

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6317772号
(P6317772)

(45) 発行日 平成30年4月25日 (2018. 4. 25)

(24) 登録日 平成30年4月6日 (2018. 4. 6)

(51) Int. Cl.

F I

G06K 9/62 (2006.01)

G06K 9/62 620B

G06F 17/28 (2006.01)

G06F 17/28 690

G06K 9/22 (2006.01)

G06K 9/22

請求項の数 1 (全 32 頁)

(21) 出願番号 特願2015-562415 (P2015-562415)
 (86) (22) 出願日 平成26年3月12日 (2014. 3. 12)
 (65) 公表番号 特表2016-519797 (P2016-519797A)
 (43) 公表日 平成28年7月7日 (2016. 7. 7)
 (86) 国際出願番号 PCT/IB2014/001371
 (87) 国際公開番号 W02014/162211
 (87) 国際公開日 平成26年10月9日 (2014. 10. 9)
 審査請求日 平成29年3月13日 (2017. 3. 13)
 (31) 優先権主張番号 61/791, 584
 (32) 優先日 平成25年3月15日 (2013. 3. 15)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 515254183
 トランスレート アブロード、インコーポ
 レイテッド
 アメリカ合衆国、ロードアイランド州 O
 2903、プロビデンス、フロアー 3、
 95 チェスナット ストリート
 (74) 代理人 100114775
 弁理士 高岡 亮一
 (74) 代理人 100121511
 弁理士 小田 直
 (74) 代理人 100202751
 弁理士 岩堀 明代
 (74) 代理人 100191086
 弁理士 高橋 香元

最終頁に続く

(54) 【発明の名称】 外国語の文字セットおよびそれらの翻訳を資源に制約のあるモバイル機器上にリアルタイムで表示するためのシステムおよび方法

(57) 【特許請求の範囲】

【請求項 1】

ビデオカメラ、プロセッサ、メモリおよびディスプレイを含むモバイル機器を使用して、ビデオフィードをリアルタイム拡張現実で、第1の言語から第2の言語に翻訳するための方法であって、前記方法が、

(a) 前記ビデオカメラを使用して、翻訳する必要のある前記第1の言語の1つ以上の語の前記ビデオフィードからフレームをリアルタイムで捕捉して、捕捉されたフレームを生成するステップと、

(b) 前記捕捉されたフレームを画像処理境界ボックスの内部に適合するように切り取って、切り取られたフレームを生成するステップと、

(c) 前記切り取られたフレームを前処理して、前処理されたフレームを生成するステップと、

(d) 前記前処理されたフレームに関して文字セグメント認識を実行して、複数の文字セグメントを生成するステップと、

(e) 前記文字セグメントに関して文字マージを実行して、複数のマージされた文字セグメントを生成するステップと、

(f) 前記マージされた文字セグメントに関して文字認識を実行して、複数の認識された文字を有する認識されたフレームを生成するステップと、

(g) 前記第1の言語の前記認識された文字の前記第2の言語の1つ以上の語への翻訳を生成するために、前記認識されたフレームを翻訳エンジンで処理して翻訳されたフレー

ムを生成し、同時に、各翻訳されたフレームに対して、前記認識された文字がどのようにうまく翻訳されているかを表す翻訳品質も計算するステップと、

(h) 前記翻訳されたフレームを、現在の翻訳されたフレームとして前記メモリに格納するステップであって、以前の翻訳されたフレームおよび以前の翻訳品質も前記メモリに格納される、前記翻訳されたフレームを、現在の翻訳されたフレームとして前記メモリに格納することステップと、

(i) 前記境界ボックスが、前記現在の翻訳されたフレームおよび前記以前の翻訳されたフレームに対して、同じ文字セット上に継続してあることを、前記現在の翻訳されたフレームと前記以前の翻訳されたフレームとの間で重複している類似した文字の割合を判断することにより確認するステップであって、より高い割合は、前記境界ボックスが、前記現在の翻訳されたフレームおよび前記以前の翻訳されたフレームに対して前記同じ文字セット上に継続してあることを示す、前記境界ボックスが、前記現在の翻訳されたフレームおよび前記以前の翻訳されたフレームに対して、同じ文字セット上に継続してあることを確認するステップと、

(j) 前記現在の翻訳されたフレームに対して前記翻訳エンジンによって判断された前記翻訳品質を、前記以前の翻訳されたフレームに対する前記以前の翻訳品質と比較するステップと、

(k) 前記メモリから削除すべき、前記以前の翻訳されたフレームおよび前記現在の翻訳されたフレームのうちの1つを、より低い翻訳品質を有するフレームに基づき選択するステップと、

(l) 前記以前の翻訳されたフレームおよび前記現在の翻訳されたフレームから最適な翻訳されたフレームを表示するステップであって、前記最適な翻訳されたフレームがより高い翻訳品質を有し、前記第2の言語の前記語が、前記モバイル機器の前記ディスプレイ上に拡張現実で、翻訳されている前記第1の言語の前記語の上に重ねられているか、またはその隣にある、前記以前の翻訳されたフレームおよび前記現在の翻訳されたフレームから最適な翻訳されたフレームを表示するステップとを含む、方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の参照

本出願は、非仮出願であり、2013年3月15日に出願された「Recognition System」という名称の米国仮出願第61/791,584号からの優先権を主張し、その全体が参照により本明細書に組み込まれる。

【0002】

著作権およびトレードドレスの表示

本特許に関連した文書の開示の一部は、著作権保護の対象となる材料を含む。本特許関連文書は、所有者のトレードドレスであるか、またはトレードドレスになり得る内容を示し、かつ/または記述し得る。著作権およびトレードドレスの所有者は、特許開示が、米国特許商標局の特許ファイルまたはレコード内に出現するとおり、あらゆる人による複製に異議はないが、それ以外ではその如何に係わらず全ての著作権およびトレードドレス権を留保する。

【0003】

発明の分野

本発明は、一般に、アジア文字セットを翻訳するためのシステムおよび方法に関する。より詳細には、本発明は、資源に制約のあるモバイル機器上でのアジア文字セットの画像処理および認識の後に、アジア文字セットおよびそれらの翻訳をリアルタイムで表示するためのシステムおよび方法に関する。本発明は、その国の現地語で書かれたテキストを読んで理解することが困難な外国の旅行者を支援するために使用され得る。より一般的に

は、本発明は、任意の２言語間の翻訳にも適用可能である。

【背景技術】

【０００４】

本節の記述は、本開示に関連した背景情報を提供するだけであり、従来技術を構成するものではない。

【０００５】

外国内の旅行者は、しばしば、レストランの名前もしくは住所、レストランのメニュー、道路標識、本、地図、列車の時刻表、または新聞などの、外国語で書かれた何らかのテキストを読んで理解できる必要がある。慣例的に、旅行者は、外国語の翻訳本を使用するか、ガイドを雇うか、または現地の人々に助けを求め得る。これらのアプローチは厄介であり、世界中で、外国人が彼らの第２言語として次第に英語を使うようになって、この言語バリアは終わりそうにない。

10

【０００６】

複雑な画像処理および光学式文字認識（OCR）ソフトウェアを使用する、翻訳装置が知られている。OCRは、１９９０年代初めにおけるその開始以来、著しく改善されてきて、インターネット上で使用されるが；外国人旅行者は、一般に、外国でのインターネット接続を備えたモバイル機器を有していない。従って、旅行者用の翻訳装置は、オフラインで、すなわち、インターネットへの接続およびオンラインサーバーへのアクセスによって提供される資源なしで、十分に機能する必要がある。

【０００７】

20

モバイルカメラ装置用のオフラインOCRアプリケーションは、プログラムコードのサイズに関してサイズ制限がある。画像処理速度およびオフラインのOCRアルゴリズムにも制限がある。モバイルカメラ装置内のプロセッサのタイプおよびメモリ資源における制限がある。オフラインのモバイル翻訳装置は、翻訳精度および再現性の欠如にも悩まされる。一般に、モバイル翻訳装置は、翻訳する外国語テキストの単一の画像フレームを捕捉するために使用される。OCRは、外国語テキストの捕捉された画像フレームに関して実行されて、その外国語テキストを、旅行者によって読むことができる言語に翻訳する。しかし、スマートフォンなどのハンドヘルドモバイルカメラ装置を使用した外国語テキストの画像捕捉中に、カメラ移動、不十分なテキスト画像のフォーカス、および不適切な外国語テキストの照明を含む画像捕捉の問題がある。OCRは、正確で安定した外国語テキストの翻訳のために、明瞭に区別できるテキスト画像を必要とするため、不明瞭なテキスト画像がOCRソフトウェアの判断を誤らせ、その結果、不完全な言語翻訳を生じるであろう。従って、スマートフォンなどのモバイルカメラ装置用のオフライン翻訳アプリは、しばしば、正確で安定した翻訳を実行しないことが知られている。翻訳は、変動するか、揺らぐか、または全く意味をなさない可能性さえある。

30

【０００８】

このような理由で、翻訳において改善された速度、精度、および意味をもたらすために、モバイルカメラ装置のための現在の翻訳技術に関連したこれらの問題に対する解決策の重要な必要性がある。外国でのより良い旅行経験を可能にするために、リアルタイムで、言語学的文法に基づく翻訳に対する必要性がある。必要とされるものは、インターネット接続を必要とすることなく、資源に制約のあるモバイル機器上で、外国語テキストの迅速で意味のある翻訳をリアルタイムで行うための方法、システムおよび装置である。

40

【０００９】

従って、既存の解決策の欠点を解決するために、リアルタイムで正確に、外国語を迅速で正確に翻訳するための方法を提供することは、最高水準の技術における進歩であろう。旅行者に意味のある情報を自動的に提供するために、インターネット接続を必要とすることなく、外国語をリアルタイムで翻訳できる、モバイル機器でのこの翻訳方法を提供することも、最高水準の技術における進歩であろう。かかる翻訳が、費用効率が高く、翻訳者もしくは辞書、またはテキストのモバイル機器への手入力を必要としないことは、さらに進歩であろう。本発明の様々な実施形態が開発されたのは、こういう背景においてである

50

。

【発明の概要】

【課題を解決するための手段】

【0010】

本発明の実施形態は、モバイルカメラ装置を使用して、第1の言語の1つ以上の語を第2の言語の1つ以上の語に翻訳するための方法およびシステムを含む。

【0011】

それに応じて、一実施形態によれば、本発明は、モバイルカメラ装置を使用して、第1の言語から第2の言語に翻訳するための方法であり、本方法は：(a) 翻訳する必要のある第1の言語の1つ以上の語のビデオ画像を表示するためにモバイルカメラ装置を位置付けて、モバイルカメラ装置が、第1の言語の1つ以上の語のビデオフィードのフレームを翻訳のために捕捉できるようにすること；(b) ビデオフィードのフレームを画像処理のために画像処理境界ボックスの内部に適合するように切り取ること；(c) ビデオフィードの切り取られたフレームをメモリ装置に格納すること；(d) 画像処理境界ボックス内のビデオフィードの切り取られたフレームを前処理すること；(e) 画像処理境界ボックス内のビデオフィードの前処理されたフレームに関して文字セグメント認識を実行すること；(f) 画像処理境界ボックス内のビデオフィードの文字セグメント認識されたフレームに関して認識フィードバックを用いて水平マージ(horiz ontal merg i ng)を実行すること；(g) 画像処理境界ボックス内のビデオフィードの水平マージされた文字セグメント認識済みフレームに関して二値またはグレースケール文字認識を実行すること；(h) 第1の言語の1つ以上の語の第2の言語の1つ以上の語への翻訳を生成するために画像処理境界ボックス内のビデオフィードの文字認識されたフレームを処理すること；(i) 第2の言語の1つ以上の翻訳された語を、言語翻訳ビデオ画像の現在のフレームとしてメモリ装置内の位置に格納すること；(j) 画像処理境界ボックスが、言語翻訳ビデオ画像の現在のフレームおよび以前のフレームについて、同じ第1の言語のテキスト文字上に継続してあることを確認すること；(k) 言語翻訳ビデオ画像の現在のフレーム内の情報品質を、言語翻訳ビデオ画像の以前のフレーム内の情報品質と比較すること(言語翻訳ビデオ画像の現在のフレームと言語翻訳ビデオ画像の以前のフレームの両方がメモリ装置内に保存されている)；(l) メモリ装置内のストレージから削除すべき、言語翻訳ビデオ画像の1つ以上の低品質のフレームを選択すること；および(m) 第2の言語の1つ以上の語の言語翻訳ビデオ画像の1つ以上のより高品質のフレームを表示し、同時に、翻訳されている第1の言語の1つ以上の語のビデオ画像も表示するために、モバイルカメラ装置を使用すること；の各ステップを含む。

【0012】

本発明の別の実施形態は、第2の言語の1つ以上の語の言語翻訳ビデオ画像の1つ以上のより高品質のフレームをリアルタイム拡張現実で表示するための方法も含む。

【0013】

本発明の別の実施形態は、中国語、韓国語、日本語、ベトナム語、クメール語、ラオ語、タイ語、英語、フランス語、スペイン語、ドイツ語、イタリア語、ポルトガル語、ロシア語、ヒンディー語、ギリシャ語、ヘブライ語、およびアラビア語から成るグループから選択された第1の言語を翻訳するための方法も含む。いくつかの実施形態では、プロセスは、ユーザーが1つを選択する必要なく、どの言語がビデオフィード内に存在しているかを自動検出することができる。

【0014】

本発明の別の実施形態は、第1の言語を第2の言語に翻訳する前に、第1の言語の方言を第1の言語のさらに少ない数の方言に変換するための変換テーブルを使用するための方法も含む。

【0015】

本発明の別の実施形態は、第1の言語を第2の言語に翻訳する前に、全ての繁体字中国語テキスト文字を簡体字中国語テキスト文字に変換するために変換テーブルを使用するた

10

20

30

40

50

めの方法も含む。

【 0 0 1 6 】

本発明の別の実施形態は、中国語、韓国語、日本語、ベトナム語、クメール語、ラオ語、タイ語、英語、フランス語、スペイン語、ドイツ語、イタリア語、ポルトガル語、ロシア語、ヒンディー語、ギリシャ語、ヘブライ語、およびアラビア語から成るグループから選択された第2の言語への翻訳を得るための方法も含む。

【 0 0 1 7 】

本発明の別の実施形態は、第1の言語のビデオ画像を表示するモバイルカメラ装置上の境界ボックスのサイズを変更することにより、第2の言語への翻訳のために、第1の言語の単一行または第1の言語の複数行を選択するための方法も含む。

10

【 0 0 1 8 】

本発明の別の実施形態は、モバイルカメラ装置が移動すると、翻訳を再計算することなく、画面上の第2の言語の翻訳を自動的に移動させるための方法も含む。

【 0 0 1 9 】

本発明の別の実施形態は、表示されている言語翻訳を変更することなく、モバイルカメラ装置の移動を可能にするために、モバイルカメラ装置上に表示されている言語翻訳を一時停止することも含む。

【 0 0 2 0 】

本発明の別の実施形態は、第1の言語および第1の言語の第2の言語への翻訳を含む、一時停止された言語翻訳を、後で再検討するためにメモリ装置に格納することも含む。

20

【 0 0 2 1 】

本発明の別の実施形態は、言語翻訳ビデオ画像の現在のフレーム内の情報品質を、言語翻訳ビデオ画像の以前のフレーム内の情報品質と比較するための方法も含み、言語翻訳ビデオ画像の情報品質は、第1の言語の文字列がいかにうまく翻訳されるかによって判断できる。

【 0 0 2 2 】

本発明の別の実施形態は、画像処理境界ボックスが、言語翻訳ビデオ画像の現在のフレームおよび以前のフレームについて、同じ第1の言語のテキスト文字上に継続してあることを確認するための方法も含み、本方法は：(a)現在の言語テキスト翻訳画像文字列内および以前の言語翻訳画像文字列内の、類似した言語テキスト文字数をカウントすること；ならびに(b)これらの類似した言語テキスト文字のどれくらいの割合が、現在および以前の言語翻訳画像文字列内で重複しているかを計算すること(その割合が高ければ、処理境界ボックスが、現在および以前の言語翻訳テキスト画像について、同じ言語テキスト上に留まっている度合いが大きい)；の各ステップを含む。

30

【 0 0 2 3 】

本発明の別の実施形態は、翻訳されている第1の言語の1つ以上の語の発音を表示するための方法も含む。

【 0 0 2 4 】

本発明の別の実施形態は、モバイルカメラ装置上で外国語を翻訳するためのコンピュータシステムであって、本システムは：第1の言語の1つ以上の語のビデオ画像を、第1の言語テキストの翻訳のために捕捉するためのモバイルカメラ；プログラムコード；プログラムコードを処理するためのプロセッサ；プログラムコードを格納するための、プロセッサに接続された1つ以上のメモリを含み、そのプログラムコードは、プロセッサによって実行される際に、プロセッサにプロセスを実行させ、プロセスは：(a)翻訳する必要のある第1の言語の1つ以上の語のビデオ画像を表示するためにモバイルカメラ装置を位置付けて、モバイルカメラ装置が、第1の言語の1つ以上の語のビデオフィードのフレームを翻訳のために捕捉できるようにすること；(b)ビデオフィードのフレームを、画像処理のために画像処理境界ボックスの内部に適合するように切り取ること；(c)ビデオフィードの切り取られたフレームをメモリ装置に格納すること；(d)画像処理境界ボックス内のビデオフィードの切り取られたフレームを前処理すること；(e)画像処理境界ボ

40

50

ックス内のビデオフィールドの前処理されたフレームに関して文字セグメント認識を実行すること；(f)画像処理境界ボックス内のビデオフィールドの文字セグメント認識されたフレームに関して認識フィードバックを用いて水平マージを実行すること；(g)画像処理境界ボックス内のビデオフィールドの水平マージされた文字セグメント認識済みフレームに関して二値またはグレースケール文字認識を実行すること；(h)第1の言語の1つ以上の語の第2の言語の1つ以上の語への翻訳を生成するために画像処理境界ボックス内のビデオフィールドの文字認識されたフレームを処理すること；(i)第2の言語の1つ以上の翻訳された語を、言語翻訳ビデオ画像の現在のフレームとしてメモリ装置内の位置に格納すること；(j)画像処理境界ボックスが、言語翻訳ビデオ画像の現在のフレームおよび以前のフレームについて、同じ第1の言語テキスト文字上に継続してあることを確認すること；(k)言語翻訳ビデオ画像の現在のフレーム内の情報品質を、言語翻訳ビデオ画像の以前のフレーム内の情報品質と比較すること(言語翻訳ビデオ画像の現在のフレームと言語翻訳ビデオ画像の以前のフレームの両方がメモリ装置内に保存されている)；(l)メモリ装置内のストレージから削除すべき、言語翻訳ビデオ画像の1つ以上の低品質のフレームを選択すること；および(m)第2の言語の1つ以上の語の言語翻訳ビデオ画像の1つ以上のより高品質のフレームを表示し、同時に、翻訳されている第1の言語の1つ以上の語のビデオ画像も表示するために、モバイルカメラ装置を使用すること；の各ステップを含む。

10

【発明の効果】

【0025】

20

本発明は、実行され得る本発明の他の方法を含む、関連したシステム実施形態も含む。かかるシステムは、モバイルカメラ装置に内蔵されたコンピュータシステムとして実装され得る。本発明の様々な実施形態の他の特徴および利点は、添付の図に例示されるように、本発明の実施形態の以下のさらに具体的な記述からさらに明らかであろう。

【0026】

前述の概要は、本発明の好ましい実施形態の以下の詳細な記述に加えて、添付の図と併せて読むと、より良く理解されるであろう。本発明の例示を目的として、現在好ましい実施形態が図に示されている。しかし、本発明は、示されている詳細な構成および手段に制限されないことが理解されるべきである。

【図面の簡単な説明】

30

【0027】

【図1A】本発明の一実施形態に従って、言語の単一行を翻訳するためのプロセスの流れ図のプロセスステップ100～120を示す。

【図1B】本発明の一実施形態に従って、言語の単一行を翻訳するための図1Aのプロセスの流れ図のプロセスステップ122～136を示す。

【図2A】本発明の一実施形態に従って、ビデオフィールドの切り取られたフレームを前処理するためのプロセスの流れ図を示す。

【図2B】本発明の一実施形態に従って、文字セグメント認識を実行するためのプロセスの流れ図を示す。

【図2C】本発明の一実施形態に従って、水平マージされた文字セグメント認識済みフレームに関して二値文字認識を実行するためのプロセスの流れ図を示す。

40

【図3A】本発明の一実施形態に従って、言語の複数行を翻訳するためのプロセスの流れ図のプロセスステップ302～318を示す。

【図3B】本発明の一実施形態に従って、言語の複数行を翻訳するための図3Aからのプロセスの流れ図のプロセスステップ352～382を示す。

【図4】本発明の一実施形態に従い、図3Bのブロック368および376におけるサブルーチンとしての、画像処理境界ボックス内のビデオフィールドの切り取られたフレームの複数行認識のためのプロセスの流れ図を示す。

【図5A】本発明の一実施形態に従った、明るい背景と暗いテキストをもつ二値画像の複数行認識のため、および暗い背景と明るいテキストをもつ二値画像の複数行認識のための

50

プロセスの流れ図を示す。

【図 5 B】本発明の一実施形態に従い、明るい背景と暗いテキストをもつ二値画像に関してサブルーチンとしての複数行認識のため、および暗い背景と明るいテキストをもつ二値画像に関する認識のための、代替プロセスの流れ図を示す。

【図 6 A】本発明の一実施形態に従い、重複する文字のある二値画像タイプに関して認識の後、複数行テキスト取消しを実行するためのプロセスの流れ図のプロセスステップ 6 0 2 ~ 6 1 2 を示す。

【図 6 B】本発明の一実施形態に従い、重複する文字のある二値画像タイプに関して認識の後、テキスト取消しを実行するための図 6 A のプロセスの流れ図のプロセスステップ 6 5 2 ~ 6 6 0 を示す。

【図 7 A】本発明の一実施形態に従い、各二値閾値タイプに対して複数行テキストグループ化を実行するためのプロセスの流れ図のプロセスステップ 7 0 2 ~ 7 2 2 を示す。

【図 7 B】本発明の一実施形態に従い、各二値閾値タイプに対して複数行テキストグループ化を実行するための図 7 A のプロセスの流れ図のプロセスステップ 7 5 2 ~ 7 8 4 を示す。

【図 8】本発明の一実施形態に従い、複数行言語翻訳および単一行言語翻訳を必要とする中国語文字で書かれた例示的な中国レストランのメニューを示す。

【図 9】本発明の一実施形態に従い、境界ボックスの下部にあるタブアイコンを指先で触れて、指先を下方にスライドすることにより、境界ボックスのサイズを拡大するために使用されているモバイルカメラ装置のユーザーインターフェースの一例を示す。

【図 1 0】本発明の一実施形態に従い、図 9 で示された動作の結果の一例を示し、境界ボックスのサイズが、図 9 と比較して図 1 0 で拡大されている。

【図 1 1】本発明の一実施形態に従い、境界ボックス内に第 1 の言語の中国語文字のアルゴリズム生成された文字を表示し、境界ボックスの下に、第 1 の言語の中国語文字の第 2 の言語への翻訳を表示している、モバイルカメラ装置のユーザーインターフェースの一例を示す。

【図 1 2】本発明の一実施形態に従い、境界ボックスの内部に、中国語文字（フェードされた）の翻訳の複数行を英語の翻訳（太字）とともに表示するモバイルカメラ装置のユーザーインターフェースの一例を示す。

【図 1 3】ユーザーが本発明の一実施形態を実施する場合のように、第 1 の言語の中国語文字がより容易に見られる、図 1 2 の一部を示す。

【図 1 4】本発明の別の実施形態に従って、第 1 の言語の中国語文字の発音を表示している、図 1 2 の一部を示す。

【図 1 5】スマートフォンおよびウェアラブルコンピュータを含む、本発明の実施形態を利用し得る、様々な代替エンドユーザー装置を示す。

【発明を実施するための形態】

【0 0 2 8】

本発明は、一実施形態では、第 1 の言語の第 2 の言語への翻訳をリアルタイムで提供するためにモバイルカメラ装置を使用するための方法およびシステムである。本発明は、一実施形態では、スマートフォン上で動作するアプリケーションであり、ある言語で印刷されたオブジェクトテキストに焦点を合わせるためにスマートフォンのカメラ要素およびソフトウェアを使用し、そのテキストが次いで、スマートフォンのディスプレイ内で見られて、ある言語のオブジェクトテキストを別の言語に翻訳し得る。翻訳されたテキストが、同じディスプレイ内で、オブジェクトテキストの表示に隣接して、ユーザーに対して表示される。一実施形態では、翻訳されたテキストは、表示されたオブジェクトテキストの上に浮くように見える。代替実施形態では、本発明は、スマートフォン以外のデジタル装置上で動作し得る。例えば、いくつかの実施形態は、i P a d、ラップトップコンピュータ、および他のコンピュータ制御の器具に適合し得る。一実施形態では、デジタル装置は、コンピュータ制御の眼鏡であり得、眼鏡の着用者は、ある言語でテキストを見ながら、元のテキストに近接して重ねられた別の言語でテキストを見得る。いくつかの実施形態では

10

20

30

40

50

、機能は、デジタル装置に完全にローカルであり得、装置は、本発明の実施形態をオフラインで実施するように動作し得る。他の実施形態では、本発明を実施可能なデジタル装置は、オンラインサーバーとのデータ接続をオープンし得、何らかの機能がオンラインサーバーにおけるソフトウェアおよびデータによって提供され得る。

【0029】

第1の言語の1つ以上の行が翻訳のために選択されている場合、本発明の処理システムは、選択された第1の言語のテキストに焦点を合わせる。これは、ユーザーが、翻訳される第1の言語テキストのモバイルカメラのターゲットボックスビューをさらに容易に位置付けるのを可能にする。いくつかの実施形態では、ターゲットボックス内の第1の言語テキストの焦点合わせは自動プロセスである。モバイルカメラ装置の位置をタップすることを含む、追加の焦点合わせ方法が随意にあり得る。いくつかの実施形態では、第1の言語のテキストを照らすために光源が使用されて、その焦点合わせ、処理、および翻訳を支援する。いくつかの実施形態では、選択されたテキストをターゲットボックスに適合するように縮小できる、ディスプレイ上での縮小のためのズーム制御がある。ズームは、ターゲットボックス内のテキストを、テキスト処理に必要な最小平均サイズまで拡大して、翻訳につなげるためにも使用され得る。一旦、第1の言語のテキストがターゲットボックス内に置かれると、テキストは、処理および第2の言語テキストへの翻訳のために利用可能にされる。モバイルカメラ装置の境界ボックス内に表示される第1の言語の単語が、第2の言語に翻訳される単語である。

【0030】

いくつかの実施形態では、ターゲットボックスは、第1の言語テキストの単一行を含むようにサイズ調整される。この場合、第2の言語テキストへの翻訳は、ターゲットボックスの外側に表示される。別の実施形態では、ユーザーインタフェースは、第1の言語テキストの発音を表示する。境界ボックス内の第1の言語の画像が暗すぎる場合、モバイルカメラ装置上のライト、または別の照明源が、より良い翻訳を実行するために使用できる。

【0031】

本発明の方法およびシステムは、正確で、変動の少ない翻訳を作成する高レベルのアルゴリズム処理を有する。本発明の実施において含まれ得る第1および第2の言語の予期される例は、中国語、韓国語、日本語、ベトナム語、クメール語、ラオ語、タイ語、英語、フランス語、スペイン語、ドイツ語、イタリア語、ポルトガル語、ロシア語、ヒンディー語、ギリシャ語、ヘブライ語、およびアラビア語から成るグループから選択された言語を含む。本発明の実施において含まれる好ましい言語は、アジアの言語、とりわけ中国語、韓国語、および日本語を翻訳することを含む。本発明の特に好ましい実施は、中国語を英語に翻訳するための方法およびシステムを伴う。ここにリストされていない他の人間の言語も、当業者によって認識され得るように、本発明の範囲内であると考えられる。

【0032】

本発明のいくつかの実施形態に対して、翻訳処理のためのコンテキスト情報が、翻訳処理速度に影響を及ぼさない程度まで、使用される。例えば、食べ物の翻訳の場合には、食べ物に関する用語が、翻訳において間違っただけを防ぐために、近くの接頭辞または接尾辞と一緒にまとめることができるように要素を抽出することにより、まとめられ得る。本発明のアルゴリズムは、いくつかの実施形態では、文字列が食料品と判断される場合、食べ物の用語ではない単一文字の翻訳を回避する。かかるプログラミングは、OCR結果の特質により、単一文字が間違っているかも知れない可能性を制御する。複数文字の単語は、間違っている可能性ははるかに低い。第1の言語における1つの単語は、その単語が使用されるコンテキストのため、とりわけ、その単語が第1の言語において複数の意味を有する場合、または第2の言語への翻訳の流動性のために、第2の言語において複数の翻訳を有し得る。1つの好ましい実施形態では、本発明プロセスは、食べ物の翻訳に重点を置き、次いで、標識および旅行翻訳に重点を置く。

【0033】

以下の記述では、説明のために、多数の具体的詳細が、本発明の完全な理解を提供する

ために記載されている。しかし、当業者には、本発明はこれらの具体的詳細なしで実施できることが明らかであろう。他の場合には、本発明を曖昧にするのを避けるために、構造、装置、活動、および方法が、略図、使用事例、および／またはフロー図を使用して示されている。

【0034】

本明細書内での「1つの実施形態」または「一実施形態」への参照は、実施形態に関連して記述される特定の特徵、構造、または特性が、本発明の少なくとも1つの実施形態内に含まれることを意味する。本明細書の様々な場所で使用される句「一実施形態では」は、必ずしも、全て同じ実施形態を参照しているとは限らず、また、他の実施形態と相互に排他的な別個または代替実施形態でもない。その上、いくつかの実施形態によって示され得るが、他によっては示されない、様々な特徴が記述される。同様に、いくつかの実施形態に対する要件であり得るが、他の実施形態に対する要件ではない、様々な要件が記述される。

10

【0035】

以下の記述は、例示を目的として多数の詳細を含むが、当業者は、提案された詳細に対する多数の変形および／または代替は、本発明の範囲内であることを理解するであろう。同様に、本発明の特徵の多くは、相互に関して、または相互に組み合わせて、説明されるが、当業者は、これらの特徵の多くは、他の特徴とは関係なく提供できることを理解するであろう。それに応じて、本発明の本記述は、本発明に対して、いかなる一般性も失うことなく、また本発明に制限を課すことなく、記載される。

20

【0036】

定義

本明細書では、用語「第1の言語」は、本発明の一実施形態を使用して、モバイルカメラ装置によって翻訳される言語を指す。翻訳される第1の言語の単語または複数の単語は、翻訳が生じ得る前に、モバイルカメラ装置のターゲットボックス内に焦点が合って現れる必要がある。

【0037】

本明細書では、用語「第2の言語」は、本発明の一実施形態を使用して、翻訳がモバイルカメラ装置によって表示される言語を意味する。第2の言語での翻訳は、拡張現実画像としてモバイルカメラ装置上に表示される。

30

【0038】

本明細書では、用語「翻訳」は、言語の翻訳、より具体的には、第1の言語テキストから第2の言語テキストへの翻訳を伴う、言語テキストの翻訳を指す。このコンテキストで、用語「翻訳」は、第1の言語の単語テキストを、同じ意味を有する第2の言語の単語テキストにするためのプロセスを意味する。前述したように、第1の言語の単語、または句が、レストランメニュー、本、列車の時刻表、道路標識、店舗の看板、および同様のものの印刷された単語を含む、様々な場所および形態で、本発明のユーザーの前に出現し得る。第2の言語のテキスト伝達が、図11～図14に示すように、モバイルカメラ装置のディスプレイ上でユーザーによって読まれ得る。

【0039】

本明細書では、「拡張現実」は、ウェアラブルコンピュータまたはスマートフォンなどのハンドヘルド装置の使用を通じたコンピュータ介在現実 (computer-mediated reality) を意味し、コンピュータは、人の現実認識に情報を追加するか、もしくは情報を差し引くか、または他の方法で人の現実認識を操作するために使用される。通常、介在されるのは、ユーザーによる環境の視覚認識である。これは、現実の世界とユーザーが認識するものとの間で視覚フィルタとして機能できる、スマートフォンなどの、ある種の電子装置の使用を通して行われる。ウェアラブルコンピュータの例は、GOOGLE GLASS (商標) および同様のものを含む。

40

【0040】

本明細書では、用語「ターゲットボックス」は、モバイルカメラ装置のユーザーインタ

50

フェース上のビューファインダーボックスである。ターゲットボックスの高さは、図 1 1 にターゲットボックス 1 1 0 4 で示すように、第 1 の言語のテキストの単一行のみを表示して翻訳するのを可能にするように設定できる。ターゲットボックスの高さは、図 1 3 にターゲットボックス 1 3 0 6 で示すように、第 1 の言語のテキストの複数行を表示して翻訳するのを可能にするように拡張できる。本発明は、ターゲットボックス内に出現する第 1 の言語の単語を翻訳のために処理する。

【 0 0 4 1 】

本発明は、言語テキスト翻訳をリアルタイムで実行できる。本明細書では、「リアルタイム」は、リアルタイム、またはほぼリアルタイムで、を意味し、ユーザーは、著しい時間遅延なしで、翻訳を見ることができる。リアルタイムは、必ずしも、数学的または物理的な意味での即時を意味するのではなく、ユーザーの前に即座に現れるだけである。

10

【 0 0 4 2 】

本明細書では、「文字」は、字 (l e t t e r)、複数の字、単語、複数の単語、文字 (c h a r a c t e r)、複数の文字、文字セット、複数の文字セット、または言語テキストに関連する任意の他の用語として、視覚的に認識され得る、第 1 の言語テキストの従来型のテキスト特徴を意味する。

【 0 0 4 3 】

本明細書では、「ビデオフィールド」は、ビデオ画像のフレームを意味する。

【 0 0 4 4 】

本明細書では、「モバイルカメラ装置」は、プロセッサ、メモリ装置、およびプログラムコード (アプリケーション) を用いてシステムとして、および本発明を使用するための方法を達成するために機能するカメラを有する携帯型ハードウェア装置を意味する。

20

【 0 0 4 5 】

本明細書では、「格納されたフレーム」は、ビデオカメラからの複数の捕捉された画像 (すなわち、フレーム) のメモリ装置内に保存されたデジタル情報を意味する。

【 0 0 4 6 】

本明細書では、「グレースケール」は、各画素の値が、単一サンプルである、すなわち、光度情報のみを保持する、画像である、グレースケールまたはグレースケール二値画像を意味する。また、この種の画像は、白黒として知られており、最も弱い輝度の黒から最も強い輝度の白まで変動する、グレーの色調から構成されるとして知られている。

30

【 0 0 4 7 】

本明細書では、「カラースケール」は、コンピューティング装置上で使用され得る画像カラースケールを意味する。パーソナルコンピュータは、通常、24ビットの色深度を有することが知られているが、色深度は、装置の機能によって異なるであろう。

【 0 0 4 8 】

本明細書では、「翻訳エンジン」は、プログラムコードを格納するメモリ装置を備えたプロセッサを伴うシステムを意味し、プロセッサは、翻訳を実行するプログラムを実行するためのプログラムコードを実行する。

【 0 0 4 9 】

本明細書では、「連結成分分析 (C C A) 」は、画像をそのセグメントに分割するために、画像処理アプリケーションで使用される分析を意味する。画像は、連結された成分のセットから成るセグメントを有し、連結された成分は、全て黒または全て白のいずれかである画素フィールドを有する画像内の領域である。連結された成分では、画素フィールドは境界によって分離されていない。

40

【 0 0 5 0 】

本明細書では、「ノイズ除去」は、全て黒または全て白のいずれかである画素フィールドを含む、連結された成分と何の関係もないランダムな画素を除去することを意味する。このノイズ除去は、連結成分分析に続いて起こり、連結成分分析は、全て黒または全て白のいずれかである画素フィールドを識別する。

【 0 0 5 1 】

50

本明細書では、「現在のフレーム」は、2つの処理されたビデオ画像フレームのうちの2番目であり、最も直前に翻訳されたビデオフレームである、処理されたビデオ画像フレームを意味する。

【0052】

本明細書では、「以前のフレーム」は、2つの処理されたビデオ画像フレームのうちの1番目であり、処理されている現在のフレームとしてメモリ装置内に格納されたビデオフレームである、処理されたビデオ画像フレームを意味する。

【0053】

本明細書では、「情報品質」は、翻訳される第1の言語テキスト内の単語数に関して、翻訳として第2の言語テキスト内に出現する単語の評価を指す。

10

【0054】

本明細書では、「低品質フレーム」は、翻訳される第1の言語テキスト内の単語数に関して、翻訳として第2の言語テキスト内に出現する単語の低い評価を意味する。

【0055】

本明細書では、「高品質フレーム」は、翻訳される第1の言語テキスト内の単語数に関して、翻訳として第2の言語テキスト内に出現する単語の高い評価を意味する。

【0056】

本明細書では、「画像文字列」は、本発明のプロセスを通じて、第1の言語テキストのビデオフレーム画像の一節を意味する。

【0057】

20

本明細書では、「空白文字列」は、第2の言語テキスト翻訳をもたらさない、本発明のアルゴリズムを通じて、第1の言語テキストのビデオフレーム画像の一節を意味する。

【0058】

本明細書では、「水平方向に重複する」は、2つの別個のテキストプレカーソル (text precursor) が、ビデオ画像フレームの中心水平線に関して、異なる垂直座標を有するが、共通の水平座標を有する部分を有することを意味する。

【0059】

本明細書では、「垂直方向にマージする」は、水平方向に重複しているテキストプレカーソルを結合することを意味する。

【0060】

30

本明細書では、「翻訳テキスト」は、1つの単語、複数の単語、1つの言語文字、複数の言語文字、1つの文字セット、または複数の文字セットとして存在する第2の言語のコンテンツを指す。第2の言語のコンテンツは、モバイルカメラ装置上に拡張現実画像テキストとして表示される。

【0061】

本明細書では、「繁体字中国語文字」は、より多くの字画数を含み得、ほとんどの外国人が簡体字中国語と区別できない、中国語文字の一形式を意味する。

【0062】

本明細書では、「簡体字中国語文字」は、翻訳のプロセスステップで、本発明によって使用される中国語文字の形式を指す。本発明は、繁体字中国語文字であり得る第1の言語テキストから認識された全ての中国語文字をそれらの対応する簡体字中国語文字に変換して、翻訳のステップ中にソートする必要のある中国語文字の数を少なくとも半分に減らす。

40

【0063】

本明細書では、「変形変換」は、翻訳を行う前に、全ての中国語文字を簡体字中国語文字に変換することを意味する。本発明のために、第1の言語の第2の言語への翻訳中に検索される必要のあり得る辞書のサイズを半分にするために変換テーブルが作成され、その結果、翻訳速度が2倍になるであろう。また、繁体字中国語文字の簡体字中国語文字への変換、次いで簡体字中国語文字の第2の言語テキストへの変換は、両方の形式の中国語を直接第2の言語テキストに変換するよりも正確であり得る。

50

【 0 0 6 4 】

本明細書では、「アスペクト比」は、高さと横幅との間の比を意味する。中国語文字はほぼ正方形であるので、中国語文字のアスペクト比は通常、1に近い。

【 0 0 6 5 】

本明細書では、「平均文字サイズ」は、テキスト文字の大部分が、第2の言語への翻訳前に有するサイズとして推定され得る。このサイズは、文字の寸法（高さおよび水平方向の長さ）、および面積（高さ×水平方向の長さ）に関して推定できる。

【 0 0 6 6 】

本明細書では、「正規化」は、画像処理の分野に関し、正規化は、同じクラスの画像の変形を削減するため、画像の形状を固定サイズに調整するために使用される。正規化プロセスは、同じ一定寸法を作成するのに役立ち得、そのため、異なる条件下の2つの画像が同じ特長を有する。

【 0 0 6 7 】

本明細書では、「特徴抽出」は、入力データを特徴のセットに変換することを意味する。これは、アルゴリズムに対する入力データが大きい場合に有用である。次いで、入力データは、削減された特徴の代表となるセットに変換されるであろう。特徴セットは、ビデオストリームの様々な特徴を検出して分離するために、関連情報を入力データから抽出して、本発明のアルゴリズム内で満足のいくように実行できる。

【 0 0 6 8 】

本明細書では、「次元縮退」は、分類の前に、特徴の数をさらに管理可能な数に減らすためのパターン認識処理を指す。

【 0 0 6 9 】

本明細書では、「クラスタリングでの分類」は、いくつかのタイプの凝縮型階層的クラスタリングを実行することを意味する。このプロセスは、クラスタのペアを見つけることにより機能して、経路が相互に類似したクラスのパアで終了するまで、クラスタの分類グラフ内の経路に従うことにより統合する。

【 0 0 7 0 】

本明細書では、「翻訳スコア」は、より多くの単語が翻訳されたことを意味する、より良い翻訳を表す数学関数を指す。

【 0 0 7 1 】

単一行翻訳実施形態の詳細な説明

図は本発明の実施形態のためのプロセス例を提供するのに過ぎない。アルゴリズム例は、第1の言語が中国語で、英語への翻訳である場合に有用な翻訳プロセスを対象とするが、本発明人は、任意の2言語間での両方向への翻訳を予期する。図1Aおよび図1Bは、本発明の一実施形態に従い、第1の言語の単一行を第2の言語に翻訳するために、ビデオモードで実行するアルゴリズムまたはプロセスの流れ図150を示す。図1Aで、プロセス150はステップ100から始まる。プロセス150はビデオモードで実行する。プロセスが終了するたびに、プロセスはトップに戻って、新しいフレームをビデオから捕捉して、再度プロセスを実行する。このプロセスは、認識された処理文字列（process string）を作成して、対応する翻訳がモバイルカメラ装置のディスプレイ画面上に現れる。ステップ102で、本発明のモバイルカメラ装置のディスプレイ上の画像に焦点が合っているかを判断するために、プロセスによって判断が実行される。プロセスは、ユーザー装置上のカメラがオートフォーカス機能を操作するのを可能にする。カメラが焦点を合わせている間、プロセスは、カメラが焦点合わせを終了するまで、いかなる処理も行わないことなく、ステップ102を繰り返してチェックする。次いで、プロセスはステップ104に進んで、処理を行う。カメラは、時々、既に焦点が合っていると考えるため、焦点を合わせようとすることなく、ぼやけた画像を処理する。その結果、プロセスは、ユーザーがそれに再度焦点を合わせるようにさせるための、タップしてフォーカス（tap-to-focus）機能を提供する。ステップ104で、プロセスは、ユーザーが、翻訳するためにテキストの単一行またはテキストの複数行を選択しているかを判断する。ユ

10

20

30

40

50

ーザーが複数行を選択している場合には、プロセスはステップ１０６に進み、そこで、図３で説明する複数行翻訳プロセスが呼ばれ；そうでない場合、プロセスは単一行翻訳に進む。

【００７２】

ステップ１０８で、プロセスは、ユーザーによって選択された画像処理境界ボックスから画像を切り取る。画像を切り取ることは、文字を強調するために、境界ボックス内の第１の言語文字の画像の外側部分を取り除くことを指す。ステップ１１０で、図２Ａに関して以下で説明するように、文字検出のための前処理が生じる。ステップ１１２で、前処理によって、テキスト文字を示唆する何らかの予備的なテキスト情報があることを示し得るテキストブレカーソルが明らかにされているか否かの判断が行われる。ステップ１１０でテキストブレカーソルに対する指標がない場合、プロセスは、ステップ１１４で、黒いボックスを示して、リセットするか、または、同様のものを示して、開始ステップ１００に戻る。

10

【００７３】

テキストブレカーソルがステップ１１２で識別されている場合、ステップ１１６のプロセスは、以下の図２Ｂでさらに詳細に説明するように、文字セグメント認識を実行する。ステップ１１８で、プロセスは、以前のステップで認識された文字が小さすぎるかを判断する。認識された文字が小さすぎるとプロセスが判断する場合、プロセスは、ステップ１２０に進み、そこで、「拡大するか、または近づいてください」または同様のメッセージをユーザーに表示して、プロセスは開始ステップ１００に戻る。ステップ１１８で、認識された文字が十分に大きいと判断される場合には、プロセスは、図１Ｂのステップ１２２に進む。ステップ１２２で、プロセスは、非中国語文字を除去してステップ１２４に進み、そこで、処理文字列内に翻訳の対象となる言語文字が見つからないことを意味する、処理文字列が空白であるかに関して判断が行われて、プロセスはステップ１２６に進み、そこで、「画像が不明瞭」、「フラッシュを使用してください」、または同様のメッセージがユーザーに対して表示される。プロセスは、次いで、ステップ１２６から、プロセス１５０の初めの開始ステップ１００に戻る。

20

【００７４】

ステップ１２４の判断で、文字が見つまっていることを示している場合、プロセスはステップ１３０に進み、そこで、プロセスは変形変換を実行する。ステップ１３０での変形変換は、繁体字中国語文字を簡体字中国語文字に変換することにより、辞書内の用語数を減らす。全ての中国語文字を簡体字中国語文字に変換することは、時々、翻訳すべき中国語テキストが簡体字および繁体字中国語文字の組合せであるために実行される。繁体字を簡体字に変換することは、簡体字を繁体字に変換することよりもずっと簡単である。ほとんどの外国人は、簡体字と繁体字中国語文字を区別できない。ステップ１３０のプロセスは、ステップ１３２の文字の翻訳処理でスキャンする必要がある中国語文字辞書のサイズを減らす。処理およびメモリ容量が、いくつかのモバイルカメラ装置に対する処理速度制限であり得るので、小さい中英辞書は、大幅に処理量を減らし、従って、モバイルカメラ装置内の単一行アルゴリズムの処理速度を向上させる。ステップ１３２で、プロセスは、簡体字中国語文字を英単語に翻訳するために、光学式文字認識（OCR）プロセスからの結果を使用する。ステップ１３２の翻訳プロセスが完了すると、プロセスはステップ１３４に進む。

30

40

【００７５】

ステップ１３４で、プロセスは、画像処理境界ボックスが、以前の文字列と比べて現在の文字列内の同じテキスト上に留まっているかをチェックする。[ステップ１３４のプロセスは：a)現在の文字列内の文字特徴の類似点を、重複について、以前の処理文字列内の文字特徴と比較すること、またはb)現在の画像および以前の画像の安定性をチェックするためのトラッキング方法、のいずれかのプロセスによってこれをチェックする]。ステップ１３４のプロセスは、境界ボックスが同じテキスト上に留まっていることを確認するために、一致した文字数が総文字数に対して十分に多いかを計算する。プロセスは１３

50

4 からステップ 1 3 6 に進み、そこで、現在の翻訳が以前の翻訳と比較される。より良い翻訳が保存されて、劣った翻訳がステップ 1 3 6 でプロセスによって削除される。流れ図 1 5 0 が処理文字列を終えるたびに、プロセスは、流れ図 1 5 0 の初めに戻って、新しいフレームをビデオから捕捉する。このプロセスは、認識された文字列を生成して、対応する翻訳が、モバイルカメラ装置のディスプレイ上に示される。

【 0 0 7 6 】

図 2 A は、ステップ 2 0 2 から始まる文字認識のための前処理に対する流れ図を示す。ステップ 2 0 2 で、切り取られたグレースケール画像がプロセスステップ 1 1 0 から入力されている。ステップ 2 0 2 はステップ 2 0 6 に進み、そこで、切り取られたグレースケール画像が固定サイズにアップサンプルされ、次いで、プロセスはステップ 2 0 8 に進む。

10

【 0 0 7 7 】

ステップ 2 0 8 で、プロセスは、グレースケール画像の二値化における閾値タイプの判断を実行する。テキストおよび背景の輝度値が、閾値タイプが明るいブレカーソル文字をもつ暗い背景か、または暗いブレカーソル文字をもつ明るい背景かを判断するために利用される。閾値タイプを判断するために、プロセスは、各行内の画素の輝度値を判断する。プロセスは、次いで、輝度値の線形結合を比較して、閾値タイプを判断する。二値化における閾値タイプを判断した後、ステップ 2 0 8 のプロセスは、次いで、適応閾値二値化処理に進んで、画像の領域上の照明状態における変化、例えば、強い照明または影の結果として生じるもの、を制御するために、テキストおよび背景の輝度値を比較する。二値化および二値化プロセスにおける閾値タイプを判断した後、プロセスはステップ 2 1 0 に進む。

20

【 0 0 7 8 】

図 2 A での処理およびそれ以後に対して、本発明の代替実施形態として、ステップ 1 1 0 からの切り取られた画像の処理が、グレースケールではなくカラースケールであり得る。モバイルカメラ装置のユーザーインタフェース上の英語に翻訳された単語が、赤、オレンジ、黄、緑、青、ピンク、紫、および任意の他の色の組合せから成るグループから選択されたフォント色で提示され得る。

【 0 0 7 9 】

ステップ 2 1 0 で、処理文字列の二値化画像をそのセグメントに分割するために連結成分分析 (C C A) が実行される。連結された成分は、全て黒または全て白のいずれかの画素のフィールドを有する。プロセスが連結成分分析 (C C A) を完了した後、プロセスはステップ 2 1 2 に進む。ステップ 2 1 2 で、プロセスは、連結成分のサイズおよび形状情報を検査することにより、画素の個々の小さいクラスタを除去することにより二値化された連結成分をノイズ除去し、次いで、プロセスは、ステップ 2 1 4 に進み、そこで流れ図 2 0 0 のプロセスが終了して処理文字列を図 1 A のステップ 1 1 2 に返す。

30

【 0 0 8 0 】

図 2 B は、流れ図 2 0 0 の続きであり、図 2 B のプロセスは、ステップ 2 1 6 で文字セグメント認識のプロセスを開始する。プロセスはステップ 2 1 8 に進み、そこで、テキストブレカーソル情報がメモリ装置に格納される。ステップ 2 1 8 から、プロセスは、次いで、ステップ 2 2 0 に進んで、水平方向に重複しているテキストブレカーソルを識別して結合することにより、垂直マージ (v e r t i c a l m e r g i n g) を実行する。水平方向に重複しているテキストブレカーソルは、画像フレームの中心水平線に関して、異なる垂直座標をもつ部分を有するが、共通の水平座標を共有する、別個のテキストブレカーソルである。この場合、重複している水平座標を有していない、近接しているが、別個のテキストブレカーソルは、この段階では同じ中国語テキスト文字の副次成分として処理されないであろう。テキストブレカーソルの垂直マージのプロセスの後、プロセスはステップ 2 2 2 に進んで、単一行内で処理されているテキストブレカーソルの明らかな単一行に対する外れ値であるアーチファクトのテキストブレカーソルを除外する。

40

【 0 0 8 1 】

プロセスステップ 2 2 2 で、処理は、単一行の中国語テキストの 3 つの共通特性によ

50

て誘導される。第 1 に、中国語テキスト文字は、文字内の全てのセグメント部分の外側縁において正方形のようなアスペクト比を有する。第 2 に、中国語テキスト文字は、同様の垂直方向の高さを有する。第 3 に、中国語文字の単一行は、常に、文字の真っ直ぐな単一行であり、そのため、テキストの単一行内に別の中国語文字よりも高い中国語文字はない。従って、単一行としてのステップ 222 のプロセスは、中国語テキストの単一行を、中心水平線から延びる領域の外側のテキストプレカーソルについて、任意のデータを削除するための処理で処理し、この領域では、あらゆる行が、現在の画像文字列内の少なくとも 1 つのテキストプレカーソルに重なる。その結果、プロセスが、ステップ 222 で、画像フレームの中心水平線から垂直方向に延びる領域の外側のアーチファクトのテキストプレカーソルを除去した後、プロセスはステップ 224 に進む。ステップ 224 で、プロセスは、テキスト文字を左から右への順番でソートし、次いで、プロセスはステップ 226 に進み、そこで、中国語文字は全体として四角い箱の形状を有すると仮定して、文字サイズが検査される。ステップ 226 から、プロセスは、判断ステップ 228 に進み、そこで、プロセスは、平均文字サイズが小さいかを判断する。プロセスがステップ 228 で、平均文字サイズが小さすぎると判断すると、プロセスは文字セグメント認識を終了して、図 1A の流れ図 150 のステップ 120 に戻る。プロセスがステップ 228 で、平均文字サイズが小さすぎないと判断する場合には、プロセスはステップ 230 に進む。

【0082】

ステップ 230 で、プロセスは、ステップ 228 からの二値化された垂直マージ済みテキストプレカーソルに関する文字認識フィードバックを使用した水平マージでの処理を実行する。水平マージプロセスは、単一行内の左端のテキストプレカーソルについて始まる。プロセスは、テキストプレカーソル全体でのテキストの組合せについて境界ボックス形状をチェックして、それらの各々に対する形状スコアを取得する。画像処理境界ボックスが言語プロファイルに適合するアスペクト比を有する場合、組合せの距離スコアを判断するために、組合せが、図 2C（以下で説明する）に示される文字認識フィードバックプロセス内で処理される。プロセスは、最高の形状スコアおよび距離スコアを有するテキストプレカーソルの最善の組合せを選択して、この「オブジェクト」を除外し、次いで、単一行の画像文字列内で右端のオブジェクトがなくなるまで、最も近い右側のオブジェクトについて文字認識フィードバック処理を用いて水平マージを繰り返す。統合された形状および距離スコアのいずれも文字であると十分に確信しない場合、1 つのオブジェクトだけが除外される。多くの中国語文字は、いくつかの他の文字から成り、そのため、文字自体およびその副次成分の距離スコアが類似している場合、形状情報を使用すると、処理が最も可能性の高い文字を見つけるのに役立つ。これは、文字列内の文字がともに近接していて、分割するのが困難な場合の問題も解決する。所与の文字列を分割するためのこの「欲張り」アルゴリズムは、大域的最適解を計算する必要なく、モバイル機器上での計算要件を減らす。ステップ 230 のプロセスが完了すると、プロセスはステップ 232 に進み、そこで、プロセスは、図 1B 内のステップ 122 に戻るように指示される。

【0083】

図 2C は、水平マージ処理されたテキストプレカーソルの結合の適合性をチェックするサブルーチンとして機能する文字認識フィードバックプロセスを示し、テキストプレカーソルの結合は、図 2B のステップ 230 から供給されている。文字認識プロセスのための図 2B のステップ 230 からのテキストプレカーソルの結合の処理が、図 2C に示すステップ 234 で始まる。ステップ 234 の二値文字認識プロセスはステップ 238 に進む。ステップ 238 で、処理は、処理文字列の二値画像上で関心領域 (ROI) を判断する。ステップ 238 での二値画像上の関心領域 (ROI) は、連結成分の集合を含む。ステップ 238 のプロセスは、ステップ 240 に進み、そこでは、画像正規化を引き起こすための処理がある。二値画像の正規化は、同じクラスの画像のばらつきを削減するために、ROI 内の画像の形状を固定サイズに正規化するプロセスである。ステップ 240 のプロセスが完了すると、プロセスは 242 に進んで、特徴抽出処理を実行する。特徴抽出のプロセスは、入力データを特徴のセットに変換する。正規化された画像データである処理文字

列の入力データは非常に大きい。従って、特徴抽出は、アルゴリズムの後続のプロセスステップにおいてデータのサイズを削減するために重要である。正規化された関心領域の特徴抽出を実行するための処理の後、文字認識プロセスはステップ244に進む。ステップ244で、プロセスは、次元縮退を実行する。次元縮退処理は、分類の前に、特徴の数をさらに管理可能な数に減らすために、ステップ244で使用する。次元縮退の後、プロセスは、削減された文字特徴のクラスタリング処理での分類のためにステップ246に進む。クラスタ処理での分類のプロセスは、クラスタの分類グラフ内の経路を、その経路が類似したクラスのペアで終了するまで、辿ることにより、マージする、クラスタのペアを見つける凝縮型階層的クラスタリングを引き起こす。ステップ246が完了すると、プロセスはプロセスステップ248に進み、ステップ248は、認識された二値文字データをもつ処理文字列に、図2Bのステップ230に戻るよう指示する。

10

【0084】

複数行翻訳実施形態の詳細な説明

本発明は、トラッキングおよびサブサンプリングされた画像処理を使用する、複数行テキスト翻訳プロセスを提供する。複数行テキスト認識処理は、フレームを処理するために、単一行テキスト認識処理よりも多くのコンピュータ使用を必要とする。リアルタイムのユーザーエクスペリエンスを提供するために、言語翻訳がテキスト上に拡張現実画像として重ね合わされ、トラッキング方法が、テキストの移動を検出するために使用される。トラッキングは、2つの目的を果たす：1つは、カメラが同じテキストに焦点を合わせているかどうかを確認する。画像文字処理は、2つの連続した動かない画像が捕捉される場合に限り行われる。トラッキングのもう1つの目的は、テキストの移動方向を取得して、テキスト位置がそれに応じて調整できるようにすることである。リアルタイムトラッキングのために、処理がトラッキングを行う前に、画像をサブサンプリングする。トラッキングは、xおよびy移動をもつベクトルを取得するために、現在のフレームおよび以前のフレームについて実行される。ユーザーは、通常、平坦な領域上のテキストに焦点を合わせるので、処理は、ベクトルが一貫性があるか小さいかどうかを確認することにより、ユーザーがスマートフォンをしっかりと保持しているかを判断する。ベクトルが一貫性があるか小さい場合、文字認識が、捕捉されたビデオフレームを使用するか、または画面上のテキスト位置を調整することにより、実行できる。

20

【0085】

翻訳を画像内のテキスト上にリアルタイムで重ね合わせるために、マルチスレッドプログラミングが使用される：テキスト位置のトラッキングが、1つのスレッド内で行われ、その間に文字認識が別のスレッド内で行われる。トラッキングを経由するループは、認識と比較して非常に高速であり、従ってテキスト位置がリアルタイムで調整できる。認識結果が準備できると、認識結果が抽出され、画面上で更新されて、テキスト位置で更新される。必要ならば、別の認識が別のスレッド内で行われる。この更新で、以前の結果が検査され、より良い結果が各テキスト位置に対して保存される。

30

【0086】

複数行文字認識方法は、2方法の二値化；水平ぼかし(horizontal blurring)；疑わしい文字部分を有するビデオフレームの認識の回避；テキスト取消しおよび水平行グループ化：を実行する。複数行プロセスは、2つのタイプの二値化：暗いテキスト/明るい背景および明るいテキスト/暗い背景を使用する。次いで、水平方向のテキストを検出するために、水平ぼかし処理が二値化画像について使用される。これは、効果的に、かつ、おそらくはテキスト位置を失うことなく、行われ得る。水平ぼかしの後、テキスト認識がこれらの領域について行われる。適切なサイズまたはアスペクト比を有していない領域が、処理速度を向上するためにスキップされる。次いで、重複する異なるタイプの二値化から文字列の1つを取り消すために、テキスト取消しが行われる。2つの文字列が重複する場合、多くの文字を有する方が保存される。最後に、文字が、同じ領域内になく、バラバラに離れている場合、テキストのグループ化が行われる。文字の2つの領域が、間隔および位置情報に従って、一緒にグループ化される。

40

50

【 0 0 8 7 】

図に従った複数行プロセスのプロセス説明がここから始まる。図 1 A のステップ 1 0 4 での判断プロセスが、翻訳のための複数行テキストの第 1 のフレームをステップ 1 0 6 に送り、そこで、プロセスは、複数行のテキストの第 1 のフレームを図 3 A のステップ 3 0 2 に送り、そこで、複数行テキスト翻訳プロセス 3 0 0 の処理が開始する。ステップ 3 0 2 のプロセスはステップ 3 0 4 に進み、そこで、プロセスは、画像処理境界ボックス内の画像フレームの外側部分を除去するために、複数行テキストの第 1 のフレームを切り取る。プロセスは、次いで、複数行テキストの第 1 の切り取られたフレームをステップ 3 0 6 に送る。ステップ 3 0 6 で、プロセスは、切り取られたフレームが、複数行のテキストの第 1 のフレームであるかを確認する。複数行テキストの切り取られたフレームが第 1 のフレームであれば、プロセスは、複数行テキストの切り取られた第 1 のフレームをステップ 3 0 8 に送る。ステップ 3 0 8 のプロセスは、複数行テキストの第 1 のフレームをメモリ装置に保存する。ステップ 3 0 8 における処理文字列は、次いで、ステップ 3 0 4 に戻る。ステップ 3 0 4 で、プロセスは、複数行テキストの第 2 のフレームを切り取って、その切り取られた第 2 のフレームをステップ 3 0 6 に送る。ステップ 3 0 6 のプロセスは、切り取られたフレームが複数行テキストの第 1 のフレームかを判断する。ステップ 3 0 6 における現在の切り取られたフレームが第 1 の切り取られたフレームではない場合、ステップ 3 0 6 のプロセスは、複数行テキストの切り取られた第 1 のフレームおよび第 2 のフレームをステップ 3 1 0 に送る。判断ステップ 3 1 0 で、プロセスは、モバイルカメラ装置のユーザーインタフェース上の一時停止ボタンが押されているかどうかを確認する。ユーザーインタフェース上で一時停止が押されていない場合、ステップ 3 1 0 の判断プロセスは、第 1 の切り取られたフレームおよび第 2 の切り取られたフレームをステップ 3 1 2 に送る。ユーザーインタフェース上で一時停止ボタンが押されている場合、ステップ 3 1 0 の判断プロセスは、ステップ 3 8 0 に進み、そこで、プロセスは、図 3 B に示すステップ 3 7 8 からの画像の処理を一時停止する。

【 0 0 8 8 】

ステップ 3 1 2 で、プロセスは、以前および現在のフレームの両方に対して切り取られた画像のサイズ変更を実行してから、切り取られて、サイズ変更された以前のフレームおよび現在のフレームについてトラッキングを実行する。ステップ 3 1 2 で、プロセスは、現在および以前のフレームのトラッキングを実行する。トラッキング位置の各々で、プロセスは、以前のフレームから現在のフレームへの画像の位置における変化を計算し、プロセスは、各トラッキング位置における以前のフレームから現在のフレームへの移動を、 X および Y 値を持つベクトルに関して定義する。プロセスは、以前のフレームを基準トラッキングフレームとして使用して、処理は最終的に現在の複数行テキストフレームのみを継続する。プロセスは、ステップ 3 1 2 からステップ 3 1 4 に進み、そこで、ステップ 3 1 4 でのトラッキングからのベクトル結果が検査される。

【 0 0 8 9 】

プロセスは判断ステップ 3 1 6 に進み、そこで、トラッキングからのベクトル結果が、モーションブラーのある不安定な画像を処理するのを防ぐために使用される。ステップ 3 1 6 の処理は、まず、(a) トラッキング位置のベクトルが類似しているか否かを判断する。ステップ 3 1 6 の処理は、(b) 認識結果更新間で累算されたトラッキング位置の平均ベクトルが小さいか否かも判断する。加えて、ステップ 3 1 6 の処理は、(c) 現在の画像サイズが以前のフレームの画像サイズと一致するか否かを判断する。ステップ 3 1 6 の処理が判断 (a ~ c) に基づき、以前のフレームに比べて現在のフレームの著しい動きがあることを示す場合、ステップ 3 1 6 の処理は、複数行テキストの現在および以前のフレームの両方をステップ 3 1 8 に送る。ステップ 3 1 8 で、プロセスは、複数行テキストの現在および以前のフレームをメモリから削除して、処理文字列を開始ステップ 3 0 2 に戻す。従って、プロセスが、(1) トラッキングアルゴリズムのベクトルが一貫して類似している；(2) 認識結果更新間で累算されたトラッキングされた位置の平均ベクトルが小さい；かつ(3) 現在のフレームの画像サイズが以前のフレームの画像サイズと一致す

10

20

30

40

50

る；と判断する場合に限り、現在のフレームが、図3Bに示すように、ステップ352に進められる。ステップ316のプロセスが現在のフレームの処理文字列をステップ352に送る場合、316ステップのプロセスは、以前のフレームを廃棄する。

【0090】

判断ステップ352で、ステップ368において以前の複数行認識結果があるか否かについて判断が行われる。プロセスが、ステップ368において複数行認識結果がなかったと判断すると、プロセスは、複数行テキスト画像フレームをステップ356でのオートフォーカスのために送る。プロセスは、焦点合わせが完了するまで、ステップ358で待機する。プロセスはステップ366に進み、そこで、プロセスは、フル解像度の画像を得るために、複数行テキストフレームを切り取る。ステップ366の後、プロセスはステップ368に進み、そこで、焦点が合って、切り取られた複数行テキストフレームが、後に説明する、図4のステップ402から始まる複数行テキスト認識プロセスに進む。ステップ368、次いでステップ352において複数行認識結果がある場合、判断は、現在のフレームが複数行認識処理に対する第1のフレームではないということであり、従って、現在のフレームは、別のスレッドが実行しているか否かをプロセスが判断する、判断ステップ354に送られるであろう。

【0091】

ステップ354のプロセスが、スレッドカウントがゼロであると判断すると、処理は複数行テキスト画像の現在のフレームをステップ362に送る。ステップ362のプロセスは、モバイルカメラ装置上に表示された翻訳テキスト結果を更新する。ステップ362の処理は、複数行テキスト画像のフレーム内でテキストの各個々の行を別々に確認し；テキストが境界ボックス内に留まっているかを確認し；かつ以前の翻訳結果が現在の翻訳結果よりも良いかどうかを確認し、そうである場合は、以前のフレーム翻訳結果を使用する。プロセスステップ362の後、プロセスはステップ372に進み、そこで、プロセスは、ベクトルデータの過剰累算を処理するのを防ぐために、累算された小さいトラッキングベクトルをゼロにリセットする。プロセスはステップ374に進み、そこで、プロセスは、より高い解像度の画像を得るために画像フレームを切り取り、次いで、処理はステップ376に進み、そこで、図4に示すように、複数行認識処理のためのスレッドが作成される。ステップ378で、現在の画像フレームが保存される。現在のフレームはステップ304で切り取られたことに留意されたい。

【0092】

ステップ354で、判断が、スレッドカウントがゼロではないという場合、ステップ354のプロセスは、ステップ360に進んで、モバイルカメラ装置の画像ディスプレイ上に現れる翻訳されたテキストの以前のフレームからのテキスト位置を調整する。調整では、現在のフレームを以前のフレームと比較するためのベクトル計算プロセスを使用して、モバイルカメラ装置の画像ディスプレイ上に現れる翻訳されたテキストを移動させるが、以前のフレームはステップ312においてトラッキング結果を提供した。ステップ360の後、プロセスはステップ364に進み、そこで、シフトベクトルがステップ360から累算される。

【0093】

ステップ368、364、および376からの複数行テキスト画像処理のためのプロセス経路がステップ378で集束され、そこで、現在の画像フレームが切り取られた複数行テキスト画像として保存される。ステップ38で一時停止ボタンが押されている場合、プロセスは画像処理を一時停止する。翻訳を失うことなくモバイルカメラ装置を移動するために、適切な翻訳が得られている場合に一時停止ボタンが頻繁に押される。ステップ380のプロセスは、次いで、ステップ382に進み、そこで、処理が、画面の位置の座標を、処理が生じる画像パuffa（メモリ）の位置と一致させる。図3Bに示すステップ368および376のプロセスは、プロセス400内の図4に示す複数行認識サブルーチンプロセスである。

【0094】

より詳細には、プロセス400はステップ402から開始する。ステップ404で、プロセスは、明るい背景と暗いテキストがある、第1のタイプの二値化を実行する。ステップ406のプロセスは、次いで、後に説明する図5Aに示すサブルーチンプロセス500を使用するか、または後に説明する図5Bに示すサブルーチンプロセス550のいずれかを使用して、第1のタイプの二値複数行画像について認識処理を実行する。プロセス500または550内の処理が完了した後、プロセスは図4内のステップ408に戻り、そこで、プロセスは、暗い背景と明るいテキストがある、第2のタイプの二値化を実行する。ステップ410のプロセスは、次いで、図5Aに示すサブルーチンプロセス500を使用するか、または図5Bに示すサブルーチンプロセス550のいずれかを使用して、第2のタイプの二値複数行画像について認識処理を実行する。プロセス500または550内の処理が完了した後、プロセスは図4内のステップ412に戻り、そこで、プロセスは、図6Aおよび6Bに示すサブルーチンプロセス600を使用して、重複したテキストを検出するためにテキスト取消しを実行する。プロセス600内の処理が完了した後、プロセスは図4内のステップ414に戻り、そこで、プロセスは、第1のタイプ(「タイプ1」)の二値閾値に対するテキストのグループ化および第2のタイプ(「タイプ2」)の二値閾値に対するテキストのグループ化を実行する。ステップ414が、後に説明する図7Aおよび7B内のサブルーチンプロセス700で処理される。二値閾値の各タイプに対するテキストのグループ化が完了した後、プロセスは、図4内のステップ416に戻る。ステップ416内の翻訳プロセスは、光学式文字認識からの結果を使用して、第1の言語のテキストである中国語文字を、第2の言語のテキストである英単語に翻訳する。翻訳の出力が、モバイルカメラ装置上に、リアルタイム拡張現実画像として表示される。

【0095】

翻訳エンジンはステップ416で、翻訳エンジン結果スコアを計算する。翻訳エンジンスコアは、中国語文字が良好に認識されて、中国語文字の英語への良好な翻訳の確率が高い兆候がある場合に、高い。現在のフレームのテキスト行が、以前のフレームに対してそのままであって、動いていないことを確認した後、現在および以前のフレームの翻訳に対する行の複数行翻訳エンジンスコアが比較されて、以前の翻訳エンジンスコアの方が高い場合、行の以前のフレームの翻訳が保持されて、現在の翻訳は保存されない。翻訳エンジン結果は、翻訳結果の認識された有効性を計算し、次いで、結果として生じた語全体にわたってそれらを合計することにより、得点される。複数行での翻訳プロセスは、複数行テキスト画像の処理および翻訳において一度に1行を処理する。

【0096】

別の実施形態では、垂直テキスト認識がプロセスとして、水平テキスト認識プロセスのプロセスに追加され得る。処理は、第1のフレーム認識において、テキスト方向を、垂直および水平方向の両方で自動的に決定するであろう。翻訳スコアは、各方向で合計され得、高い方のスコア翻訳は、その方向でより多くの意味のある句を有し得るので、高い方の翻訳スコアがさらなる処理のために選択されるであろう。時々、プロセスは、翻訳のために認識される多数の文字がないことを示す低翻訳スコアを有し得る。いくつかの翻訳スコアは単にノイズであり得る。従って、ユーザーが焦点を合わせ始め得るのは画像の中心からなので、プロセスは、画像の中心周囲のテキスト行の翻訳スコアの合計に焦点を合わせるであろう。

【0097】

前述のように、ステップ410のプロセスは、次いで、図5Aに示すサブルーチンプロセス500を使用するか、または図5Bに示すサブルーチンプロセス550のいずれかを使用して、第2のタイプの二値複数行画像について認識処理を実行する。図5Aのステップ502は、二値複数行画像の1つのタイプについて認識処理のプロセスを開始する。このプロセスは、ステップ504に進み、そこで、プロセスは連結成分を見つけて、複数行テキスト画像のノイズ除去を行う。プロセスは次いで、プロセスステップ506に進み、そこでテキスト位置を見つけるために水平ぼかしが実行される。

【0098】

10

20

30

40

50

プロセスは次いで、ステップ508に進み、そこで、処理文字列の二値化画像をそのセグメントに分割するために、連結成分分析が、水平方向にぼやけた画像について実行される。連結成分は、全部黒または全部白のいずれかの画素フィールドを有する。プロセスが連結成分分析を完了した後、プロセスはステップ510に進み、そこで、プロセスは、水平ぼかしによって画定された各連結成分の単一行領域の内部に位置する連結成分について、光学式文字認識(OCR)を実行する。ステップ510での処理は、二値認識された文字のサイズが小さすぎる場合、またはアスペクト比が水平方向のテキスト行を形成する可能性が低い場合、単一行領域についてOCRを行わない。

【0099】

二値画像複数行画像についての認識に対する代替経路550が図5Bに示されており、ステップ560から始まる。ステップ562で、二値画像複数行画像がノイズ除去される。プロセスはステップ564に進み、そこで水平ぼかしが実行され、ステップ566で、元の画像内の対応する領域の切り取りを用いた連結成分分析が続く。連結成分分析は、処理文字列の二値化画像をそのセグメントに分割するために実行される。連結成分は、全て黒または全て白のいずれかの画素のフィールドを有する。プロセスが連結成分分析を完了した後、処理はステップ566からステップ568に進み、そこで、各連結成分の領域に対して単一行のように、プロセスは、二値認識された文字のサイズが小さすぎるか、またはアスペクト比が水平方向のテキスト行を形成する可能性が低い場合に限り、OCRを実行する。次いで、プロセスは、複数行テキスト画像の各単一行をさらに細かい解像度にサイズを変更して、二値化を繰り返す。認識されている画像がより良い解像度を有し、従って、文字のさらなる細部を含むので、図5Bのプロセス550は、図5Aのプロセスよりも精度が良いが、処理速度は、図5Aのプロセス500における処理よりも遅い可能性がある。

【0100】

前述のように、図4のプロセスはステップ412で、部分6Aおよび6Bを有する図6に示されるサブルーチンプロセス600を使用して、重複したテキストを検出するためにテキスト取消しを実行する。図6Aのプロセス600で、文字認識後のテキスト取消しがステップ602で開始して、ステップ604に進み、そこで、プロセスが、タイプ1の二値化およびタイプ2の二値化における文字数をカウントする。定義により、タイプ1の二値化は白い背景上に出現する黒いテキストから成り、タイプ2の二値化は逆の二値化、すなわち、黒い背景上に出現する白いテキストである。ステップ604のプロセスは、判断ステップ606に進み、そこで、タイプ1の二値化を通る処理ループが終了されているかに関して判断する。タイプ1の二値化を通る処理ループが終了されている場合、処理文字列が、図7A内のプロセス700の開始ステップ702でテキストのグループ化を開始するために送られる。タイプ1の二値化を通る処理ループが終了されていない場合、プロセスは判断ステップ608に進んで、複数行テキスト画像フレームの単一行内に中国語文字があるかどうかを判断する。

【0101】

ステップ608で、ステップ608における行内に中国語文字がないと判断される場合、処理文字列はステップ606に戻されて、複数行テキスト画像の別の単一行内についてタイプ1の二値化を繰り返す。ステップ608で、複数行テキスト画像フレームの単一行内に中国語文字があると判断する場合、プロセスは610に進んで、その単一行の高さおよび長さの境界を計算する。ステップ610のプロセスは次いで、判断ステップ612に進み、そこで、プロセスは、タイプ2の二値化を通る処理ループが終了されているかを判断する。タイプ2の二値化を通る処理ループがステップ612で終了されている場合、処理文字列がステップ606に戻って、複数行テキスト画像の別の単一行を処理する。タイプ2の二値化を通る処理ループがステップ612で終了されていない場合、処理文字列は図6B内のステップ652に進む。ステップ652は、複数行テキスト画像フレームの単一行内に中国語文字があるかどうかを判断する判断ステップである。複数行テキスト画像フレームのこの単一行内に中国語文字がない場合、処理は、処理文字列をステップ612

に送って、複数行テキスト画像フレームの別の単一行についてタイプ2の二値化を繰り返す。複数行テキスト画像フレームの単一行内に中国語文字がある場合、プロセスはステップ654に進んで、複数行テキスト画像フレームの特定の単一行の長さおよび高さの境界を識別する。

【0102】

ステップ654のプロセスは、判断ステップ656に進み、そこで、ステップはタイプ1およびタイプ2の二値化の文字列が重複するかどうかを判断する。タイプ1およびタイプ2の二値化が重複しない場合、プロセスは判断ステップ612に戻って、複数行テキストの別の単一行を、タイプ2の二値化プロセスを経由して処理する。タイプ1およびタイプ2の二値化が重複する場合、プロセスはステップ658で、文字数が少ない方の二値化のタイプの文字列を除去する。ステップ658のプロセスは、2つの重複するテキスト行の文字数が等しい場合に限り、図6A内のステップ604からの二値化のタイプの各々に対する文字数カウントを使用する。ステップ658のプロセスは、判断ステップ660に進んで、タイプ1の二値化の文字列またはタイプ2の二値化の文字列が除去されたかどうかを判断する。タイプ1の二値化が除去された場合、処理文字列が判断ステップ606に送られて、タイプ1の二値化を通る処理ループが終了されているかを判断する。タイプ2の二値化が除去された場合、処理文字列が判断ステップ612に送られて、タイプ2の二値化を通る処理ループが終了されているかを判断する。タイプ1の文字列が全てループを通過すると、テキスト取消しが図6Aおよび図6B内で完了し、プロセスは図7Aのステップ702でのプロセス700に進む。前述のように、図4のプロセスはステップ414で、各二値閾値タイプに対してテキストのグループ化プロセスを実行し、この処理は、部分7Aおよび7Bを有し、以下で説明する図7で実行される。1つの中国語テキスト文字が単一の英単語に翻訳され得る。2つの中国語テキスト文字または3つの中国語テキスト文字は、それらが幅広く離れている場合、翻訳のためにグループ化される必要があり得る。図7に示すプロセス700は、各複数行二値化テキストビデオ画像フレーム内で一度に1つの単一テキスト行についてテキストのグループ化を判断する。

【0103】

図7Aのプロセス700が、タイプ1の二値化画像のテキストグループ化を開始するために、ステップ702から始まる。ステップ702での処理が、ステップ704に進み、そこで、プロセスがタイプ1におけるテキストの行数をカウントする。ステップ704のプロセスが、判断ステップ706に進んで、処理文字列がタイプ1のテキストのグループ化を終了しているかどうかを判断する。タイプ1のテキストのグループ化がステップ706で終了している場合、処理文字列がタイプ2を実行するために送られる。ステップ706のプロセスが、タイプ1のテキストのグループ化を終了していない場合、処理はステップ708を継続する。

【0104】

ステップ708で、プロセスは、複数行テキスト画像フレームの単一のタイプ1のテキスト行の垂直範囲に対する上限および下限を計算する。プロセスは判断ステップ710に進んで、タイプ1テキストのテキストグループ化が実行されるべきかどうかを判断する。ステップ710のプロセスが、タイプ1テキストのテキストグループ化が可能でないと判断する場合、ステップ710のプロセスは処理文字列をステップ706に戻して、タイプ1のテキストグループ化を通るループが終了しているかどうかを判断する。ステップ710のプロセスが、タイプ1テキストのテキストグループ化が可能であると判断する場合、プロセスはステップ712に進み、そこで、プロセスは、単一のテキスト行内のタイプ1のテキスト文字に対して位置情報の全てを計算する。プロセスはまず、タイプ1テキスト行の水平位置の中心点 X_i を判断し、次いで、タイプ1テキスト行の2つの左端の文字の左間隔および2つの右端の文字の右間隔を判断する。図7Aでは、左間隔は、ステップ712の右に示す、テキスト文字716とテキスト文字718との間の距離であることに留意されたい。右間隔は、ステップ712の右に示す、テキスト文字720とテキスト文字722との間の距離であることに留意されたい。712のプロセスは、判断ステップ71

4に進み、そこで、プロセスは、残りの行 j までの文字列プロセスループが終了されているかを判断する。

【0105】

ステップ714の判断プロセスが、残りの行 j までの文字列プロセスループが終了されていると判断した場合、プロセスはステップ752に進み、そこで、処理は、行「 i 」の中心点 X_i の右側のテキストをグループ化する場合、行 i の位置情報を修正する。次いで、プロセスは、ステップ752からステップ756に進み、そこで、処理は、中心点 X_i の左側のテキストをグループ化する場合、行 i の位置情報を修正する。プロセスは次いで、ステップ760に進んで、グループ化されているテキスト行（複数可）を除去し、次いで、プロセスは判断ステップ762に進み、そこで、ステップ762の処理は、現在の行 i のテキストのグループ化が生じたかどうかを判断する。行 i のテキストのグループ化が生じなかった場合、行 i のテキストのグループ化が、偽の可能なテキストグループ化としてメモリ内に設定される。行 i のテキストのグループ化が生じた場合には、タイプ1のテキストグループ化の処理文字列がステップ710に戻り、可能なテキストグループ化がまだ真に設定されているので、ステップ712に進んで、テキスト行（複数可）を現在の行 i にグループ化するさらなる可能性を見つけるであろう。（本発明人は、シンボル「 i 」および「 j 」を、我々が焦点を合わせているテキスト行を指し示すために使用する。プロセスは、行「 j 」を行「 i 」にグループ化する可能性を検査し、そのため、行「 i 」から始めて、残りの行（「 j 」）までループし、その後、次の行「 i 」に進んで、残りの行（「 j 」）までループする、などと続く）。

【0106】

ステップ714の判断ステップが、残りの行 j までの文字列プロセスループが終了されていないと判断した場合、プロセスはステップ754に進んで、行 j の中心点 X_j を計算する。（判断ステップは、行 j の中心点 Y_j が行 i の上限と下限との間にあるかどうかを判断する。そうである場合には754に進む。そうでない場合は、714に戻る）。プロセスは次いで、判断ステップ758に進み、そこで、プロセスは、行 j の中心点 X_j が、行 i の中心点 X_i 未満であるかどうかを判断する。判断ステップ758が、行 j の中心点 X_j が、行 i の中心点 X_i 未満であると判断した場合、プロセスはステップ766に進んで、行 i の左境界と行 j の右境界との間の左の距離を計算する。

【0107】

判断ステップ758が、行 j の中心点 X_j が行 i の中心点 X_i 未満ではないと判断した場合、プロセスはステップ764に進んで、行 i の右境界と行 j の左境界との間の右の距離を計算する。

【0108】

ステップ766のプロセスが、行 i の左境界と行 j の右境界との間の左の距離を計算した後、次いでプロセスは、判断ステップ768に進み、そこで、プロセスは、（1）行 i が単一の中国語文字であるか、または左の距離が行 i の左間隔の2倍未満であるかのいずれか；および（2）行 i の左境界と行 j の右境界との間の左の距離が最小の格納値未満である、かどうかを判断する。

【0109】

ステップ768のプロセスが、前述の（1）または（2）のいずれかの条件が維持されていないと判断すると、処理文字列はステップ714に戻って、残りの行 j までの処理文字列ループが終了しているかどうかに関する別の判断を実行する。

【0110】

ステップ768のプロセスが、前述の（1）および（2）の条件が両方維持されていると判断すると、プロセスはステップ776に進んで、行 i の左境界と行 j の右境界との間の左の距離の最小の格納値を更新する。次いで、ステップ776からの処理文字列が、ステップ714に戻って、残りの行 j までの処理文字列ループが終了しているかどうかに関する別の判断を実行する。

【0111】

判断ステップ758で X_j が X_i 未満でないとは判断される場合、プロセスはステップ764に進み、そこで、ステップ764のプロセスが、行 i の右境界と行 j の左境界との間の右の距離を計算し、次いで、プロセスは判断ステップ770に進み、そこで、プロセスは、(3)行 i が単一の中国語文字であるか、または右の距離が行 i の右間隔の2倍未満であるかのいずれか；および(4)行 i の右境界と行 j の左境界との間の右の距離が最小の格納値未満である、かどうかを判断する。

【0112】

770のプロセスが、前述の(3)または(4)のいずれかの条件が維持されていないと判断すると、処理文字列はステップ714に戻って、残りの行 j までの処理文字列ループが終了しているかどうかに関する別の判断を実行する。

【0113】

ステップ770のプロセスが、前述の(3)および(4)の条件が両方維持されていると判断すると、プロセスはステップ774に進んで、行 i の右境界と行 j の左境界との間の右の距離の最小の格納値を更新する。次いで、ステップ774からの処理文字列が、ステップ714に戻って、残りの行 j までの処理文字列ループが終了しているかどうかに関する別の判断を実行する。

【0114】

図7Aの判断ステップ706で、タイプ1のテキストグループ化が終了していると判断されている場合、タイプ2のテキストグループ化が、図7Aおよび図7Bのプロセス700の同じ種類のプロセスステップに従って実行される。しかし、タイプ2のテキストグループ化のためのプロセスステップは、図7Aおよび図7Bを考慮して、本発明が、タイプ2のテキストグループ化を実行する方法に関係するものがどのようなものであり得るかは当業者には明らかであるので、図には具体的には示されない。

【0115】

ユーザーインタフェース実施形態および使用事例の詳細な説明

図8は、中華料理メニューの一例を示す。ラベル802で囲まれているのは、英語への翻訳を必要とする4行の中国語文字である。ラベル804で囲まれているのは、英語への翻訳を必要とする単一行の中国語文字である。

【0116】

図9は、本発明の一実施形態の動作プロセス中である、モバイルカメラ装置の一例上のユーザーインタフェース900を示す。ターゲットボックス902の一実施形態が、ユーザーインタフェースディスプレイ上に示されている。ターゲットボックスは、任意のサイズにでき、モバイルカメラ装置上のどこにでも配置できる。指先906でタッチされているとして描かれている、ターゲットボックスのプルダウンスプレイアイコン904は、矢印の方向にスライドできて、ターゲットボックスのサイズを拡大させる。「ライト」とラベルの付けられた光照射スイッチアイコン908は、カメラが、第2の言語への翻訳のために選択された第1の言語テキストに焦点を合わせている間に、光照射を加えるためにタップできる。第1の言語テキストは、ターゲットボックス内に表示されるであろう。「固定する(freeze)」とラベルの付けられた一時停止ボタン910は、翻訳の表示を一時停止または固定するためにタップできる。プルダウンアイコン904上を指先でスライドすることによって生じるターゲットボックスのサイズへの効果例については図10を参照する。

【0117】

図10は、本発明の一実施形態の動作プロセス中である、モバイルカメラ装置の一例上のユーザーインタフェース1000を示す。ここでは、1002とラベル付けされたターゲットボックスの実施形態が、ユーザーインタフェースディスプレイ上に示されている。図9のターゲットボックス902と比較すると、図10のターゲットボックスのサイズの方が大きい。ターゲットボックスの上に、「テキストをボックスの中心に置いて、翻訳が現れるまで待つ」という指示がある。スライダーアイコン1004は、ボタンを有し、ターゲットボックス内の画像を必要に応じて拡大するためにスライドできる。

【 0 1 1 8 】

図 1 1 は、本発明の一実施形態の動作プロセス中である、モバイルカメラ装置の一例上のユーザインタフェース 1 1 0 0 を示す。ターゲットボックスのサイズが、本発明による翻訳のために焦点を合わされた中国語テキスト文字 1 1 0 4 の単一行に適合する。これらは、図 8 に示す中国語メニュー上でラベル 8 0 4 の内部に示されたのと同じ中国語文字である。ターゲットボックスの上は、中国語文字の単一行の処理された画像 1 1 0 2 である。ターゲットボックスの下は、英語翻訳 1 1 0 6 である。プルダウンタブアイコンは、1 1 0 8 とラベル付けされている。一時停止アイコンは 1 1 1 0 とラベル付けされている。光照射アイコン 1 1 1 2 は、ライトが「オフ」であることを示している。ユーザインタフェースの第 2 のビューは、ターゲットボックスのプルダウンアイコン 1 1 1 4 上の指を示す。

10

【 0 1 1 9 】

図 1 2 は、本発明の一実施形態の動作プロセス中である、モバイルカメラ装置の一例上のユーザインタフェース 1 2 0 0 を示す。ターゲットボックスのサイズは、ターゲットボックスの内部に非常に薄いグレーで示された、焦点を合わされている中国語テキスト文字の 4 行に適合する。英語翻訳 1 2 0 2 は、本発明のこの実施形態でターゲットボックスの内部に、中国語文字の各行およびその英語翻訳を重ね合わせて表示されている。一実施形態では、英語翻訳は、中国語文字の上である。一実施形態では、英語翻訳は、リアルタイムで拡張現実として表示される。

【 0 1 2 0 】

20

図 1 3 は、ユーザインタフェース 1 3 0 0 の拡大図を示しており、図 1 2 の拡大した部分を示す。本発明のこの実施形態例では、ターゲットボックスは 1 3 0 6 とラベル付けされ、プルダウンアイコンタブは 1 3 1 2 とラベル付けされている。第 1 の言語テキストはここでは中国語テキスト文字である。それらの第 2 の言語への翻訳が英語テキスト文字として表示されている。複数行翻訳の 4 行の各々が英語テキストであり、各行がここでは、例えば、行 1 3 0 2、1 3 0 4、1 3 0 8、および 1 3 1 0 と番号付けされている。一実施形態では、英語テキストは、黒のフォントであり、他方、中国語テキストの 4 行は、本発明の一実施形態に従い、グレー（元の色）のフォントであるが、フォントの色として任意の色が使用され得る。他の実施形態では、英語テキストは、白のフォント色である。外国語および翻訳されたテキストの色は、本発明の精神または範囲から逸脱することなく変更され得る。

30

【 0 1 2 1 】

図 1 4 は、本発明の別の実施形態による単一行翻訳動作を表示する、ユーザインタフェース 1 4 0 0 の拡大図を示す。中国語テキスト文字 1 4 0 4 の中国語発音 1 4 0 6 がターゲットボックスの下に表示されている。ターゲットボックスのサイズは、本発明による翻訳のために焦点を合わされた中国語テキスト文字 1 4 0 4 の単一行に適合する。ターゲットボックスの上は、中国語文字の単一行処理された画像 1 4 0 2 である。ターゲットボックスの下は、英語翻訳 1 4 0 8 である。

【 0 1 2 2 】

図 1 5 は、本発明がその上で実施され得る、様々なモバイル機器 1 5 0 2、1 5 0 4、1 5 0 6、および 1 5 0 8 を示す。示されているのは、モバイルスマートフォン 1 5 0 2、1 5 0 4、および 1 5 0 6、ならびに G O O G L E G L A S S（商標）などであるが、それに制限されない、ウェアラブルコンピュータ 1 5 0 8 である。本発明は、様々なモバイルおよびウェアラブル機器上で実施され得、そのいくつかの実例が本明細書で提供されている。しかし、本発明の適用可能性は、本明細書に示すか、または説明する、モバイル機器またはウェアラブルコンピュータに決して制限されない。かかるモバイル機器およびウェアラブルコンピュータは、1 つ以上のメモリに結合された、1 つ以上のプロセッサを有し、メモリは、プログラムコードを格納し、図に示して説明するように、本発明のプロセスを実行するためのプログラムコードを格納するために使用され得ることが知られている。

40

50

【 0 1 2 3 】

日本語の実施形態

日本語には3つの異なる書記体系（平仮名、片仮名、および漢字）があるという事実のために、（ヘ - ヘ）、（ロ - ロ）など、これらの体系にわたる2、3の文字が文字レベルで区別するのが困難であり得る。従って、プロセスは、それらを区別するためにコンテキスト情報を使用する。プロセスは、より高い精度を達成するために、言語モデルおよびいくつかの発見的規則を利用する。プロセスは、最も可能性の高い文字列を評価するために、文字の形状類似性情報を翻訳スコアと一緒に組み合わせることもできる。

【 0 1 2 4 】

結論

本発明は、ハードウェアおよび/またはソフトウェアで実装され得る。例えば、ネットワークインタフェースなどの、システムの多くの構成要素は、本発明を曖昧にしないために、示されていない。しかし、当業者は、システムは、必ずこれらの構成要素を含むことを理解するであろう。ユーザー装置は、メモリに結合された少なくとも1つのプロセッサを含むハードウェアである。プロセッサは、1つ以上のプロセッサ（例えば、マイクロプロセッサ）を表し得、メモリは、ハードウェアの主記憶装置、ならびに任意の補足レベルのメモリ、例えば、キャッシュメモリ、不揮発性またはバックアップメモリ（例えば、プログラム可能またはフラッシュメモリ）、読取り専用メモリなど、を含むランダムアクセスメモリ（RAM）装置を表し得る。加えて、メモリは、ハードウェア内のどこにでも、例えば、プロセッサ内の任意のキャッシュメモリに、物理的に配置されたメモリ記憶、ならびに、例えば、大容量記憶装置上に格納されるような、仮想メモリとして使用される任意の記憶容量、を含むと考えられ得る。

【 0 1 2 5 】

ユーザー装置のハードウェアは、通常、外部と情報をやりとりするために、いくつかの入力および出力も受信する。ユーザーとのインタフェースのために、ハードウェアは1つ以上のユーザー入力装置（例えば、キーボード、マウス、スキャナ、マイクロホン、ウェブカメラなど）およびディスプレイ（例えば、液晶ディスプレイ（LCD）パネル）を含み得る。追加の記憶装置として、ハードウェアは、1つ以上の大容量記憶装置、例えば、とりわけ、フロッピーもしくは他の取外し可能ディスクドライブ、ハードディスクドライブ、直接アクセス記憶装置（DASD）、光学式ドライブ（例えば、コンパクトディスク（CD）ドライブ、デジタル多用途ディスク（DVD）ドライブなど）および/またはテープドライブも含み得る。さらに、ハードウェアは、ネットワークに結合された他のコンピュータとの情報の伝達を可能にするために、1つ以上のネットワーク（例えば、とりわけ、ローカルエリアネットワーク（LAN）、ワイドエリアネットワーク（WAN）、無線ネットワーク、および/またはインターネット）とのインタフェースを含み得る。ハードウェアは通常、プロセッサ間の適切なアナログおよび/またはデジタルインタフェースを含むことが理解されるべきである。

【 0 1 2 6 】

ハードウェアは、オペレーティングシステムの制御下で動作して、前述のプロセス技術を実行するための、参照番号によって全体として示される、様々なコンピュータソフトウェアアプリケーション、構成要素、プログラム、コード、ライブラリ、オブジェクト、モジュールなどを実行する。

【 0 1 2 7 】

一般に、本発明の実施形態を実装するために実行される方法は、オペレーティングシステムの一部、または「コンピュータプログラム（複数可）」もしくは「コンピュータコード（複数可）」と呼ばれる特定のアプリケーション、構成要素、プログラム、オブジェクト、モジュールもしくは命令の連続として実装され得る。コンピュータプログラムは、通常、1つ以上の命令セットをコンピュータ内の様々なメモリおよび記憶装置内に様々な時に含み、それは、コンピュータ内の1つ以上のプロセッサによって読み取られて実行される場合に、コンピュータに、本発明の様々な態様を伴う要素を実行するために必要な動作

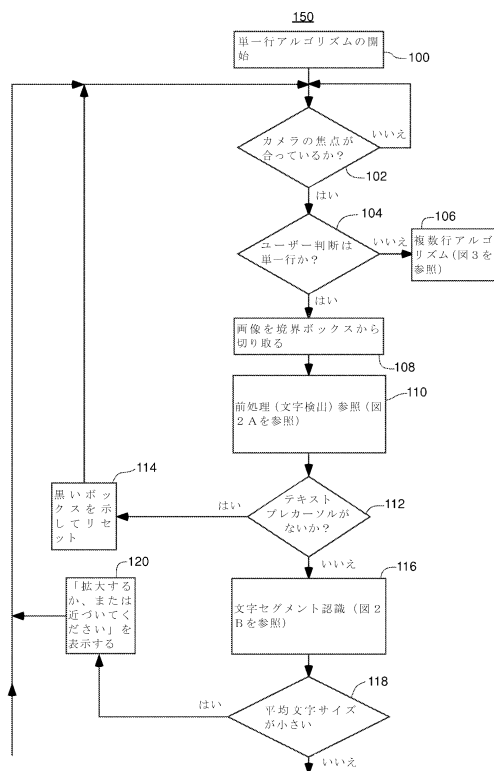
を実行させる。その上、本発明は、完全に機能するコンピュータおよびコンピュータシステムのコンテキストで説明されているが、当業者は、本発明の様々な実施形態は、プログラム製品として様々な形式で分散されることが可能であり、本発明は、実際に分散を達成するために使用されるマシンまたはコンピュータ可読媒体の特定のタイプに関わらず、均等に適用されることを理解するであろう。コンピュータ可読媒体の例は、揮発性および不揮発性のメモリ装置、フロッピーおよび他の取外し可能ディスク、ハードディスクドライブ、光ディスク（例えば、コンパクトディスク読取り専用メモリ（CD ROM）、デジタル多用途ディスク（DVD）など）などの記憶可能タイプ媒体、および、時々、モバイル機器に対する「アプリストア」として知られる、オンラインストアを通じた無線媒体経由を含む、デジタルおよびアナログ通信媒体を含むがそれらに制限されない。

10

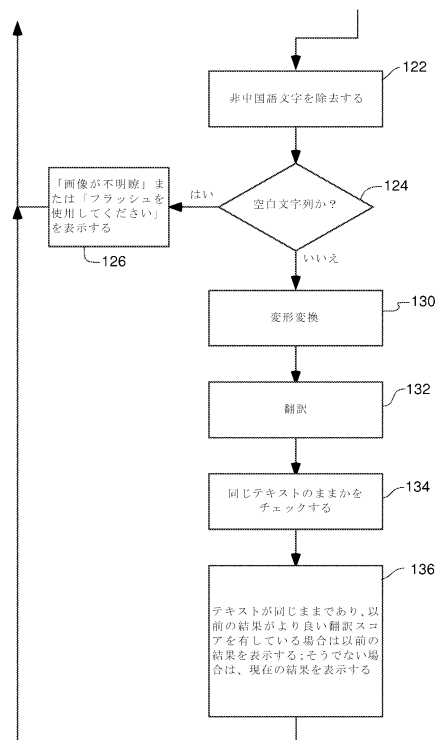
【0128】

本発明は特定の例示的な実施形態を参照して説明されているが、様々な実施形態および変更がこれらの実施形態に対して、本発明のより広範な精神から逸脱することなく、行われ得ることが明らかであろう。それに応じて、明細書および図は、制限的な意味ではなく、例示的な意味で考えられるべきである。前述の実施形態は、教示される個別の記述のいずれよりも大きい範囲を有し得る単一のより幅広い発明の特定例であることも当業者には明らかであろう。本発明の精神および範囲から逸脱することなく、多くの変更が記述において行われ得る。

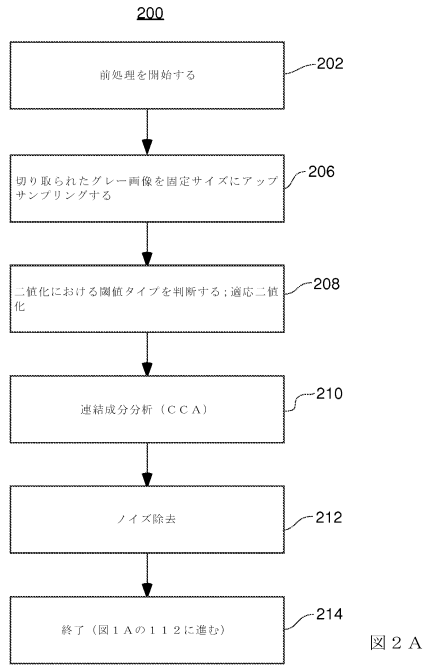
【図1A】



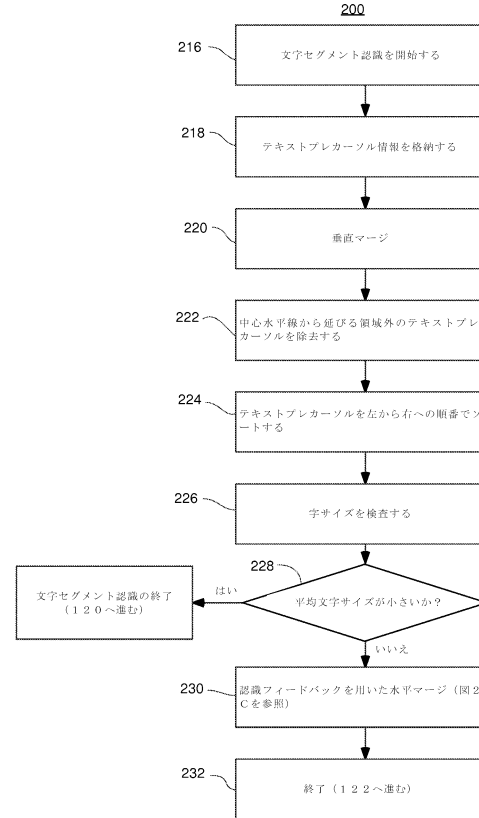
【図1B】



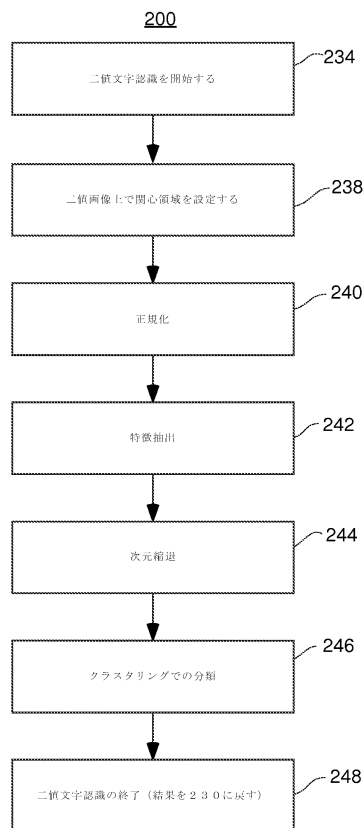
【図 2 A】



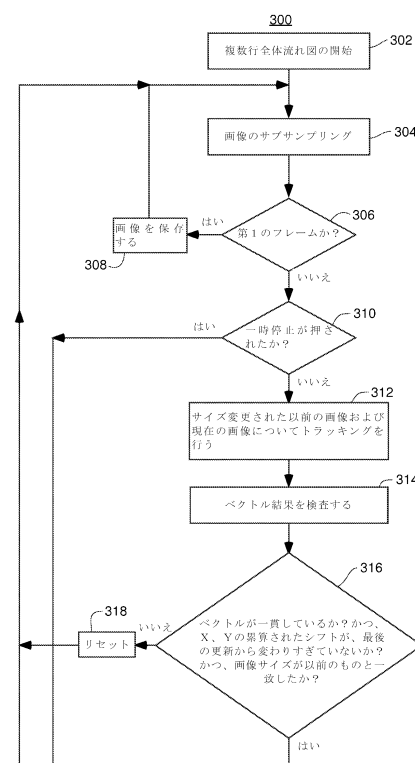
【図 2 B】



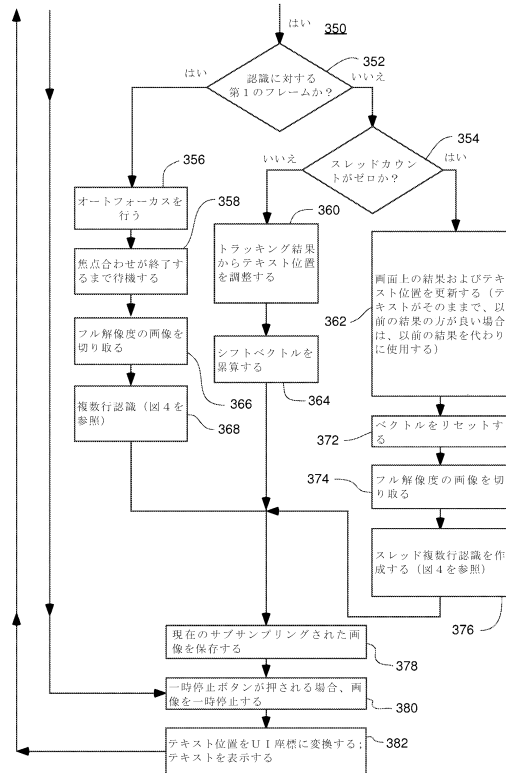
【図 2 C】



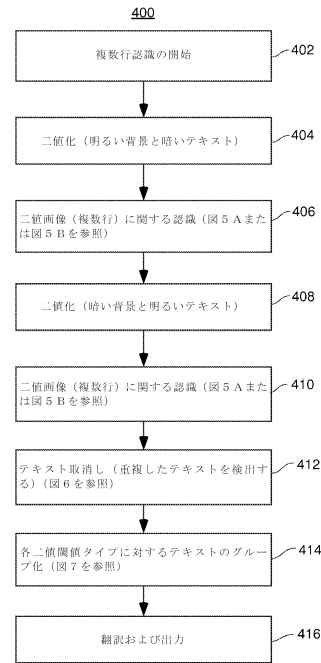
【図 3 A】



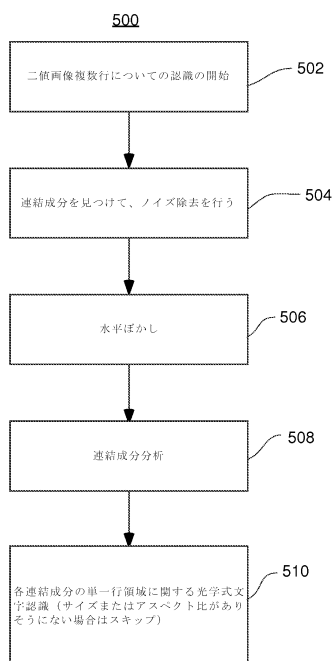
【図 3 B】



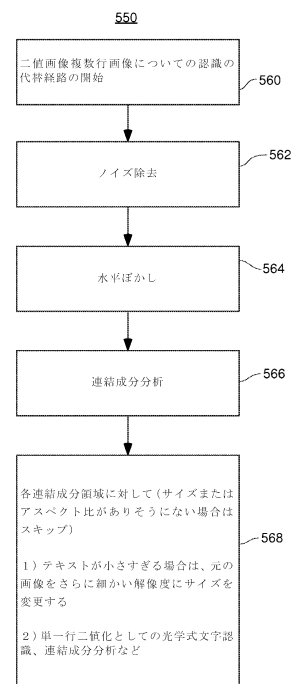
【図 4】



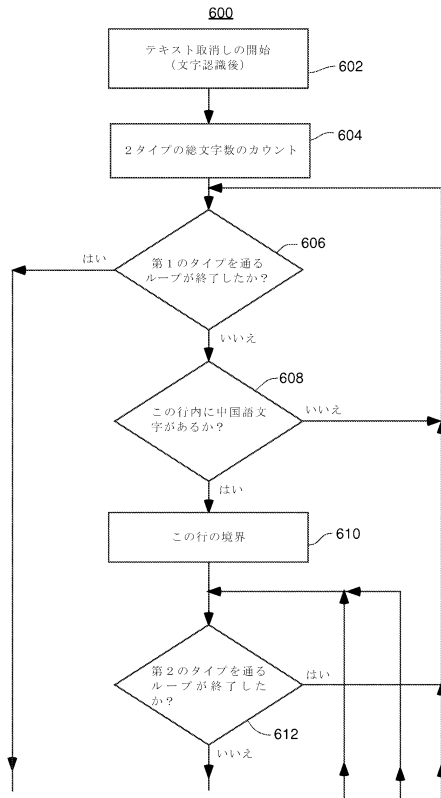
【図 5 A】



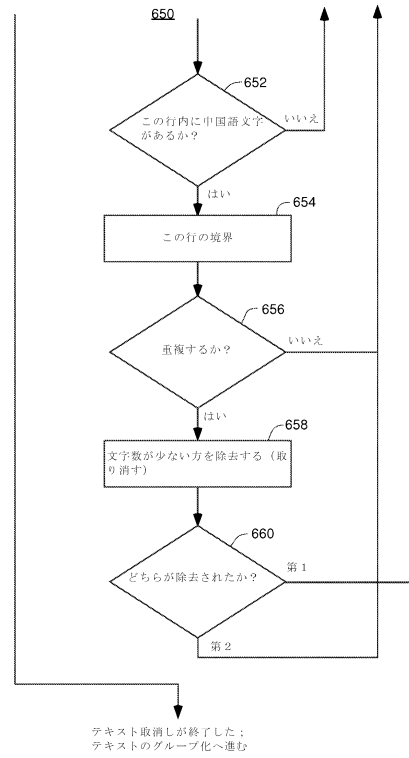
【図 5 B】



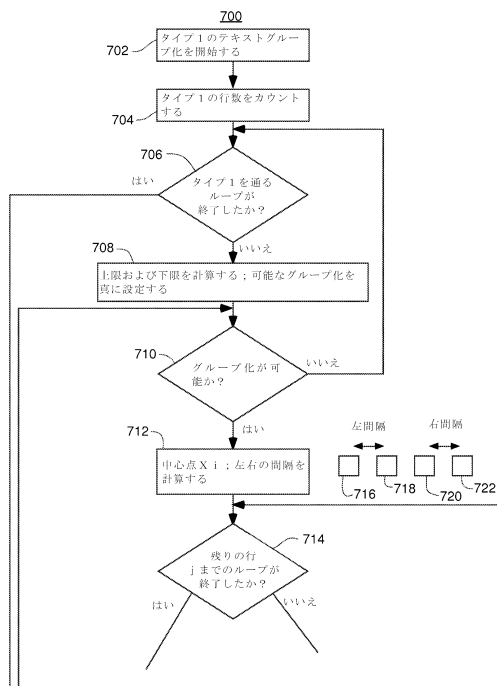
【図 6 A】



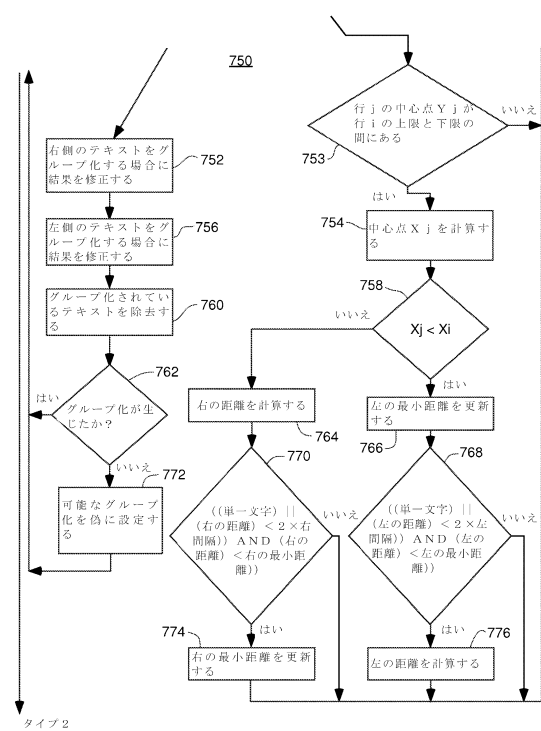
【図 6 B】



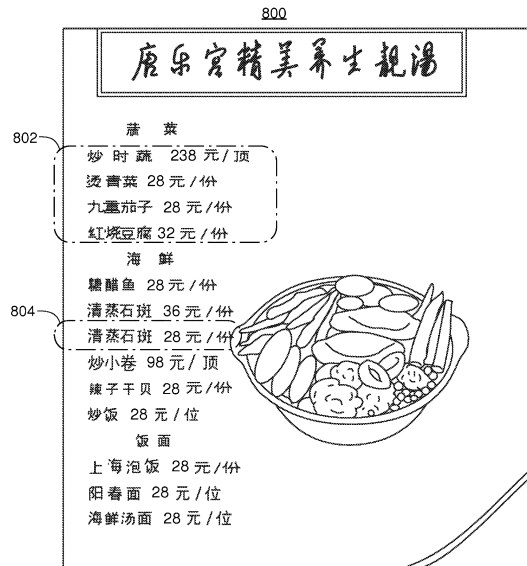
【図 7 A】



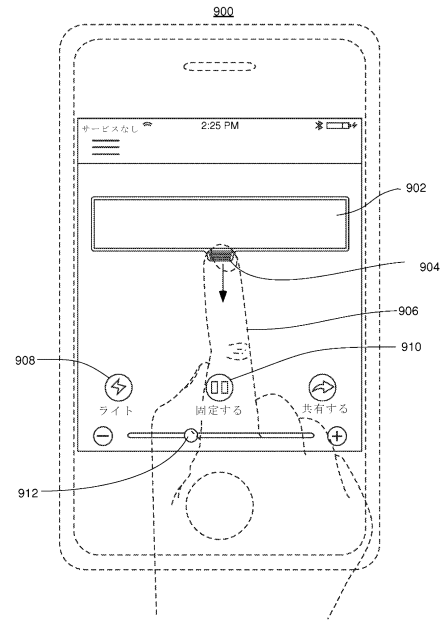
【図 7 B】



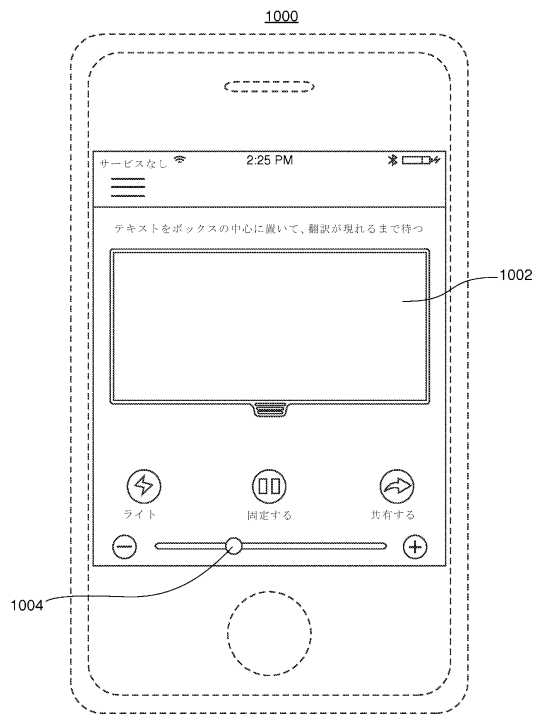
【図 8】



【図 9】



【図 10】



【図 11】

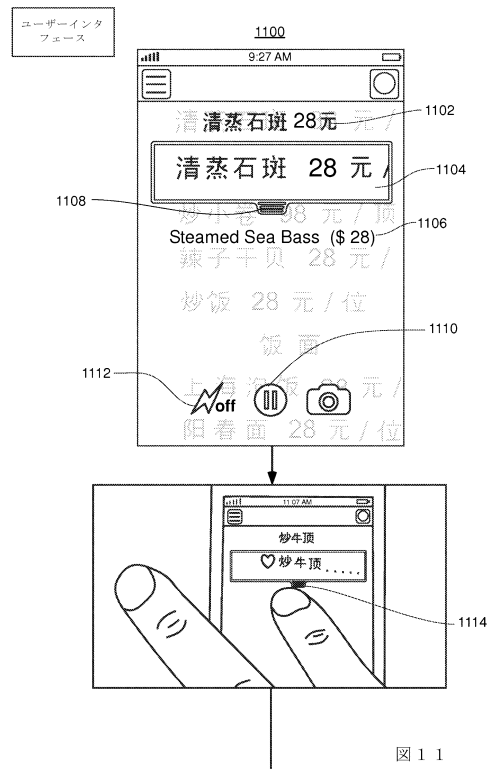
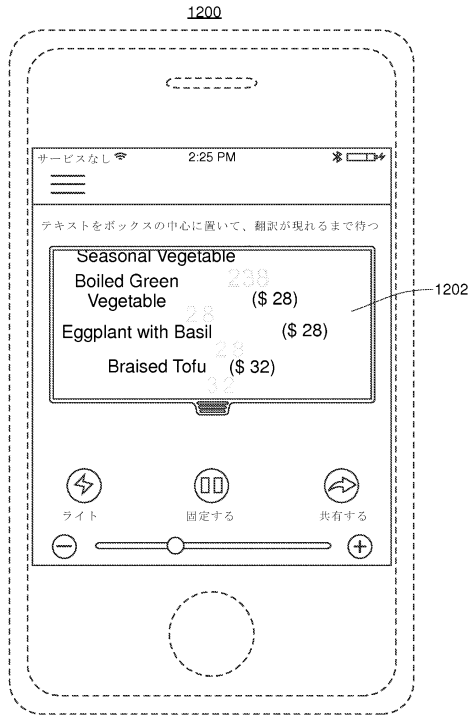
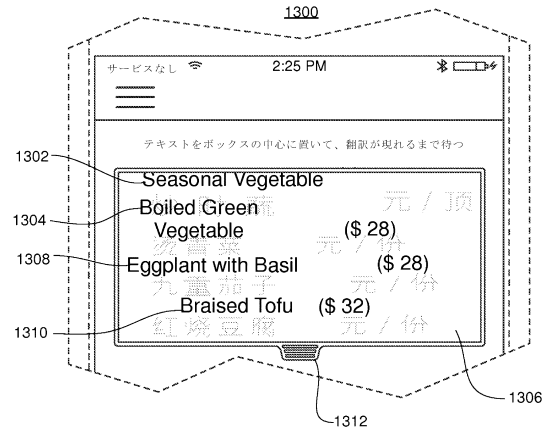


図 11

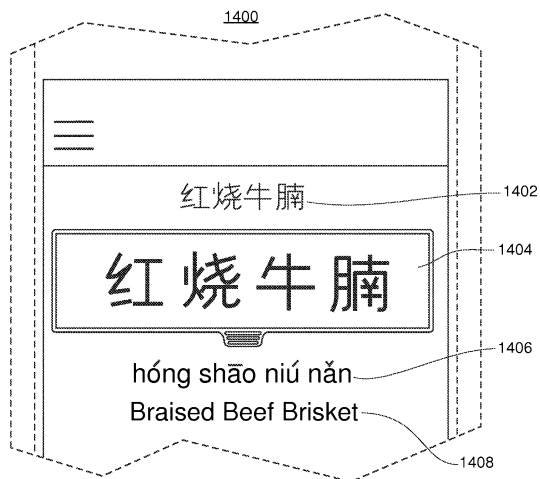
【図 12】



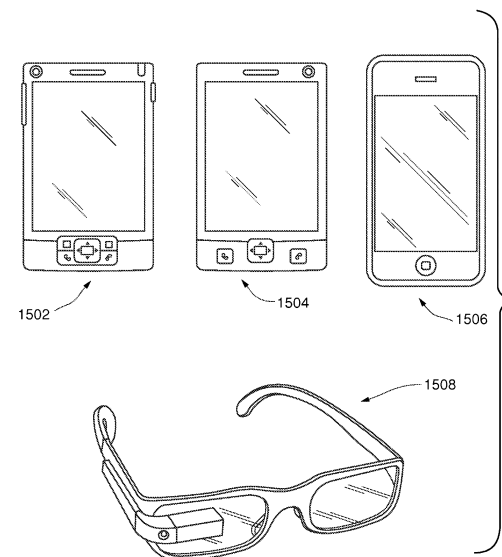
【図 13】



【図 14】



【図 15】



フロントページの続き

- (72)発明者 ロゴスキー, ライアン, リオン
アメリカ合衆国, イリノイ州 60564, ネイパービル, 2816 フェアホーサー コート
- (72)発明者 ウー, ファン ユ
台湾, 11569, タイペイ, 南港区, ドンシン ストリート, レーン 180, 6階 9番
- (72)発明者 クラーク, ケビン, アンソニー
アメリカ合衆国, ロードアイランド州 02886, ワーリック, 410 ボタンウッズ アベニ
ュー

審査官 松野 広一

- (56)参考文献 国際公開第2013/003242(WO, A1)
特開2006-302091(JP, A)
特開2001-56446(JP, A)
特開2000-23012(JP, A)
特開2006-146454(JP, A)
特開2011-134144(JP, A)
特開2003-323693(JP, A)
特開2000-207487(JP, A)
特開2006-85461(JP, A)
米国特許出願公開第2011/0090253(US, A1)
米国特許出願公開第2010/0331043(US, A1)
米国特許出願公開第2012/0330643(US, A1)
戸津 弘貴, 独断ゴメン、今月の一押しはこれだ! iPad/iPhoneアプリ勝手にラン
キング, Mac Fan, 日本, 株式会社毎日コミュニケーションズ, 2011年 3月 1日
, 第19巻 第3号, p.194

(58)調査した分野(Int.Cl., DB名)

G06K 9/62
G06F 17/28
G06K 9/22