば、記載されないこともある。具体的には、本明細書に記載されたさまざまな技術的特徴を実装するためのコンピューティング・システムおよび特定のコンピュータ・プログラムのさまざまな態様は、よく知られている。したがって、簡略にするために、多くの従来の実装に関する詳細は、本明細書では、既知のシステムまたは工程あるいはその両方の詳細を提供することなく、簡潔にのみ述べられるか、または全体的に省略される。

【0023】

本開示にはクラウド・コンピューティングに関する詳細な説明が含まれているが、本明細書において示された内容の実装は、クラウド・コンピューティング環境に限定されないと理解されるべきである。本発明の実施形態は、現在既知であるか、または今後開発される任意のその他の種類のコンピューティング環境と組み合わせて実装できる。

20

【0024】

クラウド・コンピューティングは、構成可能な計算リソース（例えば、ネットワーク、ネットワーク帯域幅、サーバ、処理、メモリ、ストレージ、アプリケーション、仮想マシン、およびサービス）の共有プールへの便利なオンデマンドのネットワーク・アクセスを可能にするためのサービス提供モデルであり、管理上の手間またはサービス・プロバイダとのやりとりを最小限に抑えて、これらのリソースを迅速にプロビジョニングおよび解放することができる。このクラウド・モデルは、少なくとも5つの特徴、少なくとも3つのサービス・モデル、および少なくとも4つのデプロイメント・モデルを含むことができる。

30

【0025】

特徴は、次のとおりである。

オンデマンドのセルフ・サービス：クラウドの利用者は、サーバの時間、ネットワーク・ストレージなどの計算能力を一方的に、サービス・プロバイダとの人間的なやりとりを必要とせず、必要に応じて自動的にプロビジョニングすることができる。

幅広いネットワーク・アクセス：クラウドの能力は、ネットワークを経由して利用可能であり、標準的なメカニズムを使用してアクセスできるため、異種のシン・クライアントまたはシック・クライアント・プラットフォーム（例えば、携帯電話、ラップトップ、およびPDA）による利用を促進する。

40

リソース・プール：プロバイダの計算リソースは、プールされ、マルチテナント・モデルを使用して複数の利用者に提供される。さまざまな物理的および仮想的リソースが、要求に従って動的に割り当ておよび再割り当てされる。場所に依存しないという感覚があり、利用者は通常、提供されるリソースの正確な場所に関して管理することも知らないが、さらに高い抽象レベルでは、場所（例えば、国、州、またはデータセンター）を指定できる場合がある。

迅速な順応性：クラウドの能力は、迅速かつ柔軟に、場合によっては自動的にプロビジョニングされ、素早くスケールアウトし、迅速に解放されて素早くスケールインすることができる。プロビジョニングに使用できる能力は、利用者には、多くの場合、任意の量を

50

いつでも無制限に購入できるように見える。

測定されるサービス：クラウド・システムは、計測機能を活用することによって、サービスの種類（例えば、ストレージ、処理、帯域幅、およびアクティブなユーザのアカウント）に適した抽象レベルで、リソースの使用を自動的に制御および最適化する。リソースの使用量は監視、制御、および報告することができ、利用されるサービスのプロバイダと利用者の両方に透明性が提供される。

IaaS（Infrastructure as a Service）：利用者に提供される能力は、処理、ストレージ、ネットワーク、およびその他の基本的な計算リソースのプロビジョニングであり、利用者は、オペレーティング・システムおよびアプリケーションを含むことができる任意のソフトウェアをデプロイして実行できる。利用者は、基盤になるクラウド・インフラストラクチャを管理することも制御することもないが、オペレーティング・システム、ストレージ、およびデプロイされたアプリケーションを制御することができ、場合によっては、選択されたネットワーク・コンポーネント（例えば、ホスト・ファイアウォール）を限定的に制御できる。

【0026】

デプロイメント・モデルは、次のとおりである。

プライベート・クラウド：このクラウド・インフラストラクチャは、ある組織のためにのみ運用される。この組織またはサード・パーティによって管理することができ、オンプレミスまたはオフプレミスに存在することができる。

コミュニティ・クラウド：このクラウド・インフラストラクチャは、複数の組織によって共有され、関心事（例えば、任務、セキュリティ要件、ポリシー、およびコンプライアンスに関する考慮事項）を共有している特定のコミュニティをサポートする。これらの組織またはサード・パーティによって管理することができ、オンプレミスまたはオフプレミスに存在することができる。

パブリック・クラウド：このクラウド・インフラストラクチャは、一般ユーザまたは大規模な業界団体が使用できるようになっており、クラウド・サービスを販売する組織によって所有される。

ハイブリッド・クラウド：このクラウド・インフラストラクチャは、データとアプリケーションの移植を可能にする標準化された技術または独自の技術（例えば、クラウド間の負荷バランスを調整するためのクラウド・バースト）によって固有の実体を残したまま互いに結合された2つ以上のクラウド（プライベート、コミュニティ、またはパブリック）の複合である。

【0027】

クラウド・コンピューティング環境は、ステートレス、疎結合、モジュール性、および意味的相互運用性に重点を置いたサービス指向の環境である。クラウド・コンピューティングの中心になるのは、相互接続されたノードのネットワークを含んでいるインフラストラクチャである。

【0028】

ここで図1を参照すると、例示的なクラウド・コンピューティング環境50が示されている。図示されているように、クラウド・コンピューティング環境50は、クラウドの利用者によって使用されるローカル・コンピューティング・デバイス（例えば、PDA（Personal Digital Assistant）または携帯電話54A、デスクトップ・コンピュータ54B、ラップトップ・コンピュータ54C、または自動車コンピュータ・システム54N、あるいはその組み合わせなど）が通信できる1つまたは複数のクラウド・コンピューティング・ノード10を備える。ノード10は、互いに通信してよい。ノード10は、1つまたは複数のネットワーク内で、本明細書において前述されたプライベート・クラウド、コミュニティ・クラウド、パブリック・クラウド、またはハイブリッド・クラウド、あるいはこれらの組み合わせなどに、物理的または仮想的にグループ化されてよい（図示されていない）。これによって、クラウド・コンピューティング環境50は、クラウドの利用者がローカル・コンピューティング・デバイス上でリソースを維持する必要のないインフラ

10

20

30

40

50

トラクチャ、プラットフォーム、またはSaaS、あるいはその組み合わせを提供できる。図1に示されたコンピューティング・デバイス54A～Nの種類は、例示のみが意図されており、コンピューティング・ノード10およびクラウド・コンピューティング環境50は、任意の種類のネットワークまたはネットワーク・アドレス可能な接続（例えば、Webブラウザを使用した接続）あるいはその両方を經由して任意の種類のコンピュータ制御デバイスと通信できるということが理解される。

【0029】

ここで図2を参照すると、クラウド・コンピューティング環境50（図1）によって提供される機能的抽象レイヤのセットが示されている。図2に示されたコンポーネント、レイヤ、および機能は、例示のみが意図されており、本発明の実施形態がこれらに限定されないということが、あらかじめ理解されるべきである。図示されているように、次のレイヤおよび対応する機能が提供される。

【0030】

ハードウェアおよびソフトウェア・レイヤ60は、ハードウェア・コンポーネントおよびソフトウェア・コンポーネントを含む。ハードウェア・コンポーネントの例としては、メインフレーム61、RISC（Reduced Instruction Set Computer）アーキテクチャベースのサーバ62、サーバ63、ブレード・サーバ64、ストレージ・デバイス65、ならびにネットワークおよびネットワーク・コンポーネント66が挙げられる。一部の実施形態では、ソフトウェア・コンポーネントは、ネットワーク・アプリケーション・サーバ・ソフトウェア67およびデータベース・ソフトウェア68を含む。

【0031】

仮想化レイヤ70は、仮想サーバ71、仮想ストレージ72、仮想プライベート・ネットワークを含む仮想ネットワーク73、仮想アプリケーションおよびオペレーティング・システム74、ならびに仮想クライアント75などの仮想的実体を提供できる抽象レイヤを備える。

【0032】

一例を挙げると、管理レイヤ80は、以下で説明される機能を提供することができる。リソース・プロビジョニング81は、クラウド・コンピューティング環境内でタスクを実行するために利用される計算リソースおよびその他のリソースの動的調達を行う。計測および価格設定82は、クラウド・コンピューティング環境内でリソースが利用される際のコスト追跡、およびそれらのリソースの利用に対する請求書の作成と送付を行う。一例を挙げると、それらのリソースは、アプリケーション・ソフトウェア・ライセンスを含むことができる。セキュリティは、クラウドの利用者およびタスクのID検証を行うとともに、データおよびその他のリソースの保護を行う。ユーザ・ポータル83は、クラウド・コンピューティング環境へのアクセスを利用者およびシステム管理者に提供する。サービス・レベル管理84は、必要なサービス・レベルを満たすように、クラウドの計算リソースの割り当てと管理を行う。サービス水準合意（SLA：Service Level Agreement）計画および実行85は、今後の要求が予想されるクラウドの計算リソースの事前準備および調達を、SLAに従って行う。

【0033】

ワークロード・レイヤ90は、クラウド・コンピューティング環境で利用できる機能の例を示している。このレイヤから提供されてよいワークロードおよび機能の例としては、マッピングおよびナビゲーション91、ソフトウェア開発およびライフサイクル管理92、仮想クラスルーム教育の配信93、データ分析処理94、データベースからのビデオの取得95、およびビデオ・シーンへの注釈付与96が挙げられる。

【0034】

図3を参照すると、本明細書における内容を実装するための処理システム100の実施形態が示されている。この実施形態では、システム100は、1つまたは複数の中央処理装置（プロセッサ）101a、101b、101cなど（総称して、または一般的に、プロセッサ101として参照される）を含む。1つまたは複数の実施形態では、各プロセッ

10

20

30

40

50

サ 1 0 1 は、縮小命令セット・コンピュータ (R I S C : reduced instruction set computer) マイクロプロセッサを含んでよい。プロセッサ 1 0 1 は、システム・バス 1 1 3 を介して、システム・メモリ 1 1 4 およびさまざまなその他のコンポーネントに接続される。読み取り専用メモリ (R O M : Read only memory) 1 0 2 は、システム・バス 1 1 3 に接続され、システム 1 0 0 の特定の基本機能を制御する基本入出力システム (B I O S : basic input/output system) を含んでよい。

【 0 0 3 5 】

図 3 は、システム・バス 1 1 3 に接続された入出力 (I / O : input/output) アダプタ 1 0 7 およびネットワーク・アダプタ 1 0 6 をさらに示している。I / O アダプタ 1 0 7 は、ハード・ディスク 1 0 3 またはテープ・ストレージ・デバイス 1 0 5 あるいはその両方、あるいは任意のその他の類似するコンポーネントと通信する小型コンピュータ・システム・インターフェイス (S C S I : small computer system interface) アダプタであってよい。I / O アダプタ 1 0 7、ハード・ディスク 1 0 3、およびテープ・ストレージ・デバイス 1 0 5 は、本明細書では、総称してマス・ストレージ 1 0 4 として参照される。処理システム 1 0 0 上で実行するためのオペレーティング・システム 1 2 0 は、マス・ストレージ 1 0 4 に格納されてよい。ネットワーク・アダプタ 1 0 6 は、バス 1 1 3 を外部ネットワーク 1 1 6 と相互接続し、データ処理システム 1 0 0 が他のそのようなシステムと通信できるようにする。画面 (例えば、表示モニタ) 1 1 5 は、ディスプレイ・アダプタ 1 1 2 によってシステム・バス 1 1 3 に接続され、ディスプレイ・アダプタ 1 1 2 は、グラフィックス重視のアプリケーションおよびビデオ・コントローラの性能を向上するために、グラフィックス・アダプタを含んでよい。一実施形態では、アダプタ 1 0 7、1 0 6、および 1 1 2 は、1 つまたは複数の I / O バスに接続されてよく、I / O バスは、中間バス・ブリッジ (図示されていない) を介してシステム・バス 1 1 3 に接続される。ハード・ディスク・コントローラ、ネットワーク・アダプタ、およびグラフィックス・アダプタなどの周辺機器を接続するのに適した I / O バスは、通常、P C I (Peripheral Component Interconnect) などの一般的なプロトコルを含む。ユーザ・インターフェイス・アダプタ 1 0 8 およびディスプレイ・アダプタ 1 1 2 を介してシステム・バス 1 1 3 に接続されているように、その他の入出力デバイスが示されている。キーボード 1 0 9、マウス 1 1 0、およびスピーカ 1 1 1 は、すべてユーザ・インターフェイス・アダプタ 1 0 8 を介してバス 1 1 3 に相互接続されており、例えばユーザ・インターフェイス・アダプタ 1 0 8 は、複数のデバイス・アダプタを 1 つの集積回路に統合するスーパー I / O チップを含んでよい。

【 0 0 3 6 】

実施形態例では、処理システム 1 0 0 は、グラフィックス・プロセッシング・ユニット 1 3 0 を含む。グラフィックス・プロセッシング・ユニット 1 3 0 は、ディスプレイへの出力を対象にしたフレーム・バッファ内の画像の作成を高速化するためにメモリを操作して変更するように設計された特殊な電子回路である。一般に、グラフィックス・プロセッシング・ユニット 1 3 0 は、コンピュータ・グラフィックスおよび画像処理の操作では非常に効率的であり、データの大きいブロックの処理が並列に実行されるアルゴリズムに関して汎用 C P U よりも効果的にする高度な並列構造を有する。

【 0 0 3 7 】

したがって、図 3 で構成されているように、システム 1 0 0 は、プロセッサ 1 0 1 の形態での処理能力と、システム・メモリ 1 1 4 およびマス・ストレージ 1 0 4 を含んでいるストレージ能力と、キーボード 1 0 9 およびマウス 1 1 0 などの入力手段と、スピーカ 1 1 1 およびディスプレイ 1 1 5 を含んでいる出力能力とを含む。一実施形態では、システム・メモリ 1 1 4 およびマス・ストレージ 1 0 4 の一部は、オペレーティング・システムを集散的に格納し、図 3 に示されたさまざまなコンポーネントの機能を調整する。

【 0 0 3 8 】

本発明の 1 つまたは複数の実施形態は、ビデオ・シーンに注釈を付け、データベースからビデオ・シーンを取得するためのシステム、方法、およびコンピュータ・プログラム製

10

20

30

40

50

品を提供する。本発明の態様は、ビデオ・シーン内の物体、実体、動作、概念、および情緒を識別するための手法を利用して、ビデオ・コンテンツを解析し、ビデオ・シーンに関連付けられたラベルを提供することを含む。それらのラベルは、キーワードのド形態または自然言語記述（すなわち、ビデオ・シーンを説明する文章）であることができる。キーワードのラベルに加えて、抽出されたキーワードに関連付けられた拡張キーワードが作成される。例えば、「野球」などのキーワードが、フィールド、ピッチャー、スタジアム、マウンド、ベースなどの拡張キーワードに関連付けられ得る。それらの拡張キーワードも、ビデオ・シーンに注釈付けされ得る。それらの注釈付きのビデオ・シーンは、検索のためにデータベースに格納され得る。

【0039】

本発明の1つまたは複数の実施形態では、ユーザは、データベース内のビデオ・シーンを検索できる。ユーザは、データベースに格納されたビデオ・シーンに関する検索要求を送信できる。検索要求は、特定のビデオ・シーンに関するユーザからの音声入力またはユーザからのテキスト入力であることができる。検索要求からキーワードを抽出し、ビデオ・シーンの注釈と比較して、一致するシーンのリストを決定できる。さらに、検索要求の抽出されたキーワードから、拡張キーワードが作成され得る。拡張キーワードをビデオ・シーンの注釈と比較して、一致するシーンのリストを決定できる。

【0040】

図4は、1つまたは複数の実施形態に従って、デジタル・ビデオを移動するためのシステム400のブロック図を示している。システム400は、サーバ401、ビデオ・シーン用の参照データベース402、外部ライブラリ405、およびクライアント406を含んでいる。サーバ401は、画像認識モジュール403、自然言語処理（NLP：natural language processing）モジュール404、および通信モジュール408を含んでいる。サーバ401は、概念タグ付け、感情解析、情緒解析、および関係抽出を含んでいる解析モジュール420も含む。クライアント406は、クライアント406と電子的に通信するユーザ入力モジュール407、表示モジュール410、およびセンサ430を含んでいる。

【0041】

本発明の1つまたは複数の実施形態では、サーバ401、クライアント406、画像認識モジュール403、自然言語処理モジュール404、通信モジュール408、および解析モジュール420は、図3にある処理システム100上に実装され得る。さらに、クラウド・コンピューティング・システム50は、システム400の要素のうちの1つまたはすべてと、有線または無線で電子的に通信できる。クラウド50は、システム400の要素の機能の一部または全部を補完するか、支援するか、または置き換えることができる。さらに、システム400の要素の機能の一部または全部は、クラウド50のノード10（図1および2に示されている）として実装され得る。クラウド・コンピューティング・ノード10は、適切なクラウド・コンピューティング・ノードの一例に過ぎず、本明細書に記載された本発明の実施形態の使用または機能の範囲に関して、いかなる制限を示唆することも意図されていない。

【0042】

本発明の1つまたは複数の実施形態では、システム400は、ビデオ・データに注釈を付けることに利用され得る。ビデオ・データに注釈を付けることは、対応するビデオ・シーンに関連付けられた意味的特性の形態（例えばタグ、説明などの、単語の形態または文章の形態）で注釈を適用することを含む。システム400は、サーバ401およびビデオ・データを格納する参照データベース402を含んでいる。ビデオ・データは、映画、テレビ番組、インターネット・ビデオなどを含むことができる。ビデオ・データは、さまざまな長さのビデオ・シーンに分割され得る。シーンの長さは、テーマまたは概念と相関性があるが、結婚式のシーン全体は、数分にわたって発生することがあるが、結婚式のシーン全体は、発生する動作あるいは通路を歩くことまたは結婚披露宴などの設定に基づいて、より短いシーンにさらに分割され得る。ビデオ・データは、ビ

10

20

30

40

50

デオ・シーンに関する概念、感情、および情緒に基づいてさらに分割され得る。

【 0 0 4 3 】

サーバ 4 0 1 は、概念タグ付け、感情解析、情緒解析、および関係抽出を利用する解析モジュール 4 2 0 を利用して、参照データベース 4 0 2 上のビデオ・シーンに注釈を付ける（「タグ」または「タグ付け」と呼ばれることもある）。サーバ 4 0 1 は、画像認識モジュール 4 0 3 も利用して、ビデオ・シーン内の物体を識別する。NLPモジュール 4 0 4 は、ビデオ・シーンの概念、感情、情緒、関係の抽出の解析および注釈のために、ビデオ・シーンの各々に含まれる音声の解析および識別に利用される。

【 0 0 4 4 】

情緒解析手法は、テキストおよび音声で表された意見を（NLPを介して）識別して分類し、話題、製品に対する話者またはその他の対象者の態度、あるいは物体、相互作用、またはイベントに対する全体的な文脈上の両極性または感情的反応を決定することを含む。抽出され得る情緒は、肯定的、否定的、および中立的を含むが、これらに限定されない。態度は、判断または評価、感情状態（すなわち、感情の状態）、または意図された感情的コミュニケーション（すなわち、話者によって意図された感情的効果）であることができる。前述した情緒解析を介して抽出された感情に加えて、感情解析は、表情認識などの手法を介して個人の顔を解析し、個人の 1 つまたは複数の感情を決定することを含むこともできる。感情解析は、表情認識に少なくとも部分的に基づいて、個人の感情の変化を決定することもできる。

【 0 0 4 5 】

画像認識モジュール 4 0 3 は、参照データベース 4 0 2 から取得されたさまざまなビデオ・シーン内の物体を決定するために利用される。画像認識は、1 つのシーン内の複数の画像に含まれる形状および物体を認識して識別するために実行される。画像認識の実行時に使用される特定の画像認識アルゴリズムは、特定のアプリケーションまたは処理の制約に利用できる任意の適切な画像認識アルゴリズムまたはパターン認識アルゴリズムであることができる。画像認識アルゴリズムは、既知の物体に対する 1 つまたは複数のシーン内の物体の照合を実行するために利用可能なデータベースによって、制限されることがある。1 つの例として、画像認識アルゴリズムは、画像の前処理を含むことができる。前処理は、画像のコントラストの調整、グレースケールまたは白黒あるいはその両方への変換、トリミング、サイズ変更、回転、ならびにこれらの組み合わせを含むことができるが、これらに限定されない。特定の画像認識アルゴリズムによれば、例えば、色、サイズ、または形状などの特徴を区別することが、特定の物体を検出するために使用するために選択され得る。物体の特徴を区別することを可能にする複数の特徴が、使用され得る。ビデオ・シーン内の物体のエッジを決定するために、エッジ検出が実行され得る。ピクセルのセットに対して、不要な成分の除去を含む動作を実行するために、画像認識アルゴリズムにおいて、モルフォロジーが実行されてよい。加えて、ノイズ低減または領域の塗りつぶしあるいはその両方が実行されてよい。さらに、画像認識アルゴリズムでは、画像内で 1 つまたは複数の物体（および関連する特性）が発見 / 検出された後に、1 つまたは複数の物体の位置が、ビデオ・シーン内でそれぞれ特定され、その後、分類され得る。特定された物体は、際立った特徴に関連する特定の指定に従って、特定された物体を評価することによって、分類され得る（すなわち、特定の形状または物体として識別され得る）。特定の指定は、数学的計算または関係を含むことができる。別の例では、ビデオ・シーン内の認識可能な物体の位置を特定する代わりに、または特定することに加えて、パターン照合が実行され得る。照合は、画像内の要素または物体あるいはその両方を、「既知」の（以前に識別または分類された）物体および要素（例えば、ラベル付きのトレーニング・データ）と比較することによって、実行され得る。画像認識モジュール 4 0 3 は、ビデオ・シーン内で識別された物体をラベル付きのトレーニング・データと比較することによって、機械学習を利用して、識別の精度を検証できる。画像認識モジュール 4 0 3 は、ニューラル・ネットワーク（NN：neural networks）およびその他の学習アルゴリズムを利用できる。画像認識モジュール 4 0 3 の識別プロセスは、信頼性しきい値などの、識別の信頼度を

10

20

30

40

50

含むことができる。このしきい値を下回る信頼性を有する物体のすべての識別が、破棄され得る。物体の識別が、特定のしきい値を超える信頼性を有する場合、シーンの物体に注釈（ラベル）を付けることができる。例えば、シーンは、背景に車両を含むことができ、画像認識モジュール403は、その車両をオートバイとして識別することができ、そのシーンの場合、ラベルまたは注釈は、「オートバイ」というラベルを含むことができる。画像認識モジュール403は、例えば、色、位置、運転中または駐車中のどちらであるか、ブランドなどの、オートバイに関する特徴を識別することもできる。説明された実施形態および例が画像認識を参照できるということを確信しているが、このことが、説明された実施形態および例を画像のみに限定していると解釈されるべきではないということが、理解されるべきである。例えば、本発明の1つまたは複数の実施形態に従って説明されているように、ビデオ信号がシステム400によって受信され、自動タグ生成プロセスが適用され得る。参照データベース402から1つまたは複数のビデオ・フレームを受信することができ、ビデオ・フレームは画像を含むことができ、画像認識が実行され得る。

【0046】

本発明の1つまたは複数の実施形態では、画像認識モジュール403は、ビデオ・シーン内の人、物体、実体、およびその他の特徴を識別するために利用される。解析モジュール420は、ビデオ・シーンの注釈用に、それらの人、物体、実体、およびその他の特徴の間の関係を決定するために利用される。例えば、法廷内の人物の識別は、関係の抽出と共に、その人物を裁判官として識別することができる。このシーンの注釈は、訴訟手続きまたは判決などに関連するキーワードでタグ付けされ得る。この人物は、識別された後に、外部ライブラリ405と相互参照され得る。外部ライブラリ405は、ビデオ・シーンに関連するインターネット・ムービー・データベース（IMDB：internet movie database）、電子番組ガイド（EPG：electronic program guide）、およびその他の類似する外部ライブラリを含むが、これらに限定されない。この人物の識別は、外部ライブラリ405内の人物の説明を相互参照し、上の例では、この人物が裁判官であるということを確認することによって、さらに改善され得る。

【0047】

NLPモジュール404を介したテキストおよび音声の解析は、ビデオ・シーンに関連付けられた字幕／サブタイトルおよび会話を解析するために利用される。解析モジュール420は、情緒、実体、動作、および概念を決定するために利用される。例えば、ビデオ・シーンに関連付けられた音声データが、NLPモジュール404によって解析され、音声テキスト変換（STT：speech to text）などの手法を使用して、音声データをテキストに変換することができる。ビデオ・シーンの注釈用のキーワードが、テキスト形式で抽出され得る。

【0048】

1つまたは複数の実施形態では、サーバ401が、通信モジュール408を介してクライアント406と通信することができる。クライアント406は、クライアント406と電子的に通信する1つまたは複数のセンサ430を含んでいる。クライアント406は、表示モジュール410を介してビデオ・シーンを視聴者に表示するための、任意の種類のコンピュータまたはインターフェイスであることができる。例えば、クライアント406は、スマートフォンの周りの個人のグループ（すなわち、視聴者）が見るために、ビデオ・シーンをスマートフォン上に表示しているスマートフォンであることができる。センサ430は、カメラまたはマイクロホンあるいはその両方を含むことができる。この例では、センサ430は、ビデオ・シーンに対する視聴者の反応を記録する、スマートフォンに組み込まれたマイクロホンおよびカメラであることができる。感情、情緒などを含む視聴者の反応は、サーバ401上の解析モジュール420を利用して解析され得る。それらの感情、情緒などは、参照データベース402上で、ビデオ・シーンにさらに注釈を付けることができる。

【0049】

視聴者は、映画館にいる個人、個人的なビデオを見ている家族などを含む、任意の種類

10

20

30

40

50

の視聴者であることができる。クライアント406は、センサ430と通信し、解析モジュール420によって解析されるセンサ・データをセンサ430からサーバに送信することができる、スマート・テレビおよびその他のシステムを含むことができる。システム400は、ビデオ・シーンを見ているときに視聴者によって行われた言葉による指示およびコメントを記録し、NLPモジュール404を利用して解析用の語句およびキーワードを抽出する。加えて、システム400は、センサ430を利用して、ビデオ・シーンを見ている間の視聴者の表情および身振りを記録する。センサ・データは、情緒解析および感情解析を利用して、解析モジュール420によって解析され得る。記録され得る反応の例としては、例えば、驚く、怖がる、泣く、熱狂するなどの反応が挙げられる。1つまたは複数の実施形態では、視聴者の反応に少なくとも部分的に基づいて、システム400は、視聴者に見られているビデオ・シーンに注釈を付けることができる。注釈は、視聴者の感情的な反応を含むことができる。

10

【0050】

視聴者の反応をビデオ・シーンに含めることの利点は、シーンの情緒解析を確認することを含む。情緒解析および感情解析は、例えば機械学習などの、任意の適切な学習アルゴリズムを使用して、実行され得る。視聴者の反応は、学習アルゴリズムによって作成されたラベルを確認または拒否し、学習アルゴリズムが学習するのを支援することができる。例えば、学習アルゴリズムは、情緒解析を使用して、シーン内の実体、音声コンテンツ、および関係に基づいて、情緒を識別することができる。識別された情緒は、それに関連付けられた信頼度を有することができる。信頼度に少なくとも部分的に基づいて、視聴者の反応が取得され、シーンの情緒を確認または拒否することができる。シーンの情緒に、「悲しい」シーンというラベルが付けられたが、視聴者の反応が「喜び」および「笑い」を含んでいる場合、そのラベルは、視聴者の反応に一致し、機械学習アルゴリズムをトレーニングするように、調整され得る。

20

【0051】

本発明の1つまたは複数の実施形態では、システム400は、デジタル・ビデオを移動することに利用され得る。ユーザは、クライアント406上のユーザ入力モジュール407を介して、ビデオ内の特定のシーンに関する要求を行うことができる。クライアント406は、個人のコンピューティング・デバイス、テレビ、スマートフォン、またはその他のスマート・デバイスなどであることができる。ユーザ入力モジュール407は、ユーザによる音声入力、テキスト入力、またはユーザからのグラフィックス入力を介して、ビデオ内のシーンに関するユーザの照会を受信するように構成される。例えば、ユーザは、キーボードまたは画面上の文字を選択するマウスなどの、クライアント406の周辺機器で、要求をテキスト入力として入力することができる。また、ユーザは、クライアント406と電子的に通信するマイクロホンなどの、クライアント406に通信によって接続された周辺機器を介して、言葉で要求を伝えることができる。ユーザは、ユーザ入力モジュール407を介して、照会のグラフィック表現を選択することもできる。例えば、照会は、ビデオ内のアクション・シーンまたはビデオ内のラブ・シーンを表すアイコンを含むことができる。その場合、ユーザは、アイコンを選択して照会をサーバ401に送信できる。

30

【0052】

クライアント406は、ユーザ入力モジュール407を介してユーザ入力を受信し、サーバ401に照会して、ユーザの要求に最も良く一致する1つまたは複数のビデオ・シーンを見つけることができる。サーバ401は、参照データベース402に格納されたビデオ・データを含む。ビデオ・データは、対応する注釈（「タグ」または「キーワード」とも呼ばれる）付きのビデオ・シーンを含む。注釈は、ビデオ・シーン内の物体、情緒、動作、実体、および概念を説明するラベルまたはタグを含む。物体は、俳優、車両、場所、建物、動物などを含むことができる。概念は、ロマンス、喜び、幸福などのシーンに対する感情を含むことができる。概念は、アクション、スリル、恐怖などのシーンに関するカテゴリを含むこともできる。

40

【0053】

50

本発明の１つまたは複数の実施形態では、ユーザ入力モジュール４０７は、ユーザから照会を受信して、通信モジュール４０８に送信する。この照会は、解析のためにＮＬＰモジュール４０４に送信される。ＮＬＰモジュール４０４は、例えば音声テキスト変換（ＳＴＴ）などの手法を利用して、音声の照会をテキスト形式に変換することができる。ＮＬＰモジュール４０４は、テキスト形式の照会を解析して、照会の言語からキーワードを抽出する。例えば照会は、「ビルがホームランを打ったシーン」を要求する。抽出されたキーワードは、「ビル」、「ホームラン」、および「ヒット」を含むであろう。抽出されたキーワードは、拡張キーワードを識別するために利用される。前の例では、拡張キーワードは、「野球」、「野球場」、「バット」、「ユニフォーム」、「スコア」などを含むことができる。拡張キーワードは、俳優の名前を含むこともでき、キーワード「ビル」が、この俳優の氏名または映画の登場人物の氏名に拡張されるであろう。キーワードおよび拡張キーワードは、キーワードおよび拡張キーワードに最も良く一致する１つまたは複数の一致するシーンを見つけるために、シーン照会モジュール４０９によって参照データベース４０２内の注釈付きのシーンと比較される。１つまたは複数の一致するシーンが、クライアント４０６の表示モジュール４１０で、ユーザに提示され得る。

【００５４】

ＮＬＰモジュール４０４を使用してユーザ入力を受信することの利点は、ユーザがビデオ・コンテンツにアクセスするための使いやすい方法を作ることを含む。クライアント４０６は、スマートフォンを含む任意の種類の電子デバイスを含むことができる。スマートフォンは、通常、検索要求を入力するためのキーボードを備えているコンピュータとは異なり、検索要求を入力する簡単な方法を有していない。音声の要求を受信し、キーワードを抽出できることは、クライアント４０６の種類に関する優位性および柔軟性を可能にする。

【００５５】

情緒解析および感情解析を採用することは、ビデオ・シーンにラベルを付けて、検索のためのさらに多くの選択肢をユーザに提供するという優位性を作り出す。通常、ユーザは、シーンを探すときに、場所または登場人物を説明しようとする。代わりに、感情および情緒を含める能力によって、特に、ユーザが登場人物の名前もシーンの説明も覚えていない場合に、ユーザは、シーンのより広い検索を投じることができるようになる。さらに、ユーザは、特定のシーンを探していないことがある。代わりに、ユーザは、特定の感情を伝えるシーンを求めていることがある。例えば、ユーザがプレゼンテーションを準備しており、プレゼンテーションに含めるために「元気が出る」シーンを見つけない場合、ユーザは「元気が出るシーン」を検索することができ、システム４００は、この説明に一致する複数のシーンを返すことができる。

【００５６】

図５は、本発明の１つまたは複数の実施形態に従ってデジタル・ビデオを移動するための例示的なシステムを示している。システム５００は、「公園内でボブがアリスに求婚しているときのシーンを見たい」と述べるユーザ入力５０１を入力しているユーザを含んでいる。ＮＬＰモジュール４０４は、シーンの説明に関連していないユーザ入力５０１のコンテンツを除去してユーザ・ステートメント５０３を作成するために、利用される。キーワード５０２および拡張キーワードが、ユーザ・ステートメント５０３から抽出される。抽出されたキーワード５０２は、「ボブ」、「アリス」、「求婚」、および「公園」を含む。「公園」に関する拡張キーワードは、「木、草、および池」を含む。「求婚」に関する拡張キーワードは、「指輪、ダイヤモンド、および花」を含む。（図４の）シーン照会モジュール４０９は、キーワードおよび拡張キーワードを参照データベース４０２内のビデオ・シーンと比較して、キーワードおよび拡張キーワードに一致する１つまたは複数のビデオ・シーンを決定する。図に示されているように、参照データベース４０２内のビデオ・シーンは、シーンのタグおよびシーンの説明５０５の形態で、注釈を含んでいる。これらの注釈は、前述した手法を利用して適用されている。この事例では、シーンのタグは「ボブ、アリス、バラ、木、草、および求婚」を含んでいる。さらに、シーンの説明５０

5は「公園内でボブがアリスに求婚している」を含んでいる。ビデオ・シーンの注釈とのユーザ・ステートメントのキーワードの比較に少なくとも部分的に基づいて、シーン照会モジュール409によってビデオ・シーンが選択され、表示のためにクライアント406に提示される。

【0057】

本発明の1つまたは複数の実施形態では、シーン照会モジュール409は、ユーザ入力キーワードとビデオ・シーンの注釈の比較に基づいて、識別されたビデオ・シーンの信頼値を決定できる。この信頼値は、一致する可能性のあるシーンのリストを提示するときに、クライアント・デバイス406上に表示され得る。ユーザが特定のシーンを選択することに少なくとも部分的に基づいて、サーバ401は、ビデオ・シーンの注釈の信頼度を増やし、ビデオ・シーンの注釈を更新することができる。例えば、5つのキーワードのうちの4つがビデオ・シーンの注釈に一致し、その後、ユーザが特定のシーンを選択した場合、ユーザによって確認されたとおりに注釈を改善するために、5番目のキーワードを使用してビデオ・シーンが更新され得る。本発明の1つまたは複数の実施形態では、システムの複数のユーザによる照会の言語およびその後の選択に基づいて、ビデオ・シーンの注釈が継続的に更新され得る。

【0058】

本発明の1つまたは複数の実施形態では、ビデオ・シーン内の物体を識別するために、画像認識モジュール403が使用可能であり、この識別を検証するために、機械学習技術が利用され得る。例えば図5では、シーンがバラを含んでおり、画像認識モジュール403は、例えば「花」、「バラ」、または「赤い花」などのタグを返すことができる。画像認識モジュール403は、これらのタグを、花、バラ、または赤い花、あるいはその組み合わせの既知の（ラベル付きの）画像と比較して、シーン内で識別されたこの物体のタグを検証することができる。

【0059】

ここで図6を参照すると、本発明の1つまたは複数の実施形態に従ってデータベースからシーンを取得するための方法600のフロー図が示されている。方法600は、ブロック602に示されているように、ユーザから、データベースに格納されたシーンに関する検索要求を受信することを含んでおり、このデータベースが注釈付きのビデオ・コンテンツを含んでいる。ブロック604で、方法600は、1つまたは複数のキーワードを検索要求から抽出することを含んでいる。方法600は、ブロック606で、1つまたは複数のキーワードの各々に関して、1つまたは複数の拡張キーワードを生成することを含んでいる。ブロック608で、方法600は、1つまたは複数の拡張キーワードおよび1つまたは複数のキーワードを、注釈付きのビデオ・コンテンツと比較して、1つまたは複数の対象のシーンを識別することを含んでおり、1つまたは複数の対象のシーンは、それぞれ対象のシーンの注釈を含んでいる。方法600は、ブロック610に示されているように、1つまたは複数の対象のシーンの各々の対象のシーンの注釈と、1つまたは複数の拡張キーワードおよび1つまたは複数のキーワードとの間の相関関係に少なくとも部分的に基づいて、信頼性評価を1つまたは複数の対象のシーンの各々に割り当てることを含んでいる。ブロック612で、方法600は、1つまたは複数の対象のシーンのうちの少なくとも1つをユーザに表示することを含んでおり、1つまたは複数の対象のシーンのうちの少なくとも1つが、信頼性評価に基づいて決定される。

【0060】

追加のプロセスが含まれてもよい。図6に示されたプロセスが例を表しており、本開示の範囲および思想を逸脱することなく、他のプロセスが追加されてよく、または既存のプロセスが削除、変更、または再配置されてよいということが、理解されるべきである。

【0061】

ここで図7を参照すると、本発明の1つまたは複数の実施形態に従ってビデオ・シーンに注釈を付けるための方法700のフロー図が示されている。方法700は、ブロック702に示されているように、プロセッサによって1つまたは複数のビデオを受信すること

を含んでいる。ブロック 704 で、方法 700 は、1 つまたは複数のビデオの各々をシーンのセットに分割することを含んでいる。方法 700 は、ブロック 706 で、シーンのセット内の第 1 のシーンを解析して、第 1 のシーンのラベルのセットを決定することを含んでいる。ブロック 708 で、方法 700 は、第 1 のシーンにラベルのセットで注釈を付けることを含んでいる。ブロック 710 で、方法 700 は、第 1 のシーンをデータベースに格納することを含んでいる。

【0062】

追加のプロセスが含まれてもよい。図 7 に示されたプロセスが例を表しており、本開示の範囲および思想を逸脱することなく、他のプロセスが追加されてよく、または既存のプロセスが削除、変更、または再配置されてよいということが、理解されるべきである。

10

【0063】

本発明は、システム、方法、またはコンピュータ・プログラム製品、あるいはその組み合わせであってよい。コンピュータ・プログラム製品は、プロセッサに本発明の態様を実行させるためのコンピュータ可読プログラム命令を含んでいるコンピュータ可読記憶媒体を含んでよい。

【0064】

コンピュータ可読記憶媒体は、命令実行デバイスによって使用するための命令を保持および格納できる有形のデバイスであることができる。コンピュータ可読記憶媒体は、例えば、電子ストレージ・デバイス、磁気ストレージ・デバイス、光ストレージ・デバイス、電磁ストレージ・デバイス、半導体ストレージ・デバイス、またはこれらの任意の適切な組み合わせであってよいが、これらに限定されない。コンピュータ可読記憶媒体のさらに具体的な例の非網羅的リストは、ポータブル・フロッピー（R）・ディスク、ハード・ディスク、ランダム・アクセス・メモリ（RAM：random access memory）、読み取り専用メモリ（ROM：read-only memory）、消去可能プログラマブル読み取り専用メモリ（EPROM：erasable programmable read-only memoryまたはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（SRAM：static random access memory）、ポータブル・コンパクト・ディスク読み取り専用メモリ（CD-ROM：compact disc read-only memory）、デジタル多用途ディスク（DVD：digital versatile disk）、メモリ・スティック、フロッピー（R）・ディスク、パンチカードまたは命令が記録されている溝の中の隆起構造などの機械的にエンコードされるデバイス、およびこれらの任意の適切な組み合わせを含む。本明細書において使用されるとき、コンピュータ可読記憶媒体は、それ自体が、電波またはその他の自由に伝搬する電磁波、導波管またはその他の送信媒体を伝搬する電磁波（例えば、光ファイバ・ケーブルを通過する光パルス）、あるいはワイヤを介して送信される電気信号などの一過性の信号であると解釈されるべきではない。

20

30

【0065】

本明細書に記載されたコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体から各コンピューティング・デバイス／処理デバイスへ、またはネットワーク（例えば、インターネット、ローカル・エリア・ネットワーク、広域ネットワーク、または無線ネットワーク、あるいはその組み合わせ）を介して外部コンピュータまたは外部ストレージ・デバイスへダウンロードされ得る。このネットワークは、銅伝送ケーブル、光伝送ファイバ、無線送信、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ、またはエッジ・サーバ、あるいはその組み合わせを備えてよい。各コンピューティング・デバイス／処理デバイス内のネットワーク・アダプタ・カードまたはネットワーク・インターフェイスは、コンピュータ可読プログラム命令をネットワークから受信し、それらのコンピュータ可読プログラム命令を各コンピューティング・デバイス／処理デバイス内のコンピュータ可読記憶媒体に格納するために転送する。

40

【0066】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ISA：instruction-set-architecture）命令、マシン命

50

令、マシン依存命令、マイクロコード、ファームウェア命令、状態設定データ、あるいは、Smalltalk、C++などのオブジェクト指向プログラミング言語、および「C」プログラミング言語または同様のプログラミング言語などの従来の手続き型プログラミング言語を含む1つまたは複数のプログラミング言語の任意の組み合わせで記述されたソース・コードまたはオブジェクト・コードであってよい。コンピュータ可読プログラム命令は、ユーザのコンピュータ上で全体的に実行すること、ユーザのコンピュータ上でスタンドアロン・ソフトウェア・パッケージとして部分的に実行すること、ユーザのコンピュータ上およびリモート・コンピュータ上でそれぞれ部分的に実行すること、あるいはリモート・コンピュータ上またはサーバ上で全体的に実行することができる。後者のシナリオでは、リモート・コンピュータは、ローカル・エリア・ネットワーク（LAN：local area network）または広域ネットワーク（WAN：wide area network）を含む任意の種類のネットワークを介してユーザのコンピュータに接続されてよく、または接続は、（例えば、インターネット・サービス・プロバイダを使用してインターネットを介して）外部コンピュータに対して行われてよい。一部の実施形態では、本発明の態様を実行するために、例えばプログラマブル論理回路、フィールドプログラマブル・ゲート・アレイ（FPGA：field-programmable gate arrays）、またはプログラマブル・ロジック・アレイ（PLA：programmable logic arrays）を含む電子回路は、コンピュータ可読プログラム命令の状態情報を利用することによって、電子回路をカスタマイズするためのコンピュータ可読プログラム命令を実行してよい。

【0067】

本発明の態様は、本明細書において、本発明の実施形態に従って、方法、装置（システム）、およびコンピュータ・プログラム製品のフローチャート図またはブロック図あるいはその両方を参照して説明される。フローチャート図またはブロック図あるいはその両方の各ブロック、ならびにフローチャート図またはブロック図あるいはその両方に含まれるブロックの組み合わせが、コンピュータ可読プログラム命令によって実装され得るということが理解されるであらう。

【0068】

これらのコンピュータ可読プログラム命令は、コンピュータまたはその他のプログラム可能なデータ処理装置のプロセッサを介して実行される命令が、フローチャートまたはブロック図あるいはその両方のブロックに指定される機能／動作を実施する手段を作り出すべく、汎用コンピュータ、専用コンピュータ、または他のプログラム可能なデータ処理装置のプロセッサに提供されてマシンを作り出すものであってよい。これらのコンピュータ可読プログラム命令は、命令が格納されたコンピュータ可読記憶媒体がフローチャートまたはブロック図あるいはその両方のブロックに指定される機能／動作の態様を実施する命令を含んでいる製品を備えるように、コンピュータ可読記憶媒体に格納され、コンピュータ、プログラム可能なデータ処理装置、または他のデバイス、あるいはその組み合わせに特定の方式で機能するように指示できるものであってもよい。

【0069】

コンピュータ可読プログラム命令は、コンピュータ上、その他のプログラム可能な装置上、またはその他のデバイス上で実行される命令が、フローチャートまたはブロック図あるいはその両方のブロックに指定される機能／動作を実施するように、コンピュータ、その他のプログラム可能なデータ処理装置、またはその他のデバイスに読み込まれてもよく、それによって、一連の動作可能なステップを、コンピュータ上、その他のプログラム可能な装置上、またはコンピュータ実装プロセスを生成するその他のデバイス上で実行させる。

【0070】

図内のフローチャートおよびブロック図は、本発明のさまざまな実施形態に従って、システム、方法、およびコンピュータ・プログラム製品の可能な実装のアーキテクチャ、機能、および動作を示す。これに関連して、フローチャートまたはブロック図内の各ブロックは、規定された論理機能を実装するための1つまたは複数の実行可能な命令を備える、

命令のモジュール、セグメント、または部分を表してよい。一部の代替の実装では、ブロックに示された機能は、図に示された順序とは異なる順序で発生してよい。例えば、連続して示された2つのブロックは、実際には、含まれている機能に応じて、実質的に同時に実行されるか、または場合によっては逆の順序で実行されてよい。ブロック図またはフローチャート図あるいはその両方の各ブロック、ならびにブロック図またはフローチャート図あるいはその両方に含まれるブロックの組み合わせは、規定された機能または動作を実行するか、または専用ハードウェアとコンピュータ命令の組み合わせを実行する専用ハードウェアベースのシステムによって実装され得るということにも注意する。

【符号の説明】

【0071】

- 400 システム
- 401 サーバ
- 402 参照データベース
- 403 画像認識モジュール
- 404 自然言語処理（NLP）モジュール
- 405 外部ライブラリ
- 406 クライアント
- 407 ユーザ入力モジュール
- 408 通信モジュール
- 409 シーン照会モジュール
- 410 表示モジュール
- 420 解析モジュール
- 430 センサ

10

20

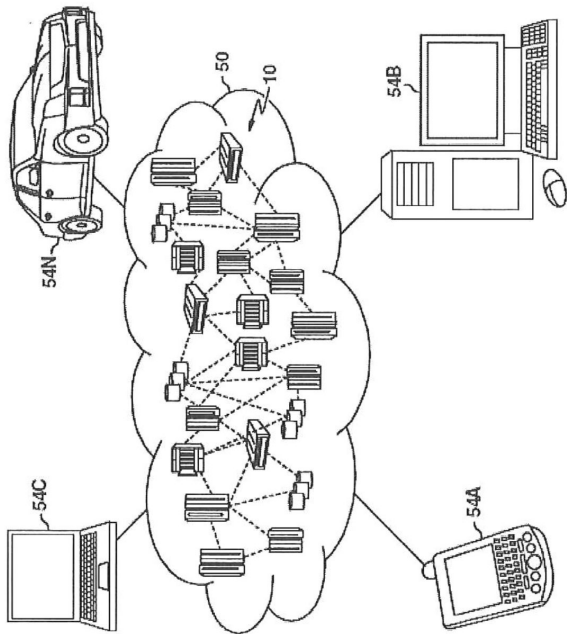
30

40

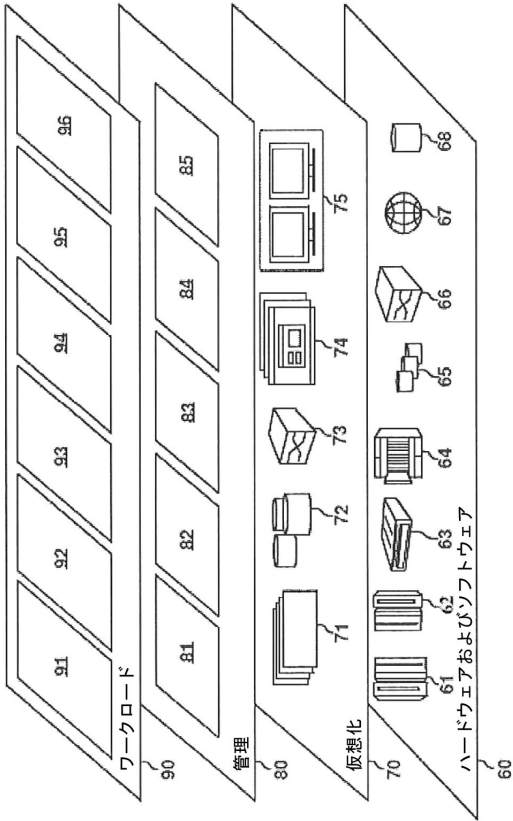
50

【図面】

【図 1】



【図 2】



10

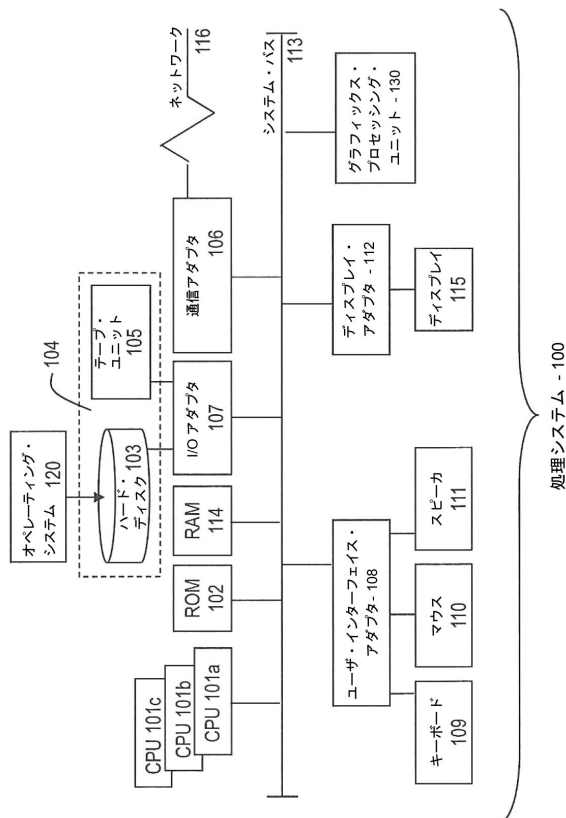
20

30

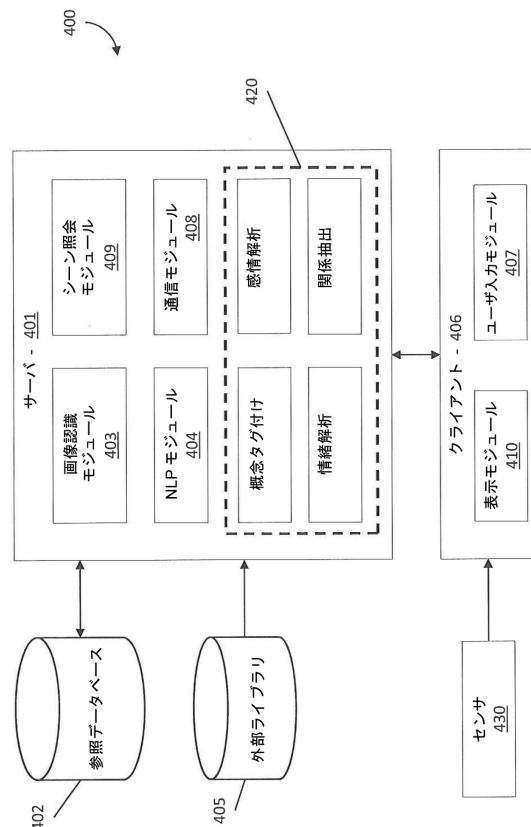
40

50

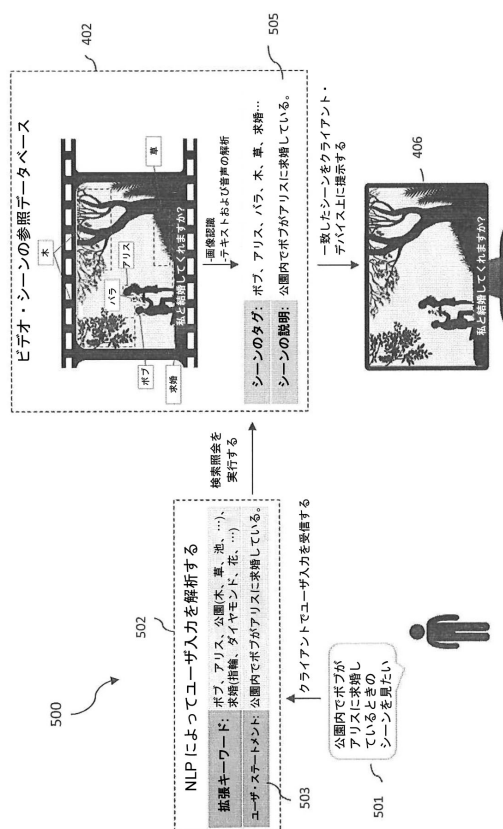
【図 3】



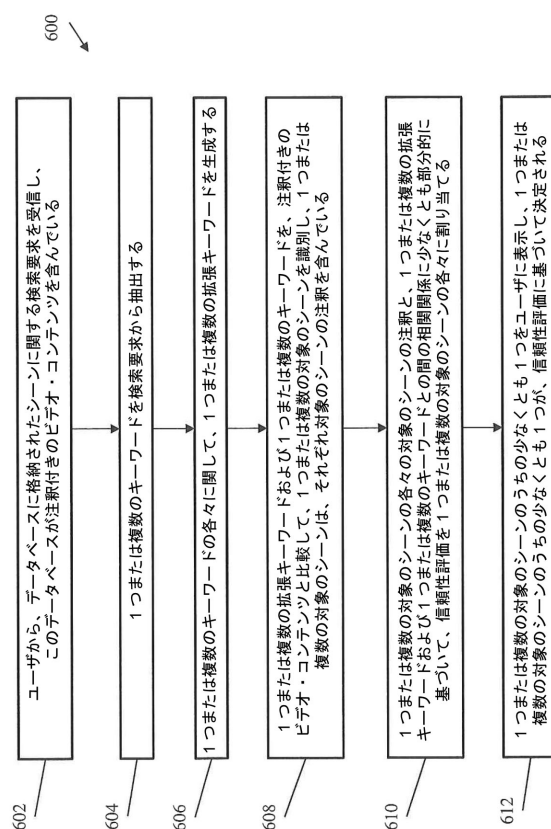
【図 4】



【図 5】

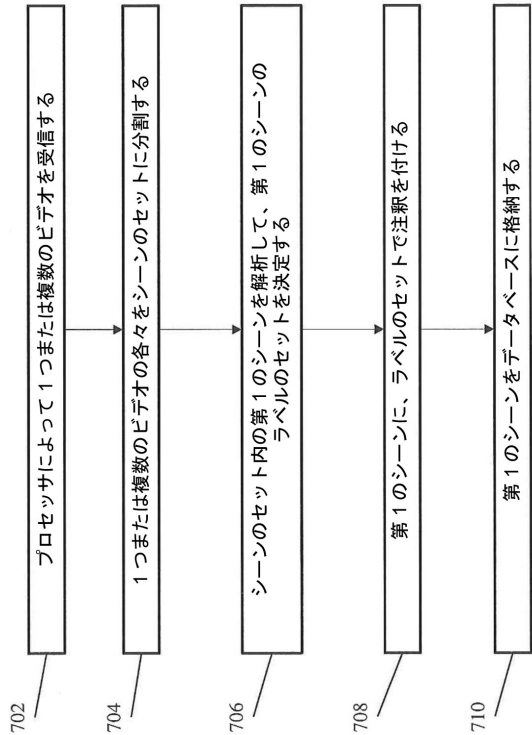


【図 6】



【図 7】

700
↓



10

20

30

40

50

フロントページの続き

- (72)発明者 シャー、イン
中華人民共和国 11 100193 ベイジン 28号棟 1-3階 ドンベイワン ウェスト・ロード8 ジョングァンツン・ソフトウェア・パーク28
- (72)発明者 ツァイ、ユー リン
中華人民共和国 11 100193 ベイジン 28号棟 1-3階 ドンベイワン ウェスト・ロード8 ジョングァンツン・ソフトウェア・パーク28
- (72)発明者 ボス、グレゴリー
アメリカ合衆国 48642-5885 ミシガン州ミッドランド リッジウッド・ドライブ1717
- 審査官 長谷川 素直
- (56)参考文献 米国特許出願公開第2011/0047163(US, A1)
特開2002-335473(JP, A)
特開2004-029906(JP, A)
特開2011-049707(JP, A)
特開2005-080000(JP, A)
特開2004-297548(JP, A)
特開2003-274320(JP, A)
特開2009-272816(JP, A)
米国特許出願公開第2015/0318020(US, A1)
- (58)調査した分野 (Int.Cl., DB名)
H04N 21/00-21/858
G06F 16/70