

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6338595号  
(P6338595)

(45) 発行日 平成30年6月6日(2018.6.6)

(24) 登録日 平成30年5月18日(2018.5.18)

(51) Int.Cl.

F I

G06T 7/33 (2017.01)  
G06K 9/62 (2006.01)G06T 7/33  
G06K 9/62 620Z

請求項の数 15 (全 22 頁)

(21) 出願番号 特願2015-551677 (P2015-551677)  
 (86) (22) 出願日 平成25年11月22日 (2013.11.22)  
 (65) 公表番号 特表2016-502218 (P2016-502218A)  
 (43) 公表日 平成28年1月21日 (2016.1.21)  
 (86) 国際出願番号 PCT/US2013/071518  
 (87) 国際公開番号 W02014/107246  
 (87) 国際公開日 平成26年7月10日 (2014.7.10)  
 審査請求日 平成28年11月4日 (2016.11.4)  
 (31) 優先権主張番号 61/749,248  
 (32) 優先日 平成25年1月4日 (2013.1.4)  
 (33) 優先権主張国 米国 (US)  
 (31) 優先権主張番号 14/021,337  
 (32) 優先日 平成25年9月9日 (2013.9.9)  
 (33) 優先権主張国 米国 (US)

(73) 特許権者 507364838  
 クアルコム、インコーポレイテッド  
 アメリカ合衆国 カリフォルニア 921  
 21 サン ディエゴ モアハウス ドラ  
 イブ 5775  
 (74) 代理人 100108453  
 弁理士 村山 靖彦  
 (74) 代理人 100163522  
 弁理士 黒田 晋平  
 (72) 発明者 マイケル・ゲルヴォーツ  
 アメリカ合衆国・カリフォルニア・921  
 21-1714・サン・ディエゴ・モアハ  
 ウス・ドライブ・5775

最終頁に続く

(54) 【発明の名称】 モバイルデバイスベースのテキスト検出および追跡

(57) 【特許請求の範囲】

【請求項 1】

移動局 (MS) 上のテキスト認識および追跡のための方法であって、

第1のテキストブロックを位置特定して認識するために、前記MS上のカメラによってキャプチャされた画像フレーム上で光学式文字認識 (OCR) を実行することによって第1の参照フレームを取得するステップと、

選択された後続の画像フレームと関連付けられたパラメータに基づいて、後続の画像フレームのセットから後続の画像フレームを選択するステップと、

第2のテキストブロックを認識するために、前記選択された後続の画像フレーム上でOCRを実行することによって第2の参照フレームを取得するステップと、

前記第1のテキストブロックと前記第2のテキストブロックとの間の幾何学的関係を判断するステップであって、前記幾何学的関係は、前記第2の参照フレーム内の前記第1のテキストブロックの位置と、前記第2の参照フレームと関連付けられたカメラ姿勢とに少なくとも部分的に基づいて判断される、ステップとを含む、方法。

【請求項 2】

前記第1のテキストブロックと前記第2のテキストブロックとの間の前記幾何学的関係に基づいて前記第1のテキストブロックと前記第2のテキストブロックとを連続的にアセンブルするステップをさらに含む、請求項1に記載の方法。

【請求項 3】

10

20

前記第1のテキストブロックと前記第2のテキストブロックとの間の前記幾何学的関係が、前記第1のテキストブロックおよび前記第2のテキストブロックが現れる媒体と関連付けられた基準系に少なくとも部分的に基づく、請求項2に記載の方法。

【請求項4】

前記アセンブルされた一連の第1および第2のテキストブロックを入力としてテキスト音声アプリケーションに与えるステップをさらに含む、請求項2に記載の方法。

【請求項5】

前記後続の画像フレームを選択するステップが、

後続の画像フレームの前記セットに対するカメラ姿勢を計算するステップであって、各カメラ姿勢が別個の後続の画像フレームと関連付けられ、前記関連付けられた後続の画像フレームを前記第1の参照フレームと整合させることに少なくとも部分的に基づいて判断される、計算するステップと、

後続の画像フレームの前記セット内の対応する画像フレームと関連付けられたパラメータを、前記計算されたカメラ姿勢に少なくとも部分的に基づいて判断するステップとをさらに含む、請求項1に記載の方法。

【請求項6】

前記整合させることが、効率的2次最小化(ESM)を使用して実行され、かつ、

前記ESMが、前記関連付けられた後続の画像フレームのより低い分解能バージョン上で動作する、請求項5に記載の方法。

【請求項7】

後続の画像フレームの前記セットに対するカメラ姿勢を計算するステップが、

前記第1の参照フレーム内の複数の特徴点を識別することによって取得された画像パッチを含む追跡ターゲットを生成するステップと、

前記第1の参照フレームと前記後続の画像フレームとの間の画像パッチの対応に基づいて、前記セット内の後続の画像フレーム内で前記追跡ターゲットのロケーションを判断するステップと、

前記後続の画像フレーム内の前記追跡ターゲットの前記ロケーションに少なくとも部分的に基づいて前記後続の画像フレームと関連付けられたカメラ姿勢を計算するステップとをさらに含む、請求項5に記載の方法。

【請求項8】

前記特徴点が、前記第1の参照フレーム内の自然特徴に基づくか、または、

個別の特徴点が重みを割り当てられ、かつ、前記第1のテキストブロックの上の特徴点が、前記第1の参照フレーム内の他の場所に位置する特徴点に対してより大きい重みを割り当てられる、請求項7に記載の方法。

【請求項9】

前記追跡ターゲットの生成が、前記関連付けられた後続の画像フレームを前記第1の参照フレームと整合させることと実質的に並列して実行される、請求項7に記載の方法。

【請求項10】

前記第1の参照フレームおよび後続の画像フレームの前記セットがマーカーレスである、請求項1に記載の方法。

【請求項11】

前記パラメータが、

前記選択された後続の画像フレームと前記第1の参照フレームとの間のオーバーラップ領域のパーセンテージが、

前記選択された後続の画像フレーム内で可視の前記第1のテキストブロックの割合であって、前記第1のテキストブロックの全領域に対する前記選択された後続の画像フレーム内の前記第1のテキストブロックの可視の部分を含む領域の比として決定される、割合か、

前記第1の参照フレームに対する前記選択された後続の画像フレームの回転の大きさが、または

10

20

30

40

50

前記第1の参照フレームに対する前記選択された後続の画像フレームの並進の大きさのうちの少なくとも1つを含む、請求項1に記載の方法。

【請求項 1 2】

前記カメラ姿勢が6自由度(6-DoF)で判断され、前記カメラが前記第1のテキストブロックおよび前記第2のテキストブロックを含む平坦な媒体に正対する位置に配置される、請求項1に記載の方法。

【請求項 1 3】

前記方法が、拡張現実(AR)アプリケーションによって起動され、  
仮想オブジェクトが、前記第1のテキストブロックおよび前記第2のテキストブロックの上に前記ARアプリケーションによって配置され、かつ、

10

前記仮想オブジェクトが、前記第1のテキストブロックおよび前記第2のテキストブロックから翻訳されたテキストを含み、前記翻訳されたテキストが、前記第1のテキストブロックおよび前記第2のテキストブロックを表現するために使用された言語と異なる言語におけるものである、請求項1に記載の方法。

【請求項 1 4】

一連の画像フレームをキャプチャするための撮像手段と、  
第1のテキストブロックを位置特定して認識するために、前記一連の画像フレーム内の画像フレーム上で光学式文字認識(OCR)を実行することによって第1の参照フレームを取得するための手段と、

20

選択された後続の画像フレームと関連付けられたパラメータに基づいて、前記一連の画像フレームから後続の画像フレームを選択するための手段と、

第2のテキストブロックを認識するために、前記選択された後続の画像フレーム上でOCRを実行することによって第2の参照フレームを取得するための手段と、

前記第1のテキストブロックと前記第2のテキストブロックとの間の幾何学的関係を判断するための手段であって、前記幾何学的関係は、前記第2の参照フレーム内の前記第1のテキストブロックの位置と、前記第2の参照フレームと関連付けられた前記撮像手段の姿勢とに少なくとも部分的に基づいて判断される、手段とを含む、装置。

【請求項 1 5】

プロセッサによって実行されると請求項1乃至13のいずれか1項に記載の方法を実行する命令を記憶した非一時的コンピュータ可読記録媒体。

30

【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、2013年9月9日に出願した「Mobile Device Based Text Tracking and Detection」と題する米国特許出願第14/021,337号に対する利益および優先権を主張し、同様に、2013年1月4日に提出した「Mobile Device Based Text Tracking and Detection」と題する米国特許仮出願第61/749,248号に対する利益および優先権を主張するものであり、両出願は参照によりその全体を本明細書に組み入れられる。

40

【0002】

本明細書で開示する主題は、テキスト検出および追跡に関する。

【背景技術】

【0003】

スマートフォンなどのモバイルデバイスは、ますます、カメラなどの1つまたは複数の光学センサを様々な他のタイプのセンサとともに組み入れている。モバイルデバイス上のカメラは、しばしば、静止画像とビデオ画像の両方に対応できる。しかしながら、テキストの走査を伴うアプリケーションに対するモバイルデバイスの使用は、多くの課題のために限定されていた。いくつかの課題の例には、ユーザの手の動きによる被写体ぶれ(motion blur)、走査されるテキストに対するカメラの向きの変動に起因する射影におけるゆが

50

み(distortions in perspective)、およびモバイルデバイスの光学サブシステムの限界が含まれる。たとえば、モバイルデバイスカメラによって撮られたビデオ画像は、スキャナに対して低分解能であり、加えて被写体ぶれおよび射影ゆがみの影響を受けることがある。速やかでシームレスなテキスト検出および追跡を容易にすることは、テキスト検出および追跡アプリケーションの出力を利用することができる広範なアプリケーションを可能にすることによって、モバイルデバイスの有用性を高めることができる。

【発明の概要】

【発明が解決しようとする課題】

【0004】

それゆえ、テキストの走査、検出および/または認識、ならびに/あるいは追跡を容易にするための装置、システム、および方法が必要である。

【課題を解決するための手段】

【0005】

一態様では、移動局(MS)上の方法は、第1のテキストブロックを位置特定して認識するためにMS上のカメラによってキャプチャされた画像フレーム上で光学式文字認識(OCR)を実行することによって第1の参照フレームを取得するステップと、選択された後続の画像フレームと関連付けられたパラメータに基づいて後続の画像フレームのセットから後続の画像フレームを選択するステップと、第2のテキストブロックを認識するために選択された後続の画像フレーム上でOCRを実行することによって第2の参照フレームを取得するステップと、第2の参照フレーム内の第1のテキストブロックの位置および第2の参照フレームと関連付けられたカメラ姿勢に少なくとも部分的に基づいて第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係を判断するステップとを含み得る。

【0006】

別の態様では、MSは、第1の画像フレームおよび後続の画像フレームのセットをキャプチャするように構成されたカメラと、カメラに結合されたプロセッサとを備え得る。プロセッサは、第1のテキストブロックを位置特定して認識するために第1の画像フレーム上で光学式文字認識(OCR)を実行することによって第1の参照フレームを取得することと、選択された後続の画像フレームと関連付けられたパラメータに基づいて後続の画像フレームのセットから後続の画像フレームを選択することと、第2のテキストブロックを認識するために選択された後続の画像フレーム上でOCRを実行することによって第2の参照フレームを取得することとを行うように構成された単語認識モジュールを備え得る。プロセッサは、第2の参照フレーム内の第1のテキストブロックの位置および第2の参照フレームと関連付けられたカメラ姿勢に少なくとも部分的に基づいて、第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係を判断するように構成されたテキストアセンブラモジュールをさらに備え得る。

【0007】

さらなる態様では、装置は、一連の画像フレームをキャプチャするための撮像手段と、第1のテキストブロックを位置特定して認識するために一連の画像フレーム内の画像フレーム上で光学式文字認識(OCR)を実行することによって第1の参照フレームを取得するための手段と、選択された後続の画像フレームと関連付けられたパラメータに基づいて一連の画像フレームから後続の画像フレームを選択するための手段と、第2のテキストブロックを認識するために、選択された後続の画像フレーム上でOCRを実行することによって第2の参照フレームを取得するための手段と、第2の参照フレーム内の第1のテキストブロックの位置および第2の参照フレームと関連付けられた撮像手段の姿勢に少なくとも部分的に基づいて第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係を判断するための手段とを含み得る。

【0008】

別の態様では、非一時的コンピュータ可読媒体は命令を含み得、命令は、プロセッサで実行されると移動局(MS)上の方法を実行し得、方法は、第1のテキストブロックを位置特定して認識するためにMS上のカメラによってキャプチャされた画像フレーム上で光学式文

10

20

30

40

50

字認識(OCR)を実行することによって第1の参照フレームを取得するステップと、選択された後続の画像フレームと関連付けられたパラメータに基づいて後続の画像フレームのセットから後続の画像フレームを選択するステップと、第2のテキストブロックを認識するために選択された後続の画像フレーム上でOCRを実行することによって第2の参照フレームを取得するステップと、第2の参照フレーム内の第1のテキストブロックの位置および第2の参照フレームと関連付けられたカメラ姿勢に少なくとも部分的に基づいて第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係を判断するステップとを含み得る。

【0009】

開示する実施形態はまた、ソフトウェア、ファームウェア、および非一時的コンピュータ可読媒体またはコンピュータ可読メモリを使用してプロセッサによって作成、記憶、アクセス、読み取り、または修正されるプログラム命令に関する。

【0010】

これらおよび他の実施形態について、下記の図面を参照して以下でさらに説明する。例示によって様々な態様が図示および記載された以下の詳細説明から、その他の態様も、当業者に容易に明らかになるであろうことを理解されたい。図面および詳細な説明は、本質的に例示的なものと見なされるべきであって、限定的なものとは見なされるべきではない。

【図面の簡単な説明】

【0011】

【図1】テキスト追跡および検出を実行することが可能な例示的なMSのブロック図である。

【図2】テキストブロックの位置を追跡および/または維持するために使用され得る表面および座標系における、単語に対するカメラの位置を示す図である。

【図3】開示する実施形態に合致したテキスト検出および追跡の例示的な方法のフローチャートである。

【図4】開示する実施形態に合致したテキスト追跡のためのシステムの例示的なアーキテクチャを示す図である。

【図5】開示する実施形態に合致した、移動局上で実行され得る例示的な方法のフローチャートである。

【発明を実施するための形態】

【0012】

本明細書で開示する実施形態について、図面を参照して単に例として説明する。添付の図面とともに以下に示す詳細な説明は、本開示の様々な態様の説明を意図しており、本開示が実行され得る唯一の態様を表すことを意図していない。本開示で説明される各態様は、本開示の単なる例または説明として与えられ、他の態様よりも好適または有利であるとは、必ずしも解釈されるべきではない。詳細な説明は、本開示の完全な理解をもたらす目的で、具体的な詳細を含んでいる。しかしながら、本開示は、これら具体的な詳細なしで実施され得ることが当業者に明らかになるであろう。場合によっては、本開示の概念を曖昧にするのを回避するために、周知の構造およびデバイスがブロック図の形式で示されている。単に利便性および明確性のために、頭文字および他の記述的な用語が使用され得るが、本開示の範囲を限定することは意図されていない。

【0013】

モバイルデバイスカメラは、しばしば、それらの適用性を制限する欠陥の影響を受ける。たとえば、カメラの低分解能および他の光学サブシステムの制約が、従来の走査および光学式文字認識(OCR)関連アプリケーションにおいてモバイルデバイスカメラの有用性を制限することがある。走査プロセス中のユーザの手振れによる被写体ぶれおよびカメラの傾きの変動による射影ゆがみが、従来のテキスト走査およびOCRアプリケーションにおいてモバイルデバイスカメラを使用することに対する付加的な困難を提起する。加えて、走査中、より大きい相対的テキストサイズおよび正確なテキストキャプチャを確実にするために、カメラは、しばしば、テキストにより近く保持され、それによって単一の画像また

はフレーム内でキャプチャされ得る文字または単語の数が制限される。加えて、従来のOCRは、大幅な計算上のオーバーヘッドを負担することがあるので、フレームごとにOCRを使用することは、容認できない遅延を招くことによって多くのアプリケーションに対して非効率であり実行が困難であることがある。

【0014】

本明細書で開示する装置、システムおよび方法を含む実施形態は、効率的で連続的なテキストの走査、追跡、および認識を容易にする。本明細書で使用する移動局(MS)という用語は、携帯電話、ゲーミングデバイス、撮像デバイス、携帯電話もしくは他のワイヤレス通信デバイス、パーソナル通信システム(PCS)デバイス、パーソナルナビゲーションデバイス(PND)、パーソナル情報マネージャ(PIM)、携帯情報端末(PDA)、ラップトップなどを

10

【0015】

さらに、「移動局」は、通信要素および/または通信機能を省略することがある。「移動局」という用語はまた、ネットワークに接続すること、または場合によっては、ワイヤレスもしくはワイヤード接続のいずれかで別のデバイスと通信することを行うように構成されていないデバイスをも対象とする。たとえば、本明細書で説明する実施形態は、ゲーミングデバイス、ハンドヘルドカメラ、あるいはネットワークに接続することまたは場合によってはワイヤレスもしくはワイヤード接続のいずれかで別のデバイスと通信することを行うように構成されていない別のデバイスなど、スタンドアロンデバイス内に実装され得る。

20

【0016】

いくつかの実施形態では、移動局は、ワイヤレス通信および/またはナビゲーション信号を受信すること、ならびに短距離のワイヤレス、赤外線、および/または配線の接続を介することを含めて他のデバイスと通信することが可能なモバイルデバイスの形態をとることができる。また、「移動局」は、インターネット、WiFi、または別のネットワークを介するなどしてサーバと通信することが可能である、ワイヤレス通信デバイス、コンピュータ、ラップトップなどを含む、すべてのデバイスを含むように意図される。

【0017】

図1は、例示的なMS 100のブロック図を示す。図1に示すように、MS 100は、カメラ110、慣性測定ユニット(IMU)130、ディスプレイ145、プロセッサ150、メモリ160、および/またはトランシーバ170を含み得、それらは、接続120を介して動作可能に結合され得る。接続120は、バス、線、光ファイバー、リンクなど、またはそれらの何らかの組合せを備え得る。

30

【0018】

トランシーバ170は、たとえば、1つまたは複数のタイプのワイヤレス通信ネットワークを介して1つまたは複数の信号を送信するように対応した送信機と、1つまたは複数のタイプのワイヤレス通信ネットワークを介して送信される1つまたは複数の信号を受信するための受信機とを含み得る。トランシーバ170は、限定はしないが、IEEE 802.11ファミリー規格に基づき得るWiFiネットワークまたはワイヤレスローカルエリアネットワーク(WLAN)、Bluetooth(登録商標)、近距離無線通信(NFC)、LTEやWiMAXなどのワイヤレス広域ネットワーク(WWAN)など、多様な技術に基づいてワイヤレスネットワークとの通信を可能にする。いくつかの実施形態では、MS 100はまた、ワイヤードネットワークを介する通信のための1つまたは複数のポートを含み得る。いくつかの実施形態では、トランシーバ170および/またはMS 100上の1つまたは複数の他のポートが省略されることがある。

40

【0019】

カメラ(複数可)110は、キャプチャされた画像をプロセッサ(複数可)150に送信し得る。いくつかの実施形態では、カメラ110は、前向きカメラおよび/または後ろ向きカメラを含み得、および/またはCMOS/光学センサをも組み込み得る。一実施形態では、前向きカメラは、デバイスの通常のユーザ動作の間にユーザの方を向き、一方、後ろ向きカメラは、デバイスの通常の動作の間にユーザと反対の方を向く。いくつかの実施形態では、カメラ11

50

0は、静止画像とビデオ画像の両方をキャプチャ可能であってよい。カメラ110は、ビデオ画像および/または一連の静止画像をプロセッサ150に送信し得る。いくつかの実施形態では、カメラ110によってキャプチャされた画像は、生の非圧縮フォーマットであってよく、処理され、および/またはメモリ160に記憶される前に圧縮されてよい。いくつかの実施形態では、画像圧縮は、無損失圧縮または損失圧縮の技法を使用してプロセッサ150によって実行され得る。

#### 【0020】

いくつかの実施形態では、プロセッサ150はまた、IMU 130からの入力を受信し得る。いくつかの実施形態では、IMU 130は、3軸加速度計(複数可)、3軸ジャイロスコープ(複数可)、および/または磁力計(複数可)を備え得る。IMU 130は、速度、向き、および/または他の姿勢関連情報をプロセッサ150に与え得る。いくつかの実施形態では、IMU 130は、カメラ110による各画像フレームのキャプチャと同期して、測定された情報を出力し得る。

10

#### 【0021】

MS 100内に含まれるすべてのモジュールが図1に示されているとは限らない。例示的なMS 100はまた、図示の機能ブロックのうちの1つまたは複数を追加すること、組み合わせること、または省略することによるなど、本開示に合致した様々な方法で修正され得る。たとえば、いくつかの構成では、MS 100はIMU 130を含まない。さらに、いくつかの例示的な実装形態では、MS 100は、周辺光センサ、マイクロフォン、音響センサ、超音波センサなど、多様な他のセンサ(図示せず)を含み得る。いくつかの実施形態では、MS 100は、カメラ110によってキャプチャされた画像フレームをライブで閲覧することを可能にするディスプレイを含み得る。いくつかの実施形態では、MS 100の一部は、1つまたは複数のチップセットなどの形態をとることができる。

20

#### 【0022】

プロセッサ150は、ハードウェア、ファームウェアおよびソフトウェアの組合せを使用して実装され得る。プロセッサ150は、テキスト検出および追跡に関連するコンピューティングプロシージャまたはプロセス、光学式文字認識(OCR)ならびに/あるいはコンピュータビジョンおよび画像処理のうちの少なくとも一部を実行するように構成可能な1つまたは複数の回路を表し得る。プロセッサ150は、メモリ160から命令および/またはデータを検索し得る。プロセッサ150は、1つまたは複数のグラフィック処理ユニット(GPU)、特定用途向け集積回路(ASIC)、デジタル信号プロセッサ(DSP)、デジタル信号処理デバイス(DSPD)、プログラマブル論理デバイス(PLD)、フィールドプログラマブルゲートアレイ(FPGA)、コントローラ、マイクロコントローラ、マイクロプロセッサ、組み込みプロセッサコア、および本明細書で説明する機能を実行するように設計された他の電子ユニット、またはそれらの組合せを使用して実装されてもよい。

30

#### 【0023】

メモリ160は、プロセッサ150の中および/またはプロセッサ150の外に実装され得る。本明細書で使用する「メモリ」という用語は、長期メモリ、短期メモリ、揮発性メモリ、不揮発性メモリ、または他のメモリのいずれかのタイプを指し、メモリの任意の特定のタイプもしくはメモリの数、またはメモリが記憶される物理的媒体のタイプに限定されない。いくつかの実施形態では、メモリ160は、画像処理、テキスト検出および追跡、ならびにプロセッサ150で実行される他のタスクを容易にするためのコードを保持し得る。たとえば、メモリ160は、データ、キャプチャされた静止画像、ビデオフレーム、プログラム結果、ならびにIMU 130および他のセンサによって与えられたデータを保持し得る。一般に、メモリ160は、任意のデータ記憶機構を表し得る。たとえば、メモリ160は1次メモリおよび/または2次メモリを含み得る。1次メモリは、たとえば、ランダムアクセスメモリ、読み取り専用メモリなどを含み得る。図1ではプロセッサ150とは別個のものとして示されているが、1次メモリの全部または一部がプロセッサ150内に設けられるか、あるいはプロセッサ150と共設および/または結合され得ることを理解されたい。

40

#### 【0024】

2次メモリは、たとえば、1次メモリと同じもしくは同様のタイプのメモリ、および/ま

50

たは、たとえばフラッシュ/USBメモリドライブ、メモリカードドライブ、ディスクドライブ、光ディスクドライブ、テープドライブ、固体メモリドライブなど、1つまたは複数のデータ記憶デバイスまたはシステムを含み得る。いくつかの実装形態では、2次メモリは、MS 100に結合されたリムーバブルメディアドライブ(図示せず)内の非一時的コンピュータ可読媒体を動作可能に受容し得るか、または場合によってはそれに結合するように構成可能であり得る。いくつかの実装形態では、非一時的コンピュータ可読媒体は、メモリ160および/またはプロセッサ150の一部を形成することができる。

【0025】

MS 100はまた、カメラ110によってキャプチャされた画像を処理するためのソフトウェアを含み得る。いくつかの実装形態では、プロセッサ150は、キャプチャされた画像における特徴を検出して解釈するためにカメラ110によってキャプチャされた1つまたは複数の画像を処理するためにソフトウェアを実行することができる。たとえば、プロセッサ150は、画像内のテキストを認識するために画像上で光学式文字認識(OCR)を実行することができる。

【0026】

いくつかの実装形態では、プロセッサ150は、画像を登録してカメラ姿勢を判断するために自然特徴追跡技法(natural features tracking technique)を使用し得る。コンピュータビジョンでは、検出は、キャプチャされた画像フレーム内のターゲットオブジェクトを位置特定し、オブジェクトに対するカメラ姿勢を計算するプロセスを指す。追跡は、時間的に連続する画像フレームにわたる、オブジェクトに対するカメラ姿勢の推定を指す。特徴追跡では、特徴は一度検出されると、前の情報またはカメラ姿勢についての仮定を使用して後続の画像内で直接追跡され得る。追跡では、連続するフレーム間のターゲットオブジェクト位置における差は一般的に小さい。カメラの姿勢は、基準系に対するカメラの位置と向きとを指す。登録(registration)、画像登録または画像整合(alignment)は、異なるデータセットを単一の座標系に変換および/または統合するプロセスを指す。したがって、たとえば、画像登録または画像整合は、画像ペア間の画素単位の差を最小化しながら2つの画像の座標系を整合させる幾何変換を決定するために使用され得る。画像登録では、整合は、任意のまたは比較的大きい基線を有する2つの画像を伴うことがある。

【0027】

一般的に、特徴ベースの追跡方法では、特徴は、オブジェクトを表すために使用される。多くの従来の特徴ベースの追跡方法では、追跡は人工的な外部のマーカーに基づいており、そのマーカーは、次いで位置特定および追跡のために使用される。しかしながら、外部からのマーカーを使用する従来の方法は、道路標識などの自然界のオブジェクト、オブジェクト上のラベル、または様々なメディア上のテキストが走査されるときには実行不可能である。それゆえ、本明細書で開示する実施形態は、追跡を実行するために、テキストベースの特徴、テキストが現れる媒体についての特徴、テクスチャ、およびキャプチャされた画像フレーム内の他の特有の様相のような、固有の特徴を使用し得る。外部からのマーカーを使用しない追跡はまた、「マーカーレス」追跡と呼ばれることもある。本明細書で開示する実施形態は、マーカーレス追跡を容易にする。

【0028】

特徴ベースのオブジェクト追跡は、特徴抽出および特徴対応を含み得る。特徴抽出は画像内の特徴の検出を指す一方で、特徴対応は、2つの画像内の特徴の間の対応を照合または判断するプロセスである。それゆえ、異なる撮像条件(たとえば、スケール、照明、向きなど)のもとで複数のビューにわたって顕著である可能性のある顕著な特徴の使用が、特徴対応判断を容易にする。

【0029】

したがって、特徴ベースの追跡方法では、1つまたは複数のオブジェクトが画像フレーム内で識別され、オブジェクトは、1つまたは複数の基準系に対するカメラの姿勢を判断するために、複数の画像フレームにわたって追跡され得る。追跡されるオブジェクトは、ターゲットまたは追跡ターゲットを構成し得る。いくつかの実装形態では、ターゲットは

10

20

30

40

50



、画像内のオブジェクトから抽出されるかまたは画像内のオブジェクトと関連付けられた特徴点を使用して追跡され得る。たとえば、画像フレームは複数のオブジェクトを含み、関心点は画像フレーム、オブジェクトのグループおよび/または画像フレーム内の個別のオブジェクトと関連付けられ得る。「自然」という用語は、人工のまたは外部のシーンマーカーなしにキャプチャされた画像フレーム内で通常発生する特徴点を指すために使用される。

#### 【0030】

いくつかの実施形態では、開示する方法は、実時間または略実時間で実行され得る。いくつかの実施形態では、テキスト認識プロセスは、初期の現在画像フレームを使用して開始され得る。画像フレームのコンテキストにおいて本明細書で使用する「現在」という用語は、カメラ110によってキャプチャされ、ユーザによる同時閲覧のために利用可能な画像フレームを指す。「初期の」、「前の」および「後続の」という用語は、互いに対する、または第3の画像フレームに対する画像フレームの時間的な位置を指す。いくつかの実施形態では、テキストが、たとえばOCRプロセスによって識別されかつ認識されている初期の現在画像フレームが記憶され、画像ターゲット生成プロセスのためのベースとして使用され得る。テキストが識別および認識されている画像フレームは、参照フレームと呼ばれる。

10

#### 【0031】

いくつかの実施形態では、画像ターゲット生成プロセスは、次いで、参照フレームに基づいて追跡モデルを構築するために使用され得る。いくつかの実施形態では、ビジュアル SLAM(VSLAM)など、同時位置推定およびマッピング(SLAM:Simultaneous Localization And Mapping)技法が、テキストを含む参照フレームおよび1つまたは複数の後続の画像フレーム内の様々な自然特徴を使用してテキストを追跡するために使用され得る。いくつかの実施形態では、現在画像と参照フレームとの間の特徴対応が、テキストを追跡するために使用され得る。いくつかの実施形態では、並列追跡およびマッピング(PTAM:Parallel Tracking and Mapping)など、SLAMの変形が、追跡のために使用され得る。

20

#### 【0032】

VSLAMでは、画像パッチの顕著な自然特徴が、特徴の位置とカメラの運動との両方を判断するために複数のビデオフレームにわたって追跡される。SLAMベースの技法は、周囲の3-D構造を同時に判断しながら、3次元(3D)位置(たとえば、x、y、z座標)およびカメラ110の向き(たとえば、角度方向のロール、ピッチおよびヨー)を含む6自由度(6-DOF)のカメラ姿勢を追跡することを可能にする。

30

#### 【0033】

いくつかの実施形態では、スケール不変特徴変換(SIFT:Scale Invariant Feature Transform)技法が、画像フレーム内のコーナー点、エッジなどの特徴を識別するために使用され得る。様々な計算効率の良い修正されたSIFT記述子を含むSIFTベースの記述子およびその変形は、比較的大きい視点変更の面内でロバストな追跡および画像間の特徴対応を容易にする。いくつかの実施形態では、画像フレーム内のキーポイントが識別され、個別のキーポイントに対応するローカルな記述子が構築され、画像特徴を追跡するために使用され得る。「パッチ」または「画像パッチ」は、追跡のために使用され得る、特徴点周りの画素の領域である。

40

#### 【0034】

上記で概説した技法は単なる例であり、いくつかの実施形態では、様々な他の技法が特徴検出のために使用され得る。たとえば、加速セグメントテストからの特徴(「FAST:Features from Accelerated Segment Test」)またはその変形、高速化ロバスト特徴(「Speeded-Up Robust Feature」)またはその変形、ハイブリッド点および/またはエッジ検出技法などに基づく技法が、当業者には明らかである適切な修正とともに特徴検出に使用され得る。

#### 【0035】

さらに、いくつかの実施形態では、画像整合技法が、参照画像フレームと1つまたは複

50

数の後続の画像フレームとの間の相対運動を計算するために使用され得る。画像整合プロセスは、オブジェクトの運動を推定するために、参照画像フレームと運動パラメータに対する現在の後続の画像フレームとの間の整合の尺度を計算する。たとえば、効率的2次最小化(ESM:Efficient Second-order Minimization)は、参照フレームと現在の後続のフレームとの間の差を反復して最小化することによって、参照フレームに対する現在の後続のフレームの相対運動および/または整合を計算するために使用され得る。

【0036】

追跡モデルおよび/または画像整合プロセスは、それぞれ、認識されたテキストの位置ならびに参照フレームに対する現在のおよび/または後続の画像フレームの姿勢を得ることができる。

10

【0037】

いくつかの実施形態では、次いで、パッチ追跡器が開始され、追跡モデルから取得された認識されたテキストの位置および/または画像整合プロセスから取得された参照フレームに対する現在の後続の画像フレームの姿勢を使用して初期化され得る。画像パッチという用語は、一般に、画像内の特徴点周りの領域に対応する画像の一部を指す。パッチ追跡アルゴリズムは、テキストに対するカメラ姿勢を計算するために、参照画像パッチと現在画像から抽出された現在画像パッチとの間の特徴点を照合し得る。

【0038】

多くの拡張現実(AR)アプリケーションでは、たとえば、テキストまたはグラフィック要素などの1つまたは複数の仮想オブジェクトが、ライブのカメラビュー上に重ね合わされてディスプレイ145上に表示され得る。いくつかの実施形態では、ディスプレイ145は、カメラ110、AR画像、グラフィカルユーザインターフェース(GUI)、プログラム出力などによってキャプチャされたライブ画像を表示するために使用され得る。ディスプレイ145は、スクリーンと呼ばれることもある。仮想オブジェクトの正確な登録は、仮想オブジェクトを描画するときにディスプレイ145上のライブビデオ撮像において、仮想テキストおよび/またはグラフィックオブジェクトの正確な配置を確実にすることができる。したがって、いくつかの実施形態では、カメラ110によって走査されているテキストは、テキストブロックがテキスト認識に続く1つまたは複数のアプリケーションによって認識および処理され得るように追跡され得る。

20

【0039】

「テキストブロック」という用語は、限定はしないが、単語、一連の単語、および/またはテキストの1行または複数行など、一連の文字を指すために使用される。「文字(character)」という用語は、一般に、文字(letter)、数字、および言語にかかわらず様々な他の記号を指すために使用される。いくつかの実施形態では、1つまたは複数の仮想オブジェクトは、テキスト検出および追跡の結果に基づいてテキストブロックの上にオーバーレイされ得る。たとえば、ARインサイチュテキスト翻訳アプリケーションでは、媒体上の文字が、OCRプロセスを介して追跡および認識され、別の言語に翻訳され、元のテキストが、翻訳されたテキストを表す文字でオーバーレイされ、それによって翻訳されたテキストのライブカメラビューを容易にすることができる。ARアプリケーションを実行しているとき、たとえば、プロセッサ150は、仮想オブジェクトをライブカメラビュー内に配置し、仮想オブジェクトは、カメラの姿勢と一致するようにキャプチャされた画像内で回転および/または変位され得る。

30

40

【0040】

いくつかの実施形態では、画像内の1つまたは複数の自然特徴および/またはテキストベースの特徴が、画像に対するMS 100および/またはカメラ110の姿勢を判断するために使用され得る。いくつかの実施形態では、高いコントラスト差を有する点、エッジ、および/またはコーナー点など、画像内の特定の特徴点がフレーム内で検出され、フレームにわたってカメラ110の姿勢を追跡するために使用され得る。いくつかの実施形態では、プロセッサ150は、特徴点に関係する画素関連情報を記憶し得る。たとえば、輝度勾配値および/または特徴点画素と関連付けられた輝度勾配値の関数が、メモリ160内に記憶され得る。

50

いくつかの実施形態では、追跡は、認識されたテキストの画像内で発生するコーナー点など、自然特徴に部分的に基づくことができる。

【0041】

いくつかの実施形態では、画像に対するカメラ110の姿勢は、IMU 130からの入力に部分的に基づいて判断または修正され得る。いくつかの実施形態では、レンズの焦点距離、カメラの最短撮影距離など、カメラ110の知られている固有パラメータおよび特性、ならびに/あるいは他のセンサからの入力、カメラ姿勢の判断を支援および/または改良するために、IMU入力および/または他のテキスト追跡方式とともに使用され得る。

【0042】

いくつかの実施形態では、テキスト追跡は、前に走査されたテキストブロックの幾何学的位置と現在走査されている要素との間の相関を容易にすることができる。テキスト追跡は、走査されたテキストブロック間の幾何学的関係を判断するために使用され得、それによって走査された画像/フレームシーケンスから走査されたテキストの構造の復元が可能になる。

【0043】

図2は、媒体上の単語に対するカメラ110の位置、およびテキストブロックの位置を追跡および/または維持するために使用され得る座標系を示す。たとえば、カメラ110は、カメラ位置230-1からテキストブロック-1 210-1を含む初期の画像フレームをキャプチャし得る。カメラ位置230-1におけるカメラ110の視野は、一点鎖線250内の領域である。

【0044】

いくつかの実施形態では、カメラ位置230-1においてキャプチャされた画像フレームは、テキストブロック210-1内の1つまたは複数の文字または単語を識別するため、および基準系275に対するテキストブロック210-1の初期姿勢260を判断するためにOCRを使用することによって処理され得る。いくつかの実施形態では、カメラ位置230-1における基準系275は、直交軸のセットを含み得、軸のうちの1つまたは複数がMS 100の本体および/またはカメラ110の平面と整合され得る。たとえば、一実施形態では、カメラ位置230-1においてキャプチャされた画像フレームは、記憶され、参照画像フレームとして使用され得る。

【0045】

いくつかの実施形態では、テキストブロック210の姿勢はまた、「ページ座標系」を表し得る基準系285を使用して維持され得る。たとえば、テキストブロック210がその上に存在する媒体に対してフロントパラレルにユーザがカメラを保持する状況において、別個のバウンディングボックスが、各テキストブロック210と関連付けられ得る。いくつかの実施形態では、認識されたテキストブロック210-1の左下の点および/またはテキストブロック210-1と関連付けられたバウンディングボックスが、ページ座標系の原点として設定され、テキストブロック210-1の基線および/または関連付けられたバウンディングボックスの基線がx軸として設定され得る。y軸は、カメラ平面内でx軸に垂直であるように定義され得る。いくつかの実施形態では、基準系285は、テキストブロック210-1が認識され、後続の画像フレームに対するページ座標系として働き得る第1の画像フレームに基づいて定義され得る。

【0046】

いくつかの実施形態では、追跡モデルは、参照フレームから開始し、1つまたは複数の自然特徴ならびに/または参照画像フレームおよび後続のフレームの中の特徴点を使用して構築され得る。追跡モデルが構築されると、追跡モデルは、後続のフレーム内の認識されたテキストの位置を判断するために使用され得る。たとえば、カメラ110は、走査プロセス中に、破線255内の領域で示す視野および基準系280を有するカメラ位置230-2に移動することができる。

【0047】

図2に示すように、カメラ位置230-2は、テキストブロック210が存在する媒体に対するカメラ110の位置および向きにおける変化を反映する。いくつかの実施形態では、自然特徴追跡が、基準系275に対するカメラ110の6-DOFの姿勢の推定を取得するために、部分的

10

20

30

40

50

に使用され得る。基準系275におけるカメラ110の推定された6-DOFの姿勢は、基準系280における認識されたテキストブロック210-1の位置を判断するために使用され得る。

【0048】

さらに、いくつかの実施形態では、テキストブロック1 210-1内で認識された文字またはそれらの一部に部分的に基づいて、画像整合技法が、カメラ位置230-1の参照フレームとカメラ位置230-2の参照フレームとの間の画像の相対運動を計算するために使用され得る。たとえば、ESMなどの画像整合技法が、認識されたテキストブロック210-1の知られている推定された位置を使用することによって、カメラ位置230-1において取得された参照フレームに対するカメラ位置230-2における現在画像の運動を判断するために使用され得る。

10

【0049】

次いで、パッチ追跡器が、画像整合プロセスによって測定された画像の相対運動および/または(参照フレームが前に生成されている場合は)前の参照フレームを使用して初期化され得る。パッチ追跡器は、テキストに対する改良されたカメラ姿勢を計算するために、参照画像パッチと現在画像から抽出された現在画像パッチとの間で特徴点を照合することができる。

【0050】

いくつかの実施形態では、たとえば、位置230-1における画像フレームとカメラ位置230-2においてキャプチャされた画像フレームとの間に十分なオーバーラップがあるとき(または参照フレームが未だ生成されていないとき)、位置230-2においてキャプチャされた画像フレームが、テキストブロック210-1およびテキストブロック210-2における1つまたは複数の文字または単語を識別するためにOCRを使用することによって処理され得る。基準系280に対するテキストブロック210-2の姿勢270は、テキストブロック1 210-1の知られている姿勢265に基づいて、およびESMまたは別の画像整合技法によって判断される位置230-1と230-2との間のカメラ110の相対的な運動および向きに基づいて判断され得る。いくつかの実施形態では、ESMは、後続の画像フレームのより低い分解能バージョンにおいて動作することができる。さらに、位置230-1と230-2との間のカメラ110の計算された相対的な運動および向きもまた、基準系285に関してテキストブロック210-2の姿勢を判断するために使用され得る。したがって、第2のテキストブロックに対するカメラの位置は、第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係を判断するために使用され得る。したがって、いくつかの実施形態では、テキストブロック210の位置は、(i)基準系275および/または280などのカメラ座標系内の相対的姿勢と、(ii)ページ座標系285内の相対的姿勢との2通りの方法で表され得る。

20

30

【0051】

図3は、開示する実施形態に合致したテキスト検出および追跡の例示的な方法300のフローチャートを示す。図3を参照すると、いくつかの実施形態では、方法300は、MS 100上で動作しているコンピュータビジョンおよび/またはARアプリケーションによって起動され得る。いくつかの実施形態では、例示的な方法300は、様々な初期化ルーチンおよび準備ルーチンが実行され得る305において開始され得る。次に、310において、第1のまたは次の走査された画像フレームが取得され得る。たとえば、ユーザは、ページなどの比較的平坦な表面上のテキストにわたってMS 100上でカメラ110をスワイプすることによってテキストの走査および/または画像フレームのキャプチャを行うことができ、走査されたテキストに対応する1つまたは複数の画像またはフレームがもたらされる。たとえば、走査プロセスの間、ユーザは、時間的に連続するいくつかの画像フレームを含むビデオをキャプチャすることができ、すなわちキャプチャされたフレームがビデオシーケンスの一部を形成することができる。

40

【0052】

320において、トリガ事象(320における「Y」)が、325におけるOCRプロセスを初期化するために使用され得る。トリガには、たとえば、カメラビュー内で前に走査されていない領域の検出が含まれ得る。いくつかの実施形態では、未走査領域は、前の参照画像に対す

50

る第2の走査された画像フレームの位置と第1の追跡ターゲットとの間にオーバーラップがないことから検出され得る。たとえば、オーバーラップの欠如がしきい値を超える場合、325におけるOCRがトリガされ得る。たとえば、320において、現在画像フレームと第1の参照フレームとの間のオーバーラップ領域のパーセンテージがしきいのパーセンテージより低い場合、方法は325に進む。

【0053】

いくつかの実施形態では、325におけるOCRは、第1の参照フレームと後続の画像フレームとの間の相対運動がしきい値を超えるとトリガされ得る。いくつかの実施形態では、第1の参照フレームと後続の画像フレームとの間の相対運動が、後続の画像フレーム内で前に認識されたテキストブロックの位置に部分的に基づいて、および/またはESMまたは別の画像整合技法を使用して計算され得る。たとえば、325におけるOCRは、第1の参照フレームに対する現在画像フレームの回転の大きさ、および/または第1の参照フレームに対する現在画像フレームの並進(translation)の大きさが何らかのしきい値を超える場合にトリガされ得る。いくつかの実施形態では、参照フレームがない場合、たとえば、新しい画像が走査されているとき、および/またはテキスト走査もしくはARアプリケーションが最初に開始されるとき、325におけるOCRが自動的にトリガされ得る。

【0054】

いくつかの実施形態では、最後の参照フレームに対して「新しい」カメラ画像内の領域が、新しい領域内に追跡される特徴が不在であることに基づいて判断され得る。いくつかの実施形態では、新しい領域が、画像フレームのしきいのパーセンテージまたは割合を超えるときはいつでも、方法は325に進む。別の実施形態では、参照カメラフレームが、ページ座標系285上に投影されて現在のカメラ姿勢に基づいて整合され、ページ平面上の「新しい」領域が、現在の参照フレームによってカバーされるページ平面上の領域のしきいの領域またはパーセンテージを超えるときはいつでも、325における新しいテキスト/OCR認識事象がトリガされ得る。

【0055】

さらなる実施形態では、325は、最後に追跡された単語または前に認識されたテキストブロックのうちの1つの何らかの部分が現在のカメラフレームの外にあるときにトリガされ得る。たとえば、一実施形態では、325は、現在の画像フレーム内で可視の第1のテキストブロックの割合が何らかのしきい値より小さいときにトリガされ得る。第1のテキストブロックの可視の割合は、第1の参照フレーム内の第1のテキストブロックの全領域に対する現在の画像フレーム内の第1のテキストブロックの可視部分を含む領域の比として決定され得る。

【0056】

いくつかの実施形態では、トリガが検出されない(320において「N」の場合、アルゴリズムは355に進むことができ、追跡ターゲットの生成を開始/継続するため、または既存のターゲットを使用してテキスト追跡プロセスを継続するために、判断がなされ得る。

【0057】

トリガ事象が320において検出される(320において「Y」の場合、325において、画像フレームは、画像フレーム内の走査されたテキストに対応する新しい単語/文字を発見し認識するために処理され得る。たとえば、例示的な方法300に関連するOCRアプリケーションが、画像フレームを処理するために起動され得る。いくつかの実施形態では、325におけるOCRアプリケーションは、テキストブロック210-1および/または210-2などのテキストブロック上で動作することができる。いくつかの実施形態では、テキストブロックのサイズは、画像フレーム内で可視のテキストの量、利用可能なメモリ160、OCRアプリケーションの速度、ARアプリケーションのタイプ、プロセッサ150の数および速度など、様々なパラメータに基づいて決定され得る。いくつかの実施形態では、OCRアプリケーションは、ユーザの手振れによる被写体ぶれおよび焦点はずれ画像を補償し得る。

【0058】

330において、OCRプロセスが成功したかどうか判断され得る。OCRプロセスが不成功

10

20

30

40

50

であった(330において「N」の)場合、方法は310に戻り、次の画像フレームが取得される。いくつかの実施形態では、エラーの可視表示または可聴表示が、MS 100を用いてテキストを走査しているユーザに与えられ得る。いくつかの実施形態では、いくつかの画像/フレームにわたってOCRアプリケーションによって単語が検出されない場合、カメラ110のオートフォーカスがトリガされ得る。

【0059】

いくつかの実施形態では、OCR処理が成功した(330において「Y」の)場合、335において、それに対してOCRが330において成功裏に実行され、認識されたテキストブロックを含む特定のカメラ画像/フレームが、参照フレームとして記憶され得る。いくつかの実施形態では、参照フレームは、メモリ160内に記憶され得る。いくつかの実施形態では、参照フレームは、追跡システムの他のモジュールを初期化するために後で利用され得る。

10

【0060】

340において、現在のカメラ位置に対する新しい単語および/またはテキストブロックの姿勢が判断され得る。たとえば、基準系280に対するテキストブロック210-2の相対位置が導出され得る。いくつかの実施形態では、部分的に、新しいテキストブロック210-2の姿勢が、テキストブロック1 210-1などの少なくとも1つの他の前に認識されたテキストブロックを検出することによって判断され得る。たとえば、いくつかの実施形態では、テキストブロック210-1において認識された文字に部分的に基づいて、ESMなどの画像整合技法が、画像の相対運動を計算するために使用され得る。いくつかの実施形態では、ESMは、後続の画像フレームのより低い分解能バージョンにおいて動作することができる。したがって、現在のカメラ位置に対する前に認識されたテキストブロックの姿勢が、整合に基づいて判断され得る。したがって、現在のカメラ位置に対する新しく認識された単語の姿勢が、前に認識されたテキストブロックの判断された位置に基づいて判断され得る。したがって、第2のテキストブロックに対するカメラの位置は、第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係を判断するために使用され得る。既知のテキストブロックおよび新しく認識されたテキストブロックの判断された姿勢に基づいて、ページ座標系285における新しい単語の姿勢が導出され得、それによって走査された画像/フレームシーケンスから走査されたテキストの構造を復元することが容易になる。

20

【0061】

次に、345において、新しい単語が、それらの姿勢に従ってテキスト表現に挿入され得る。たとえば、一実施形態では、認識されたテキストブロックの各々が、バウンディングボックスおよびページ座標系285内の座標と関連付けられ、ページ座標系285内の認識されたテキストブロックの座標に基づいてテキスト表現に挿入され得る。次いで、方法は、375に進むことができる。いくつかの実施形態では、適切なデータ構造が、テキスト表現を記憶するために使用され得る。

30

【0062】

いくつかの実施形態では、トリガが検出されない(320において「N」の)場合、アルゴリズムは355に進むことができ、自然特徴ターゲットが構築されて完了しているかどうかの判断がなされる。自然特徴ターゲットが構築されて完了している(355において「Y」の)場合、方法は370に進む。

40

【0063】

追跡ターゲット、たとえば自然特徴追跡ターゲットが構築されておらず、未完了である(355において「N」の)場合、360において、追跡ターゲット(たとえば、自然特徴追跡ターゲット)を生成するプロセスが開始または継続され得る。いくつかの実施形態では、自然特徴追跡ターゲットの生成は、1つまたは複数の画像フレーム内のテキストベースの特徴を含む1つまたは複数の自然特徴を抽出するステップを含み得る。画像に対して抽出された関心点の数が、何らかのしきい値を超えるかまたは何らかの所望の信頼度を満足すると判断されたとき、ターゲットの構築が完了したと考えられ得る。いくつかの例では、識別される特徴の数が増加することで、検出された特徴を使用する後の追跡の正確さが向上するので、そのような自然特徴追跡ターゲットの生成は、2つ以上のカメラフレームを取る

50

ことがある。いくつかの実施形態では、自然特徴追跡ターゲットの生成は、方法300における他の要素と同時に実行され得る。

【0064】

いくつかの実施形態では、高いコントラスト差を有する点、および/またはコーナ点など、画像内の特定の特徴点がフレーム内で検出され、フレームにわたってカメラ110の姿勢を追跡するために使用され得る。いくつかの実施形態では、プロセッサ150は、特徴点に関する画素関連情報を記憶し得る。たとえば、勾配値および/または特徴点画素と関連付けられた勾配値の関数が、メモリ160内に記憶され得る。いくつかの実施形態では、参照フレーム内のキーポイントの特徴ベクトルが、特徴データベース内に記憶され得る。いくつかの実施形態では、自然特徴追跡は、複数の特徴点に個別の重みを割り当てる重み付け方式を使用することができる。いくつかの実施形態では、各特徴点に対する重みは、1つまたは複数のテキストブロックに対する特徴点の位置に部分的に基づくことができる。たとえば、重み付け方式は、テキストブロックに重点を置かれ、テキストブロック上に位置する特徴を、カメラ画像フレーム内の他の特徴に優先して選好する。

10

【0065】

いくつかの実施形態では、抽出された特徴は、カメラ平面に平行な平面上にあるものと仮定され得る。たとえば、テキストブロックが平坦な、略平坦な、または比較的平坦な表面(雑誌のページ、本、新聞、製品容器、または道路標識など)の上に現れ、カメラがテキストに対してフロントパラレルに保持されており、そのことでより良好なテキスト認識が容易になる例では、自然特徴追跡は、抽出された特徴がカメラ平面に平行な平面上にあると仮定され得るモデルを構築するために使用され得る。いくつかの実施形態では、平坦なターゲットに対して、方法は、変化する照明条件下で高いロバスト性を有して、6-DOFにおいてターゲットを追跡することができる。

20

【0066】

いくつかの実施形態では、360の完了後、または自然特徴追跡ターゲットが存在して完了している場合、方法は370に進むことができる。370において、ESMなどの画像整合技法が、カメラ位置230-1において取得された参照画像に対する230-2における現在画像の運動を判断するために使用され得る。いくつかの実施形態では、ESMなどの画像整合/登録技法が、テキスト認識をトリガするために使用された参照画像に対するカメラの相対運動の測定を容易にする。いくつかの実施形態では、カメラによってキャプチャされたより低い分解能バージョンの画像が、ESMへの入力として使用され、フレーム間回転が時間一定方式で(in a time-constant manner)計算され得る。

30

【0067】

375において、参照フレームに対して現在フレーム内で認識された単語の新しいまたは改良された姿勢が計算され得る。いくつかの実施形態では、パッチ追跡器は、既存のまたは新しく生成された画像ターゲットによって開始および初期化され、画像整合370によって計算された現在の姿勢によって初期化され得る。いくつかの実施形態では、パッチ追跡器は、(i)追跡された特徴を有する参照フレームおよび/または(ii)参照画像に対するカメラの相対運動を、入力として受信し得る。いくつかの実施形態では、パッチ追跡器は、参照フレームに対する現在フレーム内で認識された単語の姿勢を計算し得る。たとえば、パッチ追跡器は、テキストに対するカメラ位置を計算するために、参照画像パッチと現在画像から抽出された現在画像パッチとの間で特徴点を照合することができる。たとえば、パッチ追跡器は、基準系280に対するテキストブロック1 210-1の姿勢265を判断することができる。いくつかの実施形態では、記憶されている参照内のキーポイントの特徴ベクトルが、現在フレーム内の対応する特徴ベクトルと照合され得る。したがって、たとえば、新しい単語の姿勢が、座標系275および/またはページ座標系285に基づいてカメラに対して判断され得る。

40

【0068】

380において、ユーザは、表示されたカメラ画像内で現在のテキスト表現に関するフィードバックを与えられ得る。たとえば、認識されたテキストは、ハイライトされ、フォン

50

ト変換され、読み出され、翻訳され、走査されたテキストの上にオーバーレイされ得るか、または場合によってはディスプレイを介してユーザに表示され得る。いくつかの実施形態では、可視フィードバックまたは可聴フィードバックが与えられ得る。たとえば、認識された単語が、テキスト音声アプリケーションによって読み出され得る。いくつかの実施形態では、フィードバックはまた、エラーが発生するときにも与えられ得る。

【0069】

いくつかの実施形態では、385において、入力されている追加のフレームが存在する(385において「Y」の場合、方法は、別の反復を開始するために310に戻る。追加のフレームがない(385において「N」の場合、プロセスは390において終了する。

【0070】

いくつかの実施形態では、方法300は、ARアプリケーションによって起動され得る。いくつかの実施形態では、画像から姿勢情報を判断するための計算は、追加の画像フレームがキャプチャされている間にバックグラウンドで(すなわち、バックグラウンド計算プロセスとして)実行され得る。いくつかの実施形態では、カメラ移動または姿勢調整の判断は、同時位置推定およびマッピング(SLAM)方法のプロシーダを使用して実行され得る。たとえば、SLAMプロシーダは、テキストの一部を追跡するため、および/またはカメラもしくはテキストの別の部分に対するテキストの相対位置を判断するために、ブロック360および/または375のうちの1つまたは複数において使用され得る。

【0071】

いくつかの実施形態では、方法300は、MS 100上で、カメラ110、プロセッサ150、メモリ160、および/またはIMU 130の何らかの組合せを使用して実行され得る。他の実施形態では、方法300の一部が、MS 100にワイヤレスに結合されたサーバまたは他のコンピューティング上で実行され得る。いくつかの実施形態では、サーバは、MS 100から画像を受信し、画像を処理し、かつ処理結果を例示的な方法300を起動するアプリケーションによって指定された移動局のセットに送信することができる。たとえば、インサイチュテキスト翻訳シナリオでは、ユーザはテキストを走査することができ、テキストは、OCRを使用して検出され、追跡され、別の言語に翻訳され、次いで、テキスト追跡に基づいて元のテキストの上端にオーバーレイされ、何人かのユーザが「ライブカメラ」ビュー内で翻訳されたテキストを閲覧/共有し得るように1人または複数人のユーザに送信され得る。別の実施形態では、テキストは、MS 100を使用して走査され、例示的な方法300を使用して検出および追跡され、テキスト音声アプリケーションが、MS 100上のまたはMS 100に結合されたスピーカを使用してテキストを読み出すことができる。たとえば、テキスト音声アプリケーションは、視覚障害者によって、またはテキストが小さすぎて読むのが容易でない場合、または読むことを未だ学習していない人々によって使用され得る。いくつかの実施形態では、自然特徴追跡を使用することで、1枚の白紙の上のテキストの短い複数の単語でさえも追跡することが容易になることがある。さらに、テクスチャ付き材料を使用することで、テキスト追跡においてより大きいロバスト性が可能になることがある。

【0072】

図4は、開示する実施形態に合致したテキスト追跡のためのシステムの例示的なアーキテクチャ400を示す。例示的なアーキテクチャ400の説明は、単に例として提供され、このアーキテクチャに対する様々な変更形態が、開示した実施形態に合致した方式で作成され得ることは、当業者には明らかであろう。たとえば、図4に示す1つまたは複数のブロックおよびモジュールの機能は、様々な方法で組み合わせられ、再分配され、および/または変更され得る。

【0073】

いくつかの実施形態では、単語認識モジュール410、単語追跡モジュール420、テキストアセンブラ430、およびフィードバックピジュアライザ460のモジュールは、ハードウェア、ファームウェア、および/またはソフトウェア、あるいはそれらの何らかの組合せを使用してMS 100上に実装され得る。いくつかの実施形態では、上記のモジュールのうちの1つまたは複数の一部は、MS 100に結合されたメモリ160に記憶されたコードおよび/または

10

20

30

40

50



コンピュータ可読媒体に組み込まれたコードを使用してプロセッサ150上で実行することができる。

【0074】

いくつかの実施形態では、システム400は、単語認識モジュール410、単語追跡モジュール420、テキストアセンブラ430、およびフィールドバックビジュアライザ460を備え得る。いくつかの実施形態では、単語認識モジュール410は、カメラ110によってキャプチャされた画像内のテキストブロックを認識し得るOCRアプリケーションを含み得る。さらに、いくつかの実施形態では、単語認識モジュール410は、新しく認識されたテキストブロックの姿勢を判断し得る。いくつかの実施形態では、新しく認識されたテキストブロックの位置が、ページ座標系285を使用して判断され得る。たとえば、一実施形態では、単語認識モジュール410は、カメラ110によってキャプチャされた画像内のテキストブロックを認識し、認識されたテキストブロックをメモリ160内に記憶するためにプロセッサ150上で実行し得るOCRサブルーチンを含み得る。

10

【0075】

いくつかの実施形態では、単語認識モジュール410は、辞書、線形判別分析(LDA)ベースの分類器、および有限状態変換器を含み得る。LDA分類器は、文字を認識するために、画像フレーム内の画素の値および分布を使用することによってOCRを容易にするために多変量統計分類技法を使用し得る。有限状態変換器は、認識器のエラーを訂正するためにOCR訂正システム内で使用され得る。いくつかの実施形態では、有限状態変換器はまた、短縮文字(truncated characters)、頭字語、部分語(partial words)などによってもたらされる曖昧さを解決するために使用され得る。新しく認識されたテキストブロックを有する画像フレームおよび新しく認識されたテキストブロックの姿勢が、単語認識モジュール410によってテキストアセンブラ430および単語追跡モジュール420に送信され得る。いくつかの実施形態では、方法300内の320、325、330および340の部分は、単語認識モジュール410によって実行され得る。

20

【0076】

いくつかの実施形態では、テキストアセンブラ430は、前に認識されたテキストブロックまたは単語に対するそれらの位置を判断するために、新しく認識されたテキストブロックまたは単語の姿勢を使用することができる。いくつかの実施形態では、新しく認識された単語は、次いで、それらの相対位置に基づいて、今回認識されたテキストブロックの表示450内に配置され得る。たとえば、今回認識されたテキストブロックの表示450は、リンクリストまたはアレイの形態をとることができ、新しく認識された単語は、相対位置に基づいてリストに添付または挿入され得る。たとえば、プロセッサ150上で動作するアプリケーションまたはサブルーチンを含み得るテキストアセンブラモジュール430は、前に認識されたテキストブロックまたは単語に対するそれらの位置を判断するために新しく認識されたテキストブロックまたは単語の姿勢を使用し、今回認識されたテキストブロックの表示450をメモリ160内に記憶することができる。いくつかの実施形態では、方法300内の340および/または345の部分は、テキストアセンブラ430によって実行され得る。

30

【0077】

いくつかの実施形態では、追跡モジュール420は、参照フレームを作成するために新しく認識されたテキストブロックの姿勢を使用することができる。いくつかの実施形態では、追跡モジュール420はまた、参照画像フレームおよび1つまたは複数の後続のフレームに基づいて追跡ターゲットを生成することができる。いくつかの実施形態では、追跡モジュール420は、姿勢推定を改良するために単語認識モジュール410によって計算された姿勢を使用し得るパッチ追跡器を含み得る。いくつかの実施形態では、追跡モジュール420は、パッチ追跡器を初期化して改良された姿勢推定を取得するために使用され得る初期姿勢を取得するためにESMを使用し得る。いくつかの実施形態では、追跡モジュール420は、今回追跡された単語およびそれらそれぞれの姿勢のリスト440とともに維持し得る。いくつかの実施形態では、355、360、370および375の部分は、追跡モジュール420によって実行され得る。一実施形態では、追跡モジュール420は、プロセッサ150上で動作するアプリケー

40

50

ションまたはサブルーチンの形態をとることができる。

【0078】

いくつかの実施形態では、フィードバックビジュアライザ460は、テキストブロックの上にオーバーレイされる単語および/または拡張オブジェクトを表示するために、今回認識されたテキストブロックの表示450および今回追跡された単語およびそれらそれぞれの姿勢のリスト440とともに使用し得る。たとえば、拡張オブジェクトは、認識されたテキストの翻訳を含むことができる。いくつかの実施形態では、認識されたテキストは、MS 100上のディスプレイ上に表示され、テキスト音声アプリケーションを使用してMS 100に結合されたスピーカーを介して読み出され得る。いくつかの実施形態では、380の一部は、フィードバックビジュアライザ460によって実行され得る。たとえば、一実施形態では、フィードバックビジュアライザ460は、プロセッサ150上で動作するアプリケーションまたはサブルーチンの形態をとることができ、ディスプレイ145を使用してテキストブロックの上にオーバーレイされた単語および/または拡張オブジェクトを表示することができる。

10

【0079】

図5は、開示する実施形態に合致した、移動局上で実行され得る例示的な方法500のフローチャートを示す。たとえば、いくつかの実施形態では、方法は、テキストを検出および追跡している間に実行され得る。

【0080】

図5を参照すると、いくつかの実施形態では、方法500は、MS 100上で動作するコンピュータビジョンおよび/またはARアプリケーションによって起動され得る。いくつかの実施形態では、例示的な方法500は、様々な初期化ルーチンおよび準備ルーチンが実行され得る505において起動され得る。

20

【0081】

次に、510において、第1の参照フレームが、第1のテキストブロックを位置特定して認識するために、移動局、たとえばMS 100の上のカメラによってキャプチャされた画像フレーム上で光学式文字認識(OCR)を実行することによって取得され得る。たとえば、いくつかの実施形態では、510の一部が、単語認識モジュール410内でOCRアプリケーションによって実行され得る。

【0082】

30

520において、後続の画像フレームが、選択された後続の画像フレームと関連付けられたパラメータに基づいて、後続の画像フレームのセットから選択され得る。たとえば、いくつかの実施形態では、520の一部が、単語認識モジュール410によって実行され得る。いくつかの実施形態では、後続の画像フレームのセットに対するカメラ姿勢が、たとえば、後続の画像フレームを選択する前に計算され得る。たとえば、いくつかの実施形態では、カメラ姿勢の計算の一部が、追跡モジュール420によって実行され得る。いくつかの実施形態では、各カメラ姿勢は、別個の後続の画像フレームと関連付けられ、および/または関連付けられた後続の画像フレームを第1の参照フレームと整合させることに少なくとも部分的に基づいて判断される。選択された後続の画像フレームは後続の画像フレームのセットからであるので、選択された後続の画像フレームに対するカメラ姿勢は、このプロセスの間に計算され得る。いくつかの実施形態では、後続の画像フレームのセット内の画像フレームと関連付けられたパラメータは、それらの画像フレームと関連付けられた、計算されたカメラ姿勢に少なくとも部分的に基づいて判断される。選択された後続の画像フレームは後続の画像フレームのセットからであるので、選択された後続の画像フレームと関連付けられたパラメータは、このプロセスの間に判断され得る。いくつかの実施形態では、パラメータの判断の一部は、追跡モジュール420によって実行され得る。いくつかの実施形態では、パラメータは、以下の、選択された後続の画像フレームと第1の参照フレームとの間のオーバーラップ領域のパーセンテージと、選択された後続の画像フレーム内で可視の第1のテキストブロックの割合(割合は、第1のテキストブロックの全領域に対する選択された後続の画像フレーム内の第1のテキストブロックの可視部分を含む領域の比と

40

50

して決定される)と、第1の参照フレームに対する選択された後続の画像フレームの回転の大きさと、第1の参照フレームに対する選択された後続の画像フレームの並進の大きさと、の各々のうちの1つまたは複数またはそれらの組合せを備えるかまたは含むことができる。

#### 【0083】

530において、第2の参照フレームは、第2のテキストブロックを認識するために、選択された後続の画像フレーム上でOCRを実行することによって取得され得る。たとえば、いくつかの実施形態では、530の一部が、単語認識モジュール410内でOCRアプリケーションによって実行され得る。

#### 【0084】

次に、540において、第1のテキストブロックと第2のテキストブロックとの間の幾何学的関係が、第2の参照フレーム内の第1のテキストブロックの位置と第2の参照フレームと関連付けられたカメラ姿勢とに少なくとも部分的に基づいて判断され得る。たとえば、いくつかの実施形態では、540の一部が、テキストアセンブラモジュール430によって実行され得る。545において、制御は、呼出しプログラムならびに/あるいはコンピュータビジョンおよび/またはARアプリケーションなどのアプリケーションに戻ることができる。

#### 【0085】

開示された態様の前述の記載は、いかなる当業者であっても、本開示に合致した様々な実施形態を製造または使用できるように提供される。これらの態様への様々な変更形態は、当業者に容易に明らかとなり、本明細書で定義された一般原理は、本開示の精神または

#### 【符号の説明】

#### 【0086】

- 100 移動局 (MS)
- 110 カメラ
- 120 接続
- 130 慣性測定ユニット (IMU)
- 145 ディスプレイ
- 150 プロセッサ
- 160 メモリ
- 170 トランシーバ
- 210 テキストブロック
- 210-1 テキストブロック
- 210-2 テキストブロック
- 230-1 カメラ位置、位置
- 230-2 カメラ位置、位置
- 250 一点鎖線
- 255 破線
- 260 テキストブロック210-1の初期姿勢
- 265 テキストブロック210-1の知られている姿勢
- 270 テキストブロック210-2の姿勢
- 275 基準系、座標系
- 280 基準系
- 285 基準系、ページ座標系
- 400 テキスト追跡のためのシステムのアーキテクチャ、システム
- 410 単語認識モジュール
- 420 単語追跡モジュール、追跡モジュール
- 430 テキストアセンブラ、テキストアセンブラモジュール
- 440 今回追跡された単語およびそれらそれぞれの姿勢のリスト
- 450 今回認識されたテキストブロックの表示

10

20

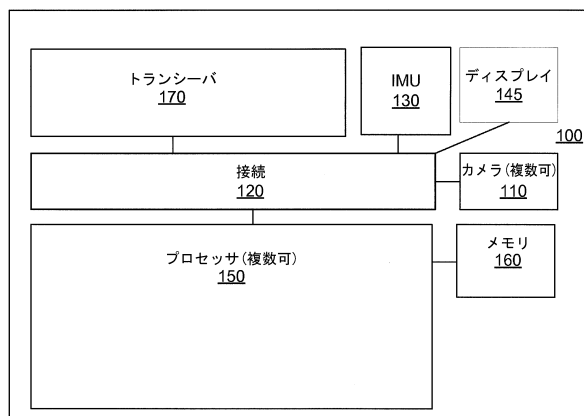
30

40

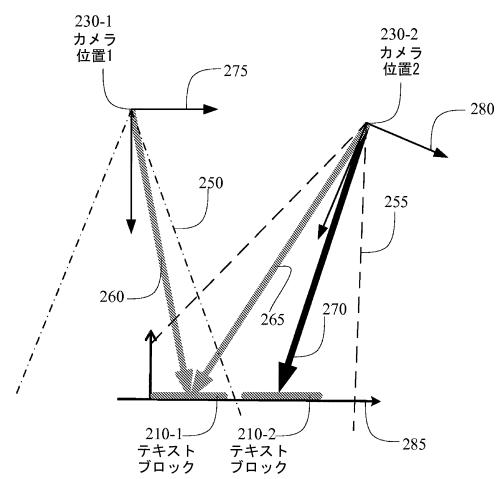
50

## 460 フィードバックビジュアライザ

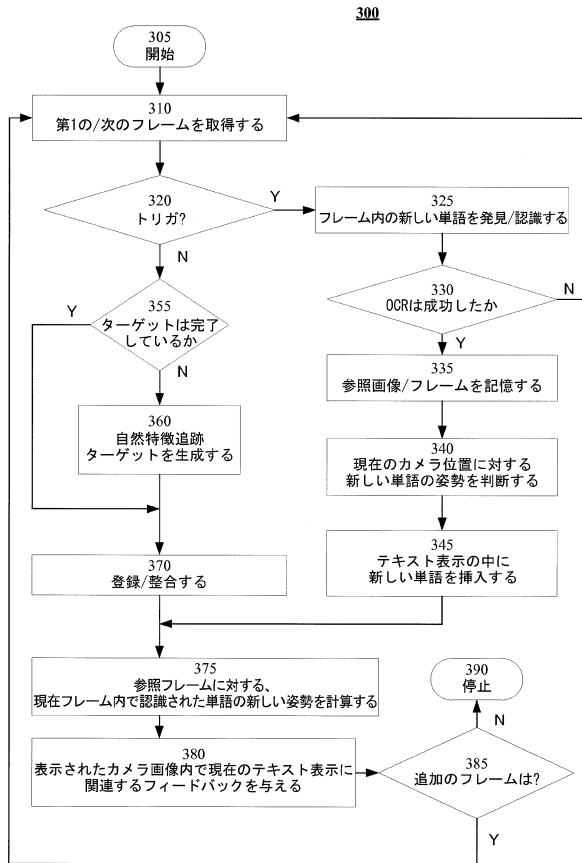
【図 1】



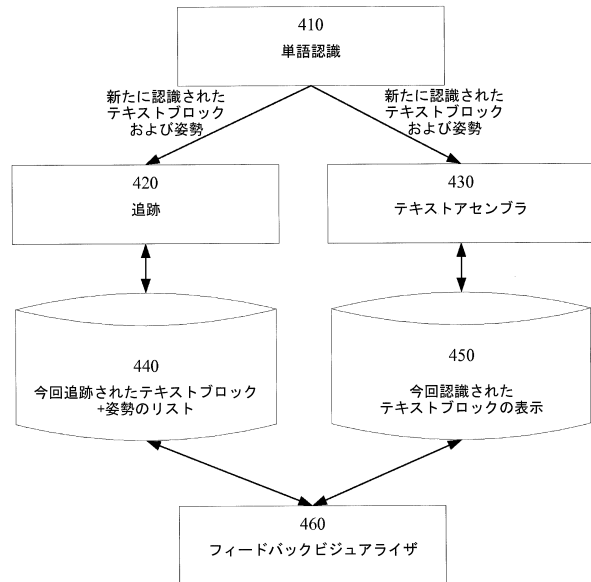
【図 2】



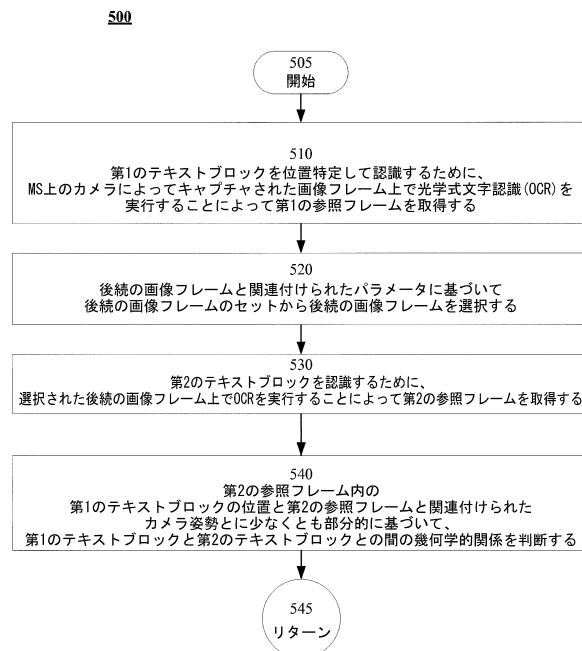
【図 3】



【図 4】



【図 5】



## フロントページの続き

- (72)発明者 ジエウン・キム  
アメリカ合衆国・カリフォルニア・92121-1714・サン・ディエゴ・モアハウス・ドライ  
ヴ・5775
- (72)発明者 ペール・オー・ニールセン  
アメリカ合衆国・カリフォルニア・92121-1714・サン・ディエゴ・モアハウス・ドライ  
ヴ・5775
- (72)発明者 ロイ・ローレンス・アショク・イニゴ  
アメリカ合衆国・カリフォルニア・92121-1714・サン・ディエゴ・モアハウス・ドライ  
ヴ・5775
- (72)発明者 チ・パン  
アメリカ合衆国・カリフォルニア・92121-1714・サン・ディエゴ・モアハウス・ドライ  
ヴ・5775
- (72)発明者 ロメイン・タロノー  
アメリカ合衆国・カリフォルニア・92121-1714・サン・ディエゴ・モアハウス・ドライ  
ヴ・5775

審査官 村松 貴士

- (56)参考文献 特開2013-250932(JP, A)  
特開2003-331217(JP, A)  
国際公開第2012/051040(WO, A1)  
米国特許出願公開第2005/0286743(US, A1)

(58)調査した分野(Int.Cl., DB名)

G06T	7/00	-	7/90
G06K	9/00	-	9/82