(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6049693号
(P6049693)
(45)発行日　平成28年12月21日(2016.12.21)　(24)登録日　平成28年12月2日(2016.12.2)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 220B
G06F 17/30 350C
G06F 17/30 320Z

請求項の数 16　(全 22 頁)

(21)出願番号　特願2014-505107(P2014-505107)
(86)(22)出願日　平成24年4月11日(2012.4.11)
(65)公表番号　特表2014-524058(P2014-524058A)
(43)公表日　平成26年9月18日(2014.9.18)
(86)国際出願番号　PCT/SG2012/000127
(87)国際公開番号　WO2012/141655
(87)国際公開日　平成24年10月18日(2012.10.18)
審査請求日　平成27年3月24日(2015.3.24)
(31)優先権主張番号　61/474,328
(32)優先日　平成23年4月12日(2011.4.12)
(33)優先権主張国　米国(US)

(73)特許権者　507335687
ナショナル　ユニヴァーシティー　オブ　シンガポール
シンガポール・１１９０７７・シンガポール・ローワー・ケント・リッジ・ロード・２１
(74)代理人　100102978
弁理士　清水　初志
(74)代理人　100102118
弁理士　春名　雅夫
(74)代理人　100160923
弁理士　山口　裕孝
(74)代理人　100119507
弁理士　刑部　俊

最終頁に続く

(54)【発明の名称】ウェブ情報マイニングを用いたビデオ内製品アノテーション

(57)【特許請求の範囲】
【請求項１】
以下の工程を含む、1人または複数のユーザにビデオ内の製品アノテーションを提供するためのコンピュータ方法:
少なくとも、
専門製品リポジトリから、製品のアノテーションされていない専門製品画像を収集する工程、
該専門製品リポジトリとは異なる複数のウェブリソースから、該アノテーションされていない専門製品画像と関連する複数のアノテーションされていない製品画像をサーチする工程、
該アノテーションされていない専門製品画像に対する類似性測度に基づいて、該複数のアノテーションされていない製品画像をフィルタリングすることにより、該複数のアノテーションされていない製品画像のサブセットを選択する工程、および
該アノテーションされていない専門製品画像と、該複数のアノテーションされていない製品画像のサブセットとから、製品視覚シグネチャを生成する工程
により、該製品の製品視覚シグネチャを生成する工程、
製品アノテーションのためのビデオを受け取る工程であって、該ビデオが複数のビデオフレームを含む、工程;
該ビデオフレームから複数のキーフレームを抽出する工程;ならびに
各キーフレームについて、

該キーフレームの視覚表現を生成する工程;
該視覚表現を、該製品視覚シグネチャを含む複数の製品視覚シグネチャと比較する工程;および
該比較に基づき、該キーフレームが該製品視覚シグネチャによって特定される該製品を含むことを判定する工程。
【請求項２】
前記ビデオから複数のキーフレームを抽出する工程が、
該複数のキーフレームの各々を該ビデオの固定点で抽出する工程
を含む、請求項1記載の方法。
【請求項３】
キーフレームの視覚シグネチャを生成する工程が、
該キーフレームから複数の視覚特徴を抽出する工程;
該複数の視覚特徴を複数のクラスタへとグループ化する工程;および
該キーフレームの該視覚シグネチャとして多次元のバッグ・オブ・ビジュアルワード（bag visual words）ヒストグラムを生成する工程
を含む、請求項1記載の方法。
【請求項４】
キーフレームの前記複数の視覚特徴が、該キーフレームのスケール不変特徴変換（Scale Invariant Feature Transform；SIFT）記述子である、請求項3記載の方法。
【請求項５】
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成する工程が、
該製品の該訓練画像に一括疎化方式を適用する工程であって、関連する製品画像に含まれる該製品と無関係な情報が、該製品の該視覚シグネチャの生成の際に低減される、工程
を含む、請求項1記載の方法。
【請求項６】
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成する工程が、
該製品の該視覚シグネチャを所定回数の反復によって繰り返し更新する工程
をさらに含む、請求項1記載の方法。
【請求項７】
製品の前記複数の専門製品画像が、該製品の様々な視点（view）における専門製品画像を含む、請求項1記載の方法。
【請求項８】
キーフレームが前記製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定する工程が、
該キーフレームの前記視覚表現と複数の該製品視覚シグネチャのうちの該製品視覚シグネチャの各々との間の製品関連性を推定する工程;および
該推定した製品関連性に基づいて、キーフレームが該製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定する工程
を含む、請求項1記載の方法。
【請求項９】
1人または複数のユーザにオンデマンドのディジタル資産ホスティングサービスを提供するための実行可能コンピュータプログラム命令を記憶した非一時的なコンピュータ可読記憶媒体であって、該コンピュータプログラム命令が、
少なくとも、
専門製品リポジトリから、製品のアノテーションされていない専門製品画像を収集すること、
該専門製品リポジトリとは異なる複数のウェブリソースから、該アノテーションされていない専門製品画像と関連する複数のアノテーションされていない製品画像をサーチすること、
該アノテーションされていない専門製品画像に対する類似性測度に基づいて、該複数

のアノテーションされていない製品画像をフィルタリングすることにより、該複数のアノテーションされていない製品画像のサブセットを選択すること、および
該アノテーションされていない専門製品画像と、該複数のアノテーションされていない製品画像のサブセットとから、製品視覚シグネチャを生成すること
により、該製品の製品視覚シグネチャを生成するための命令、
製品アノテーションのためのビデオをユーザから受け取るための命令であって、該ビデオが複数のビデオフレームを含む、命令;
該ビデオから複数のキーフレームを抽出するための命令;ならびに
各キーフレームについて、
該キーフレームの視覚表現を生成するための命令;
該視覚表現を、該製品視覚シグネチャを含む複数の製品視覚シグネチャと比較するための命令;
該比較に基づき、該キーフレームが該製品視覚シグネチャによって特定される該製品を含むことを判定するための命令
を含む、コンピュータ可読記憶媒体。
【請求項１０】
前記ビデオから複数のキーフレームを抽出するための前記コンピュータプログラム命令が、
該複数のキーフレームの各々を該ビデオの固定点で抽出するための命令
を含む、請求項9記載のコンピュータ可読記憶媒体。
【請求項１１】
キーフレームの前記視覚シグネチャを生成するための前記コンピュータプログラム命令が、
該キーフレームから複数の視覚特徴を抽出するための命令;
該複数の視覚特徴を複数のクラスタへとグループ化するための命令;および
該キーフレームの該視覚シグネチャとして多次元のバッグ・オブ・ビジュアルワード・ヒストグラムを生成するための命令
を含む、請求項9記載のコンピュータ可読記憶媒体。
【請求項１２】
キーフレームの前記複数の視覚特徴が、該キーフレームのスケール不変特徴変換（SIFT）記述子である、請求項11記載のコンピュータ可読記憶媒体。
【請求項１３】
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成するための前記コンピュータプログラム命令が、
該製品の該訓練画像に一括疎化方式を適用するための命令であって、関連する製品画像に含まれる該製品と無関係な情報が、該製品の該視覚シグネチャの生成の際に低減される、命令
を含む、請求項9記載のコンピュータ可読記憶媒体。
【請求項１４】
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成するための前記コンピュータプログラム命令が、
該製品の該視覚シグネチャを所定回数の反復によって繰り返し更新するための命令
をさらに含む、請求項9記載のコンピュータ可読記憶媒体。
【請求項１５】
製品の前記複数の専門製品画像が、該製品の様々な視点における専門製品画像を含む、請求項9記載のコンピュータ可読記憶媒体。
【請求項１６】
キーフレームが前記製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定するためのコンピュータプログラム命令が、
該キーフレームの前記視覚表現と複数の該製品視覚シグネチャのうちの各製品視覚シグ

ネチャとの間の製品関連性を推定するための命令;および
該推定した製品関連性に基づいて、キーフレームが該製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定するための命令
を含む、
請求項9記載のコンピュータ可読記憶媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願の相互参照
本出願は、2011年4月12日に出願された、「In-Video Product Annotation with Web Information Mining」という名称の米国仮出願第61/474,328号の恩典を主張するものであり、同仮出願は参照によりその全体が組み入れられる。
【背景技術】
【０００２】
背景
説明する態様は、一般にはビデオ内の製品アノテーションに関し、具体的には、ウェブ情報マイニングを使用したビデオ内製品アノテーションに関する。
【０００３】
記憶装置、ネットワークおよび圧縮技術の急速な進歩に伴って、異なるドメインからのビデオデータが爆発的な速度で増大している。ビデオアノテーション（ビデオ概念検出または高次特徴抽出としても広く知られている）は、ビデオコンテンツに記述的概念を自動的に割り当てることを目指すものであり、ここ数年にわたって研究の関心が集中している。しかし、自動ビデオアノテーションに関する大部分の既存の研究は、事象（航空機墜落やランニングなど）、場面（日没や海浜など）、物体カテゴリ（自動車や画面など）などの高次概念に焦点を当てるものであり、iPhoneビデオ広告でのiPhoneなどの特定の製品概念にアノテーションを付けることに関する研究はほとんどない。
【０００４】
製品概念のアノテーションは、ビデオのブラウジング、サーチ、および広告などの多くの用途で非常に重要なものである。ウェブ・ビデオ・サーチの問合せログの研究が示すところによれば、ユーザは、特定の問合せを一般的概念よりも頻繁に使用する。さらに、製品アノテーションは、ビデオ広告の関連性を大幅に改善することができる。しかし、製品の自動化アノテーションは、訓練データが不十分であり、適切な視覚表現を生成するのが難しいために困難な課題である。
【０００５】
自動化製品アノテーションの第1の課題は、アノテーションのための訓練データにかかっている。既存の学習ベースのビデオアノテーションの手法は、訓練データの質に大きく依存するが、手動での訓練サンプルの収集は、時間を要し、労働集約的である。特に、製品画像の複数の視点（view）の問題がある。特定の製品には普通、正面図、側面図、および背面図などの様々な図があり、これらの図は全く異なって見える可能性がある。したがって、製品の様々な視点について記述する訓練データを収集する必要がある。
【０００６】
第2の課題は、効果的な視覚表現である。バッグ・オブ・ビジュアルワード（Bag of Visual Words；BoVW）特徴は流行の手法であり、画像の分類、クラスタリング、および検索などの多くの用途でその有効性を実証している。画像のBoVW表現を生成するために、複数の検出されたキーポイント上の、または製品画像のパッチを密にサンプリングすることによるスケール不変特徴変換（Scale Invariant Feature Transform；SIFT）記述子が抽出され、ビジュアルワードへと量子化される。製品画像を記述するためにBoVWヒストグラムが生成される。しかし、画像の記述子は、画像に含まれる製品部分ではなく画像全体についてのものであり、製品アノテーションについての多くのノイズ（無関係な情報）を含む。

【発明の概要】
【０００７】
概要
本発明の態様は、ウェブマイニングからの製品訓練画像を使用して1人または複数のユーザにビデオ内の製品アノテーションを提供する。
【０００８】
一態様では、コンピュータシステムが1人または複数のユーザにビデオ内の製品アノテーションのサービスを提供する。システムはユーザからビデオを受け取り、ビデオは複数のビデオフレームを含む。システムはビデオから複数のキーフレームを抽出し、キーフレームの視覚表現を生成する。システムは、キーフレームの視覚表現を複数の製品視覚シグネチャと比較し、各視覚シグネチャは製品を特定するものである。複数の製品視覚シグネチャを生成するために、システムは、専門製品リポジトリから得られる複数の専門製品画像を含む複数の訓練画像を収集し、各専門製品画像は、複数のウェブリソースから得られる複数の製品画像と関連付けられる。キーフレームの視覚表現と製品視覚シグネチャとの比較に基づき、システムは、キーフレームが製品の視覚シグネチャで特定される製品を含むかどうか判定する。
【０００９】
本明細書で説明する特徴および利点は包括的なものではなく、特に、図面、明細書、および添付の特許請求の範囲を考察すれば当業者には多くのさらに別の特徴および利点が明らかになるであろう。またさらに、本明細書で使用する言葉は、主として読みやすさと教示とを目的として選択したものであり、開示の主題を正確に叙述し、またはその範囲を定めるために選択したものでない場合もあることに留意すべきである。
[本発明1001]
以下の工程を含む、1人または複数のユーザにビデオ内の製品アノテーションを提供するためのコンピュータ方法:
製品アノテーションのためのビデオを受け取る工程であって、該ビデオが複数のビデオフレームを含む、工程;
該ビデオフレームから複数のキーフレームを抽出する工程;ならびに
各キーフレームについて、
該キーフレームの視覚表現を生成する工程;
該視覚表現を複数の製品視覚シグネチャと比較する工程;および
該比較に基づき、該キーフレームが該製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定する工程。
[本発明1002]
前記ビデオから複数のキーフレームを抽出する工程が、
該複数のキーフレームの各々を該ビデオの固定点で抽出する工程
を含む、本発明1001の方法。
[本発明1003]
キーフレームの視覚シグネチャを生成する工程が、
該キーフレームから複数の視覚特徴を抽出する工程;
該複数の視覚特徴を複数のクラスタへとグループ化する工程;および
該キーフレームの該視覚シグネチャとして多次元のバッグ・オブ・ビジュアルワード（bag visual words）ヒストグラムを生成する工程
を含む、本発明1001の方法。
[本発明1004]
キーフレームの前記複数の視覚特徴が、該キーフレームのスケール不変特徴変換（Scale Invariant Feature Transform；SIFT）記述子である、本発明1003の方法。
[本発明1005]
複数の製品のうちの各製品について、
該製品の複数の訓練画像を収集する工程;および

該製品の該収集した訓練画像から該製品の視覚シグネチャを生成する工程
をさらに含む、本発明1001の方法。
[本発明1006]
製品の前記複数の訓練画像を収集する工程が、
専門製品リポジトリから該製品の複数の専門製品画像を収集する工程;
該製品の各専門製品画像について、
複数のウェブリソースから複数の関連する製品画像をサーチする工程;および
各関連する製品画像と該専門製品画像の間の類似性測度に基づいて、所定数の関連する製品画像を選択する工程
を含み、
該専門製品画像および該選択された関連する製品画像が、該製品の該訓練画像を構成する、
本発明1005の方法。
[本発明1007]
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成する工程が、
該製品の該訓練画像に一括疎化方式を適用する工程であって、関連する製品画像に含まれる該製品と無関係な情報が、該製品の該視覚シグネチャの生成の際に低減される、工程を含む、本発明1005の方法。
[本発明1008]
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成する工程が、
該製品の該視覚シグネチャを所定回数の反復によって繰り返し更新する工程
をさらに含む、本発明1005の方法。
[本発明1009]
製品の前記複数の専門製品画像が、該製品の様々な視点（view）における専門製品画像を含む、本発明1005の方法。
[本発明1010]
キーフレームが前記製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定する工程が、
該キーフレームの前記視覚表現と複数の該製品視覚シグネチャのうちの該製品視覚シグネチャの各々との間の製品関連性を推定する工程;および
該推定した製品関連性に基づいて、キーフレームが該製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定する工程
を含む、本発明1001の方法。
[本発明1011]
1人または複数のユーザにオンデマンドのディジタル資産ホスティングサービスを提供するための実行可能コンピュータプログラム命令を記憶した非一時的なコンピュータ可読記憶媒体であって、該コンピュータプログラム命令が、
製品アノテーションのためのビデオをユーザから受け取るための命令であって、該ビデオが複数のビデオフレームを含む、命令;
該ビデオから複数のキーフレームを抽出するための命令;ならびに
各キーフレームについて、
該キーフレームの視覚表現を生成するための命令;
該視覚表現を複数の製品視覚シグネチャと比較するための命令;
該比較に基づき、該キーフレームが該製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定するための命令
を含む、コンピュータ可読記憶媒体。
[本発明1012]
前記ビデオから複数のキーフレームを抽出するための前記コンピュータプログラム命令が、
該複数のキーフレームの各々を該ビデオの固定点で抽出するための命令

を含む、本発明1011のコンピュータ可読記憶媒体。
[本発明1013]
キーフレームの前記視覚シグネチャを生成するための前記コンピュータプログラム命令が、
該キーフレームから複数の視覚特徴を抽出するための命令;
該複数の視覚特徴を複数のクラスタへとグループ化するための命令;および
該キーフレームの該視覚シグネチャとして多次元のバッグ・オブ・ビジュアルワード・ヒストグラムを生成するための命令
を含む、本発明1011のコンピュータ可読記憶媒体。
[本発明1014]
キーフレームの前記複数の視覚特徴が、該キーフレームのスケール不変特徴変換（SIFT）記述子である、本発明1013のコンピュータ可読記憶媒体。
[本発明1015]
複数の製品のうちの各製品について、
該製品の複数の訓練画像を収集するためのコンピュータプログラム命令;および
該製品の該収集した訓練画像から該製品の視覚シグネチャを生成するためのコンピュータプログラム命令
をさらに含む、本発明1011のコンピュータ可読記憶媒体。
[本発明1016]
製品の前記複数の訓練画像を収集するための前記コンピュータプログラム命令が、
専門製品リポジトリから該製品の複数の専門製品画像を収集するための命令;
該製品の各専門製品画像について、
複数のウェブリソースから複数の関連する製品画像をサーチするための命令;および
関連する製品画像の各々と該専門製品画像の間の類似性測度に基づいて所定数の関連する製品画像を選択するための命令
を含み、
該専門製品画像および該選択された関連する製品画像が該製品の該訓練画像を構成する、
本発明1015のコンピュータ可読記憶媒体。
[本発明1017]
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成するための前記コンピュータプログラム命令が、
該製品の該訓練画像に一括疎化方式を適用するための命令であって、関連する製品画像に含まれる該製品と無関係な情報が、該製品の該視覚シグネチャの生成の際に低減される、命令
を含む、本発明1015のコンピュータ可読記憶媒体。
[本発明1018]
前記製品の前記収集した訓練画像から該製品の視覚シグネチャを生成するための前記コンピュータプログラム命令が、
該製品の該視覚シグネチャを所定回数の反復によって繰り返し更新するための命令
をさらに含む、本発明1015のコンピュータ可読記憶媒体。
[本発明1019]
製品の前記複数の専門製品画像が、該製品の様々な視点における専門製品画像を含む、本発明1015のコンピュータ可読記憶媒体。
[本発明1020]
キーフレームが前記製品視覚シグネチャのうちの一つによって特定される製品を含むかどうか判定するためのコンピュータプログラム命令が、
該キーフレームの前記視覚表現と複数の該製品視覚シグネチャのうちの各製品視覚シグネチャとの間の製品関連性を推定するための命令;および
該推定した製品関連性に基づいて、キーフレームが該製品視覚シグネチャのうちの一つ

によって特定される製品を含むかどうか判定するための命令
を含む、
本発明1011のコンピュータ可読記憶媒体。

【図面の簡単な説明】

【００１０】

【図１】クライアントにビデオ内製品アノテーションサービスを提供するように構成されたコンピューティング環境のブロック図である。

【図２】製品視覚シグネチャを生成し、ビデオストリームで検出された製品にアノテーションを付けるためのビデオ内製品アノテーションモジュールのブロック図である。

【図３】本発明の一態様によるビデオ内製品アノテーションプロセスのための訓練画像収集の例を示す図である。

【図４】製品視覚シグネチャを一括して生成するための製品画像の例を示す図である。

【図５】本発明の一態様による、製品の視覚シグネチャを生成するためのプロセスの流れ図である。

【図６】本発明の一態様による、ビデオストリームの一つまたは複数のビデオフレームにおいて製品を検出し、検出した製品にアノテーションを付けるためのプロセスの流れ図である。

【図７】本発明の一態様によるビデオ内製品アノテーションシステムの例を示す図である。

【図８】本発明の一態様によるビデオ内製品アノテーションプロセスの結果例を示す図である。

【発明を実施するための形態】

【００１１】

各図には本発明の様々な態様が描かれているが、それらは単なる例示である。当業者は、以下の考察を読めば、本明細書で説明する発明の原理から逸脱することなく本明細書で例示する構造および方法の代替の態様が用いられうることを容易に理解するであろう。

【００１２】

詳細な説明

本発明の各図および説明は、本発明の態様の明確な理解のために関連性のある要素を例示するように簡略化されており、典型的なウェブベースのビデオプレーヤおよび同ビデオプレーヤを使用する方法で見られる多くの他の要素を明確にするために省いてあることを理解すべきである。当業者は、本発明の実施に際して、他の要素および/または工程が求められ、および/または必要とされることを理解しうる。しかし、そのような要素および工程は当分野で周知であり、それらは本発明のよりよい理解を促すものではないため、そのような要素および工程の考察は本明細書では行わない。本明細書での開示は、当業者に公知のそうした要素および方法に対するすべてのそうした変形および改変を対象とするものである。

【００１３】

図1は、クライアント110にビデオ内製品アノテーションサービスを提供するように構成されたコンピューティング環境100のブロック図である。複数のユーザ/閲覧者がクライアント110A～Nを使用してビデオ内製品アノテーションサービス120にビデオストリームを提供し、ビデオ内製品アノテーションサービス120に、ビデオストリームのビデオフレームに含まれる製品にアノテーションを付けるよう要求する。製品アノテーションサービス120は、ビデオストリームを記憶し、要求に応答してクライアント110に製品検出およびアノテーションの結果を提供する。各クライアント110は、製品アノテーションサービス120からのビデオストリームおよび製品アノテーションの結果をブラウズするためのブラウザ112を実行する。他の態様は異なる構成を有するものとすることができる。

【００１４】

図1に示す個別エンティティを見ると、各クライアント110は、ユーザがビデオ内製品アノテーションサービス120によって提供されるサービスを使用するのに使用される。例え

ば、ユーザはクライアント110を使用してビデオをブラウズし、ビデオに含まれる製品のアノテーションを要求し、製品アノテーションサービス120から製品検出およびアノテーションの結果を受け取る。クライアント110は、パーソナルコンピュータ（デスクトップ、ノートブック、ラップトップなど）コンピュータなどの任意の種類のコンピュータ機器、ならびに携帯電話やビデオコンテンツを記録する機能を有する携帯情報端末などの機器とすることができる。クライアント110は通常、プロセッサ、表示装置（または表示装置への出力）、ユーザがタスクを実行する際に使用するデータをクライアント110が記憶する、ハードドライブやフラッシュ・メモリ・デバイスなどのローカルストレージ、およびネットワーク130を介してビデオ内製品アノテーションサービス120に結合するためのネットワークインターフェースを含む。

【００１５】

ネットワーク130は、クライアント110とビデオ内製品アノテーションサービス120との間の通信を可能にする。一態様では、ネットワーク130はインターネットであり、現在公知である、または後で開発される、クライアント110がビデオ内製品アノテーションサービス120と通信することを可能にする標準化されたインターネットワーキング通信技術およびプロトコルを使用する。別の態様では、ネットワーク130はクラウド・コンピューティング・ネットワークであり、ビデオ内製品アノテーションサービス120の一つまたは複数の構成要素を含む。

【００１６】

一態様では、ビデオ内の製品を検出し、製品にアノテーションを付けるための2つの段階である、製品視覚シグネチャ生成段階および実行時ビデオ処理段階がある。視覚シグネチャ生成段階は3つの構成部分、すなわち、AMAZON（商標）などのリポジトリからの製品の高品質の視覚的例の収集、インターネット製品画像サーチ結果による収集された視覚的例の拡張、ならびに製品の高品質の視覚的例およびそれらの対応するサーチ結果からの製品画像を含む訓練例からの視覚シグネチャの生成を含む。様々な公知の製品の視覚シグネチャが製品視覚シグネチャファイルに記憶される。

【００１７】

実行時ビデオ処理段階は2つの構成部分、すなわち、特徴抽出および製品アノテーションを含む。入力ビデオストリームにつき、製品アノテーションサービス120は、ビデオストリームの1組のキーフレームを特定し、各キーフレームについて、製品アノテーションサービス120は、視覚特徴（スケール不変特徴変換（SIFT）記述子など）を抽出し、抽出した特徴の視覚表現（バッグ・オブ・ビジュアルワード（BoVW）ヒストグラムなど）を生成する。製品アノテーションサービス120は、視覚シグネチャファイルに記憶された各製品の視覚シグネチャを、入力ビデオの各キーフレームのBoVWヒストグラムと比較することによって、製品アノテーションを行う。

【００１８】

図1に示す態様では、ビデオ内製品アノテーションサービス120は、ビデオ内製品アノテーションモジュール102と、ビデオサーバ104と、製品画像データベース106とを有する。ビデオ内製品アノテーションモジュール102は、製品視覚シグネチャ生成のための製品視覚シグネチャ生成モジュール200と、クライアント110からの入力ビデオを処理するためのビデオ処理モジュール300とを含む。ビデオサーバ104は、クライアント110から受け取られるビデオストリームと、ビデオストリームのアノテーション付きビデオフレームとを記憶する。製品画像データベース106は、AMAZON（商標）などの1社または複数のオンライン製品販売業者から得られた高品質の製品画像と、インターネットサーチにより収集された関連する製品画像とを記憶するための2つの下位データベース、データベース1（106A）およびデータベース2（106B）を含む。

【００１９】

公知の製品販売業者からの製品画像は一般に高画質を有するが、所与の製品についてのそうした製品画像の数は限られたものである可能性がある。所与の製品についての、GOOGLE（商標）などの様々なサーチエンジンを使用したインターネットサーチによる製品の関

連画像の数は、大きいが、ノイズの多い（例えば、製品と無関係なテキスト情報を含む）ものとなる可能性がある。製品アノテーションサービス120は、インターネットサーチ結果から得られる関連する製品画像を高品質の製品画像に基づいてフィルタリングして製品視覚シグネチャを生成し、製品視覚シグネチャを使用してビデオストリーム内の製品を検出し、製品にアノテーションを付ける。本発明の一態様を簡略化するために、公知の販売業者からの高品質の製品画像を「専門製品画像」と呼び、所与の専門製品画像について、インターネットサーチから得られた、それと関連付けられた画像を「拡張製品画像」と呼ぶ。

【００２０】

ビデオ内製品アノテーション-視覚シグネチャ生成

図2は、一態様による、製品視覚シグネチャを生成し、ビデオストリーム内で検出される製品にアノテーションを付けるためのビデオ内製品アノテーションモジュール102のブロック図である。製品アノテーションモジュール102は、製品視覚シグネチャ生成モジュール200と、ビデオ処理モジュール300とを含む。製品視覚シグネチャ生成モジュール200は、専門製品画像モジュール210と、拡張製品画像モジュール220と、視覚シグネチャ生成モジュール230とを含む。ビデオ処理モジュール300は、フレーム抽出モジュール310と、特徴抽出・量子化モジュール320と、製品アノテーションモジュール330とを含む。

【００２１】

製品視覚シグネチャ生成モジュール200は、製品視覚シグネチャを生成するように構成されている。専門製品画像モジュール210は、製品の高品質の視覚的例（正面図、側面図、および背面図などの様々な視点における専門製品画像など）を収集するように構成されている。一態様では、専門製品画像モジュール210は、ディジタルカメラ、車、およびディジタル電話などの様々な消費者製品についてのAMAZON（商標）からの専門製品画像を収集する。

【００２２】

所与の製品の専門製品画像は往々にして、当該製品の満足のいく視覚シグネチャを構築するには少なすぎる。例えば、ある製品についてAMAZON（商標）から収集される専門製品画像の数は、1～8枚で変動する。他方、様々なサイズおよび視点の製品画像がインターネット上で豊富に入手可能であり、これらには画像サーチエンジンによって簡単にアクセスすることができる。一態様では、拡張製品画像モジュール220は、1枚または複数枚の専門製品画像を有する製品の関連付けられる画像を、インターネットから収集するように構成されている。例えば、各専門製品画像について、GOOGLE（商標）サーチエンジンを使用したインターネットでの関連付けられる製品画像についてのサーチ問合せとして、製品名が使用される。このプロセスにより、ウェブ製品画像データベースを使用して専門製品画像が拡張される。

【００２３】

インターネットサーチからの画像は多くのノイズ、例えば（その多くがサーチ問合せと無関係であるタイトル周りのテキストなどの）テキスト情報を含む。シグネチャ生成モジュール230が製品の視覚シグネチャを生成する前に、シグネチャ生成モジュール230は、専門製品画像に基づいてインターネットサーチ結果からの拡張製品画像をランク付けし直す。各専門製品画像について、専門製品画像に近い所定数の拡張製品画像がフィルタリングの結果として選択される。所与の製品につき、専門製品画像およびフィルタリングされた拡張製品画像が、製品のための1組の確実な訓練画像を形成し、それらの訓練画像からシグネチャ生成モジュール230は製品のための視覚シグネチャを生成する。公知の製品の訓練画像の収集を自動化してビデオ内製品アノテーションシステムの性能を改善することができる。

【００２４】

専門製品画像と関連付けられる拡張製品画像をフィルタリングするために、シグネチャ生成モジュール230は、専門製品画像の視覚特徴およびそれと関連付けられる拡張製品画像を抽出する。一態様では、製品画像の視覚特徴はバッグ・オブ・ビジュアルワード（Bo

VW）特徴である。シグネチャ生成モジュール230は、いくつかの検出されたキーポイント上で、または各製品画像のパッチを密にサンプリングすることによって一つまたは複数のSIFT記述子を抽出し、SIFT記述子を複数のビジュアルワードへと量子化する。各画像を記述するために量子化SIFT記述子からBoVWヒストグラムが生成される。例えば、シグネチャ生成モジュール230は、視覚特徴検出抽出法、例えば、Difference-of-Gaussian法を使用して、製品画像から128次元のSIFT特徴を抽出し、階層的K平均法を用いてSIFT特徴を160,000個のクラスタへとグループ化する。製品画像は、160,000次元のBoVWヒストグラムで表される。

【００２５】

各専門製品画像について、シグネチャ生成モジュール230は、以下の式（1）で定義される類似性測度に基づいて、専門製品画像と関連付けられる拡張製品画像の中から、所定数の最も近い近隣画像を選択する：

$$sim(x,y)=\frac{\sum_{d=1}^{D}\min\{(x_d),(y_d)\}}{\min\{\sum_{d=1}^{D}x_d,\sum_{d=1}^{D}y_d\}} \qquad (1)$$

式中、xおよびyは2つのBoVWヒストグラムであり、Dはヒストグラムの長さである。このようにして、シグネチャ生成モジュール230は、所与の製品についてのkn枚の確実な訓練画像を獲得し、kは専門製品画像の数であり、nは専門製品画像の所定の最も近い近隣画像（すなわち拡張製品画像）数である。

【００２６】

ビデオ内製品アノテーション訓練画像の収集をさらに例示するために、図3に、ディジタルカメラ、Canon 40Dのための訓練データ収集プロセスの例を示す。製品視覚シグネチャ生成モジュール200は、オンライン販売業者のAMAZON（商標）から、このカメラの5枚の専門製品画像302を収集する。各専門製品画像について、製品視覚シグネチャ生成モジュール200は、GOOGLE（商標）サーチエンジンを使用してインターネットをサーチして、いくつかの関連する製品画像304を収集する。インターネットサーチから得られる製品画像はノイズの多い（例えば、製品と無関係なテキストを含む）ものである可能性があるため、製品視覚シグネチャ生成モジュール200は、専門製品画像に基づいて関連する製品画像をフィルタリングする。例えば、各専門製品画像について、製品視覚シグネチャ生成モジュール200は、後述する相関疎化を適用して、インターネットサーチの中から製品画像の所定数の最も近い近隣画像を選択することによってノイズを低減させる。関連する製品画像の選択は、関連する製品画像とその対応する専門製品画像の間の類似性測度に基づくものである。フィルタリングの結果として、製品視覚シグネチャ生成モジュール200は、ディジタルカメラ、Canon 40Dについての1組の訓練例306を獲得し、ここで製品視覚シグネチャ生成モジュール200は、ディジタルカメラ、Canon 40Dのための視覚シグネチャを生成する。

【００２７】

高次元特徴空間で表された製品画像に含まれる製品に効果的にアノテーションを付けるために、シグネチャ生成モジュール230は、製品の確実な訓練画像を平均することによってアノテーションのためのテンプレートを生成する。一態様では、シグネチャ生成モジュール230は、製品の複数の訓練画像の視覚表現をマージして製品の累積ヒストグラムを生成する。画像背景からの記述子によって生じる多くのノイズがあるため、実際、累積ヒストグラムには多くのノイズを含むビンがある。

【００２８】

ノイズを低減させる一手法は、L1正則化最小二乗最適化問題にフィットする、以下の式（2）に記述されたL1疎化を用いるものであり、

$$\arg\min_{v_i}\|v_i-\bar{v}_i\|_2^2+\lambda_1\|v_i\|_1 \qquad (2)$$

式中、

$\|\cdot\|_2$ および $\|\cdot\|_1$

は、それぞれ、2-ノルムおよび1-ノルムを示す。パラメータλ_1はL1ノルムの影響を調節し、

$\bar{v}_i$

は、第iの製品についての元の累積BoVWヒストグラムであり、v_iは学習されるべき視覚シグネチャである。式（2）の第1項は、獲得されたシグネチャを元のシグネチャに近いままに保ち、第2項は、獲得された視覚シグネチャの1-ノルム値を最小化し、シグネチャを疎にする。

【００２９】

同じクラス/カテゴリの複数の製品は近い外観を有することに留意されたい。例えば、製品、Canon 40Dと製品Nikon D90とは、非常に近い外観を有する。よって、これら2つの製品のヒストグラム表現は非常に近いはずである。同じクラスの製品の画像の近さを反映するために、シグネチャ生成モジュール230は、製品の視覚シグネチャを一括して生成する。一態様では、シグネチャ生成モジュール230は、式（2）にグラフラプラシアン項を加えることによって、式（2）で定義される視覚シグネチャ生成を以下のように変更する：

$$\arg\min_{\{v_1,v_2,\ldots,v_n\}} \sum_{i=1}^{n} \|v_i-\bar{v}_i\|_2^2+\lambda_1\sum_{i=1}^{n}\|v_i\|_1+\lambda_2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\|v_i-v_j\|_2^2 \qquad (3)$$

式中、w_{ij}は、製品iと製品jの間の類似性であり、λ_2は、グラフラプラシアン項の影響を調節するパラメータである。グラフラプラシアン項、

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\|v_i-v_j\|_2^2$$

は、全製品のシグネチャを結合する。

【００３０】

式（3）は、最適化の手法を使用して解くことができる。v_iを除くすべての視覚シグネチャが固定されているものと仮定する。式（3）で記述される問題は、以下の式（4）として書き換えることができる：

$$\arg\min_{v_i} \|v_i-\bar{v}_i\|_2^2+\lambda_1\|v_i\|_1+\lambda_2\sum_{j=1}^{n} w_{ij}\|v_i-v_j\|_2^2 \qquad (4)$$

。視覚シグネチャv_iは、式（5）によって以下のように定義される：

$$\arg\min_{v_i}\left\|\begin{pmatrix} I \\ \sqrt{\lambda_2 w_{i1}}I \\ \sqrt{\lambda_2 w_{i2}}I \\ \cdots \\ \sqrt{\lambda_2 w_{in}}I \end{pmatrix} v_i-\begin{pmatrix} \bar{v}_i \\ \sqrt{\lambda_2 w_{i1}}v_1 \\ \sqrt{\lambda_2 w_{i2}}v_2 \\ \cdots \\ \sqrt{\lambda_2 w_{in}}v_n \end{pmatrix}\right\|_2^2+\lambda_1\|v_i\|_1 \qquad (5)$$

式中、Iは、D＊Dの単位行列であり、視覚シグネチャ生成はL1正則化最小二乗最適化問題として表される。一態様では、シグネチャ生成モジュール230は、内点法を使用して式（5）で定義される問題を解く。製品の視覚シグネチャは製品のグランドトルースを表し、ビデオストリームのビデオフレームが実行時に製品を含むかどうか判定するのに使用することができる。

【００３１】

2組の製品画像の間の類似性は、式（6）によって以下のように定義される：

$$w_{ij} = \frac{1}{2|P_i|}\sum_{k=1}^{|P_i|}\max_{p\in P_j} sim(p_i^{(k)}, p) + \frac{1}{2|P_j|}\sum_{k=1}^{|P_j|}\max_{p\in P_i} sim(p_j^{(k)}, p) \quad (6)$$

式中、｜p_i｜および｜p_j｜は画像集合P_iおよびP_jについての画像の数であり、$P_i^{(k)}$は、集合P_i内の第kの製品を指示し、sim（.,.）は、異なる集合からの画像対の類似性である。式（6）で定義される類似性測度は以下の特性を有する。

（1）$w_{ij}=w_{ji}$:類似性は対称である。

（2）$P_i=P_j$の場合、w_{ij}=1:2つの製品の画像集合が同一である場合、2つの製品の類似性は1である。

（3）あらゆるp'∈P_iおよびp"∈P_jについてsim（p',p"）=0の場合に限り、w（p_i,p_j）=0:2つの画像集合によって形成されるあらゆる対がゼロの類似性を有する場合に限り、類似性は0である。

【００３２】

一態様では、異なる集合からの画像対の類似性sim（.,.）は、式（1）に記述される画像対のヒストグラム交差から計算される。類似性計算を簡略化するために、ある製品の2つの異なるサブカテゴリに属する2つの製品（例えば、同じ製品クラス「電子機器」の下にあるビデオゲームと携帯用オーディオ/ビデオ製品など）の類似性はゼロに設定される。

【００３３】

図4は、本発明の一態様による、製品視覚シグネチャを一括して生成するための3組の製品画像の例である。図4の例は、3つの製品についての3組の製品画像を含み、製品画像集合410はディジタルカメラ、Canon 40Dのものであり、製品画像集合420はディジタルカメラ、Nikon D90のものであり、製品画像集合430はビデオ・ゲーム・コンソール、Xboxのものである。製品Canon 40Dと製品Nikon D90とは、同じ製品クラスに属するために、非常に近い外観を有することに留意されたい。製品の視覚シグネチャを一括して生成することは、製品の視覚シグネチャが同じクラスの製品の画像の近さを反映することを可能にする。

【００３４】

式（5）および式（6）から、シグネチャ生成モジュール230は、各v_iを繰り返し更新することによって各v_iを解く反復プロセスを導出することができる。視覚シグネチャv_iを反復して更新するための疑似コードの例は以下の通りである。

Input:
$\overline{V} = \{\overline{v}_1, \overline{v}_2, ..., \overline{v}_n\}$; /* Original accumulated BoVW representation */
Output:
$V = \{v_1, v_2, ..., v_n\}$; /*BoVW representation after optimization */
Process:
For each product *i* and *j*
Compute their similarity w_{ij} according to Eq. (6)
End
Initialization: *t*=0;
Initialize $v_i^{(t)}$ to be $\overline{v}_i$
Iterative until t = T
For *i* = 1, 2,..., *n*
Update the signature of *ith* product according to Eq. (5), and let the result be $v_i^{(t+1)}$
End
END
Return $V = \{v_1^{(T)}, v_2^{(T)}, ..., v_n^{(T)}\}$

【００３５】
図5は、本発明の一態様による、製品の視覚シグネチャを生成するためのプロセスの流れ図である。最初に、製品視覚シグネチャ生成モジュール200は、製品のリポジトリから専門製品画像をサーチし510、専門製品画像を記憶場所（例えば、図1の製品画像データベース106のデータベース1など）に記憶する520。各専門製品画像について、製品視覚シグネチャ生成モジュール200は、ウェブマイニング（インターネットサーチなど）によって複数の関連する製品画像を収集する530。製品視覚シグネチャ生成モジュール200は、関連する製品画像をその対応する専門製品画像に基づいてフィルタリングし、フィルタリングしたものから訓練サンプル製品画像を生成する540。訓練サンプル製品画像を使用して、製品視覚シグネチャ生成モジュール200は、製品のための視覚シグネチャを生成する550。製品視覚シグネチャ生成モジュール200は、様々な製品のための視覚シグネチャを含む視覚シグネチャファイルを集約する560。

【００３６】
ビデオ内製品アノテーション-製品アノテーション

図2に戻って、ビデオ内製品アノテーションモジュール102は、ビデオストリームの一つまたは複数のビデオフレーム内の製品にアノテーションを付けるビデオ処理モジュール300を有する。ビデオ処理モジュール300は、クライアント110からビデオストリームを受け取り、ビデオストリームの一つまたは複数の選択されたビデオフレームを処理する。各選択されたビデオフレームについて、ビデオ処理モジュール300は、製品視覚シグネチャ生成モジュール200によって提供された製品視覚シグネチャを使用して、ビデオフレームが公知の製品を含むかどうか判定する。

【００３７】
一態様では、ビデオ処理モジュール300は、ビデオフレーム抽出モジュール310と、特徴抽出・量子化モジュール320と、製品アノテーションモジュール330とを含む。ビデオフレーム抽出モジュール310は、複数のビデオフレームからなるビデオストリームを受け取り、ビデオストリームからいくつかのキーフレームを抽出する。ビデオストリームからキーフレームを抽出する一つの方法は、ビデオストリームの固定点でビデオフレームを選択すること、例えば、ビデオストリームの5秒ごとにビデオフレームを抽出することである。

フレーム抽出モジュール310の他の態様は、キーフレームを獲得するのに異なる方法を使用することができ、例えば、ビデオストリームのピクチャ群（GOP）ごとの第1のフレームを選択する。

【００３８】

特徴抽出・量子化モジュール320は、ビデオストリームのキーフレームから視覚特徴を抽出するためのものであり、各キーフレームの視覚表現を生成するために抽出された視覚特徴が量子化される。一態様では、特徴抽出・量子化モジュール320は、Difference-of-Gaussian法を使用してキーフレーム内のキーポイントを検出し、各キーポイントから、モジュール320は、128次元のSIFT特徴を抽出する。モジュール320は、階層的K平均法を用いてSIFT特徴を多数のクラスタ（160,000個のクラスタなど）へとグループ化する。キーフレームは、多次元のバッグ・オブ・ビジュアルワード・ヒストグラム（160,000次元のBoVWヒストグラムなど）で表される。

【００３９】

製品アノテーションモジュール330は、製品視覚シグネチャをキーフレームの視覚表現（160,000次元のBoVWヒストグラムなど）と比較することによって、ビデオストリームのキーフレームが公知の製品を含むかどうか判定する。製品視覚シグネチャとキーフレームの視覚表現との比較は、以下の式（7）で定義される製品関連性尺度で測られる：

$$s(f, v_i) = \frac{\sum_{d=1}^{D} \min\{(f_d), (v_{i,d})\}}{\min\{\sum_{d=1}^{D} f_d, \sum_{d=1}^{D} v_{i,d}\}} \qquad (7)$$

式中、fはキーフレームの視覚BoVWヒストグラムであり、v_iは、第iの製品のための視覚シグネチャファイルである。キーフレームの製品関連性尺度と公知の製品の視覚的製品特徴とに基づき、製品アノテーションモジュール330は、キーフレームが公知の製品を含むかどうか判定する。一態様では、推定される製品関連性尺度が、キーフレームが公知の製品を含むかどうか判定するための閾値と比較される。

【００４０】

図6は、本発明の一態様による、ビデオストリームの一つまたは複数のビデオフレームにおいて製品を検出し、検出した製品にアノテーションを付けるためのプロセスの流れ図である。ビデオ処理モジュール300は、クライアント110からビデオストリームを受け取り610、ビデオストリームから複数のキーフレームを抽出する620。各キーフレームについて、ビデオ処理モジュール300は、キーフレームの視覚特徴（SIFT特徴など）を抽出し630、キーフレームの視覚表示（多次元BoVWヒストグラムなど）を生成する620。ビデオ処理モジュール300は、キーフレームの視覚表現を公知の製品の各視覚シグネチャと比較する650。比較に基づき、ビデオ処理モジュール300は、キーフレームが公知の製品を含むかどうか判定する660。

【００４１】

図7は、本発明の一態様によるビデオ内製品アノテーションシステムの例である。ビデオ内製品アノテーションシステム700は、ディジタルカメラCanon G9 702の視覚シグネチャをオフラインで生成するための製品視覚シグネチャ生成サブシステム701Aと、実行時の製品アノテーションのためにビデオストリーム712の選択されるビデオフレームを処理するためのビデオ処理サブシステム701Bとを有する。製品視覚シグネチャ生成サブシステム701Aは、製品の視覚的例706としてAMAZON（商標）704からCanon G9の様々な視点における1枚または複数枚の専門製品画像を収集する。Canon G9の各専門製品画像について、製品視覚シグネチャ生成サブシステム701Aは、GOOGLE（商標）710から複数の関連付けられる製品画像を収集する。インターネットサーチからのCanon G9の製品画像はノイズを含む可能性があるため、製品視覚シグネチャ生成サブシステム701Aは、ノイズを低減させるために、相関疎化法708によってインターネットサーチからの製品画像をフィルタリングする。フィルタリングされた製品画像および関連付けられる専門画像はCanon G9の1組の訓練

画像を形成し、製品視覚シグネチャ生成サブシステム701Aは、Canon G9のための視覚シグネチャを生成する。

【００４２】

ビデオ処理サブシステム701Bは、ビデオストリーム712を受け取り、ビデオストリームから複数のキーフレームを抽出する714。各キーフレームについて、複数の視覚特徴が抽出され、量子化されて716、キーフレームの視覚表現（BoVWヒストグラムなど）が生成される。各キーフレームについて、ビデオ処理サブシステム701Bは、キーフレームの視覚表現をCanon G9の視覚シグネチャと比較し、比較に基づき、ビデオ処理サブシステム101Bは、キーフレームがディジタルカメラ、Canon G9を含むかどうか判定する。例示のために、製品視覚シグネチャ生成701Aによって集約された製品視覚シグネチャファイルはCanon G9の視覚シグネチャだけを含むと仮定する。製品視覚シグネチャファイルがより多くの製品の視覚シグネチャを含む場合、ビデオ処理サブシステム701Bは、キーフレームの視覚表示を各視覚シグネチャと比較して、キーフレームが視覚シグネチャによって特定される製品を含むかどうか判定する。

【００４３】

実験

ビデオ内製品アノテーションシステムの態様の性能を評価するために、Canon 40D、Nikon D90、Canon G9、Xbox360、および他の電子機器を含む電子機器ドメインからの20製品を評価のために選択した。選択した電子機器に関連する1044件のウェブビデオをYOUTUBE（商標）から収集した。各ビデオにつき、キーフレームを5秒ごとに抽出した。計52,941個のキーフレームを抽出した。選択した電子機器のグランドトルースラベル付けにおいて3つのラベラを使用した。各製品について、投票によって製品ビデオの各フレームの関連性をラベル付けした。キーフレームの中には、製品のうちの少なくとも一つに関連性を有する16,329個のキーフレームがあり、どんな製品にも関連性のない36,162個のキーフレームがあった。特徴表現について、Difference-of-Gaussian法を使用してキーポイントを検出し、各キーポイントから、128次元のSIFT特徴を抽出した。SIFT特徴を、階層的K平均法を用いて160,000個のクラスタへとグループ化した。各製品画像を160,000次元のBoVWヒストグラムで表した。

【００４４】

性能評価メトリックのために、平均適合率（AP）方式（式（8））を用いて検索有効度を測定した：

$$AP = \frac{1}{R}\sum_{i=1}^{s}\frac{R_j}{j}I_j \qquad (8)$$

式中、Rは、サイズSの集合内の真に関連性のあるフレームの数であり、R_jは、任意の所与のインデックスjでの上位j個の結果内の関連性のあるフレームの数であり、第jのフレームが関連性を有する場合はj＝1、そうでない場合はj＝0である。平均適合率の平均値（mean average precision；MAP）は、全製品に及ぶ平均適合率の平均値である。以下の3種類の訓練データを性能評価のために考察した。

（1）AMAZON（商標）の例の使用のみ。画像は非常に少ない。

（2）上位のGOOGLE（商標）画像サーチ結果の使用のみ。各製品について300画像が使用される。

（3）AMAZON（商標）とGOOGLE（商標）画像サーチエンジンを同時に統合する提案の手法。各製品について300画像が使用される。

【００４５】

3種類の訓練データに、それぞれ、「Amazonのみ」、「Googleのみ」、および「Amazon＋Google」とラベル付けした。製品アノテーションアルゴリズムについては、アルゴリズムの以下の3つの変形をテストした。

（1）疎でない:すべての確実な訓練画像の累積BoVWヒストグラムを直接使用。

（2）1-ノルム疎化:式（2）で定義されるような、1-ノルム疎化法を使用。

（3）相関疎化:式（3）で定義されるような相関疎化法を使用。
第2および第3の方法では、パラメータλ_1およびパラメータλ_2を経験的にそれぞれ5と0.05とに設定した。訓練データソースとアノテーションアルゴリズムとの異なる組み合わせの性能結果（MAP結果など）を以下の表Iに示す。

【００４６】

（表Ｉ）視覚シグネチャファイルを生成するための異なる画像ソースの比較

MAP	Amazonのみ	Googleのみ	Amazon+Google
疎でない	0.16	0.2	0.18
1-ノルム疎化	0.17	0.31	0.35
相関疎化	0.28	0.32	0.38

【００４７】

表Iの結果は以下を示している。
－「Googleのみ」の性能は「Amazonのみ」よりもよい。これは、AMAZON（商標）の画像例が満足のいく視覚シグネチャを構築するのに十分ではないことを実証するものである。「Amazon＋Google」は「Googleのみ」より優れており、これは、提案の手法の少なくとも一つの態様の有効性を裏付けるものである。
－「1-ノルム疎化」の性能は「疎でない」よりもよい。これは、疎化の手法がBoVWヒストグラムのノイズを低減させるからである。提案の「相関疎化」は「1-ノルム疎化」をさらに改善し、これは、式（3）のグラフラプラシアン項の有効性を実証するものである。

【００４８】

表1に示す結果をさらに例示するために、図8は、製品アノテーションアルゴリズムの変形によって処理されたビデオフレームおよびそれらのアルゴリズムの対応するビデオフレームにおけるビジュアルワードの例である。左列は、異なるアノテーション法による視覚シグネチャ（810A、820Aおよび830Aを含む。視覚シグネチャ810Aは「疎でない」方法によって生成される。視覚シグネチャ820Aは「1-ノルム疎化」法によって生成され、視覚シグネチャ830Aは「相関疎化」法によって生成される。右列に、対応するビデオフレームにおけるビジュアルワード（810B、820Bおよび830B）を示す。図8の例は、疎化法が、ノイズの多い複数のビンを除去することができ、得られる視覚シグネチャがよりよくなることを示している。「相関疎化」の手法は、複数の製品の相関を探り、よりよい品質の視覚シグネチャを生成する。

【００４９】

前述のビデオ内製品アノテーションの別の態様が多モードの製品アノテーションであり、ビデオストリームと関連付けられたテキスト情報も利用する。ビデオと関連付けられたテキスト情報は、ビデオのタイトル、説明およびタグを含む。ビデオフレームの視覚情報と関連付けられたテキスト情報を統合することによって、ビデオ内製品アノテーションの性能をさらに向上させることができる。例えば、実験結果が示すところでは、単にテキストだけに基づくビデオ内製品アノテーションの方法は0.23のMAP測定値を獲得し、これは視覚情報を使用するだけで獲得されるMAP0.39に劣る。テキスト情報と視覚情報とを統合することによって、MAP尺度を0.55まで押し上げることができる。

【００５０】

本発明の態様は、有利には、製品視覚シグネチャの相関疎化を使用したビデオ内製品アノテーションを提供する。製品の疎な視覚シグネチャは、専門製品画像と、ウェブマイニングからの関連付けられる製品画像とを使用して生成される。関連付けられる製品画像のノイズは、相関疎化によって低減される。ウェブマイニングからの製品の拡張訓練画像は、専門製品画像と共に、本発明の態様が製品視覚シグネチャを生成することを可能にし、製品視覚シグネチャは、実行時にビデオストリームのビデオフレームに含まれる製品を効率よく特定するのに使用される。

【００５１】

本発明の態様の性能は計算効率がよい。例えば、特徴抽出後のビデオフレームに対する

製品のアノテーションは、実際に、視覚シグネチャのビンのゼロ以外の数で概算される。大きなデータセットにアノテーションを付けるときには、製品視覚シグネチャの疎性を調べることによって反転構造を構築することができる。したがって、製品視覚シグネチャの疎化は、アノテーション性能を向上させるのみならず、計算コストの低減も達成する。

【００５２】

本明細書で「一態様」または「ある態様」という場合、それは、各態様と関連して説明する特定の特徴、構造、または特性が本発明の少なくとも一つの態様に含まれることを意味するものである。本明細書の様々な箇所で「一態様において」または「好ましい態様」という句を使用する場合、必ずしもすべて同じ態様を指すものであるとは限らない。

【００５３】

上記のうちのある部分は、コンピュータメモリ内のデータビットに対する操作の方法および記号表現として提示されている。これらの記述および表現は、当業者により、その作業の内容を他の当業者に最も効率よく伝えるのに使用される手段である。方法とは、本明細書では、また一般に、所望の結果に導く自己矛盾のない工程（命令）のシーケンスとみなされる。工程は、物理量の物理的操作を必要とする工程である。必須ではないが、普通、これらの量は、記憶され、移動され、組み合わされ、比較され、また別様に操作されることの可能な、電気信号、磁気信号、または光学信号の形をとる。往々にして、主に一般的な用法のために、これらの信号を、ビット、値、要素、記号、文字、項、数などと呼ぶのが好都合である。さらに、往々にして、物理量の物理的操作を必要とする工程のある特定の配置を、一般性を失うことなく、モジュールまたはコードデバイスと呼ぶのも好都合である。

【００５４】

しかし、これらおよび類似の用語はすべて、適切な物理量と関連付けられるべきであり、単にこれらの量に適用される好都合なラベルにすぎないことに留意すべきである。以下の考察から明らかなように特に指示しない限り、説明全体を通して、「処理（processing）」、または「計算処理（computing）」、または「計算（calculating）」、または「判定（determining）」、または「表示（displaying）」、または「判定（determining）」などの用語を使用した考察は、コンピュータシステムのメモリまたはレジスタまたは他のそうした情報の記憶、伝送または表示の装置内の物理（電子）量として表されたデータを操作し、変換する、コンピュータシステム、または類似の電子的コンピューティング機器の動作およびプロセスを指すものであることが理解される。

【００５５】

本発明のある局面は、本明細書で方法の形で説明されるプロセスの工程および命令を含む。本発明のプロセスの工程および命令は、ソフトウェア、ファームウェアまたはハードウェアとして実施することができ、ソフトウェアとして実施されるときには、様々なオペレーティングシステムによって使用される異なるプラットフォーム上に置かれ、それらのプラットフォームから操作されるようにダウンロードすることができることに留意すべきである。

【００５６】

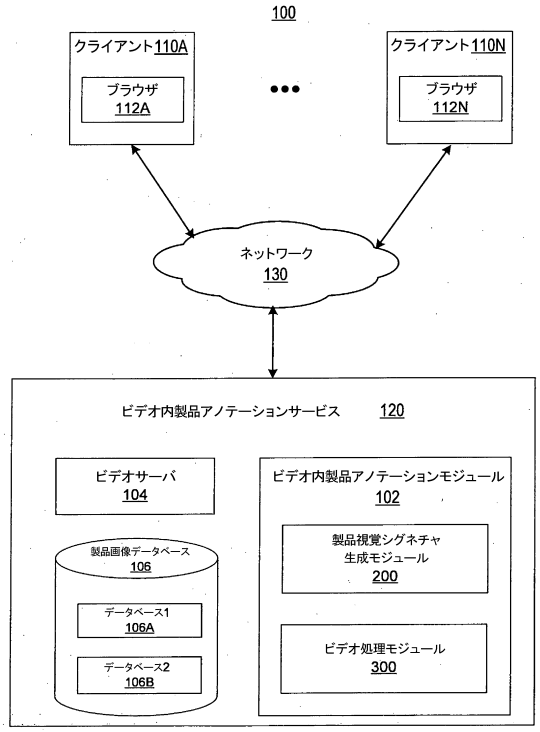
また本発明は、本明細書中の動作を実行する装置に関するものでもある。この装置は、必要とされる目的のために専用に構築されてもよく、コンピュータに記憶されたコンピュータプログラムによって選択的に活動化され、または再構成される汎用コンピュータを含んでいてもよい。そのようなコンピュータプログラムは、それだけに限らないが、フロッピーディスク、光ディスク、CD-ROM、光磁気ディスクを含む任意の種類のディスク、読取り専用メモリ（ROM）、ランダム・アクセス・メモリ（RAM）、EPROM、EEPROM、磁気カードまたは光カード、特定用途向け集積回路（ASIC）、または電子的命令を記憶するのに適した任意の種類の媒体などの、各々コンピュータ・システム・バスに結合されたコンピュータ可読記憶媒体に記憶されていてよい。さらに、本明細書でいうところのコンピュータは、一つのプロセッサを含んでいてもよく、計算処理能力を高めるための複数プロセッサ設計を用いたアーキテクチャとすることもできる。

【００５７】
本明細書で提示する方法および表示は、いかなる特定のコンピュータにも他の装置にも本質的に関連するものではない。様々な汎用システムが本明細書の教示に従ったプログラムと共に使用されてもよく、必要とされる方法工程を実行するためのより特化された装置を構築することが好都合となる場合も考えられる。様々なこれらのシステムの必要とされる構造は、以下の説明を読めば想起されるであろう。加えて、本発明は、いかなる特定のプログラミング言語を参照して説明されるものでもない。本明細書で説明した発明の教示を実施するのに様々なプログラミング言語が使用されてよく、以下で特定の言語に言及する場合、それは、本発明の実施可能要件および最良の形態を開示するために示すものであることが理解されるであろう。

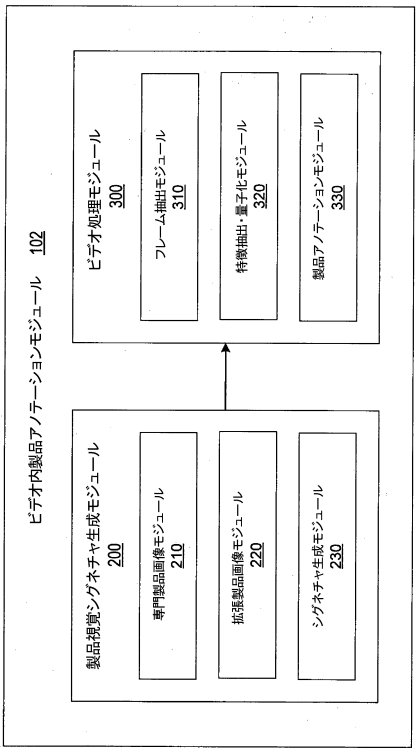
【００５８】
以上、好ましい態様およびいくつかの代替の態様を参照して本発明を詳細に図示し、説明したが、本発明の趣旨および範囲を逸脱することなく様々な形態および詳細の変更をそれらの態様において加えることができることを、当業者は理解するであろう。

【００５９】
最後に、本明細書で使用した言葉は、主として読みやすさと教示を目的として選択したものであり、本発明の主題を正確に叙述し、またはその範囲を定めるために選択したものでない場合もあることに留意すべきである。したがって、本発明の開示は、本発明の範囲の限定ではなく、その例示のためのものである。

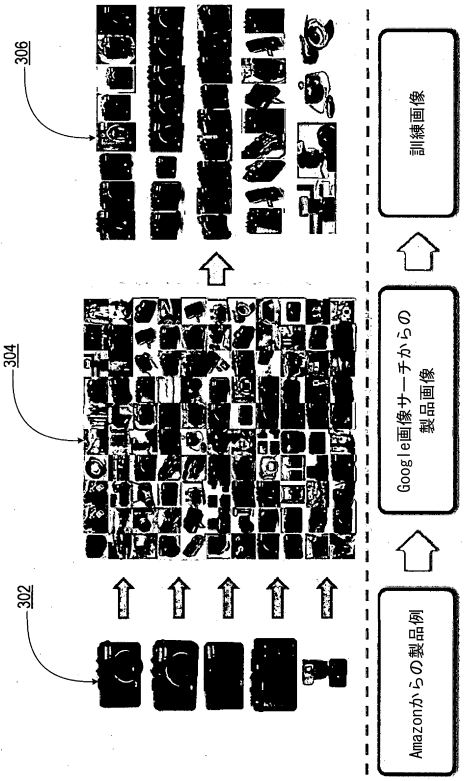
【図１】



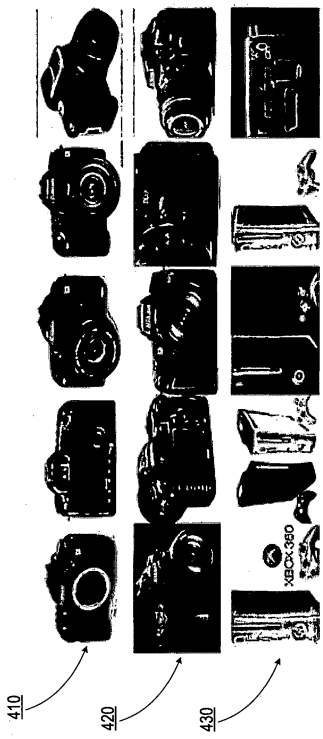
【図２】



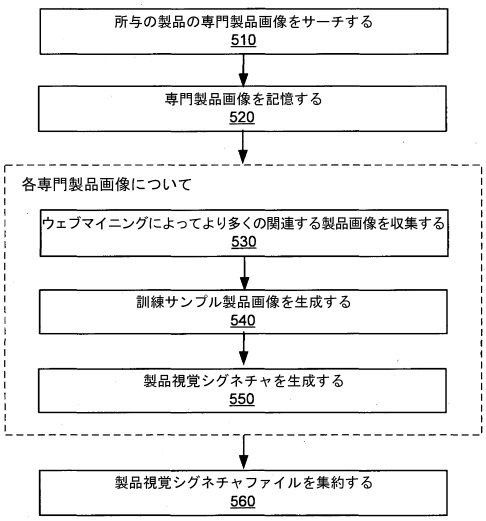
【図３】


306
304
302
訓練画像
Google画像サーチからの製品画像
Amazonからの製品例

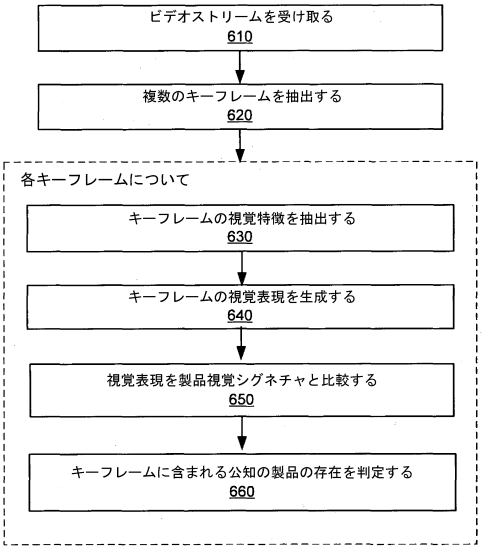
【図４】


410
420
430
XBOX360

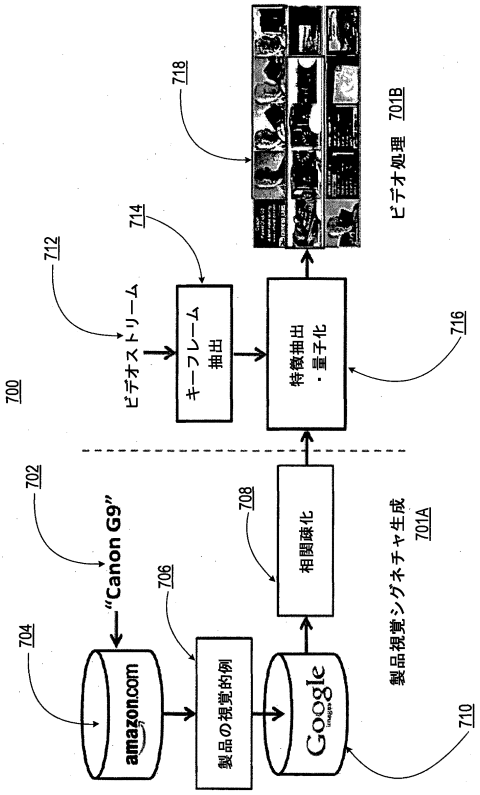
【図５】


所与の製品の専門製品画像をサーチする
510
専門製品画像を記憶する
520
各専門製品画像について
ウェブマイニングによってより多くの関連する製品画像を収集する
530
訓練サンプル製品画像を生成する
540
製品視覚シグネチャを生成する
550
製品視覚シグネチャファイルを集約する
560

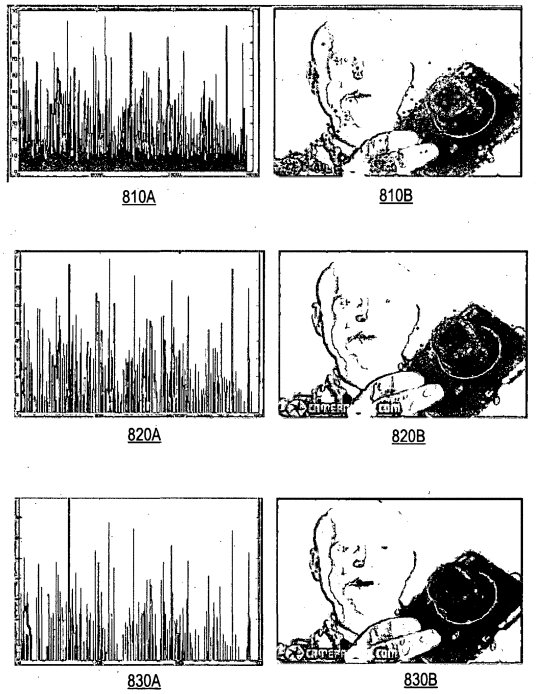
【図６】


ビデオストリームを受け取る
610
複数のキーフレームを抽出する
620
各キーフレームについて
キーフレームの視覚特徴を抽出する
630
キーフレームの視覚表現を生成する
640
視覚表現を製品視覚シグネチャと比較する
650
キーフレームに含まれる公知の製品の存在を判定する
660

【図７】


700
ビデオストリーム
712
キーフレーム抽出
714
718
ビデオ処理 701B
特徴抽出・量子化
716
"Canon G9"
702
amazon.com
704
製品の視覚的例
706
Google
710
相関確率化
708
製品視覚シグネチャ生成 701A

【図８】


810A
810B
820A
820B
830A
830B

フロントページの続き

(74)代理人 100142929
弁理士 井上 隆一
(74)代理人 100148699
弁理士 佐藤 利光
(74)代理人 100128048
弁理士 新見 浩一
(74)代理人 100129506
弁理士 小林 智彦
(74)代理人 100114340
弁理士 大関 雅人
(74)代理人 100114889
弁理士 五十嵐 義弘
(74)代理人 100121072
弁理士 川本 和弥
(72)発明者 チュア タット セン
シンガポール共和国 シンガポール ローワー ケント リッジ ロード ２１ ナショナル ユニヴァーシティー オブ シンガポール スクール オブ コンピューティング ディパートメント オブ コンピューター サイエンス内
(72)発明者 リ グアンダ
シンガポール共和国 シンガポール ローワー ケント リッジ ロード ２１ ナショナル ユニヴァーシティー オブ シンガポール エヌユーエス グラデュエート スクール オブ インタラクティブ サイエンス アンド エンジニアリング内
(72)発明者 ル ゼン
シンガポール共和国 シンガポール ローワー ケント リッジ ロード ２１ ナショナル ユニヴァーシティー オブ シンガポール スクール オブ コンピューティング ディパートメント オブ コンピューター サイエンス内
(72)発明者 ワン メン
シンガポール共和国 シンガポール ローワー ケント リッジ ロード ２１ ナショナル ユニヴァーシティー オブ シンガポール スクール オブ コンピューティング ディパートメント オブ コンピューター サイエンス内

審査官 小太刀 慶明

(56)参考文献 特開２００８－１４６４９４（ＪＰ，Ａ）
特表２０１１－５０７６０３（ＪＰ，Ａ）
特開２００５－２１５９２２（ＪＰ，Ａ）
特開２００７－１１０７０９（ＪＰ，Ａ）
特開２００３－０２３５９５（ＪＰ，Ａ）
特開２００３－０４４７１７（ＪＰ，Ａ）
米国特許出願公開第２００７／００８３８１５（ＵＳ，Ａ１）
米国特許出願公開第２００６／０２１８５７８（ＵＳ，Ａ１）
米国特許出願公開第２００８／００５９８７２（ＵＳ，Ａ１）
柳井 啓司，一般物体認識における機械学習の利用，電子情報通信学会技術研究報告，日本，社団法人電子情報通信学会，２０１０年 ６月 ７日，Ｖｏｌ．１１０ Ｎｏ．７６，ｐ.１０３～１１２

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １７／３０