

(19)日本国特許庁(JP)

**(12)特許公報(B2)**

(11)特許番号  
**特許第7329430号**  
**(P7329430)**

(45)発行日 令和5年8月18日(2023.8.18)

(24)登録日 令和5年8月9日(2023.8.9)

(51)国際特許分類	F I		
G 0 6 F	16/53 (2019.01)	G 0 6 F	16/53
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00
G 0 6 N	20/00 (2019.01)	G 0 6 N	20/00
G 0 6 F	16/33 (2019.01)	G 0 6 F	16/33
G 0 6 F	16/73 (2019.01)	G 0 6 F	16/73

請求項の数 20 (全17頁)

(21)出願番号	特願2019-222699(P2019-222699)
(22)出願日	令和1年12月10日(2019.12.10)
(65)公開番号	特開2020-109631(P2020-109631)
	A)
(43)公開日	令和2年7月16日(2020.7.16)
審査請求日	令和4年12月12日(2022.12.12)
(31)優先権主張番号	16/235,743
(32)優先日	平成30年12月28日(2018.12.28)
(33)優先権主張国・地域又は機関	米国(US)
早期審査対象出願	

(73)特許権者	504407000 パロ アルト リサーチ センター インコ ーポレイテッド アメリカ合衆国 カリフォルニア州 94 304 パロ アルト カイオーテ ヒル ロード 3333
(74)代理人	100094569 弁理士 田中 伸一郎
(74)代理人	100109070 弁理士 須田 洋之
(74)代理人	100067013 弁理士 大塚 文昭
(74)代理人	100086771 弁理士 西島 孝喜
(74)代理人	100109335

最終頁に続く

(54)【発明の名称】 深層ニューラルネットワークのアンサンブルを使用した迅速なビデオクエリ

**(57)【特許請求の範囲】****【請求項1】**

方法であつて、

ユーザによって指定されたユーザオブジェクトを受信することと、

前記ユーザオブジェクトと、データベース内の1つ以上の候補オブジェクトとの間の類似性関数を使用し、それぞれの特徴ベクトルに基づいて類似性スコアを計算することと、

前記それぞれの計算された類似性スコアに基づいて、前記1つ以上の候補オブジェクトの第1のサブセットを前記ユーザに提示することと、

前記ユーザから、前記候補オブジェクトの第1のサブセットに関する第1のフィードバックを受信することと、

前記受信した第1のフィードバックに基づいて前記類似性関数を調整することと、

前記調整された類似性関数に基づいて、候補オブジェクトの第2のサブセットを前記ユーザに提示することと、

前記ユーザから、前記候補オブジェクトの第2のサブセットに関する第2のフィードバックを受信することと、を含み、

各候補オブジェクトについて、前記類似性スコアが、現在の閾値未満であるかどうかを判定することであつて、前記現在の閾値が、前記類似性関数に基づいている、判定することと、前記類似性スコアが所定の限度以上であるかどうかを判定することと、前記類似性スコアが前記現在の閾値未満であり、かつ前記所定の限度超であると判定さ

れた場合、前記候補オブジェクトの類似性に関するユーザからの入力を要求することと、を更に含む、方法。

**【請求項 2】**

前記受信された第 2 のフィードバックに基づいて、前記調整された類似性関数を調整すること、を更に含む、請求項 1 に記載の方法。

**【請求項 3】**

前記受信された第 2 のフィードバックに基づいて、前記調整された類似性関数を調整することを反復的に繰り返すプロセスを実行し、前記プロセスに関連する所定の事象が生じたときに、前記プロセスを終了すること、を更に含む、請求項 2 に記載の方法。

**【請求項 4】**

前記所定の事象は、前記プロセスの完了の通知が前記ユーザから受信されることと、を含む、請求項 3 に記載の方法。

**【請求項 5】**

前記ユーザオブジェクトが、ビデオクリップ、画像、及びテキストオブジェクトのうちの 1 つ以上である、請求項 1 に記載の方法。

**【請求項 6】**

前記類似性スコアを計算することが、前記それぞれの特徴ベクトルに基づいて、アンサンブル平均を計算すること、を更に含む、請求項 1 に記載の方法。

**【請求項 7】**

前記特徴ベクトルが、複数の三つ組を含む、請求項 1 に記載の方法。

**【請求項 8】**

前記類似性スコアを計算することが、前記 1 つ以上の候補オブジェクト及び前記ユーザオブジェクトのそれぞれの特徴ベクトルのドット積を判定することを含む、請求項 1 に記載の方法。

**【請求項 9】**

前記類似性スコアが所定の閾値以上であるかどうかを判定することを更に含み、前記類似性スコアが、前記所定の閾値以上であると判定された場合、前記複数の候補オブジェクトをクラスタ化する、請求項 1 に記載の方法。

**【請求項 10】**

システムであって、

プロセッサと、

コンピュータプログラム命令を記憶するメモリと、を備え、前記コンピュータプログラム命令は、前記プロセッサによって実行されると、前記プロセッサに、

ユーザによって指定されたユーザオブジェクトを受信することと、

前記ユーザオブジェクトと、データベース内の 1 つ以上の候補オブジェクトとの間の類似性関数を使用し、それぞれの特徴ベクトルに基づいて類似性スコアを計算することと、

前記それぞれの計算された類似性スコアに基づいて、前記 1 つ以上の候補オブジェクトの第 1 のサブセットを前記ユーザに提示することと、

前記ユーザから、前記候補オブジェクトの第 1 のサブセットに関する第 1 のフィードバックを受信することと、

前記受信した第 1 のフィードバックに基づいて前記類似性関数を調整することと、前記調整された類似性関数に基づいて、候補オブジェクトの第 2 のサブセットを前記ユーザに提示することと、を含み、

各候補オブジェクトについて、

前記類似性スコアが、現在の閾値未満であるかどうかを判定することであって、前記現在の閾値が、前記類似性関数に基づいている、判定することと、

前記類似性スコアが所定の限度以上であるかどうかを判定することと、

前記類似性スコアが前記現在の閾値未満であり、かつ前記所定の限度超であると判定さ

10

20

30

40

50

れた場合、前記候補オブジェクトの類似性に関するユーザからの入力を要求することと、  
を更に含む、動作を実行させる、システム。

**【請求項 1 1】**

前記プロセッサが、

前記受信された第 2 のフィードバックに基づいて、前記調整された類似性関数を調整すること、を行うように、更に構成されている、請求項 1 0 に記載のシステム。

**【請求項 1 2】**

前記プロセッサが、前記受信された第 2 のフィードバックに基づいて、前記調整された類似性関数を調整することを反復的に繰り返すプロセスを実行し、前記プロセスに関連する所定の事象が生じたときに、前記プロセスを終了するように更に構成されている、請求項 1 1 に記載のシステム。 10

**【請求項 1 3】**

前記ユーザオブジェクトが、ビデオクリップ、画像、及びテキストオブジェクトのうちの 1 つ以上である、請求項 1 0 に記載のシステム。

**【請求項 1 4】**

前記プロセッサが、前記 1 つ以上の候補オブジェクトの特徴ベクトルを計算するように構成されている、請求項 1 0 に記載のシステム。

**【請求項 1 5】**

前記プロセッサが、前記 1 つ以上の候補オブジェクト及び前記ユーザオブジェクトのそれぞれの特徴ベクトルのドット積を判定することによって、前記類似性スコアを計算するように構成されている、請求項 1 0 に記載のシステム。 20

**【請求項 1 6】**

前記プロセッサが、

前記類似性スコアが所定の閾値以上であるかどうかを判定することと、

前記類似性スコアが前記所定の閾値以上であると判定される場合、前記複数の候補オブジェクトをクラスタ化することと、を行うように更に構成されている、請求項 1 0 に記載のシステム。

**【請求項 1 7】**

前記プロセッサが、前記それぞれの特徴ベクトルに基づいてアンサンブル平均を計算することと、かつ前記アンサンブル平均に基づいて前記類似性スコアを計算することと、を行うように更に構成されている、請求項 1 0 に記載のシステム。 30

**【請求項 1 8】**

コンピュータプログラム命令を記憶する非一時的コンピュータ可読媒体であって、前記コンピュータプログラム命令は、プロセッサによって実行されると、前記プロセッサに、ユーザからユーザオブジェクトを受信することと、

前記ユーザオブジェクトと、データベース内の 1 つ以上の候補オブジェクトとの間の類似性関数を使用し、それぞれの特徴ベクトルに基づいて類似性スコアを計算することと、前記それぞれの計算された類似性スコアに基づいて、前記 1 つ以上の候補オブジェクトの第 1 のサブセットを前記ユーザに提示することと、

前記ユーザから、前記候補オブジェクトの第 1 のサブセットに関する第 1 のフィードバックを受信することと、 40

前記受信した第 1 のフィードバックに基づいて前記類似性関数を調整することと、

前記調整された類似性関数に基づいて、候補オブジェクトの第 2 のサブセットを前記ユーザに提示することと、

前記ユーザから、前記候補オブジェクトの第 2 のサブセットに関する第 2 のフィードバックを受信することと、を含み、

各候補オブジェクトについて、

前記類似性スコアが、現在の閾値未満であるかどうかを判定することであって、前記現在の閾値が、前記類似性関数に基づいている、判定することと、

前記類似性スコアが所定の限度以上であるかどうかを判定することと、

10

20

30

40

50

前記類似性スコアが前記現在の閾値未満であり、かつ前記所定の限度超であると判定された場合、前記候補オブジェクトの類似性に関するユーザからの入力を要求することと、を更に含む、動作を実行させる、非一時的コンピュータ可読媒体。

**【請求項 19】**

前記コンピュータプログラム命令は、プロセッサによって実行されると、前記プロセッサに、

前記受信された第2のフィードバックに基づいて、前記調整された類似性関数を調整すること、を含む、動作を更に実行させる、請求項18に記載の非一時的コンピュータ可読媒体。

**【請求項 20】**

前記コンピュータプログラム命令は、プロセッサによって実行されると、前記プロセッサに、前記受信された第2のフィードバックに基づいて、前記調整された類似性関数を調整することを反復的に繰り返すプロセスを実行し、前記プロセスに関連する所定の事象が生じたときに、前記プロセスを終了することを含む動作を更に行わせる、請求項19に記載の非一時的コンピュータ可読媒体。

**【発明の詳細な説明】**

**【技術分野】**

**【0001】**

本開示は、ビデオ処理、特にビデオデータセットクエリを対象とする。

**【0002】**

何らかの方法で類似しているオブジェクトのデータセットを検索することは、多くの用途において有用であり得る。検索は、データセット内の全てのオブジェクトを事前に分類し、次に、関心のあるクラスを検索することによって実行できる。しかしながら、この手法は、クラスを事前に知っておく必要があり、データ及び訓練モデルをラベル付けする際に多大な事前投資を含む。

**【0003】**

本明細書に記載される実施形態は、ユーザによって指定されたユーザオブジェクトを受信することを含む方法を伴う。類似性スコアは、それぞれの特徴ベクトルに基づいて、ユーザオブジェクトとデータベース内の1つ以上の候補オブジェクトとの間の類似性関数を使用して計算される。1つ以上の候補オブジェクトの第1のサブセットは、それぞれの計算された類似性スコアに基づいてユーザに提示される。第1のフィードバックは、候補オブジェクトの第1のサブセットについてユーザから受信される。類似性関数は、受信された第1のフィードバックに基づいて調整される。

**【0004】**

実施形態は、プロセッサと、コンピュータプログラム命令を記憶するメモリとを備えるシステムを伴い、コンピュータプログラム命令は、プロセッサによって実行されると、ユーザによって指定されたユーザオブジェクトを受信することを含む動作をプロセッサに行わせる。類似性スコアは、それぞれの特徴ベクトルに基づいて、ユーザオブジェクトとデータベース内の1つ以上の候補オブジェクトとの間の類似性関数を使用して計算される。

1つ以上の候補オブジェクトの第1のサブセットは、それぞれの計算された類似性スコアに基づいてユーザに提示される。第1のフィードバックは、候補オブジェクトの第1のサブセットについてユーザから受信される。類似性関数は、受信された第1のフィードバックに基づいて調整される。

**【0005】**

実施形態は、コンピュータプログラム命令を記憶する非一時的コンピュータ可読媒体を伴い、コンピュータプログラム命令は、プロセッサによって実行されると、ユーザからユーザオブジェクトを受信することを含む動作をプロセッサに行わせる。類似性スコアは、それぞれの特徴ベクトルに基づいて、ユーザオブジェクトとデータベース内の1つ以上の候補オブジェクトとの間の類似性関数を使用して計算される。1つ以上の候補オブジェクトの第1のサブセットは、それぞれの計算された類似性スコアに基づいてユーザに提示さ

10

20

30

40

50

れる。第1のフィードバックは、候補オブジェクトの第1のサブセットについてユーザから受信される。類似性関数は、受信された第1のフィードバックに基づいて調整される。

【図面の簡単な説明】

【0006】

【図1】本明細書に記載される様々な実施形態による、迅速なオブジェクト検索のためのプロセスを示す。

【図2】本明細書に記載される様々な実施形態による、迅速なオブジェクト検索のための別のプロセスを示す。

【図3】本明細書に記載される実施形態による、迅速なオブジェクト検索のためのより詳細なプロセスを示す。

【図4】本明細書に記載の実施形態による、例示的な分析パイプラインを示す。

【図5 A】本明細書に記載される実施形態による、迅速なビデオ検索を行うための例示的なグラフィカルユーザインターフェースを示す。

【図5 B】本明細書に記載される様々な実施形態によるフィードバックのためにユーザに提示される例示的な検索結果を示す。

【図5 C】本明細書に記載される様々な実施形態によるフィードバックのためにユーザに提示される例示的な検索結果を示す。

【図5 D】本明細書に記載される様々な実施形態によるフィードバックのためにユーザに提示される例示的な検索結果を示す。

【図6】本明細書に記載される実施形態を実装することが可能であるシステムのブロック図を示す。

【発明を実施するための形態】

【0007】

本明細書に記載される実施形態を使用すると、ユーザは、訓練データの従来の人間のラベル付け及び従来のコンピュータビジョンベースの教師付き機械学習に必要な費用と時間を伴わずに、提供された1つ以上の例に類似するオブジェクトに対してデータセットの迅速で、共同的な及び／又は進化するクエリを行うことができる。代わりに、予め訓練済みの深層ニューラルネットワークを使用して、データセット内のオブジェクトの深層ニューラルネットワーク表現が計算され、次に、これらの表現に対してクエリが行われる。この説明において、「予め訓練済みの深層ニューラルネットワーク」とは、手元のタスク以外の何かのタスクで以前に訓練された深層ニューラルネットワークである。深層学習の最新の実践では、1つのコンテキストで訓練された深層ニューラルネットワークの表現が、様々なコンテキストで役立つことが多いことが示されており、本明細書に記載の実施形態は、深層ニューラルネットワークのこの特性を採用する。例示的な実施形態として、本発明は、UCF101 Action Recognition Data Setで訓練された深層ニューラルネットワークを使用して、UCF101データセット内の101クラスの1つではない動きに対するクエリを行うことができる。

【0008】

本明細書に記載される実施形態は、類似性関数を使用して、ビデオクリップ、画像、及び／又はテキストなどのユーザオブジェクトの深層ニューラルネットワーク表現を記憶し検索する構造化された方法を伴う。構造化されていないデータをこの構造化表現に変換することにより、より速く、より強力で、かつより自動化された検索が可能になる。本明細書に記載される実施形態は、予め訓練済みの深層学習ニューラルネットワークを使用して、オブジェクトの署名を計算することと、構造化データベースに署名を供給することとを含む。次に、署名を使用して、ユーザオブジェクトとの類似性スコアを計算することができる。これにより、ユーザは、類似性関数に基づいて、提供された例と同様のオブジェクトに対するデータベースクエリを行うことができる。

【0009】

本明細書に記載される実施形態は、限られた一組の例、多くの場合一例のみを事前に必要とするだけである。クエリ結果を再検討するためにユーザが費やす努力は、ユーザが特

10

20

30

40

50

定の検索概念を反復し、焦点を当て、有効化する際にのみ更に必要とされる。ユーザは、本明細書に記載される実施形態を使用して、アルゴリズムの反復中に学習された新しい洞察に基づいて検索努力に適応し、再指示することができる。本明細書に記載される実施形態は、動的かつ対話式の方法で関連するデータを発見し、焦点を合わせるために、大きなデータセットの迅速な閲覧を可能にする。

#### 【0010】

本明細書に記載される実施形態は、教師なしクラスタ化と迅速なビデオクエリを組み合わせて、現在の手法と比較して著しく少ないユーザの関与でラベル付き及びラベルなしデータから学習することができるようになる。例えば、輸送専門家、例えば、交通技術者は、限定された一組のラベル付けされた例を使用する車両追従モデルを作成し、検証することができる。

10

#### 【0011】

ユーザは、動きの所定のリストに限定されることなく、検索を実行する際に動きの1つ又は少数の参照例を提供することができる。本明細書に記載される実施形態の有益な態様は、深層ニューラルネットワークモデルが訓練されるときに、指定されない動きに対して、それらが機能することである。したがって、本明細書に記載する実施形態は、新しい動きが調査されるたびに、未処理の非構造化データ自体に対して標準的なコンピュータビジョンベースの機械学習を行う手法を回避する。したがって、大規模なビデオデータセットを用いた研究に必要なビデオの手動表示の量を大幅に低減でき、より低コストでより広範かつ詳細な研究が可能になる。

20

#### 【0012】

図1は、本明細書に記載される様々な実施形態による、迅速なオブジェクト検索のためのプロセスを示す。検索を行う前に、データセット内の複数のオブジェクトの各オブジェクトについて、深層学習特徴ベクトルが計算され、これらの特徴ベクトルは、構造化データベース内に記憶される。この文脈では、深層学習特徴ベクトルは、深層ニューラルネットワークにおける層から出力される数のベクトルである。特定の実施形態では、深層学習特徴ベクトルは、複数の三つ組を含み、三つ組は、3つの異なるデータスプリットで訓練された、深層ニューラルネットワークの3つのバージョンを表す。各予め訓練済みの深層ニューラルネットワークに関して、埋込み型特徴の計算は、例えば、新たな検索が新たなる組の動きに対して行われるたびに再計算するのではなく、データセットに対して1回行われ、データベースに記憶されてもよい。本明細書では、所与のオブジェクトについて計算された一組の特徴ベクトルは、オブジェクトの署名と呼ばれることがある。

30

#### 【0013】

いくつかの深層学習モデルは、二つ以上のストリームを含み、各ストリームは、特定のタイプのニューラルネットワークを含む。ストリームは、特定の方法で処理されたビデオデータを使用する深層ニューラルネットワークモデルである。例えば、ビデオから抽出されたRGBフレームを使用して、RGBストリームを訓練することができ、ビデオの光学フロー計算を使用して、光学フローストリームを訓練することができ、屈折した光学フローを使用して、代替の形態の光学フローストリームを訓練することができる。一般に、光学フロー計算だけでなく、ビデオの任意の前処理を使用して、別個のストリームの入力を生成することができ、他の例としては、例えば、既知の又は予想されるノイズのフィルタリング、バックグラウンドの減算、スペクトル分析、平均シフト、カムシフト、及び/又は光学フロー変異を含む。

40

#### 【0014】

いくつかの実施形態では、1つ以上の深層ニューラルネットワークストリームタイプからの表現を使用することができる。予め訓練済みの深層ニューラルネットワークストリームは、共に訓練される必要がない。いくつかの実施形態では、予め訓練済みの深層ニューラルネットワークの一部を抽出し、使用することができる。場合によっては、抽出部分は、それがより大きなモデルの一部として訓練されていても、なお、予め訓練済みのニューラルネットワークと称される。

50

**【 0 0 1 5 】**

図 1 では、ユーザオブジェクトが、ユーザインターフェースを使用してユーザから受信される(110)。ユーザオブジェクトのために、特徴ベクトルを計算する(120)。類似性スコアが、それぞれの特徴ベクトルに基づいて、ユーザオブジェクトとデータベース内の 1 つ以上の候補オブジェクトとの間で計算される(130)。様々な実施形態によれば、類似性スコアを計算することは、候補及びユーザ提供の例示的なオブジェクトのそれぞれの特徴ベクトルのドット積を判定することと、後者の適切なノルムの二乗によって、周知のユークリッドノルムである例を正規化することとを含む。候補オブジェクトのサブセットの第 1 のセットが、それぞれの計算された類似性スコアに基づいて、ユーザに提示される(140)。候補オブジェクトの第 1 のセットに関する第 1 のフィードバックが、ユーザから受信される(150)。類似性スコアを計算するために使用される関数は、受信された第 1 のフィードバックに基づいて調整される(160)。

10

**【 0 0 1 6 】**

図 1 に記載されるプロセスは、輸送業界で使用されるように構成されてもよい。この例では、ユーザが提供するビデオは、大型トラックに追従する車両の前向きのビデオクリップであってもよい。ユーザは、一致する可能性のある少数のサンプルを受信し、これらの候補を検査した後、ユーザは結果を採点することでフィードバックを提供する。ユーザは、絞り込み検索を要求することができる。絞り込み検索は、以前のラウンド内の採点を反映する候補オブジェクトの新しいサンプルを返す。ユーザの要求を満たすのに十分な反復が行われた後、最終的なオブジェクト検索基準を使用して、関心のあるデータセット内の全ての一致を検索する。ユーザは、十分な反復が完了したときを判定することができる。場合によっては、繰り返しの数は所定の値である。上記の例に統いて、ユーザは、反復的で迅速なクエリが行われている最中に、検索概念を修正することができる。例えば、最終的に返される一致は、大型トラックに追従する車両が複数車線の道路上にあり、交通がストップアンドゴーのスピードよりも速い速度で移動している場合のビデオクリップであり得る。

20

**【 0 0 1 7 】**

図 2 は、本明細書に記載される様々な実施形態による、迅速なオブジェクト検索のための別のプロセスを示す。検索を行う前に、データセット内の複数のオブジェクトの各オブジェクトについて、深層学習特徴ベクトルが計算され、これらの特徴ベクトルは、構造化データベース内に記憶される。ユーザオブジェクトがユーザから受信される(210)。220 では、ユーザオブジェクトに必要な特徴ベクトルが、予め計算された特徴ベクトルを含む構造化データベースから得られる。類似性スコアは、それぞれの特徴ベクトルに基づいて、ユーザオブジェクトとデータベース内の 1 つ以上の候補オブジェクトとの間で計算される(230)。候補オブジェクトのサブセットの第 1 のセットは、それぞれの計算された類似性スコアに基づいて、ユーザに提示される(240)。

30

**【 0 0 1 8 】**

類似性関数が十分に機能しているかどうかが判定される(260)。これは、所定の事象が起きていることに基づいてもよい。例えば、ユーザは、240 で提示されたオブジェクトの十分に大きな割合が、ユーザオブジェクトに十分に類似していることに満足しているため、類似性関数の反復改善の終了を要求する場合がある。類似性関数が更なる改善を必要とすると判定された場合、ユーザは、ユーザに提示された候補オブジェクトの最新のセットにフィードバックを与える(265)。類似性関数は、受信されたフィードバックに基づいて調整され、類似性スコアが再計算される。候補オブジェクトの新しいセットが、ユーザに提示される(240)。場合によっては、候補オブジェクトの新しいセットは、候補オブジェクトの第 1 のセットと重複する。様々な実装態様によれば、候補オブジェクトの新しいセットは、候補オブジェクトの第 1 のセットと実質的に同じであり、システムは、収束が生じたと判定する。例えば、十分に類似しているとユーザによって検証された候補オブジェクトの新しいセットの割合が、以前に選択された候補オブジェクトのセットに対する割合と同じ又はほぼ同じである場合、システムは、収束が生じたと判定するこ

40

50

とができる。収束が生じたと判定される場合、類似性関数が満足に機能していると判定されてもよい(260)。類似性関数が満足に機能していないと判定された場合(260)、プロセスは継続する。類似性関数が満足に機能していると判定されるとき(260)、システムは、260からの満足な類似性関数を使用して、データセット内の全ての一致するオブジェクトを見つけ出す(270)。一致するオブジェクトは、記憶されてもよく及び／又はユーザに提示されてもよい。ユーザによって完了される反復処理は、データセット内のオブジェクトのサブセットのみを使用してもよい。様々な実施形態によれば、システムは、データセット全体の中の全ての一致するオブジェクトを見つけ出す。場合によつては、システムは、データセットのサブセット内の全ての一致するオブジェクトを見つけ出す。

10

#### 【0019】

本明細書に記載される様々な実施形態によれば、オブジェクトは、ユーザによって検討され得る任意のデータ要素であってもよい。例えば、オブジェクトは、ビデオクリップ、画像、及びテキストオブジェクトのうちの1つ以上を含んでもよい。オブジェクトは、いくつかの実施形態では、データ要素の組み合わせを含んでもよい。例えば、オブジェクトは、ビデオクリップ及び関連付けられたテキストオブジェクト、並びに／又は関連付けられたテキストオブジェクトを有する画像を含んでもよい。本明細書に記載される実施形態は、迅速なビデオ検索と併せて説明される場合があるが、ユーザオブジェクトは、上述のようにユーザによって検討され得る任意のデータ要素であり得ることを理解されたい。図3は、本明細書に記載される実施形態による、迅速なビデオ検索のための詳細なプロセスを示す。ビデオデータセット内のビデオの少なくともサブセットに関して、ビデオデータは、検索の開始前に処理される。ビデオデータの処理は、サブセット内の各ビデオをクリップに分割する(305)ことを含む。データセット内のビデオデータの少なくとも一部は、ビデオクリップの形態であってもよい。したがって、ビデオデータをクリップに分割することは必要ない場合がある。クリップの各々についての深層学習埋込み型特徴ベクトルを計算する(310)。様々な実施形態によれば、深層学習埋込み型特徴ベクトルは、三つ組として計算される。一組の特徴ベクトルは、本明細書においてクリップ署名と称されることがある。様々な実施形態によれば、クリップ署名は、構造化データベース内に記憶される(315)。

20

#### 【0020】

検索中、ビデオデータセット内のビデオクリップの少なくともサブセットについて、アンサンブル平均化を使用して、ユーザ参照クリップに関する類似点を計算する(320)。ユーザ参照クリップは、上述のように、大型トラックに追従する車両の前向きのビデオクリップなどの、ユーザにとって関心のある動きを含むことができる。

30

#### 【0021】

ストリーム、スプリット、及び／又は特徴名にわたるアンサンブル平均を準備する第1のステップとして、方程式1を使用して個々の類似点を計算する。

#### 【0022】

#### 【数1】

$$S(V_{\text{候補}}, V_{ref} \mid \text{ストリーム, 訓練スプリット, 特徴名}) = \frac{V_{\text{候補}} \cdot V_{ref}}{\|V_{ref}\|_2^2} \dots \dots (I)$$

40

#### 【0023】

ここで、 $V_{\text{候補}}$ は、候補クリップ署名内の特定の特徴ベクトルを表し、 $V_{ref}$ は、参照クリップの署名における対応する特徴ベクトルを表し、ストリーム、訓練スプリット、及び特徴名は、式1が動作している特徴ベクトルを識別する。方程式1の分子は、候補及び参照に関する特徴ベクトルのドット積である。ドット積を理解している人には理解できるように、方程式1は、 $v_r$ の方向にある $v_c$ の部分を抽出し、次に、それが $v_{ref}$ とどの程度類似しているかを計算する。

50

**【 0 0 2 4 】**

いくつかの実施形態では、署名内の複数の訓練スプリットのための特徴ベクトルが存在し、これらの事例では、ストリームと特徴名の各対について、全ての訓練スプリットにわたる  $S(\cdot)$  のアンサンブル平均を計算することが好ましい場合がある。単一のストリーム及び单一の特徴名のみを有する実施形態では、 $= 1$  は完全な一致と考えられる。

**【 0 0 2 5 】**

複数のストリーム及び / 又は特徴名を有する実施形態では、アンサンブル平均  $R$  は、別のアンサンブル化を経験する。例えば、全てのストリームにわたる加重平均 ( $R$ ) 及び特徴名は、方程式 2 に示されるように計算されてもよい (325)。この方程式において、 $w$  は、類似性関数上の各反復において調整される、各 {ストリーム、特徴名} 対の重みを表す。

**【 0 0 2 6 】****【 数 2 】**

$$R = \frac{\sum w^2 * (1 - \sigma)^2}{\sum w^2} \quad (2)$$

**【 0 0 2 7 】**

明確にするために、方程式 2 の和は {ストリーム、特徴名} 対を上回り、任意の 1 対の及び  $w$  の値は、他の対の値とは異なり得る。方程式 3 に示されるように荷重平均を使用して、各候補クリップについて類似性スコアを計算する (330)。

**【 0 0 2 8 】****【 数 3 】**

$$\text{スコア} = 1 - \sqrt{R} \quad (3)$$

**【 0 0 2 9 】**

座標系での距離の計算に精通している人は、各 {ストリーム、特徴名} 「座標次元」がそれ自身の重み  $w$  を有する座標系において、方程式 3 が、全ての  $\sigma$  が 1 に等しい点からの「距離」を表すことを容易に理解するであろう。上述のように、各 {ストリーム、特徴名} 対について、 $\sigma = 1$  は理想値であり、全ての  $\sigma$  が 1 に等しいとき、方程式 2 からの  $R$  は 0 に等しく、したがって、方程式 3 におけるスコアは 1 に等しい。したがって、等式 2 及び 3 は、望ましい方法で挙動する可能性のあるアンサンブルを表す。方程式 3 における一組の  $w$  値は、各 {ストリーム、特徴名} 次元に対する重要性を定量化し、これらの値は、満足のいく類似性関数を学習することの一部として学習することができる。

**【 0 0 3 0 】**

ユーザフィードバックは、候補クリップの少なくともサブセット上で要求される。様々な実施形態によれば、類似性スコアに基づいてユーザフィードバックが要求される (331)。例えば、現在の閾値を上回る類似性スコアは、候補クリップが所望の概念を示すかどうかを検証又は無効化するために、ユーザフィードバック要求をトリガすることができる。場合によっては、現在の閾値よりも小さく、ある下限よりも大きい類似性スコアもまた、ユーザフィードバック要求をトリガし得る。これらの要求の目的は、アルゴリズムが「ニアミス」と考えられるクリップに関するフィードバックを取得することである。ユーザは、例えば、計算されたニアミスが、実際には所望の概念に一致するクリップであるというフィードバックを与えることができる。計算された一致及び計算されたニアミスの両方に関するフィードバックが有用であり得る。ユーザからフィードバックを獲得することは、ユーザの所望のクエリ結果のコンテキストで、計算された一致及びニアミスの有効性がどのように評価されるかである。

**【 0 0 3 1 】**

ユーザフィードバックは、候補クリップのサブセットの少なくとも一部で受信される (

10

20

30

40

50

335)。閾値は、類似性関数を調整することの一部として、受信したユーザフィードバックに基づいて調整され得る(340)。ユーザフィードバックを要求すること(331)、ユーザフィードバックを受信すること(335)、及び類似性関数を調節すること(340)は、重みw又は閾値又はその両方のいずれかを調整することによって、所定のイベントが発生するまで反復的に繰り返すことができる(350)。場合によっては、所定の事象は、プロセスを終わらせるユーザ入力を受信することを含む。所定の事象は、例えば、所定のサイクル数に到達したことでもよく、及びノ又は収束が生じたという判定であつてもよい。所定のイベントが生じたら、最終的な重み及びスコア閾値を使用して、ビデオデータセット内の全ての一一致クリップを見つけ出す。したがって、ユーザは、少数のクリップを検討してフィードバックするだけで、データセット全体のクエリ結果を獲得することができる。例えば、最終的なクエリ結果のクリップの>50%を、ユーザが、意図された検索概念に有効な一致であると考えるとき、クエリは適切に機能していると考えることができる。

#### 【0032】

様々な実施形態によれば、本明細書に記載されるプロセスの前又は後に候補クリップをフィルタるために、1つ以上の追加の基準を使用してもよい。例えば、時刻に関心がある場合、ビデオのタイムスタンプを使用して、時刻によるフィルタリングが行われ得る。場合によっては、候補クリップは、例えば、昼間対夜間ビデオ分類子、若しくは1つ以上の特定のオブジェクトの存在を探すオブジェクト分類子などの所定の分類子を使用してフィルタリングされる。このような分類子は、深層学習を含む機械学習の実践において既知である。様々な実施形態によれば、加速度の補助データが利用可能な場合、又は既知のコンピュータビジョン方法を使用して加速度が計算される場合、クリップ内の車両の加速度に従って候補クリップをフィルタリングすることができる。例えば、システムは、所定の加速度範囲内にある候補クリップのみを返してもよい。候補クリップは、2つのビデオストリーム間の相関に基づいてフィルタリングされてもよい。例えば、店舗の外側を観察する第1のカメラ、及び店舗内の顧客の動きを観察する第2カメラなど、2つの異なるカメラから取得された、ある動きの2つのタイムスタンプされたビデオストリームの場合を考えてみる。ユーザは、店舗の外側のどの動きが店舗内の望ましい又は望ましくない動きのいずれかと相關するかを学習するために、迅速なビデオクエリを行いたい場合がある。

#### 【0033】

様々な実施態様によれば、実際にユーザが2つ以上の別個の概念に関心があることをユーザフィードバックが示唆しているときに、候補クリップ検索を2つ以上の検索に分割することができる。実際にユーザが2つ以上の別個の概念に関心があることをユーザフィードバックが示唆しているかどうかを判定するために、様々な技術が使用されてもよい。例えば、KL発散、多重ガウス分析、及びノ又は階層的クラスタ化から観察された発散は、全て機械学習の実践における既知の分析方法であり、いつクエリを2つの別個のクエリに再編成する意味があるかを識別するために使用されてもよい。これらの技術は、候補クリップ上のユーザフィードバックのデータが、従属変数としての深層学習特徴ベクトルに関して、2つ以上のモードの徴候を示すかどうかを分析することができる。2つ以上のクエリモードの可能性が十分に高い場合、ユーザは、クエリを2つ以上のクエリに再編成するように促される場合がある。他の実施形態では、再編成は、ある所定の閾値を超える可能性によってトリガされるとき自動的に行われる。

#### 【0034】

クエリの最終結果が配信された後、場合によっては、一致クリップの署名を分析することによって結果を更に分析することが有用であり得る。例えば、このような分析は、一致クリップをグループ化するのに役立ち得るか、又は異常を識別するのに役立ち得る。実施形態では、最終結果は、機械学習及びデータ科学の技術分野において既知の技術を採用して、特徴空間内にクラスタ化することによって、迅速なクエリの終わりに返されたクリップの署名を使用してクラスタ化されてもよい。クラスタ化は、全ての一一致クリップの埋込み型特徴に対して行うことができる。場合によっては、ユーザは、クラスタを意味的にラ

ベル付けして、データのパターン及び区別を更に詳しく調べることができる。

【 0 0 3 5 】

本明細書に記載される実施形態は、運転者の行動のモダ化をサポートするために、ビデオ分析のための迅速な人と機械の共同手法の一部として使用することができる。例示的な分析パイプラインを図4に示す。ビデオデータセット内のビデオクリップの少なくとも一部について、深層学習埋込み型特徴ベクトルを計算する(410)。上記の車両追従の例では、ユーザ、例えば、輸送専門家は、関心のあるシナリオを描写する前方視界のビデオを提供することによって開始してもよい。この実施例は、420内の対話型検索ループをトリガし、これは、同様の前方視界クリップのサンプルを返す。専門家は、検索を絞り込むためのフィードバックを提供する。いったん、前方視界ビデオのサンプルが、専門家が念頭に置いているものと一致すると、関心のあるデータセットからの全ての一一致が見つかる。一致のタイムスタンプを使用して、第二次戦略的ハイウェイ研究プログラム(Strategic Highway Research Program 2、SHRP2)自然な運転研究(Naturalistic Driving Study、NDS)などの自然な運転研究で取得されたデータなど、関連する表形式データのクエリだけでなく、対応する車内ビデオの深層学習埋込み型特徴を使用するクエリも行うことができる。これらのクエリの結果を430に融合させることができる。このデータ融合ステップの目標は、高レベルの行動の理解を高めることである。複数のビデオビューからの埋込み型特徴に加えて、既存のコンピュータビジョンアルゴリズムを使用して容易に判定できる低レベルのビデオ行動クラスを計算し、データ融合に組み込むこともできる。

10

【 0 0 3 6 】

次に、専門家は、必要に応じて、専門家が重要であると考える高レベル拳動のサブセットを発見及びグループ化するために、対話型クラスタ化ループ（440）を行う。システムによって提示されたグループが有用ではない場合、専門家は、そのパラメータ（例えば、クラスタの数）を調整するか、又は、どのビデオを含めるか、及び／又はどのビデオを除外するかについての修正されたアイデアを用いて初期ビデオ検索に戻るかのいずれかによつて、クラスタ化アルゴリズムを再指示する。専門家が有用なカテゴリを見つけ、それらを更に分解するための特定の示唆を有する場合、専門家は、それらを半教師付きクラスタ化ループ（455）に提示して、グループ化を絞り込むことができる。半教師付き学習に関して、専門家は、関心のあるクラスタ内のサンプルにラベル付けする。クラスタ化ステップが収束すると、専門家は、各クラスタに意味的説明を提供する（450）。次に、その説明は、それぞれのクラスタに関連付けられ、記憶される。

20

【 0 0 3 7 】

例示的な実施形態として、教師なし学習（440）は、大型トラックに追従するためのデータ融合結果をクラスタ化することができ、かつ440で学習されるクラスタの1つは、頭の近くに物体を保持し、大きな間隔を保ちながら大型トラックに追従するためのものであり得る。専門家は、いくつかの例が、頭の近くに保持されている電話であり、他の例が、食べている食品であることに気付き得る。専門家は、クラスタは、電話を保持している1つのクラスタと、物を食べている別のクラスタに分解される必要がある、と判断することができる。専門家は、両方の少数の例にラベル付けし、次に、そのクラスタを、半教師付き学習のために提出することができる。他の様々な実施形態では、専門家は、より広範なクラスの行動モデルを開発することを決定することができ、したがって、440において学習されたクラスタを更に分解するための半教師付き学習は行わない。教師なしクラスタ化（440）及びその後の意味的ラベル付け（450）は、更なる半教師付きクラスタ化（455）の有無にかかわらず、例えば、運転者行動の行動モデルを開発する（460）ために使用され得る。

30

〔 0 0 3 8 〕

クラスタ化ステップは、運動学的データ、レーダーデータ、車両力学モデル、道路状況データ、及び／又は既知のコンピュータビジョン方法を使用して計算された低レベル行動などの、専門家が指定するビデオ以外のソースからの予測特徴を活用することができる。

50

場合によっては、クラスタは、ユーザによる少ないラベル付けのみを必要とする、半教師付き学習を使用して、更に分解され得る。最終的なグループ化は、専門家にとって関心のある行動クラスをモデル化する。上記の例によれば、教師なし学習が発見してモデルとする1つの行動は、電話で話しながら、トラックから大きく離れたままでトラックに追従する場合であり、一方、別の行動は、ブレーキペダルを何度も踏んでより近くで追従する場合である。

#### 【0039】

トラックに追従する車両の例では、主に頭と顔の動きに焦点を当てたものなど、異なる予め訓練済みの深層ネットを、外向きビデオに対して車内ビデオに使用できる。車両追従のプロセスは、衝突回避を考慮した運動及び車両力学の観点からモデル化される。特に、トラック追従と運転者の注意散漫が、運転者の車両追従行動に与える影響をモデル化できる。様々な条件の影響を確認することができる。例えば、運転者が二次的な作業に従事している間に、自動車に追従する自動車及び／又はトラックに追従する自動車の影響を、本明細書に記載する実施形態を使用して決定することができる。運転者の知覚・反応時間を計算することができる。次に、これらの条件を反映する運転者車両追従モデルを開発し、較正し、かつ微視的な交通シミュレーション環境に組み込んで、このような行動のネットワーク全体の影響を定量化し得る。

10

#### 【0040】

図5Aは、本明細書に記載される実施形態による、迅速なビデオ検索を行うための例示的なグラフィカルユーザインターフェースを示す。ユーザは、参照ビデオクリップが記憶されているビデオデータセットを選択する(510)。次に、ユーザは、参照ビデオを選択する(515)。次に、ユーザは、参照ビデオ又はビデオクリップにおいて関心のある動きが生じる特定の時間を選択する(520)。様々な構成によれば、システムは、選択された時間の後及び／又は前に所定の追加時間を加えた、参照ビデオにおける選択された時間を使用して、参照ユーザクリップを作成する。ユーザは、クエリ名を提供することができる(525)。次に、ユーザは、検索を行うビデオデータセットを選択することができる(530)。選択されたビデオデータセットは、参照ビデオが記憶されているデータセットと同じデータセット又は異なるデータセットであってもよい。次に、ユーザは、図1の140及び／又は図2の240のようなステップで検証するために返す、最大数の一一致を選択することができる(535)。任意選択的に、ユーザはまた、特定の検索のためのノートを提供することができる(540)。

20

#### 【0041】

図5B～図5Dは、結果を提示するために1つの可能なグラフィカルな方法を使用して、フィードバックのためにユーザに提示される例示的な検索結果を示す。図5Bは、第1の検索後の結果を示す。この例では、550内の円は不一致と考えられ、550の外側及び555内の円は一致と考えられる。次に、ユーザは、各個々の円をクリックして、対応するビデオクリップを見ることができる。次に、ユーザは、ビデオ検索システムが正しい結果を有しているかどうかを示すことができる。この例では、不一致と判定された両方のクリップが実際には一致したとユーザが判定したため、円は緑色で塗りつぶされている。ユーザ入力に基づいて、システムは新しい検索を行い、更新された結果をユーザに提示する。図5Cは、例示的な検索の第2のラウンドの返却結果を示す。この例では、示唆された一致の大部分は、実際には不一致であったと判定され、したがって、ユーザによって赤色で満たされている。この情報は、図5Dに示すように、検索の第3のラウンドを行うために使用される。この例では、混入された不一致であるとユーザが判定したのは2個だけで、11個の検証された一致があった。この時点で、ユーザは検索がかなり正確であると判定し、検索を終わらせるなどを決定することができる。図5B～図5Dは、特定の方法で示された結果を示しているが、結果は、ユーザに対して任意の合理的な方法で示され得ることを理解されたい。例えば、結果は、ビデオのリスト、又は予め定義されたスコア範囲のピンを有するヒストグラムの形態でユーザに示されてもよい。

30

#### 【0042】

40

50

上記の方法は、周知のコンピュータプロセッサ、メモリユニット、記憶デバイス、コンピュータソフトウェア及び他の構成要素を使用してコンピュータに実装することができる。このようなコンピュータの高レベルブロック図が図6に示されている。コンピュータ600は、そのような動作を定義するコンピュータプログラム命令を実行することによって、コンピュータ600の全体的な動作を制御するプロセッサ610を含む。プロセッサ610は、命令を実行することが可能である任意のタイプのデバイスを含み得ることを理解されたい。例えば、プロセッサ610は、中央処理装置(central processing unit、CPU)、グラフィック処理ユニット(graphical processing unit、GPU)、フィールドプログラマブルゲートアレイ(field-programmable gate array、FPGA)及び特定用途向け集積回路(application-specific integrated circuit、ASIC)のうちの1つ以上を含んでもよい。コンピュータプログラム命令は、記憶デバイス620(例えば、磁気ディスク)内に記憶され、コンピュータプログラム命令の実行が所望されるときにメモリ630にロードされてもよい。したがって、本明細書に記載される方法のステップは、メモリ630に記憶され、コンピュータプログラム命令を実行するプロセッサ610によって制御されるコンピュータプログラム命令によって定義され得る。様々な実装態様によれば、コンピュータは、社内サーバ又はクラウドベースのサービスの一部として、方法ステップを行うことができる。コンピュータ600は、ネットワークを介して他のデバイスと通信するための1つ以上のネットワークインターフェース650を含んでもよい。コンピュータ600はまた、コンピュータ600とのユーザ対話を可能にする他の入力/出力デバイス(例えば、ディスプレイ、キーボード、マウス、スピーカ、ボタンなど)660を含む。様々な実施形態によれば、図6は、例示目的のためのコンピュータの可能な構成要素の高レベル表現であり、コンピュータは他の構成要素を含んでもよい。

#### 【0043】

別途記載のない限り、本明細書及び特許請求の範囲で使用される特徴サイズ、量、及び物理的特性を表す全ての数は、全ての場合において、用語「約」によって改変されるものとして理解されるべきである。したがって、反対に指示されない限り、前述の明細書及び添付の特許請求の範囲に記載される数値パラメータは、本明細書に開示される教示を利用して当業者が得ようとする所望の特性に依存して変化し得る近似値である。端点による数値範囲の使用は、その範囲内の全ての数(例えば、1~5は、1, 1.5, 2, 2.75, 3, 3.80, 4及び5を含む)、及びその範囲内の任意の範囲を含む。

#### 【0044】

上記の様々な実施形態は、特定の結果を提供するために相互作用する回路及び/又はソフトウェアモジュールを使用して実装され得る。コンピューティング技術の当業者は、当技術分野で一般的に知られている知識を使用して、モジュールレベルで又は全体として、このような記載された機能を容易に実装することができる。例えば、本明細書に示されるフローチャートは、プロセッサによる実行のためのコンピュータ可読命令/コードを作成するために使用されてもよい。このような命令は、コンピュータ可読媒体に記憶され、当該技術分野において既知のように実行するためにプロセッサに転送されてもよい。上に示された構造及び手順は、上記のようなインクジェット排出器の診断を容易にするために使用することができる実施形態の代表的な例に過ぎない。

10

20

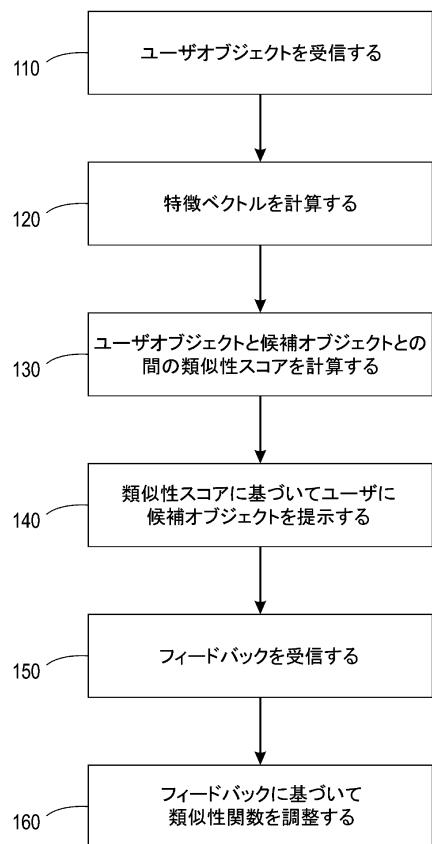
30

40

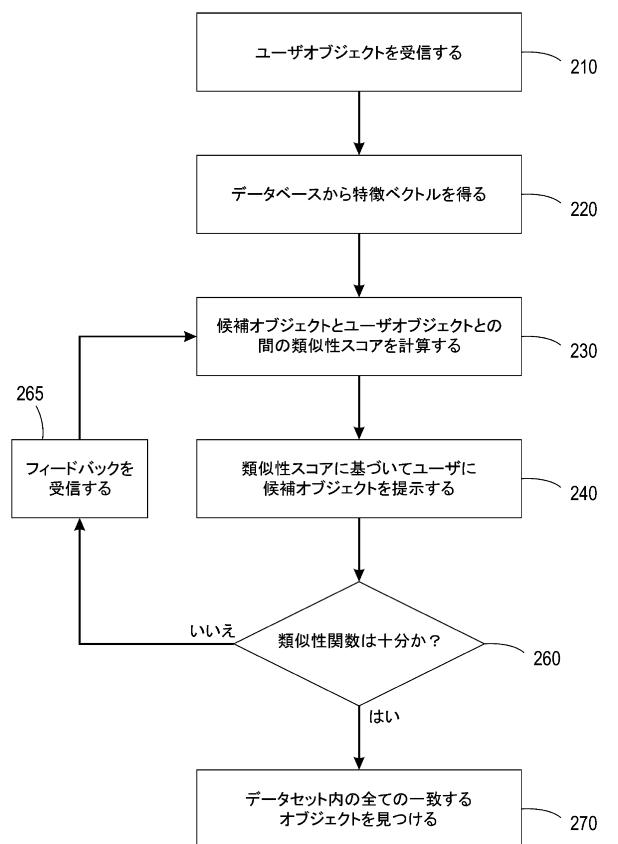
50

## 【図面】

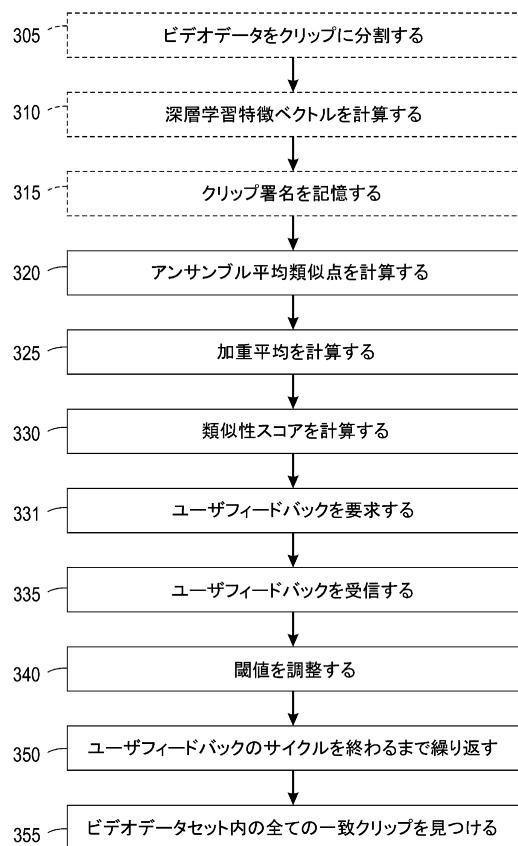
## 【図 1】



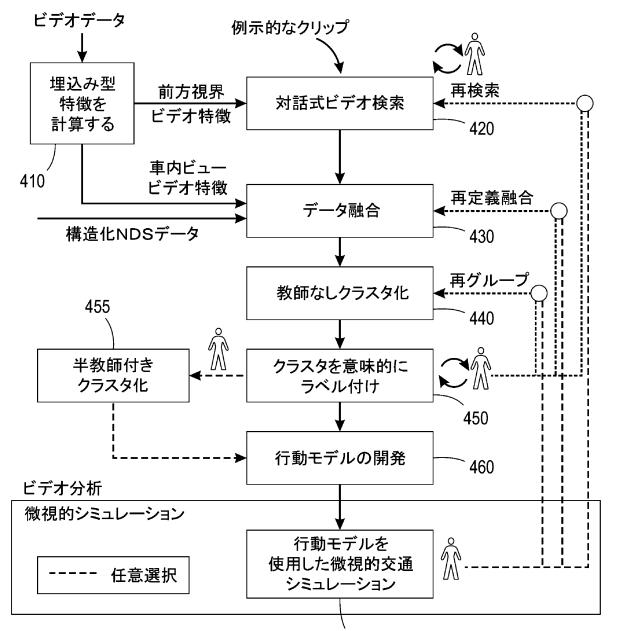
## 【図 2】



## 【図 3】



## 【図 4】



10

20

30

40

50

【図 5 A】

参照ビデオクリップ？

参照ビデオが存在するセットは？  
サンプルビデオ

参照ビデオ：  
DowntownBrooklynDrive\_480p

参照時間: 00:03:51  
0 時間 3 分 51 秒

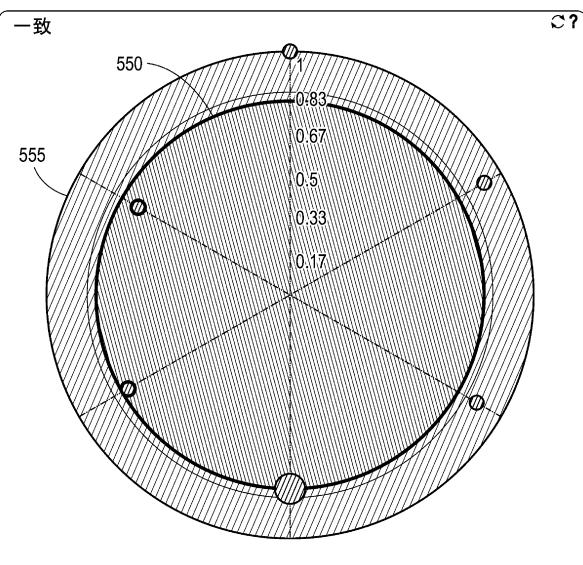
クエリ名：  
サンプルクエリ

クエリする検索のセット：？  
サンプルビデオ 最大一致：？  
24

メモ：？  
これは、ビデオクエリアプリケーションを使用する方法の例です。

提出

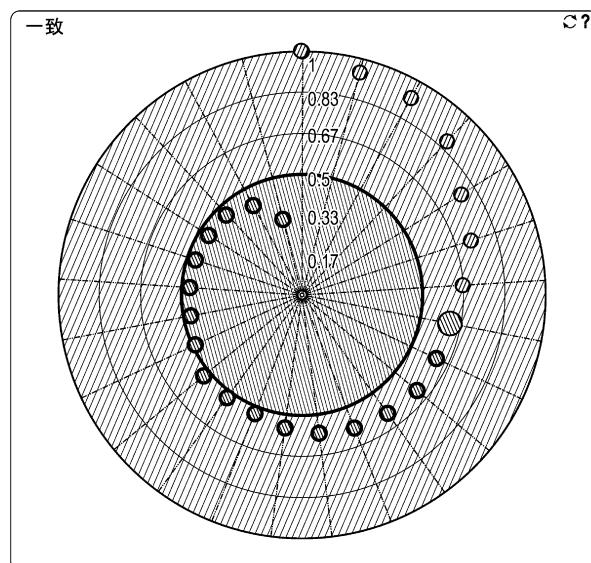
【図 5 B】



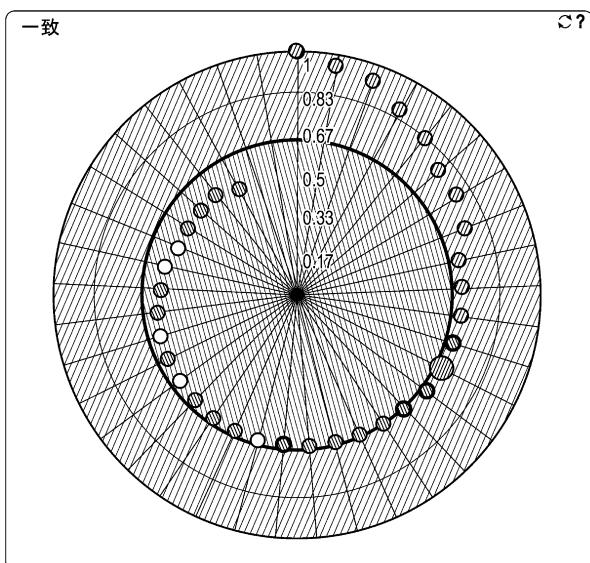
10

20

【図 5 C】



【図 5 D】

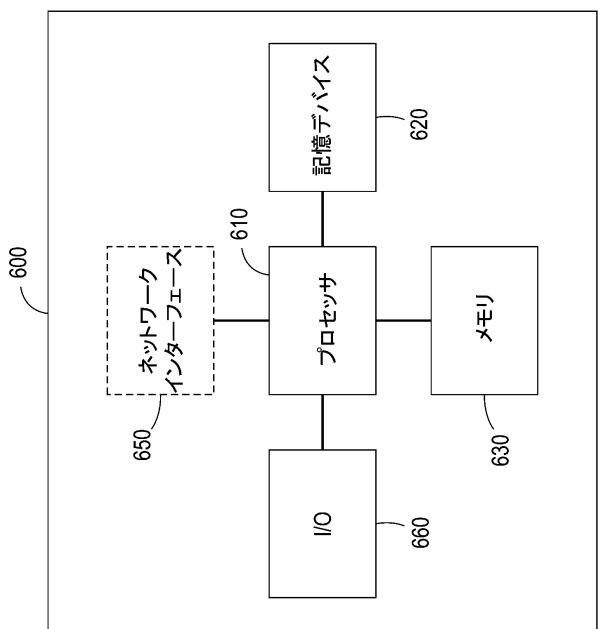


30

40

50

【図 6】



10

20

30

40

50

---

フロントページの続き

弁理士 上杉 浩  
(74)代理人 100120525  
弁理士 近藤 直樹  
(74)代理人 100139712  
弁理士 那須 威夫  
(72)発明者 フランシスコ・イー・トレス  
アメリカ合衆国 カリフォルニア州 95120 サンノゼ ミアンダー・ドライブ 5857  
(72)発明者 ホダ・エルダーディリー  
アメリカ合衆国 カリフォルニア州 94070 サン・カルロス チェスナット・ストリート 320  
(72)発明者 マシュー・シュリーブ  
アメリカ合衆国 カリフォルニア州 94043 マウンテン・ビュー ロック・ストリート 221  
0 アパートメント 27  
(72)発明者 ガウラン・ガヴァイ  
アメリカ合衆国 カリフォルニア州 94086 サニーベール アリバ・ドライブ 253  
(72)発明者 チャド・ラモス  
アメリカ合衆国 イリノイ州 60630 シカゴ ノース・マクビッカー・アベニュー 4841  
審査官 原 秀人  
(56)参考文献 特開2018-142160 (JP, A)  
特開2014-016968 (JP, A)  
特開平11-195087 (JP, A)  
特開2012-173537 (JP, A)  
特開2011-100350 (JP, A)  
米国特許出願公開第2018/0157916 (US, A1)  
(58)調査した分野 (Int.Cl., DB名)  
G 06 F 16/00 - 16/958  
G 06 T 7/00  
G 06 N 20/00