

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6626843号
(P6626843)

(45) 発行日 令和1年12月25日(2019.12.25)

(24) 登録日 令和1年12月6日(2019.12.6)

(51) Int.Cl.

F 1

G06K 9/20 (2006.01)
G06T 7/12 (2017.01)
HO4N 21/435 (2011.01)
HO4N 21/4722 (2011.01)

GO6K 9/20 340K
 GO6T 7/12
 HO4N 21/435
 HO4N 21/4722

請求項の数 20 (全 25 頁)

(21) 出願番号 特願2016-569905 (P2016-569905)
 (86) (22) 出願日 平成27年5月27日 (2015.5.27)
 (65) 公表番号 特表2017-522648 (P2017-522648A)
 (43) 公表日 平成29年8月10日 (2017.8.10)
 (86) 国際出願番号 PCT/US2015/032618
 (87) 国際公開番号 WO2015/183914
 (87) 国際公開日 平成27年12月3日 (2015.12.3)
 審査請求日 平成30年5月28日 (2018.5.28)
 (31) 優先権主張番号 14/289, 142
 (32) 優先日 平成26年5月28日 (2014.5.28)
 (33) 優先権主張国・地域又は機関
米国(US)

(73) 特許権者 501112323
 グレースノート 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
 608, エミリービル, スイート 138
 O, パウウェル・ストリート 2000
 (74) 代理人 100086771
 弁理士 西島 孝喜
 (74) 代理人 100088694
 弁理士 弟子丸 健
 (74) 代理人 100094569
 弁理士 田中 伸一郎
 (74) 代理人 100067013
 弁理士 大塚 文昭
 (74) 代理人 100109070
 弁理士 須田 洋之

最終頁に続く

(54) 【発明の名称】ビデオ内のテキスト検出

(57) 【特許請求の範囲】

【請求項1】

コンピュータにより実行される方法であって、
 メモリ及び少なくとも1つのプロセッサを有する機械によって、ビデオコンテンツの一部がテキストを有すると識別する段階であって、
 前記ビデオコンテンツのフレームに対して連結成分分析を行って、前記フレーム内の連結成分を検出する段階と、

前記連結成分を複数のテキスト行にまとめる段階と、
 前記複数のテキスト行から少なくとも1つのテキスト行を除去するために、水平及び垂直投影を用いて前記複数のテキスト行を精細化する段階と、

前記複数のテキスト行の前記少なくとも1つのサイズに基づいて、前記複数のテキスト行の少なくとも1つをフィルタ除去して、フィルタ済みのテキスト行セットを形成する段階と、

前記複数のテキスト行の前記少なくとも1つをフィルタ除去することにより形成された前記フィルタ済みのテキスト行セットを二値化する段階と、

前記テキスト行の少なくとも1つにおける成分の形状及び前記テキスト行の少なくとも1つにおける成分の位置のうちの少なくとも1つに基づいて、前記二値化されたフィルタ済みのテキスト行セットから前記テキスト行の少なくとも1つをフィルタ除去して、テキストを有する前記ビデオコンテンツの一部分を形成する段階と、
 を含む段階と、

10

20

前記ビデオコンテンツの前記識別された部分内の前記テキストを識別する段階と、

前記識別されたテキストのカテゴリを判定する段階と、

前記識別されたテキスト及び前記判定されたカテゴリに基づいて、前記ビデオコンテンツに対応する付加情報を判定する段階と、

メディアコンテンツデバイス上のソフトウェアアプリケーションに、前記付加情報を用いた機能であって、前記判定されたカテゴリに対応する機能を実行させる段階と、
を含む方法。

【請求項 2】

前記ビデオコンテンツが少なくとも 1 つの所定の条件を満たすか否かを判定する段階を
さらに含み、

10

前記ビデオコンテンツの一部分がテキストを有すると識別する段階を実行することは、
前記ビデオコンテンツが前記少なくとも 1 つの所定の条件を満たすという判定が条件とさ
れる、請求項 1 に記載の方法。

【請求項 3】

前記少なくとも 1 つの所定の条件は、最低レベルの明瞭性、最低レベルのコントラスト
、及び複数のフレームにわたる最低レベルのコンテンツ安定性のうちの少なくとも 1 つを
含む、請求項 2 に記載の方法。

【請求項 4】

前記付加情報をメディアコンテンツデバイス上に表示させる段階をさらに含む、請求項
1 に記載の方法。

20

【請求項 5】

前記ビデオコンテンツに関連して又は前記ビデオコンテンツの識別された視聴者に関連
して、前記付加情報を格納する段階をさらに含む、請求項 1 に記載の方法。

【請求項 6】

前記付加情報を メディア コンテンツデバイス上のソフトウェアアプリケーションに提供
する段階をさらに含む、請求項 1 に記載の方法。

【請求項 7】

前記付加情報は、統一資源位置指定子(U R L)、ユーザーアカウントの識別、メタデータ
タグ、又は電話番号のうちの少なくとも 1 つを含む、請求項 6 に記載の方法。

【請求項 8】

30

前記メディアコンテンツデバイスは、テレビ、ラップトップコンピュータ、デスクトップコンピュータ、タブレットコンピュータ、又はスマートフォンのうちの 1 つを含む、請
求項 6 に記載の方法。

【請求項 9】

前記ビデオコンテンツに関連して又は前記ビデオコンテンツの識別された視聴者に関連
して、前記識別されたテキストを格納する段階をさらに含む、請求項 1 に記載の方法。

【請求項 10】

テキストを有する前記ビデオコンテンツの一部分を識別する段階は、

前記ビデオコンテンツのフレームをグレースケールに変換する段階と、

前記フレームに対してエッジ検出を実行する段階と、

40

前記フレームに対してダイレーションを実行して、前記フレーム内の垂直エッジを連結
する段階と、

前記フレームを二値化する段階と、

をさらに含む、請求項 1 に記載の方法。

【請求項 11】

前記ビデオコンテンツの前記識別された部分内のテキストを識別する段階は、前記ビデ
オコンテンツの前記識別された部分に対して光学文字認識を行う段階を含む、請求項 1 に
記載の方法。

【請求項 12】

前記識別されたテキストの前記カテゴリを判定する段階は、

50

前記識別されたテキストを構文解析して、前記識別されたテキストの複数のセグメントを判定する段階と、

前記複数のセグメントの少なくとも1つと前記カテゴリとの間の格納済みの関連性に基づいて前記カテゴリを判定する段階と、
を含む、請求項1に記載の方法。

【請求項13】

前記ビデオコンテンツは、テレビ番組、非エピソード映画、ウェビソード、ビデオ共有ウェブサイト用のユーザ生成コンテンツ、又はコマーシャルの一部分を含む、請求項1に記載の方法。

【請求項14】

メモリ及び少なくとも1つのプロセッサを有する機械と、

前記機械上の少なくとも1つのモジュールと、
を備えたシステムであって、

前記少なくとも1つのモジュールは、

ビデオコンテンツの一部分がテキストを有すると識別する段階であって、

前記ビデオコンテンツのフレームに対して連結成分分析を行って、前記フレーム内の連結成分を検出する段階と、

前記連結成分を複数のテキスト行にまとめる段階と、

前記複数のテキスト行から少なくとも1つのテキスト行を除去するために、水平及び垂直投影を用いて前記複数のテキスト行を精細化する段階と、

前記複数のテキスト行の前記少なくとも1つのサイズに基づいて、前記複数のテキスト行の少なくとも1つをフィルタ除去して、フィルタ済みのテキスト行セットを形成する段階と、

前記複数のテキスト行の前記少なくとも1つをフィルタ除去することにより形成された前記フィルタ済みのテキスト行セットを二値化する段階と、

前記テキスト行の少なくとも1つにおける成分の形状及び前記テキスト行の少なくとも1つにおける成分の位置のうちの少なくとも1つに基づいて、前記二値化されたフィルタ済みのテキスト行セットから前記テキスト行の少なくとも1つをフィルタ除去して、テキストを有する前記ビデオコンテンツの一部分を形成する段階と、

を含む段階と、

前記ビデオコンテンツの前記識別された部分内の前記テキストを識別する段階と、

前記識別されたテキストのカテゴリを判定する段階と、

前記識別されたテキスト及び前記判定されたカテゴリに基づいて、前記ビデオコンテンツに対応する付加情報を判定する段階と、

メディアコンテンツデバイス上のソフトウェアアプリケーションに、前記付加情報を用いた機能であって、前記判定されたカテゴリに対応する機能を実行させる段階と、
を含む演算を実行するように構成されるシステム。

【請求項15】

前記演算は、

前記ビデオコンテンツが少なくとも1つの所定の条件を満たすか否かを判定する段階と、

前記ビデオコンテンツが前記少なくとも1つの所定の条件を満たすという判定に応答して、前記ビデオコンテンツの一部分がテキストを有すると識別する段階と、
をさらに含む、請求項14に記載のシステム。

【請求項16】

前記少なくとも1つの所定の条件は、最低レベルの明瞭性、最低レベルのコントラスト、及び複数のフレームにわたる最低レベルのコンテンツ安定性のうちの少なくとも1つを含む、請求項15に記載のシステム。

【請求項17】

前記演算は、前記付加情報をメディアコンテンツデバイス上のソフトウェアアプリケー

10

20

30

40

50

ションに提供する段階をさらに含む、請求項 1 4 に記載のシステム。

【請求項 1 8】

少なくとも 1 つのプロセッサによって実行されると、前記少なくとも 1 つのプロセッサに、以下を含む一連の演算を実行させる命令セットを有形で具現化する非一時的機械可読ストレージデバイスであって、前記演算は、

ビデオコンテンツの一部分がテキストを有すると識別する段階であって、

前記ビデオコンテンツのフレームに対して連結成分分析を行って、前記フレーム内の連結成分を検出する段階と、

前記連結成分を複数のテキスト行にまとめる段階と、

前記複数のテキスト行から少なくとも 1 つのテキスト行を除去するために、水平及び垂直投影を用いて前記複数のテキスト行を精細化する段階と、

前記複数のテキスト行の前記少なくとも 1 つのサイズに基づいて、前記複数のテキスト行の少なくとも 1 つをフィルタ除去して、フィルタ済みのテキスト行セットを形成する段階と、

前記複数のテキスト行の前記少なくとも 1 つをフィルタ除去することにより形成された前記フィルタ済みのテキスト行セットを二値化する段階と、

前記テキスト行の少なくとも 1 つにおける成分の形状及び前記テキスト行の少なくとも 1 つにおける成分の位置のうちの少なくとも 1 つに基づいて、前記二値化されたフィルタ済みのテキスト行セットから前記テキスト行の少なくとも 1 つをフィルタ除去して、テキストを有する前記ビデオコンテンツの一部分を形成する段階と、

を含む段階と、

前記ビデオコンテンツの前記識別された部分内の前記テキストを識別する段階と、

前記識別されたテキストのカテゴリを判定する段階と、

前記識別されたテキスト及び前記判定されたカテゴリに基づいて、前記ビデオコンテンツに対応する付加情報を判定する段階と、

メディアコンテンツデバイス上のソフトウェアアプリケーションに、前記付加情報を用いた機能であって、前記判定されたカテゴリに対応する機能を実行させる段階と、
を含むストレージデバイス。

【請求項 1 9】

前記付加情報は、統一資源位置指定子 (U R L) を含み、前記メディアコンテンツデバイス上の前記ソフトウェアアプリケーションに、前記 U R L が前記メディアコンテンツデバイス上のブラウザ上にロードされることを含む前記機能を実行させる請求項 1 に記載の方法。

【請求項 2 0】

前記付加情報は、電話番号を含み、前記メディアコンテンツデバイス上の前記ソフトウェアアプリケーションに、前記メディアコンテンツデバイスに前記電話番号に発呼することとのプロンプトを提供させることを含む前記機能を実行させる請求項 1 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

(優先権の主張)

本出願は、引用により内容全体が本明細書に組み込まれている 2 0 1 4 年 5 月 2 8 日出願の米国特許出願番号第 1 4 / 2 8 9 , 1 4 2 号への優先権の利益を主張する。

【0 0 0 2】

(技術分野)

本出願は、一般に、データ処理の技術分野に関し、種々の実施形態において、ビデオ内のテキストを検出する方法及びシステムに関する。

【背景技術】

【0 0 0 3】

テレビ番組、映画、及びコマーシャル等のビデオコンテンツは、テキストを含む場合が

10

20

30

40

50

ある。このテキストは、視聴者に役立つ情報を提供することができる。しかしながら、視聴者は、テキストを利用するという点では限られた選択肢を有する。例えば、視聴者は通常、後で使用するためにテキストを書き留める以外の選択肢はほとんどない。

【0004】

本開示の一部の実施形態は、例示的に示されており、同様の参照番号が類似する要素を示す添付の図面に限定されるものではない。

【図面の簡単な説明】

【0005】

【図1】一部の例示的な実施形態による、テキスト検出システムを示すブロック図である。

10

【図2】一部の例示的な実施形態による、テキスト検出システムの機能を実施することができる環境を示すブロック図である。

【図3A】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3B】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3C】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3D】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

20

【図3E】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3F】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3G】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3H】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

【図3I】一部の例示的な実施形態による、ビデオコンテンツのフレーム内のテキストを検出して解釈する段階を示す図である。

30

【図4】一部の例示的な実施形態による方法を示すフローチャートである。

【図5】一部の例示的な実施形態による方法を示すフローチャートである。

【図6】一部の例示的な実施形態による方法を示すフローチャートである。

【図7】一部の例示的な実施形態によるモバイルデバイスを示すブロック図である。

【図8】一部の例示的な実施形態による、本明細書で説明する方法論を実施可能な例示的なコンピュータシステムのブロック図である。

【発明を実施するための形態】

【0006】

ビデオ内のテキストを検出する例示的な方法及びシステムが開示される。以下の説明では、説明目的で、例示的な実施形態の完全な理解をもたらすために多くの具体的な詳細が記載されている。しかしながら、当業者であれば、本実施形態は、これらの具体的な詳細がなくても実施可能であることは明らかであろう。

40

【0007】

本明細書に開示するように、テキスト検出システムは、ビデオコンテンツ内にオーバーレイ表示されたテキストを効果的に認識するように構成及び使用することができる。テキストは、情報又は情報ソースに関係すると分析されて識別することができ、限定ではないが、統一資源位置指定子(ＵＲＬ)、ユーザーアカウントの識別(例えばTwitter(登録商標)ハンドル)、メタデータタグ(例えばハッシュタグ)、及び電話番号を含む。この分析の結果は(例えば、識別されたテキスト、情報、及び/又は情報ソースは)、データベース内に格納すること及び/又はメディアコンテンツデバイス上のソフトウェアア

50

プリケーションに提供することができる。

【0008】

一部の実施形態において、ビデオコンテンツの一部分がテキストを有すると識別される。ビデオコンテンツの識別された部分内のテキストを識別することができる。識別されたテキストのカテゴリを判定することができる。

【0009】

一部の実施形態において、ビデオコンテンツが少なくとも1つの所定の条件を満たすか否かが判定され、ビデオコンテンツが所定の条件を満たすという判定に応じて、ビデオコンテンツの一部分がテキストを有すると識別される。一部の実施形態において、所定の条件は、最低レベルの明瞭性、最低レベルのコントラスト、及び複数のフレームにわたる最低レベルのコンテンツ安定性のうちの少なくとも1つを含む。10

【0010】

一部の実施形態において、識別されたテキスト及び判定されたカテゴリに基づいて、ビデオコンテンツに対応する付加情報が判定される。付加情報は、メディアコンテンツデバイス上に表示させることができる。付加情報は、ビデオコンテンツに関連して又はビデオコンテンツの識別された視聴者に関連して格納することができる。付加情報は、メディアコンテンツデバイス上のソフトウェアアプリケーションに提供することができる。付加情報は、統一資源位置指定子（URL）、ユーザーアカウント識別、メタデータタグ、及び電話番号のうちの少なくとも1つを含むことができる。メディアコンテンツデバイスは、テレビ、ラップトップコンピュータ、デスクトップコンピュータ、タブレットコンピュータ、及びスマートフォンのうちの1つを含むことができる。識別されたテキストは、ビデオコンテンツに関連して又はビデオコンテンツの識別された視聴者に関連して格納することができる。20

【0011】

一部の実施形態において、テキストを有するビデオコンテンツの部分を識別する段階は、ビデオコンテンツのフレームをグレースケールに変換する段階と、フレームに対してエッジ検出を実行する段階と、フレームに対してダイレーションを実行してフレーム内の垂直エッジを連結する段階と、フレームを二値化する段階と、フレームに対して連結成分分析を実行してフレーム内の連結成分を検出する段階と、連結成分を複数のテキスト行にまとめる段階と、水平及び垂直投影を用いて複数のテキスト行を精細化する段階と、複数のテキスト行の少なくとも1つのサイズに基づいて、複数のテキスト行の少なくとも1つをフィルタ除去して、フィルタ済みのテキスト行セットを形成する段階と、フィルタ済みのテキスト行セットを二値化する段階と、テキスト行の少なくとも1つにおける成分の形状及びテキスト行の少なくとも1つにおける成分の位置のうちの少なくとも1つに基づいて、二値化されたフィルタ済みのテキスト行セットからテキスト行の少なくとも1つをフィルタ除去して、テキストを有するビデオコンテンツの一部分を形成する段階と、を含む。一部の実施形態において、複数のテキスト行の少なくとも1つのサイズに基づいて複数のテキスト行の少なくとも1つをフィルタ除去する段階は、処理から省略することができる。例えば、テキスト行の全てが所定の基準を満たすと判定される場合、このフィルタリング段階は省くことができる。30

【0012】

一部の実施形態において、ビデオコンテンツの識別された部分内のテキストを識別する段階は、ビデオコンテンツの識別された部分に対して光学文字認識を実行する段階を含む。

【0013】

一部の実施形態において、識別されたテキストのカテゴリを判定する段階は、識別されたテキストを構文解析して、識別されたテキストの複数のセグメントを判定する段階と、複数のセグメントの少なくとも1つとカテゴリとの間の格納済みの関係に基づいてカテゴリを判定する段階と、を含む。

【0014】

10

20

30

40

50

一部の実施形態において、ビデオコンテンツは、テレビ番組、非エピソード映画、ウェビソード、ビデオ共有ウェブサイト用のユーザ生成コンテンツ、又はコマーシャルの一部分を含む。

【0015】

本明細書に開示する方法又は実施形態は、1又は2以上のモジュール（例えば、ハードウェアモジュール又はソフトウェアモジュール）を有するコンピュータシステムとして実装可能である。このようなモジュールは、コンピュータシステムの1又は2以上のプロセッサによって実行可能である。本明細書に開示する方法又は実施形態は、1又は2以上のプロセッサによって実行されると、1又は2以上のプロセッサに命令を実行させる、機械可読媒体上に格納された命令として具現化することができる。

10

【0016】

図1は、例示的な実施形態によるテキスト検出システム100を示すブロック図である。一部の実施形態において、テキスト検出システム100は、コンテンツ選択モジュール110と、テキスト検出モジュール120と、テキスト認識モジュール130と、テキスト解釈モジュール140と、情報モジュール150と、1又は2以上のデータベース160と、を備える。

【0017】

一部の実施形態において、テキスト検出モジュール120は、ビデオコンテンツの一部分がテキストを有すると識別するように構成される。ビデオコンテンツの一部分がテキストを有すると識別する段階は、ビデオコンテンツ内のテキストの場所を識別する段階を含むことができる。ビデオコンテンツの一部分がテキストを有すると識別するための技術については、本開示において以下で議論することにする。ビデオコンテンツは、テレビ番組、非エピソード映画、ウェビソード、ビデオ共有ウェブサイト（例えば、YouTube（登録商標））用のユーザ生成コンテンツ、又はコマーシャルの一部分を含むことができる。他のタイプのビデオコンテンツも、本開示の範囲内である。ビデオコンテンツは、ビデオコンテンツの1又は2以上のフレームを含むことができる。

20

【0018】

ビデオコンテンツの一部分がテキストを有すると識別することなどによって、ビデオコンテンツのフレームを処理する段階は、計算的に及び費用的に高価となり得る。本開示は、この費用を軽減するための技術を紹介する。一部の実施形態において、コンテンツ選択モジュール110は、テキストを含む可能性が高いフレームを識別するために、フレームの事前分析を実行するように構成される。次に、本明細書に開示するテキスト検出機能は、識別されたこれらのフレームに対して実行することができるが、テキストを含む可能性が低いと判定されたフレームは、テキスト検出処理から省くことができ、これにより費用を削減する。上述の事前分析は、全域(global)及び局所コントラスト、ぼけ、全域及び局所の動き、及びフレームの他の特徴の分析を含むことができる。人間がテキストを容易に読み取るために、テキストは、通常、一定のサイズであり、良好なコントラスト（例えば、黒地に白、白地に黒、明対暗、その他）を有し、複数のフレームにわたってほとんど静止している。また、ビデオコンテンツの製作者からの意図は、人々がテキストを見て読み取ることなので、一般的には、テキストから眼をそらす動き等の、画像内の余計な局所又は全域の動きは存在しない。これらのガイドラインは、本明細書で説明するテキスト検出技術でさらに分析するための有力候補を識別するために用いることができる。

30

【0019】

一部の実施形態において、コンテンツ選択モジュール110は、ビデオコンテンツが少なくとも1つの所定の条件を満たすか否かを判定するように構成される。テキスト検出モジュール120は、ビデオコンテンツの1又は2以上の部分がテキストを有すると識別するため、コンテンツ選択モジュール110によって肯定的判定が為されたビデオコンテンツに対しては自己の動作を実行することができるが、否定的判定が為されたビデオコンテンツは、このような動作から省くことができる。一部の実施形態において、所定の条件は、最低レベルの明瞭性、最低レベルのコントラスト、及び複数のフレームにわたる最低

40

50

レベルのコンテンツ安定性のうちの少なくとも 1 つを含む。他の所定の条件も、本開示の範囲内である。

【 0 0 2 0 】

テキストを含む可能性が高いフレームが識別されると、複数のフレームにわたって平均化することなどによって、残りの動き及びノイズを除去するためにこれらのフレームを使用することができる。一部の実施形態において、連続フレームの補間によって解像度を上げることができ、これにより、テキストのより良好な識別が可能になる。

【 0 0 2 1 】

多くの文字言語では、特定の構成的方向が支配的である。例えば、ラテン文字では、これらの方針は水平及び垂直線である。この概念をテキスト検出モジュール 120 によって利用して、対象領域を更に狭めることができる。また、フレーム内のテキストの有無の判定に関して本明細書で議論した同じルール及び条件（例えば、コントラスト、ぼけ、動き）をここで用いて、フレーム内の対象領域を識別することができる。10

【 0 0 2 2 】

一部の実施形態において、テキスト認識モジュール 130 は、ビデオコンテンツの識別された部分内のテキストを識別するように構成される。ビデオコンテンツの識別された部分内のテキストを識別する段階は、ビデオコンテンツの識別された部分に対して OCR を実行する段階を含むことができる。ビデオコンテンツの識別された部分内のテキストを識別する他の技術も同様に本開示の範囲内である。

【 0 0 2 3 】

一部の実施形態において、テキスト認識モジュール 130 は、識別されたテキストのカテゴリを判定するように構成される。カテゴリの例は、限定ではないが、URL、ユーザーアカウント識別、メタデータタグ、及び電話番号を含む。他のタイプのカテゴリも同様に本開示の範囲内である。識別されたテキストのカテゴリを判定する段階は、識別されたテキストを構文解析して、識別されたテキストの複数のセグメントを判定する段階と、次に、セグメントの少なくとも 1 つとカテゴリとの間の格納済みの関連性に基づいてカテゴリを判定する段階と、を含むことができる。関連性は、データベース 160 内に格納することができる。一部の実施形態において、関連性は、テキストのセグメントが対応するカテゴリを判定するためのルールを含むことができる。1 つの実施例において、テキスト「www.examplewebsite.com」を、テレビ番組のビデオコンテンツの上に表示することができる。テキスト認識モジュール 130 は、「www.」又は「.com」の使用に基づいて、このテキストがウェブサイトを指していると判定することができ、或いは、テキスト「www.examplewebsite.com」が、データベース 160 内に格納された同じテキスト（例えば、「www.examplewebsite.com」）又は類似するテキスト（例えば、「examplewebsite.com」）に一致することに基づいて、このテキストがウェブサイトを指していると判定することができる。一部の実施形態において、テキスト認識モジュール 130 は、対応するテキスト及びカテゴリのルックアップテーブルへのアクセスに基づいて、カテゴリを判定することができる。識別されたテキストは、ビデオコンテンツに関連して又はビデオコンテンツの識別された視聴者に関連して格納することができる。次に、以下で更に詳細に議論するように、格納されたテキストは、アクセスして読み出して、さらなる処理で使用することができる。30

【 0 0 2 4 】

一部の実施形態において、情報モジュール 150 は、識別されたテキスト及び判定されたカテゴリに基づいて、ビデオコンテンツに対応する付加情報を判定するように構成される。付加情報は、統一資源位置指定子（URL）、ユーザーアカウントの識別（例えば Twitter（登録商標）ハンドル）、メタデータタグ（例えばハッシュタグ）、及び電話番号のうちの少なくとも 1 つを含むことができる。付加情報は、ビデオコンテンツに関連して又はビデオコンテンツの識別された視聴者に関連して格納することができる。付加情報の履歴（例えばリンク及び参照）をデータベース 160 内に又はメディアコンテンツデ40

バイス上に保持して、ユーザ（例えば視聴者）が、付加情報を用いてフォローアップ及び／又はさらなる調査を行うことができるようとする。一部の実施形態において、メディアコンテンツデバイスは、ビデオコンテンツ及び／又はオーディオコンテンツ等のメディアコンテンツを再生するように構成された何らかの電子デバイスを含むことができる。メディアコンテンツデバイスの例は、限定ではないが、テレビ、ラップトップコンピュータ、デスクトップコンピュータ、タブレットコンピュータ、スマートフォン、ゲーム用デバイス、及び／又はオーディオ、ビデオ、及び／又は他のマルチメディアコンテンツを受信して表示する能力がある他のデバイスを含む。

【0025】

付加情報は、メディアコンテンツデバイス上のソフトウェアアプリケーションに提供することができる。付加情報はまた、メディアコンテンツデバイス上に表示させることもできる。例えば、リンクに対応するテキストの検出に応答して、リンクを自動的にソフトウェアアプリケーション（例えばブラウザ）にロードして、ビデオコンテンツ又はビデオコンテンツ内の検出されたテキストに対応するページ（例えばウェブページ）を表示することができる。別の実施例では、識別されたテキスト、判定されたカテゴリ、及び／又は判定された付加情報に基づいて、特定のキーワードに関するインターネット検索又は特定のサービス検索を自動的に実行することができる。例えば、テレビ番組（又は何らかの他のビデオコンテンツ）のエンドクレジットで検出される名前に応答して、検出された名前を用いて検索を自動的に実行して、これらの名前に対応する人々（例えば、俳優、女優、他の寄与者）についての情報を集めることができる。一部の実施形態において、キーワード又は記号等の検出されたテキスト又は判定された付加情報に関して取るべきアクションを識別するために、特定の予め定義された指示を使用することができる。例えば上記の例において、ビデオコンテンツのエンドクレジットで検出された用語「キャスト」は、用語「キャスト」に続いて検出された名前にに関する文脈（context）及び意味を与えるために用いることができる。その後、判定された文脈及び意味に対応する1又は2以上の動作を実行することができる。検出されたテキストがビデオコンテンツのキャストメンバーの名前に対応するという判定に基づいて、名前を、サービス（例えばIMDb（登録商標））を用いて検索することができる。検索は、名前に対応する付加情報、例えば俳優／女優の名前に対応する架空のキャラクタ等を提示することができる。検出されたテキスト、判定されたカテゴリ、判定された付加情報の1又は2以上のいずれかの組合せは、対応するビデオコンテンツに関連して及び／又は対応するビデオコンテンツの対応する視聴者等のユーザに関連して、データベース160に格納することができる。

【0026】

図2は、例示的な実施形態による、テキスト検出システム100の機能を実施することができる環境200を示すブロック図である。テキスト検出システム100は、本明細書に開示するテキスト検出機能をビデオコンテンツプロバイダ210、第1メディアコンテンツデバイス220又は第2メディアコンテンツデバイス240等のメディアコンテンツデバイス、及び／又はメディアチューナ／プレーヤデバイス230に提供するように構成することができる。

【0027】

ビデオコンテンツプロバイダ210は、ビデオコンテンツをメディアコンテンツデバイスに提供するように構成されたコンピュータシステムを備えることができる。一部の実施形態において、ビデオコンテンツプロバイダ210は、ビデオコンテンツの放送局（例えば、ケーブルテレビ放送局、衛星ラジオ放送局、ウェブ放送局）を含む。前述のように、ビデオコンテンツは、テレビ番組、非エピソード映画、ウェビソード、ビデオ共有ウェブサイト用のユーザ生成コンテンツ、又はコマーシャルを含むことができる。他のタイプのビデオコンテンツも本開示の範囲内である。

【0028】

一部の実施形態において、第1メディアコンテンツデバイス220及び第2メディアコンテンツデバイス240は、ビデオコンテンツを再生するように構成される。一部の実施

10

20

30

40

50

形態において、第1メディアコンテンツデバイス220は、スマートテレビを含む。従つて、第1メディアコンテンツデバイス220は、インターネットベースのオンデマンドメディアコンテンツの再生を可能にするために1又は2以上のスマートTVアプリケーション222(例えば、Netflix(登録商標)又はHulu(登録商標))を含むことができる。他の構成も本開示の範囲内である。一部の実施形態において、第2メディアコンテンツデバイス240は、様々な方法で第1メディアコンテンツデバイス220と同時に使用することができる。一部の実施形態において、第2メディアコンテンツデバイス240は、第1メディアコンテンツデバイス220を制御するためのリモートコントロールとして使用することができる。一部の実施形態において、第2メディアコンテンツデバイス240は、第1メディアコンテンツデバイス220上に提示されているビデオコンテンツに対応する付加情報を表示するために使用することができる。10

【0029】

1つの実施例において、ユーザは、第2メディアコンテンツデバイス240を保持しながら第1メディアコンテンツデバイス上でテレビコマーシャルを視聴することができる。テレビコマーシャルはテキストを含むことがあり、本明細書に開示するように、このテキストをテキスト検出システム100によって検出して解釈することができる。テキスト検出システムは、テレビコマーシャルに対応するURL(例えば、コマーシャルで宣伝されている製品のウェブサイト)等の検出されたテキストに基づいて、付加情報を判定することができる。テキスト検出システム100は、第2メディアコンテンツデバイス240上への表示又はそこでのさらなる処理のために、第2メディアコンテンツデバイス240に付加情報を提供することができる。URLを含む付加情報の例では、第2メディアコンテンツデバイス240のウェブブラウザ上でURLを開くことができるので、ユーザが、コマーシャルの製品についての付加情報を閲覧できる。一部の実施形態では、ビデオコンテンツ内で電話番号をテキスト検出システム100によって検出することができ、次にテキスト検出システム100は、この電話番号にかけたいか否かをユーザに尋ねるためのプロンプトをユーザのモバイルデバイス(例えばスマートフォン)上に自動的に提供することができる。他の実施例及び構成も本開示の範囲内である。20

【0030】

一部の実施形態において、メディアコンテンツデバイスは、メディアチューナ/プレーヤデバイス230を利用することができ、このデバイス230は、コンテンツプロバイダ(例えば、ケーブル又は衛星テレビプロバイダ、ウェブテレビプロバイダ、地上波又は衛星ラジオ局、オンライン音楽サービス、その他)からビデオ及び/又はオーディオコンテンツのストリームを受信して、メディアコンテンツデバイスのディスプレイによって利用可能なストリーム及び出力情報(例えば、デジタル又はアナログ)を処理することによってコンテンツのストリームを再生して、メディアコンテンツデバイスを利用するユーザにビデオ及び/又はオーディオコンテンツを提示又は再生するように構成される。メディアチューナ/プレーヤデバイス230の例は、限定ではないが、セットトップユニット(例えばケーブルボックス)及びデジタルメディアプレーヤ(例えば、Apple TV(登録商標)プレーヤ又はRoku(登録商標)プレーヤ)を含む。一部の実施形態において、メディアチューナ/プレーヤデバイス230は、このサービスを利用しているメディアコンテンツデバイスとは別体でその外部とすることができます。他の実施形態において、メディアチューナ/プレーヤデバイス230は、このサービスを利用しているメディアコンテンツデバイス内に組み込むことができる。30

【0031】

本明細書に開示するデバイス及びシステムは、他のデバイス及びシステムと様々な方法で通信することができる。一部の実施形態において、デバイス及びシステムは、1又は2以上のネットワークを介して通信することができる。使用可能なネットワークの例は、限定ではないが、有線ネットワーク、無線ネットワーク(例えば、モバイルネットワーク又はWi-Fiネットワーク)、無線又は電気通信ネットワーク、衛星ネットワーク、ケーブルネットワーク等を含む。一部の実施形態において、ネットワークは、プライベートネ4050

ットワーク（例えば、ケーブルテレビネットワーク又は衛星ラジオネットワーク）、公衆ネットワーク（例えば、無線（over the air）放送チャンネル又はインターネット）等を構成する1又は2以上の部分を含むことができる。デバイスとシステムとの間の他の構成及び情報通信方法も本開示の範囲内である。

【0032】

一部の実施形態において、検出されたテキスト、判定されたカテゴリ、及び／又は判定された付加情報は、ユーザアカウントに関連してデータベース160内に格納することができる。各ユーザアカウントは、特定のメディアコンテンツデバイス、特定の人物、及び／又は特定の人物グループに関連付けること及びこれらに対応することができる。例えば、1つのユーザアカウントは、家族のリビングルームの特定のテレビに対応することができる。別の実施例では、別のユーザアカウントは、特定の1人の家族員がどのメディアコンテンツデバイスを使用しているかを問わず、この家族員に対応することができる。この点において、「ユーザアカウント」及び「ユーザのアカウント」という用語は、指定されたユーザのアカウント、指定されたユーザの指定されたグループ（例えば家族員）のアカウント、或いは1又は2以上の指定されたデバイス（例えば家族のキッチンのTV）のアカウントを指すことができ、特許請求の範囲に明確に列挙されていない限り、1人の特定のユーザに属するアカウントのみに限定されると解釈されるべきではない。10

【0033】

各ユーザアカウントは、ユーザアカウントに対応するメディアコンテンツデバイス上に提示されたビデオコンテンツ内でどんなテキストが検出されたか、及び／又はユーザアカウントに対応する人物が視聴したビデオコンテンツ内でどんなテキストが検出されたかの履歴を含むことができる。さらに、各ユーザアカウントは、検出されたテキストに対応するカテゴリ及び／又は付加情報の履歴を含むことができる。これらの履歴の1又は2以上のいずれの組合せも、後で参照するためにユーザがアクセスして見ることができ、或いは、検出されたテキストに対応する付加情報を、対応するユーザアカウントに関連するユーザに提示するのに用いるために、ソフトウェアアプリケーションが自動的にアクセスすることができる。例えば、ユーザは、第1メディアコンテンツデバイス220上でテレビ番組を見ることができる。テレビ番組中に、テレビ番組のウェブサイトのURLを含むテキストが表示されることがある。URLは、テキスト検出システム100によって検出されて、ユーザに関連して格納することができる。テレビ番組終了後の数日後に、ユーザは、第2メディアコンテンツデバイス240上のウェブブラウザを使用することができ、テキスト検出システム100は、ウェブブラウザにテレビ番組のウェブサイトのURLを自動的にロードさせることができる。他の実施例及び構成も本開示の範囲内である。2030

【0034】

一部の実施形態において、テキスト検出システム100、又はこれの構成要素（例えばモジュール及び／又はデータベース）の1又は2以上のいずれの組合せは、ビデオコンテンツプロバイダ210、第1メディアコンテンツデバイス220、メディアチューナ／プレーヤデバイス230、及び第2メディアコンテンツデバイス240のうちのいずれか1つの上に常駐すること又はこれらに組み込むことができる。一部の実施形態では、テキスト検出システム100の構成要素の全ては、ビデオコンテンツプロバイダ210、第1メディアコンテンツデバイス220、メディアチューナ／プレーヤデバイス230、及び第2メディアコンテンツデバイス240のうちの1つに実装することができる。他の実施形態では、テキスト検出システム100の様々な構成要素は、異なるデバイス及びシステムにわたって分散させができる（例えば、テキスト検出モジュール120はビデオコンテンツプロバイダ210内に組み込まれ、テキスト認識モジュール130、テキスト解釈モジュール140、及び情報モジュール150は、第2メディアコンテンツデバイス240内に組み込まれる）。他の構成も本開示の範囲内である。40

【0035】

一部の実施形態において、テキスト検出システム100又はこれの構成要素の一部は、前述のメディアコンテンツデバイス220、240、又はメディアプレーヤ／チューナデ50

バイス 230 等のクライアントデバイス内に組み込むことができる。一部の実施形態において、テキスト検出システム 100 又はこれの構成要素の一部は、サーバシステム内に組み込むことができ、この場合、プロードキャストフィードはリアルタイムに分析され、次にクライアントデバイスは、抽出された情報を識別方法（例えば、フィンガープリンティング、ウォーターマーキング）を用いて読み出す。

【0036】

1つの例示的な実施形態において、第 2 スクリーンデバイス（例えば、タブレット又はスマートフォン）は、オーディオフィンガープリンティングを用いて、ユーザが TV で視聴中の番組を識別して、次にこの番組に関連する情報を第 2 スクリーンデバイスに対して引き出すことができる。この情報は、番組のオープニングクレジット上のテキストオーバーレイを用いて集めることができる。このような情報の例は、ソーシャルメディアリンク（例えば、Facebook（登録商標）又は Twitter（登録商標）リンク）を含むことができる。しかしながら、他のタイプの情報も本開示の範囲内である。10

【0037】

図 3A - 3I は、一部の例示的な実施形態による、ビデオコンテンツ 310 のフレーム 300 内のテキスト 320 を検出して解釈する様々な段階を示している。図 3A - 3H の各段階は、テキストを有するビデオコンテンツ部分を識別する段階に対応し、図 3I は、ビデオコンテンツの識別された部分内のテキストを識別して、識別されたテキストのカテゴリを判定する段階に対応する。20

【0038】

図 3A は、異なる場所に様々なテキスト 320 を有するビデオコンテンツ 310 のフレーム 300 を示している。一部の実施形態では、フレーム 300 は、グレースケールに変換することができる。次に、変換されたフレーム 300 に対してエッジ検出を実行することができる。エッジ検出とは、デジタル画像内で画像の明るさが激しく変化する又は途切れる地点を識別することを目的とする一連の数学的方法の名称である。一部の実施形態において、ソーベル（Sobel）フィルタ又は演算子をエッジ検出に使用することができる。ソーベル（Sobel）演算子は、離散微分演算子であり、画像強度関数の勾配の近似値を計算する。画像の各地点において、ソーベル（Sobel）演算子の結果は、対応する勾配ベクトル又はこのベクトルのノルムのいずれかである。30

【0039】

一部の実施形態において、フレーム 300 に対してダイレーションを実行して、フレーム 300 内の垂直エッジを連結することができる。二値画像に対するダイレーションの基本的な作用は、フォアグラウンド画素の領域の境界を徐々に拡大することである。従って、フォアグラウンド画素の範囲のサイズは大きくなるが、これらの領域内のホールは小さくなる。30

【0040】

一部の実施形態において、次にフレーム 300 は二値化することができる。フレームは、Otsu の手法を用いて二値化することができる。Otsu の手法を用いて、クラスタリングベースの画像閾値化又はグレーレベル画像の二値画像への変換を自動的に実行する。アルゴリズムは、閾値化される画像が 2 つのクラスの画素又は二峰性（bi-modal）ヒストグラム（例えばフォアグラウンド及びバックグラウンド）を含むことを前提としており、次にこれらの合成された広がり（クラス内分散）が最小になるように、これら 2 つのクラスを分離する最適な閾値を計算する。40

【0041】

一部の実施形態において、次に、フレーム 300 に対して連結成分分析を実行して、フレーム 300 内の連結成分を検出することができる。連結成分分析は、グラフ理論のアルゴリズム的応用であり、所与のヒューリスティックに基づいて、連結成分の部分集合に一意にラベル付けを行う。連結成分分析は、二値デジタル画像内の連結領域を検出するためにコンピュータ・ビジョンで用いられるが、より高い次元を有するカラー画像及びデータを処理することもできる。関連する入力データから、頂点及び連結エッジを含むグラフが50

作成される。頂点は、比較ヒューリスティックによって要求される情報を含み、エッジは、連結された「隣接部」を示す。アルゴリズムがグラフを横断して、隣接部の連結性及び相対値に基づいて、頂点にラベル付けを行う。ラベル付け段階の後、グラフを部分集合に分割することができ、その後、元の情報を回復して処理することができる。

【0042】

図3Bは、連結成分分析が実行された後のフレーム300を示している。図3Bにおいて、検出された連結成分の周囲に境界ボックス330が配置されている。図3Cでは、連結成分をテキスト行340にまとめることができる。図3Dでは、テキスト行340を、水平投影を用いて精細化することができ、図3Eでは、テキスト行340を、垂直投影を用いて精細化することができる。この精細化は、複数のテキスト行を細分化するのを助け、テキストである可能性が低い行を取り除くことができる。行への二値画像の投影は、行をビンに分割して、各ビンに直交する行の上にある1画素の数を見つけることによって取得することができる。投影には多くの有用な情報が保持されるので、投影は、画像のコンパクト表現である。垂直方向及び水平方向それぞれにおいて各ビンの1画素の数を見つけることによって、水平及び垂直投影を取得することができる。

10

【0043】

図3Fにおいて、テキスト行340は、これらのエッジ情報に基づいて、テキストを含む可能性がある範囲を表す。この処理においてこれまでに識別されたテキスト行340の1又はそれ以上は、これらのサイズに基づいてフィルタ除去することができる。これらのフィルタ除去されたテキスト行は、テキスト（又は有効なテキスト）を含む可能性が低い。これにより、テキストである可能性が高い行のフィルタ済みセットを形成することができる。図3Gは、フレーム300の20個の様々なテキスト行340を示している。これらの20個の様々なテキスト行340は、これらのエッジ情報に基づいて、テキストを含む可能性がある範囲を表す。図3Hでは、テキスト行340が二値化されており、結果として得られた二値化の成分が文字である可能性が低いことに基づいて（二値化の結果、ほとんどが黒い画像になる）、テキスト行のうちの17個がフィルタ除去されている。テキスト行340はまた、これらの対応する成分の形状及び／又は位置に基づいてフィルタ除去することもでき、これによりテキストを有するビデオコンテンツ310の一部分を形成する。

20

【0044】

30

図3Iは、テキストを有するビデオコンテンツ310の一部分として1つのテキスト行340を示している。テキスト行340は、二値化することができ、これによりテキスト行340の二値化バージョン350を形成する。二値化バージョン350内のテキストは、二値化バージョン350に対してOCRを実行することによって識別することができる。次に、識別されたテキストを構文解析して、複数のテキストセグメントを判定することができる。結果として得られた図3Iの表にあるテキストセグメントは、af（OCRはMTVロゴの一部分を「af」と解釈した）、NEWS、TWEET、US、@MTVNEWS、WITNESS、YOUR、JO、BROS、QUESTIONS、#ASKJONASを含む。テキストセグメントとカテゴリとの間の格納済みの関連性に基づいて、各テキストセグメントに関して、対応するカテゴリを判定することができる。例えば、「@」の使用と「ハンドル」というカテゴリ（例えば、Twitter（登録商標）ハンドル）との間には格納済みの関連性がある可能性がある。結果として、「@MTVNEWS」は、「ハンドル」と識別することができる。同様に、「#」の使用と「ハッシュタグ」というカテゴリとの間にも格納済みの関連性がある可能性がある。結果として、「#ASKJONAS」は、「ハッシュタグ」と識別することができる。これらの関連性は、適切なカテゴリを判定するための文法又は構文規則（例えば、「www.」の使用はURLを示す）を含むことができ、及び／又はカテゴリが判定されるテキストの全て（又はほとんど）を含むことができる。（例えば、テキスト「www.examplewebsite.com」が格納されており、カテゴリを判定すべきテキストが、この格納済みのテキストと比較され、これが一致しており同じカテゴリを割り当てる必要があるか否か判定するこ

40

50

できる)。

【0045】

一部の実施形態において、テキストのカテゴリの判定を用いて、テキスト検出システム100によって実行されるさらなる動作を決定することができる。例えば、ビデオコンテンツ内のテキストがURLと識別されて分類されると、テキスト検出システム100は、カテゴリがURLであるという判定に基づいて、メディアコンテンツデバイス上へのURLのロードを行うことができる。別の実施例では、ビデオコンテンツ内のテキストが電話番号と識別されて分類されると、テキスト検出システム100は、カテゴリが電話番号であるという判定に基づいて、ユーザがこの電話番号にかけるためのプロンプトを生じさせることができる。他の実施例及び構成も本開示の範囲内である。

10

【0046】

付加情報は、サーバ側(例えば、ビデオコンテンツプロバイダ210)又はクライアント側(例えば、第1メディアコンテンツデバイス220又は第2メディアコンテンツデバイス240)のいずれかでフェッチ及び集約することができる。一部の実施形態では、対応するアプリケーションを開始することができる。検出されたテキストから抽出された付加情報は、アプリケーションに提供することができる(例えば、ビデオコンテンツ内のテキストから抽出されたTwitter(登録商標)ハンドルを、Twitter(登録商標)アプリケーションに提供することができる)。

【0047】

場合によっては、識別されたテキスト及び/又はこれの対応する付加情報は、何らかの意味ある使用に対応しない可能性がある。例えば、ビデオコンテンツ内で識別されたテキストに基づいて、(例えば付加情報として)特定のURLを判定することができる。しかしながら、ウェブブラウザ内へのURLのロードによって、対応するページが生じないことがある(例えば、URLが存在しない又は何らかの関連するコンテンツも有さない場合がある)。一部の実施形態において、識別されたテキスト及び/又はこれの対応する付加情報が、有用性の所定の閾値を満たさない(例えば、URLがウェブブラウザによって見つからない場合)という判定に応答して又はそうでなければこれに基づいて、検出されたテキスト及び/又は付加情報の修正(例えば、テキストの異なる綴り及び/又は記号の組合せ)を、テキスト検出システム100によって無作為に又は体系的に試行することができる。これらの修正は、限定ではないが、識別されたテキスト及び/又は付加情報に属さなかつたテキストを事前に引き出すか又は他の方法で使用するテキスト検出システム100に対処するためのサブストリングを含むことができる。

20

【0048】

一部の実施形態において、テキスト認識(例えばOCR)段階は、識別されたテキスト及び/又はこれの対応する付加情報が有用性の所定の閾値を満たさないという判定に応答して又はそうでなければこれに基づいて、繰り返すことができる。このテキスト認識段階は、可能性の高い他の記号、文字、及び/又は数値の組合せ、又は構文解析で起こり得るエラーを発見するために繰り返すことができる。一部の実施形態において、先のテキスト認識段階の実行中にOCR結果を補正するために先に辞書が使用された場合、この辞書機能なしでテキスト認識段階の実行の繰り返しを行うことができる。なぜならば、いくつかの文字の組合せは意図的に辞書の言葉に近いものとすることができますからである(例えば、文字を見かけが類似する数字に置き換える、「E」を「3」に、「I」を「1」に、等)。

30

【0049】

一部の実施形態において、URL、ハンドル、又は他の付加情報を、使用のためにソフトウェアアプリケーションに提供する前に、テキスト検出システム100は、付加情報に対して検証処理を実行して、テキスト認識が正しく実行されるように保証することができる。例えば、テキスト検出システム100は、Facebook(登録商標)URL又はTwitter(登録商標)ハンドルが正式なものであること又は検証済みのアカウントに対応することを確認するためにチェックを行うことができる。テキスト検出システム1

40

50

00は、付加情報をソフトウェアアプリケーションに提供するための条件として検証を用いることができる。他の実施例及び構成も本開示の範囲内である。

【0050】

一部の実施形態において、テキスト検出システム100は、現在のテキスト認識処理に訂正を加える又は他の方法で作用するために、以前に検出して認識／識別した用語（又は他のテキスト）の格納済みの履歴を用いるように構成することができる。

【0051】

図4は、一部の例示的な実施形態による方法400を示すフローチャートである。方法400の動作は、システム又はシステム（例えば、図1のテキスト検出システム100）のモジュールによって実行することができる。動作410において、ビデオコンテンツが少なくとも1つの所定の条件を満たすか否かを判定することができる。一部の実施形態において、所定の条件は、最低レベルの明瞭性、最低レベルのコントラスト、及び複数のフレームにわたる最低レベルのコンテンツ安定性のうちの少なくとも1つを含む。ビデオコンテンツが所定の条件を満たさないと判定されると、方法400は、本明細書に開示するテキスト検出手法をビデオコンテンツに対して実行することなく終了することができる。ビデオコンテンツが所定の条件を満たすと判定されると、動作420において、ビデオコンテンツの一部分がテキストを有すると識別することができる。動作430において、ビデオコンテンツの識別された部分内のテキストを識別することができる。動作440において、識別されたテキストのカテゴリを判定することができる。動作450において、識別されたテキスト及び判定されたカテゴリに基づいて、ビデオコンテンツに対応する附加情報を見つけることができる。附加情報は、統一資源位置指定子（URL）、ユーザーアカウントの識別、メタデータタグ、及び電話番号のうちの少なくとも1つを含むことができる。動作460において、附加情報をメディアコンテンツデバイスに提供することができる。附加情報は、メディアコンテンツデバイス上のソフトウェアアプリケーションに提供することができる。附加情報は、メディアコンテンツデバイス上に表示させることができる。附加情報は、ビデオコンテンツに関連して又はビデオコンテンツの識別された視聴者に関連して格納することができる。識別されたテキストは、ビデオコンテンツに関連して又はビデオコンテンツの識別された視聴者に関連して格納することができる。方法400の動作は、本明細書に開示する何らかの他の機能を組み込むことができることを理解されたい。

【0052】

図5は、一部の例示的な実施形態による方法500を示すフローチャートである。方法500の動作は、システム又はシステムのモジュールによって実行することができる（例えば、図1のテキスト検出システム100）。方法500は、テキストを有するビデオコンテンツの一部分を識別するために使用することができる。動作502において、ビデオコンテンツのフレームをグレースケールに変換することができる。動作504において、フレームに対してエッジ検出を実行することができる。動作506において、フレームに対してダイレーションを実行して、フレーム内の垂直エッジを連結することができる。動作508において、フレームを二値化することができる。動作510において、フレームに対して連結成分分析を実行して、フレーム内の連結成分を検出することができる。動作512において、連結成分を複数のテキスト行にまとめることができる。動作514において、水平及び垂直投影を用いて、複数のテキスト行を精細化することができる。動作516において、複数のテキスト行の少なくとも1つのサイズに基づいて、複数のテキスト行の少なくとも1つをフィルタ除去して、フィルタ済みのテキスト行セットを形成することができる。一部の実施形態において、動作516は、方法500から省くことができる。例えば、テキスト行の全てが所定の基準を満たすと判定される場合、このフィルタリング動作516は行わないことができる。動作518において、フィルタ済みのテキスト行セット（又はフィルタリング動作516が実行されない場合は、単に精細化されたセット）を二値化することができる。動作520において、テキスト行の少なくとも1つにある成分の形状及びテキスト行の少なくとも1つにある成分の位置のうちの少なくとも1つに

10

20

30

40

50

基づいて、二値化されたフィルタ済みのテキスト行セットからテキスト行の少なくとも1つをフィルタ除去して、テキストを有するビデオコンテンツの一部分を形成することができる。方法500の動作は、本明細書に開示する何らかの他の機能を組み込むことができるることを理解されたい。

【0053】

図6は、一部の例示的な実施形態による方法600を示すフローチャートである。方法600の動作は、システム又はシステムのモジュールによって実行することができる（例えば、図1のテキスト検出システム100）。動作610において、ビデオコンテンツの識別された部分（例えば、テキストを有すると識別された部分）に対してOCRを実行して、識別された部分内のテキストを識別することができる。動作620において、識別されたテキストを構文解析して、識別されたテキストの複数のセグメントを判定することができる。動作630において、複数のセグメントの少なくとも1つとカテゴリとの間の格納済みの関連性に基づいて、カテゴリを判定することができる。方法600の動作は、本明細書に開示する何らかの他の機能を組み込むことができることを理解されたい。10

【0054】

例示的なモバイルデバイス

図7は、例示的な実施形態によるモバイルデバイス700を示すブロック図である。モバイルデバイス700は、プロセッサ702を含むことができる。プロセッサ702は、モバイルデバイス700にとって適切な様々な異なるタイプの市販のプロセッサのいずれか（例えば、X Scaleアーキテクチャマイクロプロセッサ、Microprocessors or without Interlocked Pipeline Stages (MIPS)アーキテクチャプロセッサ、又は他のタイプのプロセッサ）とすることができる。ランダムアクセスメモリ(RAM)、フラッシュメモリ、又は他のタイプのメモリ等のメモリ704が、通常、プロセッサ702からアクセス可能である。メモリ704は、オペレーティングシステム(OS)706、並びにLBSをユーザに提供することができるモバイル位置対応のアプリケーションのようなアプリケーションプログラム708を格納するように適合することができる。プロセッサ702は、直接又は適切な中間ハードウェアを介してディスプレイ710に、及びキーパッド、タッチパネルセンサ、マイクロフォン等の1又は2以上の入出力(I/O)デバイス712に接続することができる。同様に、一部の実施形態において、プロセッサ702は、アンテナ716とインタフェースするトランシーバ714に接続することができる。トランシーバ714は、モバイルデバイス700の性質に応じて、セルラーネットワーク信号、無線データ信号、又は他のタイプの信号を、アンテナ716を介して送信及び受信の両方を行うように構成することができる。さらに、一部の実施形態において、GPS受信機718もまた、アンテナ716を利用してGPS信号を受信することができる。20

【0055】

モジュール、構成要素、及びロジック

特定の実施形態について、本明細書ではロジック、又は多くの構成要素、モジュール、又は機構を含むものとして説明する。モジュールは、ソフトウェアモジュール（例えば、機械可読媒体上又は伝送信号内に具現化されたコード）又はハードウェアモジュールのいずれかを構成することができる。ハードウェアモジュールとは、一定の動作を実行する能力がある有形のユニットで、特定の物理的方法で構成又は配置することができる。例示的な実施形態において、1又は2以上のコンピュータシステム（例えば、スタンドアロン、クライアント、又はサーバコンピュータシステム）、或いはコンピュータシステムの1又は2以上のハードウェアモジュール（例えば、プロセッサ又はプロセッサグループ）は、本明細書で説明するような特定の動作を実行するように動作するハードウェアモジュールと同様にソフトウェア（例えば、アプリケーション又はアプリケーション部分）によって構成することができる。40

【0056】

種々の実施形態において、ハードウェアモジュールは、機械的に又は電子的に実装する50

ことができる。例えば、ハードウェアモジュールは、特定の動作を実行するように永続的に構成された（例えば、フィールドプログラマブルゲートアレイ（FPGA）又は特定用途向け集積回路（ASIC）のような専用プロセッサとして）専用回路構成又はロジックを含むことができる。ハードウェアモジュールはまた、特定の動作を実行するようにソフトウェアによって一時的に構成されたプログラム可能ロジック又は回路構成（例えば、汎用プロセッサ又は他のプログラム可能プロセッサ内に包含されるものとして）も含むことができる。専用の永続的に構成された回路構成で、或いは一時的に構成された回路構成で（例えば、ソフトウェアによる構成）ハードウェアモジュールを機械的に実装するという決定は、コスト及び時間を考慮することによって決定することができる点を理解されたい。

10

【0057】

従って、用語「ハードウェアモジュール」は、特定の方法で動作するために及び／又は本明細書で説明する特定の動作を実行するために、エンティティが物理的に構築された、永続的に構成された（例えば、組み込まれた）、又は一時的に構成された（例えば、プログラムされた）もののいずれであれ、有形のエンティティを包含すると理解されるべきである。ハードウェアモジュールが一時的に構成された（例えばプログラムされた）実施形態を考慮すれば、ハードウェアモジュールの各々をどの時点においても構成又はインスタンス化する必要はない。例えば、ハードウェアモジュールが、ソフトウェアを用いて構成された汎用プロセッサを備える場合、汎用プロセッサは、異なる時点で、それぞれ異なるハードウェアモジュールとして構成することができる。従って、ソフトウェアは、例えば、ある時点において特定のハードウェアモジュールを構築するように、及び異なる時点において異なるハードウェアモジュールを構築するようにプロセッサを構成することができる。

20

【0058】

ハードウェアモジュールは、他のハードウェアモジュールに情報を提供し、そこから情報を受け取ることができる。従って、本明細書に記載のハードウェアモジュールは、通信可能に結合されるとみなすことができる。複数のこのようなハードウェアモジュールが同時に存在する場合、通信は、ハードウェアモジュールを接続する信号伝送を通じて（例えば、適切な回路及びバス上で）実現することができる。複数のハードウェアモジュールが異なる時点で構成又はインスタンス化される実施形態では、このようなハードウェアモジュール間の通信は、例えば、複数のハードウェアモジュールがアクセス権を有するメモリ構造への情報の格納及び読み出しを通じて行うことができる。例えば、あるハードウェアモジュールが動作を実行して、自己が通信可能に結合されたメモリデバイスにこの動作の出力を格納することができる。その後、別のハードウェアモジュールが、後の時点でメモリデバイスにアクセスして、格納された出力を読み出して処理することができる。また、ハードウェアモジュールは、入力又は出力デバイスとの通信を開始することもでき、リソース（例えば情報の集合）に基づいて動作することができる。

30

【0059】

本明細書で説明する例示的な方法の種々の動作は、関連する動作を実行するように（例えば、ソフトウェアによって）一時的に構成された又は永続的に構成された、1又は2以上のプロセッサによって少なくとも部分的に実行することができる。一時的又は永続的に構成されたかにかかわらず、このようなプロセッサは、1又は2以上の動作又は機能を実行するように動作するプロセッサ実施のモジュールを構築することができる。本明細書で言及するモジュールは、一部の例示的な実施形態において、プロセッサ実装モジュールを含むことができる。

40

【0060】

同様に、本明細書で説明する方法は、少なくとも部分的にプロセッサ実装とすることができます。例えば、方法の動作の少なくとも一部は、1又は2以上のプロセッサ又はプロセッサ実装のモジュールによって実行することができる。動作の一部の実行は、単一の機械内に常駐するだけでなく、多くの機械にわたって配備された1又は2以上のプロセッサの

50

間で分散させることができる。一部の例示的な実施形態において、1又は2以上のプロセッサは、単一の場所（例えば、ホーム環境、オフィス環境の中、又はサーバファームとして）に配置することができるが、別の実施形態では、多くの場所にわたってプロセッサを分散することができる。

【0061】

また、1又は2以上のプロセッサは、「クラウドコンピューティング」環境で又は「software as a service (SaaS)」として、関連動作の実行をサポートするように動作することもできる。例えば、動作の少なくとも一部は、コンピュータのグループ（プロセッサを含む機械の例として）によって実行することができ、これらの動作には、ネットワークを介して、及び1又は2以上の適切なインターフェース（例えばAPI）を介してアクセス可能である。10

【0062】

例示的な実施形態は、デジタル電子回路構成において、又はコンピュータハードウェア、ファームウェア、ソフトウェアにおいて、又はこれらの組合せにおいて実施可能である。例示的な実施形態は、データ処理装置、例えば、プログラム可能プロセッサ、コンピュータ、又は複数のコンピュータによって実行するために又はこれらの動作を制御するために、情報キャリア内、例えば機械可読媒体内に有形で具現化されたコンピュータプログラムのようなコンピュータプログラム製品を用いて実施することができる。

【0063】

コンピュータプログラムは、コンパイル又は翻訳言語を含む何らかの形式のプログラミング言語で記述することができ、スタンドアロンプログラムとして、又はモジュール、サブルーチン、又はコンピュータ環境で使用するのに適した他のユニットとして、を含む様々な形式で導入することができる。コンピュータプログラムは、1つのコンピュータ上で、或いは1つの場所にある又は複数の場所にわたって分散され通信ネットワークによって相互接続された複数のコンピュータ上で実行されるように導入することができる。20

【0064】

例示的な実施形態において、動作は、入力データに対して作用して出力を生成することによって機能を実行するために、コンピュータプログラムを実行する1又は2以上のプログラム可能プロセッサによって実行することができる。また、方法の動作は、専用論理回路構成（例えば、FPGA又はASIC）によって実行することができ、さらに例示的な実施形態の装置は、専用論理回路構成として実装することができる。30

【0065】

コンピューティングシステムは、クライアント及びサーバを含むことができる。クライアント及びサーバは、一般に、相互に離れており、通常は通信ネットワークを介して相互作用する。クライアント及びサーバの関係は、それぞれのコンピュータ上で実行され、相互にクライアント - サーバ関係を有するコンピュータプログラムによって生じる。プログラム可能コンピューティングシステムを備える実施形態では、ハードウェア及びソフトウェアの両方のアーキテクチャが考慮に値することは理解されるであろう。具体的には、永続的に構成されたハードウェア（例えばASIC）で、一時的に構成されたハードウェア（例えばソフトウェア及びプログラム可能プロセッサの組合せ）で、又は永続的に構成された及び一時的に構成されたハードウェアの組合せで特定の機能を実施するかどうかの選択は、設計上の選択となり得る。以下に、種々の例示的な実施形態において導入可能なハードウェア（例えば機械）及びソフトウェアのアーキテクチャを示す。40

【0066】

図8は、例示的な実施形態による、本明細書で議論する方法論の任意の1又は2以上を機械に実行させるための命令824をその中で実施可能なコンピュータシステム800の例示的な形態の機械のブロック図である。代替的な実施形態において、機械は、スタンドアロンデバイスとして動作し、又は他の機械に接続（例えば、ネットワーク接続）することができる。ネットワーク配置において、機械は、サーバ - クライアントネットワーク環境ではサーバ又はクライアント機械として、或いはピアツーピア（又は分散）ネットワー50

ク環境ではピア機械として動作することができる。機械は、パーソナルコンピュータ（P C）、タブレット P C、セットトップボックス（S T B）、携帯情報端末（P D A）、携帯電話、ウェブアプライアンス、ネットワークルータ、スイッチ又はブリッジ、或いはその機械によって行われるべき動作を指定する命令（シーケンシャル又はそれ以外）を実行する能力があるいずれかの機械とすることができます。さらに、唯一の機械のみが示されているが、「機械」という用語は、本明細書で議論する方法論の任意の 1 又は 2 以上を実行するための命令セット（又は複数セット）を個別に又は共同で実行する何らかの機械の集合を含むと解釈するべきである。

【0067】

例示的なコンピュータシステム 800 は、プロセッサ 802（例えば、中央処理ユニット（C P U））、グラフィックス処理ユニット（G P U）、又はその両方）と、メインメモリ 804 と、スタティックメモリ 806 と、を含み、これらはバス 808 を介して相互に通信する。コンピュータシステム 800 は、ビデオディスプレイユニット 810（例えば、液晶ディスプレイ（L C D）又はブラウン管（C R T））をさらに含むことができる。コンピュータシステム 800 はまた、英数字入力デバイス 812（例えばキーボード）と、ユーザインターフェース（U I）ナビゲーション（又はカーソル制御）デバイス 814（例えばマウス）と、ディスクドライブユニット 816 と、信号発生装置 818（例えばスピーカ）と、ネットワークインターフェースデバイス 820 と、も含むことができる。

10

【0068】

ディスクドライブユニット 816 は、本明細書で説明する方法論又は機能の任意の 1 又は 2 以上を具現化する又はこれらによって利用されるデータ構造及び命令の 1 又は 2 以上のセット 824（例えばソフトウェア）が格納された機械可読媒体 822 を含む。命令 824 はまた、完全に又は少なくとも部分的に、メインメモリ 804 内及び / 又はプロセッサ 802 内に、コンピュータシステム 800 によるこの命令の実行中に常駐することもでき、メモリ 804 及びプロセッサ 802 が、機械可読媒体を構成することもできる。命令 824 はまた、完全に又は少なくとも部分的に、スタティックメモリ 806 内に常駐することもできる。

20

【0069】

例示的な実施形態では、機械可読媒体 822 が単一の媒体であるように示されているが、「機械可読媒体」という用語は、1 又は 2 以上の命令 824 又はデータ構造を格納する単一媒体又は複数媒体（例えば、集中又は分散データベース、及び / 又は関連するキャッシュ及びサーバ）を含むことができる。「機械可読媒体」という用語はまた、機械によって実行するための命令を格納し、エンコードし、又は伝送する能力があり、本実施形態の方法論の任意の 1 又は 2 以上を機械に実行させ、又はこのような命令によって利用され又はこれに関連するデータ構造を格納し、エンコードし、又は伝送する能力がある何らかの有形媒体を含むと解釈すべきである。従って、「機械可読媒体」という用語は、限定ではないが、固体メモリ、並びに光学及び磁気媒体を含むと解釈すべきである。機械可読媒体の具体的な例は、不揮発性メモリを含み、例として半導体メモリデバイス（例えば、消去可能プログラマブル読み取り専用メモリ（E P R O M）、電気的消去可能プログラマブル読み取り専用メモリ（E E P R O M）、及びフラッシュメモリデバイス）、内部ハードディスク及びリムーバブルディスク等の磁気ディスク、光磁気ディスク、及びコンパクトディスク読み取り専用メモリ（C D - R O M）及びデジタル多用途ディスク（又はデジタルビデオディスク）読み取り専用メモリ（D V D - R O M）ディスクを含む。

30

【0070】

さらに、命令 824 は、伝送媒体を用いて通信ネットワーク 826 上で送受信することができる。命令 824 は、ネットワークインターフェースデバイス 820 及び多くの公知の転送プロトコルのいずれか 1 つ（例えばH T T P）を用いて伝送することができる。通信ネットワークの例は、L A N、W A N、インターネット、携帯電話ネットワーク、P O T S ネットワーク、及び無線データネットワーク（例えばW i F i 及びW i M A X ネットワーク）を含む。「伝送媒体」という用語は、機械によって実行するための命令を格納し、

40

50

エンコードし、又は伝送する能力がある何らかの無形媒体を含むと解釈すべきであり、このようなソフトウェアの通信を円滑にするためのデジタル又はアナログ通信信号又は他の無形媒体を含む。

【0071】

特定の例示的な実施形態を参照しながら実施形態について説明してきたが、本開示のより広い精神及び範囲から逸脱することなく、これらの実施形態には種々の修正及び変更を加えることができる点は明らかになるであろう。従って、明細書及び図面は、限定的ではなく例示的な意味でとらえるものとする。これの一部分を成す添付図面は、限定ではなく例証として、主題を実施することができる特定の実施形態を示している。例証する実施形態については、当業者が本明細書に開示された教示を実施できるようにするために十分詳細に説明する。他の実施形態を利用してそこから導出することができる所以、本開示の範囲から逸脱することなく構造上及び論理的な置換え及び変更を加えることができる。従つて、この詳細な説明は、限定的な意味でとらえるものではなく、種々の実施形態の範囲は、添付の特許請求の範囲、並びにこのような特許請求の範囲が享受する均等物の全範囲によってのみ定義される。10

【0072】

本明細書では特定の実施形態について示して説明してきたが、同じ目的を達成すると予想されるいずれの配置も、ここに示した特定の実施形態に置き換えることができる点を理解されたい。本開示は、種々の実施形態のありとあらゆる適応又は変形を網羅するものである。上記の説明を再検討することで、本明細書に特には記載されていない上記の実施形態の組合せ及び他の実施形態が当業者には明らかになるであろう。20

【符号の説明】

【0073】

- 100 テキスト検出システム
- 110 コンテンツ選択モジュール
- 120 テキスト検出モジュール
- 130 テキスト認識モジュール
- 140 テキスト解釈モジュール
- 150 情報モジュール
- 160 データベース

30

【図1】

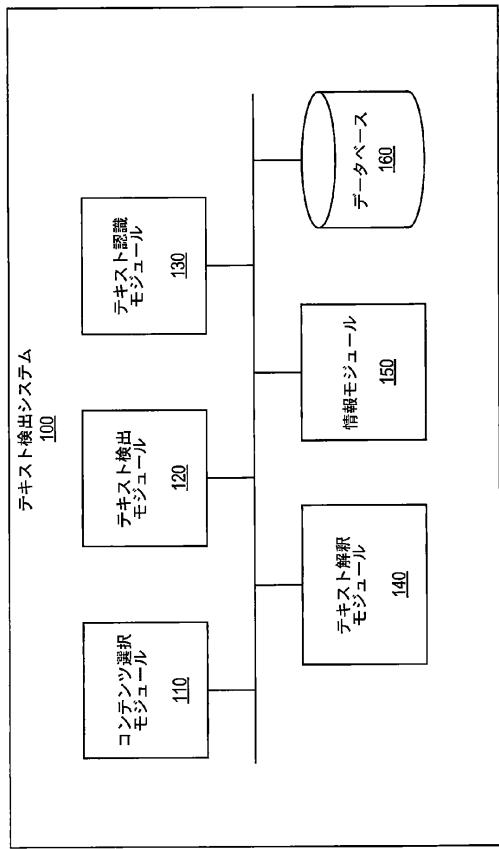


FIG. 1

【図2】

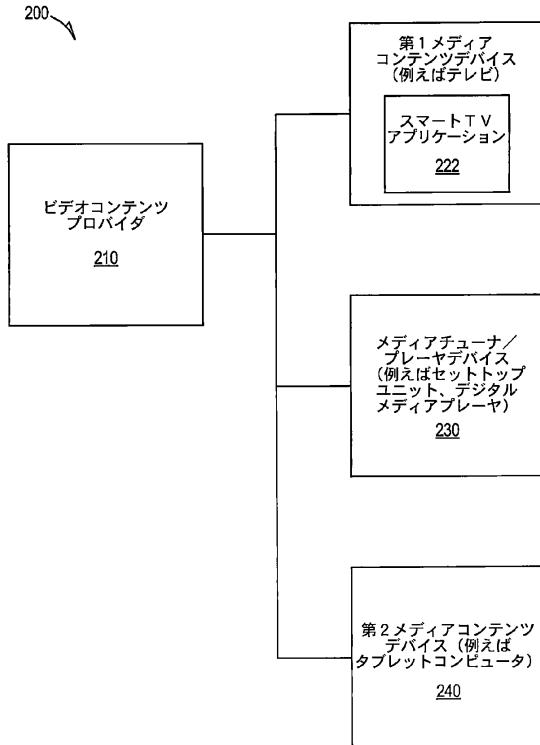


FIG. 2

【図3A】

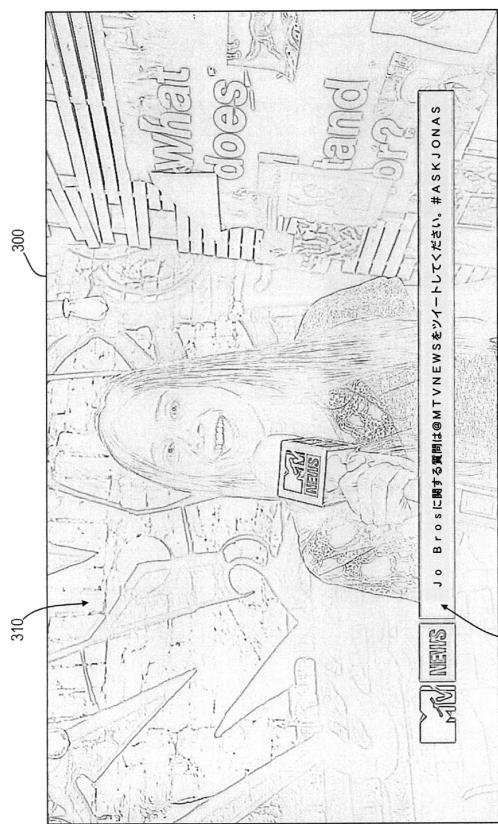


FIG. 3A

【図3B】

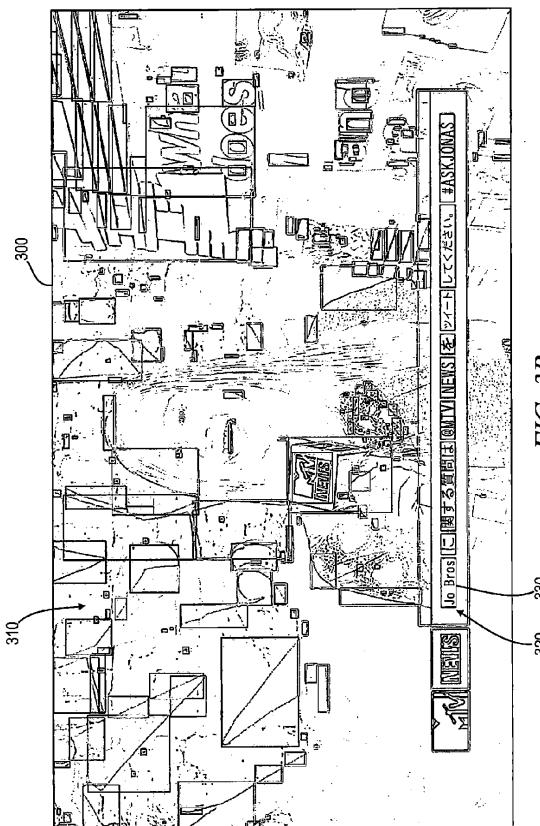


FIG. 3B

【図 3 C】

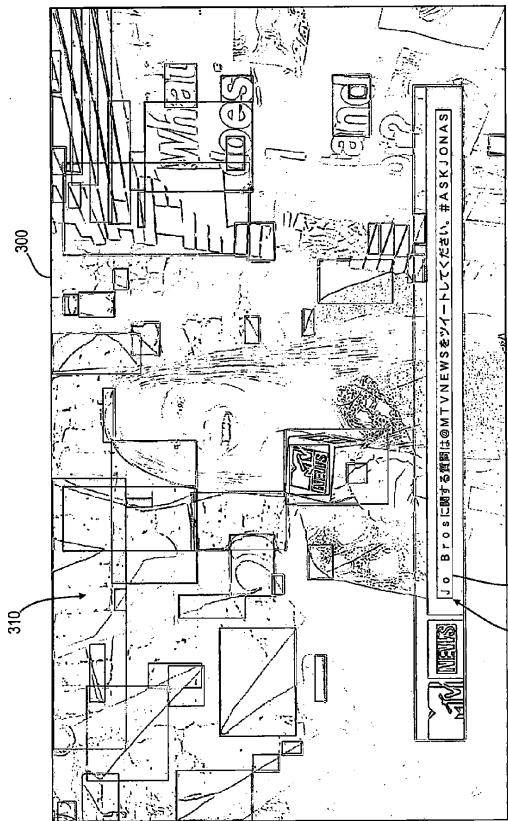


FIG. 3C

【図 3 D】

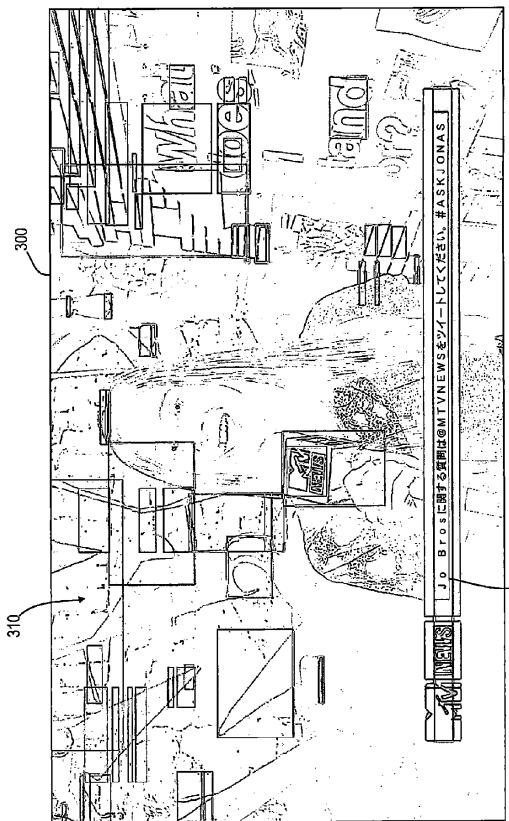


FIG. 3D

【図 3 E】

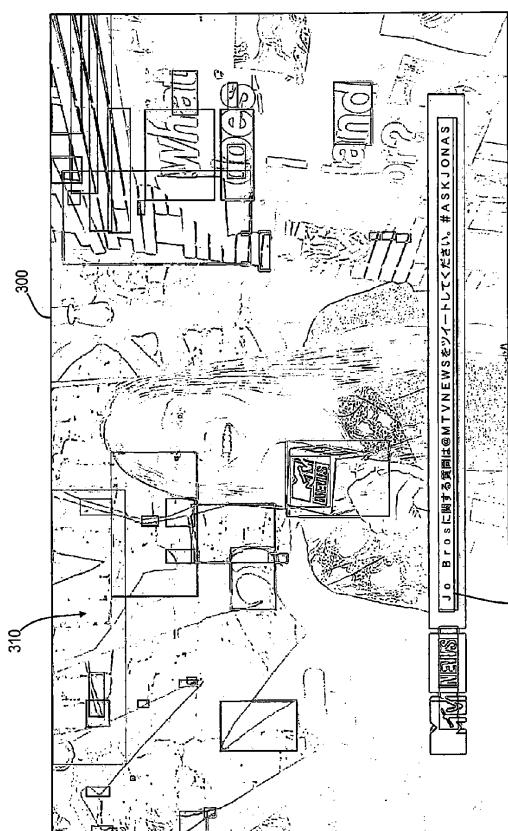


FIG. 3E

【図 3 F】

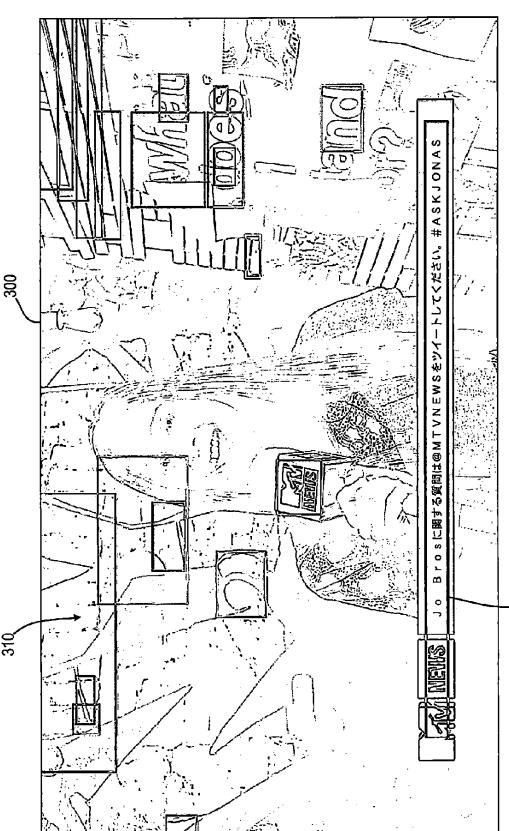


FIG. 3F

【図3G】

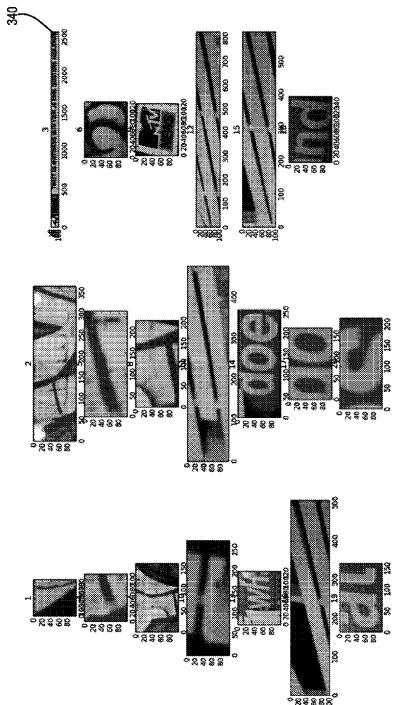


FIG. 3G

【図3H】

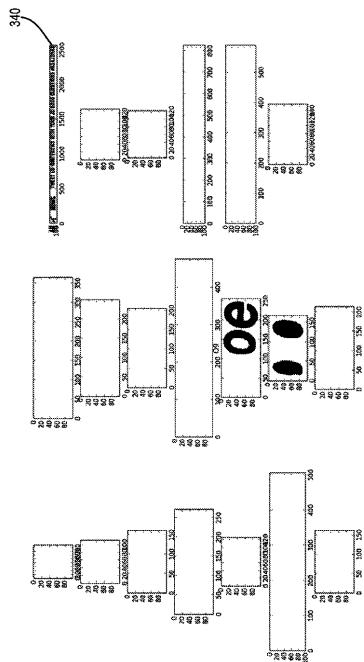


FIG. 3H

【図3I】

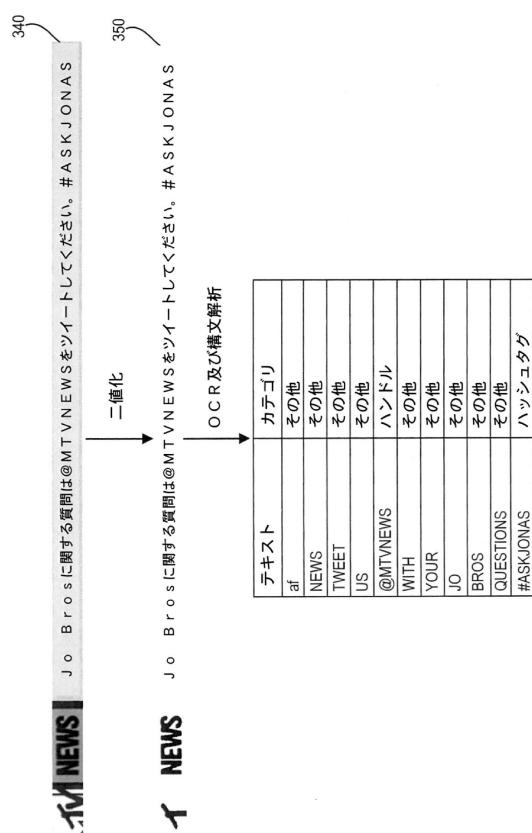


FIG. 3I

【図4】

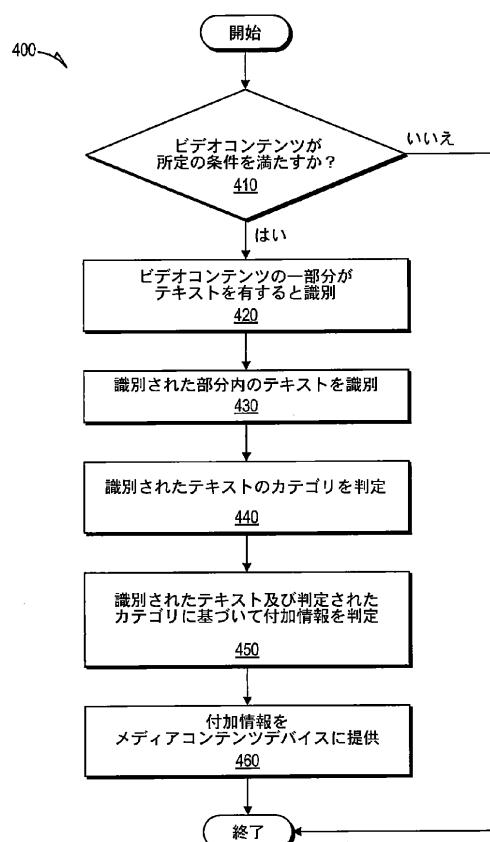


FIG. 4

【図5】

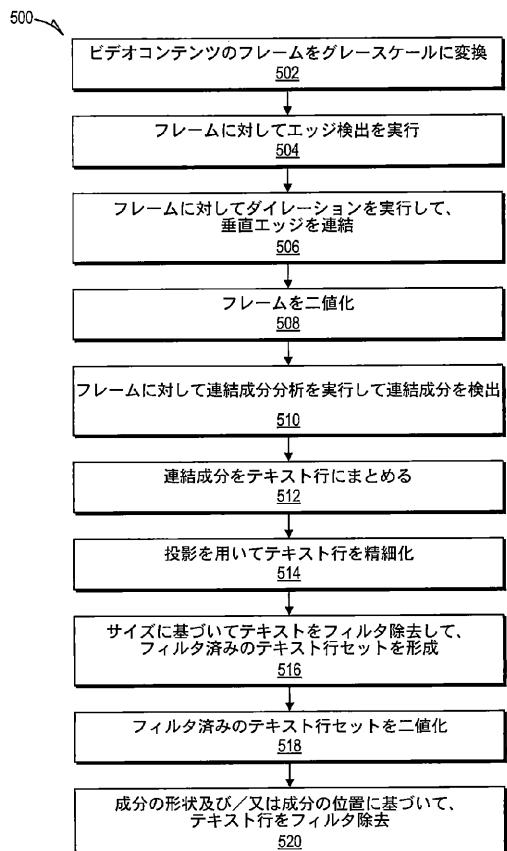


FIG. 5

【図6】

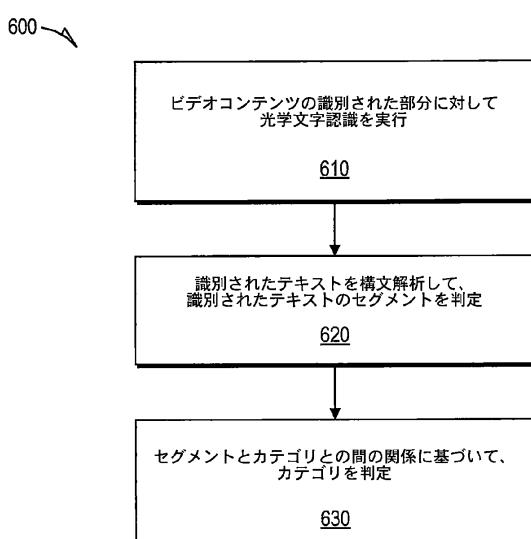


FIG. 6

【図7】

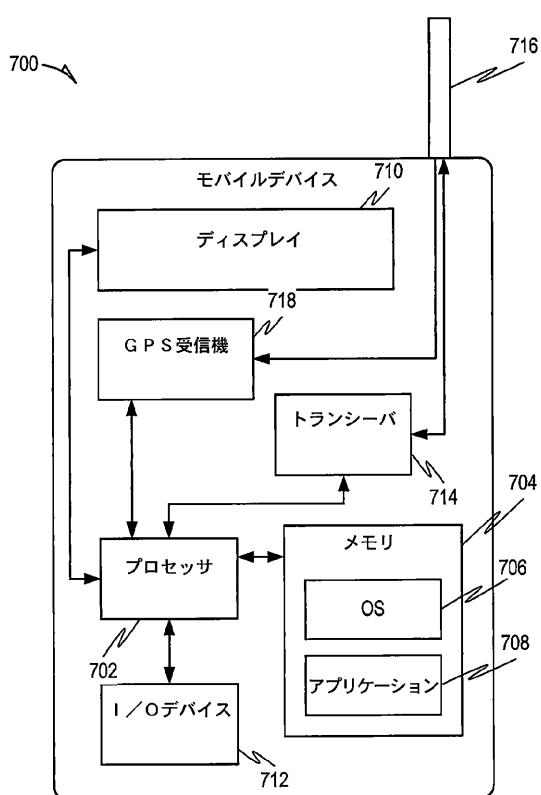


FIG. 7

【図8】

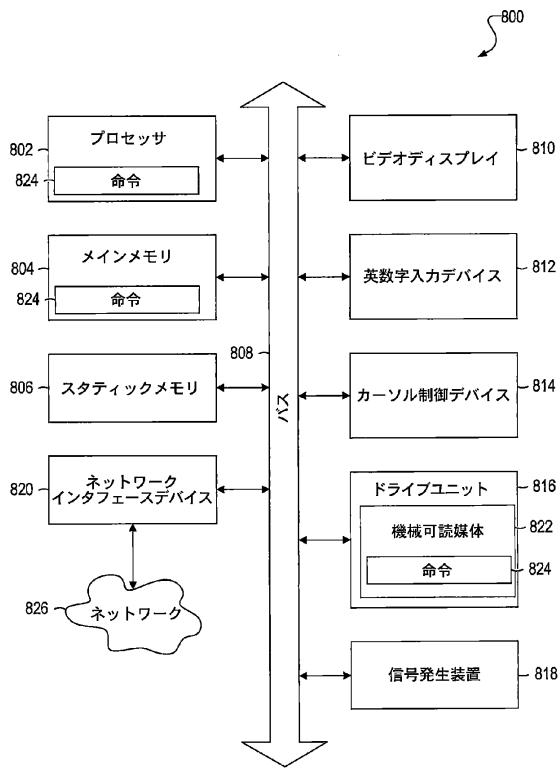


FIG. 8

フロントページの続き

(74)代理人 100109335
弁理士 上杉 浩
(74)代理人 100120525
弁理士 近藤 直樹
(74)代理人 100151987
弁理士 谷口 信行
(72)発明者 チュー アイリーン
アメリカ合衆国 カリフォルニア州 94110 サンフランシスコ トリート アヴェニュー
1138
(72)発明者 ハロン ウィルソン
アメリカ合衆国 カリフォルニア州 94705 バークリー シャタック アヴェニュー 30
32
(72)発明者 クレマー マルクス ケイ
アメリカ合衆国 カリフォルニア州 94563 オリンダ ロスト ヴァレー ドライブ 15

審査官 笠田 和宏

(56)参考文献 特開2009-130899(JP,A)
特開2000-182053(JP,A)
米国特許出願公開第2013/0330003(US,A1)

(58)調査した分野(Int.Cl., DB名)

I P C	G 0 6 K	9 / 0 0	-	9 / 8 2
	G 0 6 T	7 / 0 0	-	7 / 9 0
	H 0 4 N	7 / 1 0		
		7 / 1 4	-	7 / 1 7 3
		7 / 2 0	-	7 / 5 6
		2 1 / 0 0	-	2 1 / 8 5 8