

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6105627号  
(P6105627)

(45) 発行日 平成29年3月29日 (2017.3.29)

(24) 登録日 平成29年3月10日 (2017.3.10)

(51) Int. Cl.		F I			
<b>G06T</b>	<b>7/00</b>	<b>(2017.01)</b>	<b>G06T</b>	<b>7/00</b>	<b>300Z</b>
<b>G06K</b>	<b>9/68</b>	<b>(2006.01)</b>	<b>G06K</b>	<b>9/68</b>	<b>E</b>

請求項の数 15 (全 36 頁)

(21) 出願番号	特願2014-551409 (P2014-551409)	(73) 特許権者	507364838
(86) (22) 出願日	平成25年1月8日 (2013.1.8)		クアルコム、インコーポレイテッド
(65) 公表番号	特表2015-504220 (P2015-504220A)		アメリカ合衆国 カリフォルニア 921
(43) 公表日	平成27年2月5日 (2015.2.5)		21 サン ディエゴ モアハウス ドラ
(86) 国際出願番号	PCT/US2013/020699		イブ 5775
(87) 国際公開番号	W02013/106346	(74) 代理人	100108453
(87) 国際公開日	平成25年7月18日 (2013.7.18)		弁理士 村山 靖彦
審査請求日	平成27年12月8日 (2015.12.8)	(74) 代理人	100163522
(31) 優先権主張番号	61/584,765		弁理士 黒田 晋平
(32) 優先日	平成24年1月9日 (2012.1.9)	(72) 発明者	ヨンキー・ペク
(33) 優先権主張国	米国 (US)		アメリカ合衆国・カリフォルニア・921
(31) 優先権主張番号	13/736,527		21-1714・サン・ディエゴ・モアハ
(32) 優先日	平成25年1月8日 (2013.1.8)		ウス・ドライブ・5775
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 OCRキャッシュ更新

(57) 【特許請求の範囲】

【請求項1】

モバイルデバイスにおいて複数の画像を受信するステップと、  
 前記複数の画像の各々の中のデータを認識するステップであって、前記複数の画像に対して光学式文字認識を実施するステップを含む、認識するステップと、  
 前記認識されたデータに少なくとも部分的に基づいて、前記複数の画像のうち少なくとも2つの画像が、第1のシーンにある情報を含むと判断するステップと、  
各画像中で認識されたグラフィカルオブジェクトの数と、前記数のグラフィカルオブジェクトの各々の前記認識の信頼度とに基づいて、前記少なくとも2つの画像の各々のエネルギーを算出するステップと、  
 前記少なくとも2つの画像のうち少なくとも1つの画像、および前記少なくとも1つの画像に関連付けられた前記認識されたデータを使って、前記第1のシーンを表す最終画像を記憶するステップであって、前記少なくとも1つの画像は、前記少なくとも2つの画像の各々に関連付けられる前記エネルギーを比較することによって選択されるキー画像を含む、  
ステップと  
 を含む方法。

【請求項2】

受信する前記ステップが、カメラを使って前記複数の画像を順次キャプチャするステップを含む、請求項1に記載の方法。

【請求項3】

前記複数の画像のうち前記少なくとも1つの画像、および前記少なくとも1つの画像についての前記認識されたデータの記憶は、後続画像がキャプチャされ、前記後続画像中でデータが認識される間、維持される、請求項1に記載の方法。

【請求項4】

前記最終画像を記憶するのに先立って、前記方法が、前記少なくとも2つの画像と、前記少なくとも2つの画像に関連付けられた前記認識されたデータとを使って、前記最終画像を合成するステップをさらに含み、前記記憶は、前記合成された画像の記憶を含む、請求項1に記載の方法。

【請求項5】

合成する前記ステップが、前記少なくとも2つの画像のうち第2の画像にある要素を前記少なくとも1つの画像に組み込んで、前記最終画像を作成するステップを含む、請求項4に記載の方法。

【請求項6】

前記最終画像が前記少なくとも1つの画像を含み、合成する前記ステップが、前記少なくとも2つの画像のうち第2の画像の前記認識されたデータを、前記少なくとも1つの画像の前記認識されたデータに組み込むステップを含む、請求項4に記載の方法。

【請求項7】

判断する前記ステップが、前記少なくとも2つの画像の各々におけるグラフィカルオブジェクト値、グラフィカルオブジェクト色、背景色および前景色のうち1つまたは複数と比較するステップを含む、請求項1に記載の方法。

【請求項8】

前記複数の画像のうち1つまたは複数の画像が、前記第1のシーン以外の1つまたは複数のシーンにある情報を含むと判断するステップと、前記最終画像、および前記1つもしくは複数のシーンをそれぞれが表す1つもしくは複数の画像を表示する、または前記少なくとも1つの画像についての、および前記1つもしくは複数のシーンについての前記認識されたデータを表示するステップとをさらに含み、請求項1に記載の方法。

【請求項9】

ユーザが前記表示された画像のうち1つまたは前記1つの画像についての表示される認識データを選択したと判断するステップと、前記選択された画像または認識されたデータに関連した情報をプログラムまたはアプリケーションに入力するステップとをさらに含み、請求項8に記載の方法。

【請求項10】

前記最終画像を記憶するのに先立って、前記方法が、

前記第1のシーンにある2つ以上の画像のうちの第1の画像から、グレアのある第1の領域を識別するステップと、

前記第1のシーンにある前記2つ以上の画像のうちの第2の画像から、グレアのない第2の領域を識別するステップであって、前記第1の領域および前記第2の領域が、前記第1のシーンの同じ領域を表す、ステップと、

前記第1の画像、および前記第2の画像からのグレアのない前記第2の領域を使って、グレアが低減された前記最終画像を合成するステップとをさらに含み、請求項1に記載の方法。

【請求項11】

プロセッサによって実行可能な命令を含む非一時的コンピュータ可読記憶媒体であって、前記命令が、

複数の画像を受信するための命令と、

前記複数の画像の各々の中のデータを認識するための命令であって、前記複数の画像の各々の中のデータを認識することが、前記複数の画像に対して光学式文字認識を実施することを含み、命令と、

前記認識されたデータに少なくとも部分的に基づいて、前記複数の画像のうち少なくとも2つの画像が、第1のシーンにある情報を含むと判断するための命令と、

10

20

30

40

50

各画像中で認識されたグラフィカルオブジェクトの数と、前記数のグラフィカルオブジェクトの各々の前記認識の信頼度とに基づいて、前記少なくとも2つの画像の各々のエネルギーを算出するための命令と、

前記少なくとも2つの画像のうち少なくとも1つの画像、および前記少なくとも1つの画像に関連付けられた前記認識されたデータを使って、前記第1のシーンを表す最終画像を記憶するための命令であって、前記少なくとも1つの画像は、前記少なくとも2つの画像の各々に関連付けられる前記エネルギーを比較することによって選択されるキー画像を含む、命令とを含む、非一時的コンピュータ可読記憶媒体。

【請求項12】

前記複数の画像のうち前記少なくとも1つの画像、および前記少なくとも1つの画像についての前記認識されたデータの記憶は、後続画像がキャプチャされ、前記後続画像中でデータが認識される間、維持される、請求項11に記載の非一時的コンピュータ可読記憶媒体。

【請求項13】

前記最終画像を記憶するのに先立って、前記命令が、前記少なくとも2つの画像と、前記少なくとも2つの画像に関連付けられた前記認識されたデータとを使って、前記最終画像を合成するための命令を含み、前記記憶は、前記合成された画像の記憶を含む、請求項11に記載の非一時的コンピュータ可読記憶媒体。

【請求項14】

モバイルデバイスにおいて複数の画像を受信するための手段と、

前記複数の画像の各々の中のデータを認識するための手段であって、前記複数の画像の各々の中のデータを認識することが、前記複数の画像に対して光学式文字認識を実施することを含む、手段と、

認識されたデータに少なくとも部分的に基づき、前記複数の画像のうちの少なくとも2つの画像が第1のシーンにある情報を含むことを判断するための手段と、

各画像中で認識されたグラフィカルオブジェクトの数と、前記数のグラフィカルオブジェクトの各々の前記認識の信頼度とに基づいて、前記少なくとも2つの画像の各々のエネルギーを算出するための手段と、

前記少なくとも2つの画像のうち少なくとも1つの画像、および前記少なくとも1つの画像に関連付けられた前記認識されたデータを使って、前記第1のシーンを表す最終画像を記憶するための手段であって、前記少なくとも1つの画像は、前記少なくとも2つの画像の各々に関連付けられる前記エネルギーを比較することによって選択されるキー画像を含む、手段とを備える装置。

【請求項15】

前記複数の画像のうち前記少なくとも1つの画像、および前記少なくとも1つの画像についての前記認識されたデータの記憶は、後続画像がキャプチャされ、前記後続画像中でデータが認識される間、維持される、請求項14に記載の装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、OCRキャッシュ更新に関する。

【背景技術】

【0002】

光学式文字認識(OCR)は、手で書かれた、タイプライタで打たれた、または印刷されたテキスト、グラフィックスまたはシンボルを走査した画像を機械的に符号化されたテキストに機械的または電子的に変換することである。OCRは、コンピュータビジョンおよび拡張現実の適用における研究の成長分野である。コンピュータビジョンは、デバイスが、その近傍の環境を知覚することを可能にする。コンピュータビジョンは、エンドユーザの実世界との対話による感覚入力が仮想世界に及ぶことを可能にすることによって、拡張現実における適用を可能にする。実生活の例示的な用途では、コンピュータビジョン対応OC

10

20

30

40

50

Rは、視覚障害者が、スマートフォンなど、個人用モバイルデバイスのカメラの視野の中の書かれたテキストおよびシンボルを知覚するのを助け得る。

【発明の概要】

【課題を解決するための手段】

【0003】

本明細書に記載する技法は、拡張現実の適用においてモバイルデバイスを使ってOCRを実施しながら、OCRキャッシュを自動的および知的に作成および/または更新するための方法を提供する。モバイルデバイスに結合されたカメラを使ってキャプチャされた画像は、OCRキャッシュに記憶された前の画像と照合される。一致が見つかった場合、OCRキャッシュは、新規画像を別個のエントリとしてOCRキャッシュに記憶するのではなく、新規画像を使用して、新規および/またはより優れた情報で更新される。代替として、新規画像は、OCRキャッシュ中の前の画像と置き換わり得る。さらに、複数の画像を使って画像からグレアを低減するための技法について記載する。たとえば、画像用の領域からのグレアは、キー画像と同じ領域に対する、グレアをもたない同じシーンからの別の画像にある情報を使うことによって削除することができる。一実施形態では、グレアは、OCRを行わない場合でも、画像から低減または除去することができる。一実施形態では、グレア低減およびOCRキャッシュは、組み合わせて実施することができる。

【0004】

知的および自動OCRキャッシュ更新は、同じシーンに関連付けられた情報すべてにユーザがアクセスできるようにし、かつ/または複数のシーンに関連付けられた情報にユーザがアクセスできるようにし得るので、ユーザにとって有利である。さらに、長い時間にわたる同じシーンについての複数の画像の間に、画像情報とのオーバーラップがあるので、本明細書に記載する技法は、シーンに関連付けられた最良または新たに入手された情報を記憶すればよく、したがって、時間をかけて結果をさらに精製する。本明細書に記載する技法に関連した他の利点は、OCR結果を記憶するためのメモリ空間の削減、およびOCRキャッシュに対するヒット数の削減を含むことができ、これにより、全体的システムにおける性能がより速くなる。

【0005】

例として、ユーザが、レストランで席につき、スナップショットモードまたはプレビューモードを使って画像を取得するための従来の手段を使って、ディナーメニューに対してOCRを実施している場合、ユーザは、複数の画像を取得し、それらの画像に関連付けられた情報を別個に処理する必要がある。ユーザが、ディナーメニューから遠く離れた画像を取得するカメラを持っている場合、その画像は、画像に対してOCRを適切に実施するのに十分な解像度をもたない可能性がある。一方、ユーザが、画像を取得するカメラをあまりにも近くにもってきた場合、画像は、関心エリア全体を含めることができない。ただし、自動および知的OCRキャッシュ更新を用いるOCRキャッシングモードを使って、本明細書に記載する技法により、ユーザは、ユーザにとっての関心エリアを含む単一の画像を、適正に焦点を合わせ、良好な解像度でメニューから生成することができる。ある事例では、ユーザは、メニューからテキストを選択し、別の言語に翻訳し、またはオンラインサーチエンジンを使って探索を実施することを選ぶことができる。

【0006】

上記の例では、シーンにある画像のうち1つは、画像の領域に関連付けられたグレアを有し得る。記載する技法は、グレアのある第1の領域を第1の画像から識別し、グレアのない第2の領域を第2の画像から識別することによって、グレアを除去するのを容易にすることができ、ここで第1の領域および第2の領域は、同じシーンにある同じ領域を表す。第1の画像と、第2の画像からの、グレアのない第2の領域を使って、グレアを低減するように最終画像が合成される。

【0007】

本発明の実施形態を実施するための例示的な方法は、複数の画像を受信するステップと、複数の画像の各々中のデータを認識するステップであって、複数の画像に対して光学式

10

20

30

40

50

文字認識を実施するステップを含む、認識するステップと、認識されたデータに少なくとも部分的に基づいて、複数の画像のうち少なくとも2つの画像が第1のシーンにある情報を含むと判断するステップと、少なくとも2つの画像のうち少なくとも1つの画像および少なくとも1つの画像に関連付けられた認識データを使って、第1のシーンを表す最終画像を記憶するステップとを含み得る。一実施形態では、受信するステップは、カメラを使って複数の画像を順次キャプチャするステップを含む。一実装形態では、複数の画像のうち少なくとも1つの画像、および少なくとも1つの画像についての認識データの記憶は、後続画像がキャプチャされ、後続画像中でデータが認識される間、維持される。一態様では、複数の画像のうち少なくとも2つの画像が、第1のシーンにある情報を含むと判断するステップは、2つ以上の画像の各々の中のグラフィカルオブジェクト値、グラフィカルオブジェクト色、背景色および前景色のうち1つまたは複数と比較するステップを含み得る。

10

#### 【0008】

一実施形態では、最終画像を記憶するのに先立って、本方法は、少なくとも2つの画像および少なくとも2つの画像に関連付けられた認識データを使って最終画像を合成するステップと、次いで、その後で最終画像を記憶するステップとをさらに含む。一実装形態では、合成するステップは、少なくとも2つの画像のうち第2の画像にある要素を少なくとも1つの画像に組み込んで、最終画像を作成するステップを含む。別の実施形態では、最終画像は少なくとも1つの画像を含み、合成するステップは、少なくとも2つの画像のうち第2の画像の関連データを、少なくとも1つの画像の関連データに組み込むステップを含む。

#### 【0009】

20

別の実施形態では、最終画像を記憶するのに先立って、本方法は、少なくとも2つの画像から、最終画像として第1のシーンを表すための少なくとも1つの画像を選択するステップと、次いで、その後で最終画像を記憶するステップとをさらに含む。選択するステップは、2つ以上の画像の各々のエネルギーを比較するステップを含み得る。2つ以上の画像の各々のエネルギーの算出は、各画像中で認識されるいくつかのグラフィカルオブジェクトと、いくつかのグラフィカルオブジェクトの各々の認識についての信頼度とに基づき得る。

#### 【0010】

本方法は、複数の画像のうち1つまたは複数の画像が、第1のシーン以外の1つまたは複数のシーンにある情報を含むと判断するステップと、最終画像、ならびに1つもしくは複数のシーンまたは少なくとも1つの画像についての、および1つもしくは複数のシーンについての認識データをそれぞれが表す1つまたは複数の画像を表示するステップとをさらに含み得る。さらに、本発明の実施形態によって実施される方法はまた、ユーザが表示画像のうち1つまたは1つの画像についての表示される認識データを選択したと判断するステップと、選択された画像または認識されたデータに関連した情報をプログラムまたはアプリケーションに入力するステップとを含み得る。

30

#### 【0011】

一実装形態では、本方法は、第1のシーンにある2つ以上の画像のうちの第1の画像からグレアのある第1の領域を識別すること、第1のシーンにある2つ以上の画像のうちの第2の画像からグレアのない第2の領域を識別することであって、第1の領域および第2の領域が、第1のシーンの同じ領域を表すこと、ならびに第1の画像と、第2の画像からの、グレアのない第2の領域とを使って、グレアが低減された最終画像を合成することによって、グレアを除去または低減することもできる。

40

#### 【0012】

本発明の実施形態を実施するための例示的なデバイスは、複数の画像を受信するように構成された画像キャプチャモジュールと、光学式文字認識を使って複数の画像の各々中のデータを認識するように構成されたOCRエンジンと、認識されたデータに少なくとも部分的に基づいて、複数の画像のうち少なくとも2つの画像が、第1のシーンにある情報を含むと判断するように構成された照合モジュールと、少なくとも2つの画像のうち少なくとも1つの画像および少なくとも1つの画像に関連付けられた認識データを使って、第1のシーン

50

を表す最終画像を記憶するように構成された統合モジュールとを含み得る。画像は、デバイスに結合されたカメラを使って順次キャプチャまたは入手され得る。一実施形態では、判断することは、2つ以上の画像の各々の中の、グラフィカルオブジェクト値、グラフィカルオブジェクト色、背景色、前景色および境界ボックスのうち1つまたは複数を比較することを含む。

【0013】

統合モジュールは、少なくとも2つの画像および少なくとも2つの画像に関連付けられた認識データを使って最終画像を合成し、次いで、その後で最終画像を記憶するようにさらに構成され得る。一実施形態では、合成することは、少なくとも2つの画像のうち第2の画像にある要素を少なくとも1つの画像に組み込んで、最終画像を作成することを含む。別の実施形態では、最終画像は少なくとも1つの画像を含み、合成することは、少なくとも2つの画像のうち第2の画像の関連データを、少なくとも1つの画像の関連データに組み込むことを含む。

10

【0014】

最終画像を記憶するのに先立って、選択モジュールは、少なくとも2つの画像から、最終画像として、第1のシーンを表すための少なくとも1つの画像を選択し、次いで、その後で最終画像を記憶するように構成され得る。一実施形態では、選択することは、2つ以上の画像の各々のエネルギーを比較することを含む得る。2つ以上の画像の各々のエネルギーの算出は、各画像中で認識されるいくつかのグラフィカルオブジェクトと、いくつかのグラフィカルオブジェクトの各々の認識についての信頼度とに基づく。

20

【0015】

デバイスは、少なくとも2つの画像または少なくとも2つの画像について認識されたデータを表示するように構成されたデバイスディスプレイも備え得る。デバイスは、表示画像または1つの画像についての表示される認識データのうち1つにユーザがタッチしたと判断すること、およびタッチした画像または認識されたデータに関連した情報をプログラムまたはアプリケーションに入力することをさらに含み得る。入力することは、インターネット照会を提出すること、翻訳を判断すること、またはテキストを入れることを含む得る。

【0016】

デバイスの一例では、最終画像を記憶するのに先立って、デバイスは、第1のシーンにある2つ以上の画像のうちの第1の画像からグレアのある第1の領域を識別し、第1のシーンにある2つ以上の画像のうちの第2の画像からグレアのない第2の領域を識別し、第1の領域および第2の領域が、第1のシーンの同じ領域を表し、ならびに第1の画像と、第2の画像からの、グレアのない第2の領域とを使って、グレアが低減された最終画像を合成するようにさらに構成される。

30

【0017】

例示的な非一時的コンピュータ可読記憶媒体であって、非一時的コンピュータ可読記憶媒体は、プロセッサによって実行可能な命令を備え、命令は、複数の画像を受信するための命令と、複数の画像の各々の中のデータを認識するための命令であって、複数の画像の各々の中のデータを認識することが、複数の画像に対して光学式文字認識を実施することを含む、命令と、認識されたデータに少なくとも部分的に基づいて、複数の画像のうち少なくとも2つの画像が第1のシーンにある情報を含むと判断するための命令と、少なくとも2つの画像のうち少なくとも1つの画像および少なくとも1つの画像に関連付けられた認識データを使って、第1のシーンを表す最終画像を記憶するための命令とを含む。

40

【0018】

例示的な装置は、複数の画像を受信するための手段と、複数の画像の各々の中のデータを認識するための手段であって、複数の画像の各々の中のデータを認識することが、複数の画像に対して光学式文字認識を実施することを含む、手段と、認識されたデータに少なくとも部分的に基づいて、複数の画像のうち少なくとも2つの画像が第1のシーンにある情報を含むと判断するための手段と、少なくとも2つの画像のうち少なくとも1つの画像および少なくとも1つの画像に関連付けられた認識データを使って、第1のシーンを表す最終画

50

像を記憶するための手段とを含み得る。複数の画像は、カメラを使って順次キャプチャされ得る。複数の画像のうち少なくとも1つの画像、および少なくとも1つの画像についての認識されたデータの記憶は、後続画像がキャプチャされ、後続画像中でデータが認識される間、維持され得る。一態様では、複数の画像のうち少なくとも2つの画像が、第1のシーンにある情報を含むと判断するステップは、2つ以上の画像の各々の中のグラフィカルオブジェクト値、グラフィカルオブジェクト色、背景色および前景色のうち1つまたは複数と比較するステップを含み得る。

【0019】

一実施形態では、最終画像を記憶するのに先立って、装置は、少なくとも2つの画像および少なくとも2つの画像に関連付けられた認識データを使って最終画像を合成し、次いで、その後で最終画像を記憶するための手段をさらに備え得る。別の実施形態では、最終画像を記憶するのに先立って、装置は、少なくとも2つの画像から、最終画像として、第1のシーンを表すための少なくとも1つの画像を選択し、次いで、その後で最終画像を記憶するための手段をさらに備え得る。選択するための手段は、2つ以上の画像の各々のエネルギーを比較するための手段を備え得る。2つ以上の画像の各々のエネルギーの算出は、各画像中で認識されるいくつかのグラフィカルオブジェクトと、いくつかのグラフィカルオブジェクトの各々の認識についての信頼度とに基づき得る。

【0020】

一実施形態では、最終画像を記憶するのに先立って、装置は、第1のシーンにある2つ以上の画像のうちの第1の画像から、グレアのある第1の領域を識別するための手段と、第1のシーンにある2つ以上の画像のうちの第2の画像から、グレアのない第2の領域を識別するための手段であって、第1の領域および第2の領域が、第1のシーンの同じ領域を表す、手段と、第1の画像、および第2の画像からの、グレアのない第2の領域を使って、グレアが低減された最終画像を合成するための手段とを含み得る。

【0021】

前記の説明では、次の詳細な説明を理解しやすくするために例の特徴および技術的利点の概略をかなり広範にわたって述べている。以下で、追加の特徴および利点について説明する。本開示の同じ目的を実施するために他の構造を修正または設計する際の基礎として、開示される概念および特定の実施例が容易に利用され得る。そのような等価な構造は、添付の特許請求の範囲の趣旨および範囲から逸脱しない。構成と動作方法の両方に関する、本明細書で開示される概念の特徴であると考えられる特徴は、関連する利点とともに、添付の図面とともに検討すると以下の説明からよりよく理解されよう。しかし、図面の各々は例示および説明のみのために与えられ、特許請求の範囲の限定範囲を定めるために与えられるものではない。

【0022】

本開示の態様は、例として説明される。次の説明は、図面を参照して提示されているが、図面全体を通して、同様の符号は同様の要素を参照するために使用される。本明細書で1つまたは複数の技法の様々な詳細を説明するが、他の技法も同様に可能である。いくつかの場合には、様々な技法の説明を容易にするために、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。

【0023】

明細書の残りの部分および図面を参照することによって、本開示によって与えられる実施例の性質および利点をさらに理解することができ、同様の参照番号が、同様の構成要素を指すために、いくつかの図面全体で用いられる。いくつかの場合には、複数の同様の構成要素の1つを指すために、サブラベルが参照番号と関連付けられる。存在するサブラベルを特定することなく、参照番号に言及する場合、参照番号はそのようなすべての同様の構成要素を指す。

【図面の簡単な説明】

【0024】

【図1】本発明の実施形態を実施する際に使用されるデバイスの部品を組み込んだ例示的

10

20

30

40

50

なコンピュータデバイスを示す図である。

【図2】経時的な画像のOCR処理のための、本発明の例示的な実施形態を示すフロー図である。

【図3A】画像のOCR処理のための本発明の非限定的な例示的な実施形態を示すフロー図である。

【図3B】画像のOCR処理のための本発明の別の非限定的な例示的な実施形態を示すフロー図である。

【図4】本発明の例示的な実施形態による、本の表紙および関連シーン記述子の代表的画像を示す2つのブロック図である。

【図5】記憶された画像と同じシーンを表す複数の画像からの画像の選択を示すブロック図である。

10

【図6】記憶された画像と同じシーンから画像を検出するための非限定的な例示的な方法を示すフロー図である。

【図7】複数の画像からの、キー画像としての画像の選択を示すブロック図である。

【図8】複数の画像からの、キー画像としての画像の選択のための非限定的な例示的な方法を示すフロー図である。

【図9】同じシーンにある複数の画像のうちの、最終画像としての画像の統合を示すブロック図である。

【図10】同じシーンにある複数の画像から最終画像を合成するための非限定的な例示的な方法を示すフロー図である。

20

【図11】本発明の実施形態によって提供される方法を実施するための例示的な構成要素を示すブロック図である。

【図12】本発明の実施形態によって提供される方法を実施するための例示的な構成要素を示す別のブロック図である。

【図13A】テキストを含む画像についての例示的なシーン記述子を示す図である。

【図13B】テキストを含む画像についての例示的なシーン記述子を示す図である。

【図14】画像からグレアを除去または低減するための本発明の実施形態を示す図である。

【図15】本開示の1つまたは複数の例示的な態様による、画像中の反射グレアを低減または除去するための本発明の実施形態を示すフロー図である。

30

【図16A】自動および知的OCRキャッシュ更新を用いてOCRキャッシングモードで動作する、過度に簡略化したモバイルデバイスを使うユーザインターフェースのための例示的な実装形態を示す図である。

【図16B】自動および知的OCRキャッシュ更新を用いてOCRキャッシングモードで動作する、過度に簡略化したモバイルデバイスを使うユーザインターフェースのための例示的な実装形態を示す図である。

【図16C】自動および知的OCRキャッシュ更新を用いてOCRキャッシングモードで動作する、過度に簡略化したモバイルデバイスを使うユーザインターフェースのための例示的な実装形態を示す図である。

【図16D】自動および知的OCRキャッシュ更新を用いてOCRキャッシングモードで動作する、過度に簡略化したモバイルデバイスを使うユーザインターフェースのための例示的な実装形態を示す図である。

40

【発明を実施するための形態】

【0025】

モバイルデバイスに結合されたカメラの視野内にテキストを備える画像は、いくつかの実装形態では、スナップショットモードまたはプレビューモードを使って入手され得る。スナップショットモードにおいて、ユーザは、視野からキャプチャする際にユーザが興味を示すテキストに、カメラの焦点を合わせ、写真を撮る。得られたピクチャは、コンピュータデバイスによって分析することができ、OCR結果が生成される。一方、プレビューモードは、視野内のテキストの認識を継続的にループさせ、入手された各画像またはフレー

50



ムに対してOCRを実施する。このモードは、いくつかの事例では、カメラビューファインダーの真横にある、デバイスが認識しているものの動的なリアルタイムの表示を示す。ユーザが、画像のOCR結果を吟味するのに、より多くの時間を費やし、または結果を使用したいとき、ユーザは、画像キャプチャ/OCR認識をストップすることができ、その後、凍結キャプチャ/認識によって現在表示されている結果で続行すればよい。

【0026】

当該技術分野アプリケーションにおいて実装されるスナップショットモードおよびプレビューモードは、自動履歴アーカイブのための機構をもたない。ユーザが、大きい入力ターゲットまたは多くのターゲットに直面すると、ユーザは、スナップショットモードまたはプレビューモードにより複数の画像を撮り、各画像に対してOCRエンジンを凍結し、各画像を別個に分析することに頼らなければならない。特にプレビューモードでは、ユーザが関心エリアの上でモバイルデバイスをゆっくりと動かす間、ユーザの所では、繰り返し撮られた同じシーンの多くの画像であふれてしまう。また、現行のOCRシステムは、異なるOCR結果からの情報を集約してOCR性能を改善し、または得られたキャッシュ画像中のグレアを低減する機構をもたない。

【0027】

本発明の実施形態は、これらおよび他の問題に対処する。

【0028】

図1は、本発明の実施形態を実施する際に使用されるデバイスの部品を組み込んだ例示的なコンピュータデバイスを示す。図1に示すコンピュータデバイスは、本明細書では、どのコンピュータ化されたシステムの一部として組み込まれてもよい。たとえば、コンピュータデバイス100は、モバイルデバイスの構成要素のいくつかを表現し得る。モバイルデバイスは、カメラ150など、1つまたは複数の入力知覚ユニットまたは入力デバイス115と、ディスプレイユニットやタッチスクリーンなど、1つまたは複数の入力/出力デバイスとをもつ、どのコンピューティングデバイス100であってもよい。モバイルデバイスの例には、限定はしないが、ビデオゲーム機、タブレット、スマートフォン、ラップトップ、ノートブック、または他のポータブルデバイスがある。図1は、本明細書で説明する様々な他の実施形態によって提供される方法を実行し得、かつ/あるいは、ホストコンピュータデバイス、リモートキオスク/端末、販売時点管理(point-of-sale)デバイス、モバイルデバイス、セットトップボックスおよび/またはコンピュータデバイスとして機能し得る、コンピュータデバイス100の一実施形態の概略図を与える。図1は、様々な構成要素の一般化された図を与えるものにすぎず、それらの構成要素のいずれかまたはすべては適宜に利用され得る。したがって、図1は個別のシステム要素が比較的個別化された様式または比較的一体化された様式でどのように実装され得るかをおおざっぱに示している。

【0029】

コンピュータデバイス100は、バス105を介して電氣的に結合され得る(または適宜、他の何らかの形で通信しているものとしてよい)ハードウェア要素を備えるように図示されている。ハードウェア要素には、1つもしくは複数の汎用プロセッサおよび/または1つもしくは複数の専用プロセッサ(デジタル信号処理チップ、グラフィックス高速化プロセッサなど)を限定なしに含む1つまたは複数のプロセッサ110と、カメラ、センサ(慣性センサを含む)、マウス、キーボードなどを限定なしに含む1つまたは複数の入力デバイス115と、図11および図12のデバイスディスプレイ(1118)などのディスプレイユニット、プリンタなどを限定なしに含む1つまたは複数の出力デバイス120とがあり得る。

【0030】

コンピュータデバイス100は、限定することなく、ローカルおよび/またはネットワークアクセス可能ストレージを含むことができ、かつ/あるいは限定することなく、ディスクドライブ、ドライブアレイ、光記憶デバイス、プログラム可能、フラッシュ書換可能であるものとしてよい、ランダムアクセスメモリ(「RAM」)および/もしくはリードオンリーメモリ(「ROM」)などのソリッドステート記憶デバイス、ならびに/または同様のものを含むことができる、1つまたは複数の非一時的な記憶デバイス125を、さらに備える(かつ/また

は通信する)ことができる。そのような記憶デバイスは、限定はしないが、様々なファイルシステム、データベース構造などを含む任意の適切なデータストレージを実装するように構成され得る。図11および図12にあるデバイスストレージ1120は、上述したように、1つまたは複数の非一時的記憶デバイスを使うことができる。

#### 【0031】

コンピュータデバイス100は、限定することなく、モデム、ネットワークカード(ワイヤレスもしくは有線)、赤外線通信デバイス、ワイヤレス通信デバイスおよび/またはチップセット(Bluetooth(登録商標)デバイス、802.11デバイス、WiFiデバイス、WiMaxデバイス、セルラー通信設備など)、ならびに/あるいは同様のものを含み得る、通信サブシステム130を備えることも可能である。通信サブシステム130は、データが、ネットワーク(一例を挙げると、以下で説明するネットワークなど)、他のコンピュータデバイス、および/または本明細書で説明する任意の他のデバイスと交換されることを可能にし得る。通信サブシステム130は、いくつかの実施形態では、たとえば他のデバイスと通信するように構成されないデバイスでは、省かれてもよい。多くの実施形態では、コンピュータデバイス100は、上記で説明したRAMまたはROMを含み得る非一時的作業メモリ135をさらに備える。OCRキャッシュ(1112)は、上述したように、非一時的作業メモリ135を使って実装することができる。

#### 【0032】

コンピュータデバイス100は、本明細書で説明するように、オペレーティングシステム140、デバイスドライバ、実行可能ライブラリ、および/または様々な実施形態によって提供されるコンピュータプログラムを含み得、かつ/もしくは方法を実装し、かつ/もしくは他の実施形態によって実現されるシステムを構成するように設計され得る、1つもしくは複数のアプリケーションプログラム145などの他のコードを含む、作業メモリ135内に現在配置されているものとして図示されている、ソフトウェア要素も備えることができる。単に例としてであるが、上で説明されている方法に関して説明されている1つまたは複数の手順は、コンピュータ(および/またはコンピュータ内のプロセッサ)によって実行可能なコードおよび/または命令として実装することも可能であり、次いで、一態様において、そのようなコードおよび/または命令は、説明されている方法に従って1つまたは複数のオペレーションを実行するように汎用コンピュータ(または他のデバイス)を構成し、および/または適合させるために使用され得る。

#### 【0033】

1組のこれらの命令および/またはコードは、上述の1つまたは複数の記憶デバイス125などのコンピュータ可読記憶媒体に記憶することができる。場合によっては、記憶媒体は、コンピュータデバイス100などのコンピュータデバイス内に組み込まれ得る。他の実施形態では、記憶媒体は、コンピュータデバイスから分離し(たとえば、コンパクトディスクなどの取り外し可能な媒体)、および/またはインストールパッケージで提供されることも可能であり、これにより、記憶媒体を使用して、そこに記憶されている命令/コードで汎用コンピュータをプログラムし、構成し、かつ/または適合させることができる。これらの命令は、コンピュータデバイス100によって実行可能である、実行可能コードの形態をとることも可能であり、かつ/またはコンピュータデバイス100でコンパイルおよび/もしくはインストールした後に(たとえば、様々な一般的に利用可能なコンパイラ、インストールプログラム、圧縮/展開ユーティリティなどのいずれかを使用して)、実行可能コードの形態をとるソースコードおよび/もしくはインストール可能コードの形態をとることも可能である。

#### 【0034】

大幅な変形が、特定の要件に従って作成され得る。たとえば、カスタマイズされたハードウェアもまた、使用されることがあり、および/または特定の要素が、ハードウェア、ソフトウェア(アプレットなどのポータブルソフトウェアを含む)、または両方で実装されることもある。さらに、ネットワーク入力/出力デバイスなどの他のコンピューティングデバイスへの接続が、使用され得る。

## 【 0 0 3 5 】

いくつかの実施形態は、本開示による方法を実施するために、コンピュータデバイス(コンピュータデバイス100など)を使用し得る。たとえば、説明する方法の一部または全部の手順は、作業メモリ135に含まれた(オペレーティングデバイス140および/またはアプリケーションプログラム145などの他のコードに組み込まれ得る)1つまたは複数の命令の1つまたは複数のシーケンスを実行するプロセッサ110にตอบสนองして、コンピュータシステム100によって実施され得る。そのような命令は、記憶デバイス125のうちの1つまたは複数のような別のコンピュータ可読媒体から作業メモリ135に読み込まれてよい。単に例として、作業メモリ135内に含まれる命令のシーケンスの実行では、プロセッサ110に、本明細書で説明する方法の1つまたは複数の手順を実施させ得る。

10

## 【 0 0 3 6 】

本明細で使用する「機械可読媒体」および「コンピュータ可読媒体」という用語は、機械を特定の方式で動作させるデータを与えることに関与する任意の媒体を指す。コンピュータデバイス100を使用して実装される一実施形態では、様々なコンピュータ可読媒体が、実行のためにプロセッサ110に命令/コードを与えることに関与し、かつ/またはそのような命令/コード(たとえば、信号)を記憶および/または搬送するために使用されることがある。多くの実装形態において、コンピュータ可読媒体は物理記憶媒体および/または有形記憶媒体である。そのような媒体は、限定はしないが、不揮発性媒体、揮発性媒体、および伝送媒体を含む多くの形態を取り得る。不揮発性媒体は、たとえば、記憶デバイス125などの光ディスクおよび/または磁気ディスクを含む。揮発性媒体は、限定はしないが、作業メモリ135などの動的メモリを含む。伝送媒体は、限定はしないが、バス105、ならびに通信サブシステム130(および/または通信サブシステム130が他のデバイスとの通信を提供する媒体)の様々な構成要素を備える線を含めて、同軸ケーブル、銅線、および光ファイバーを含む。したがって、伝送媒体はまた、(限定はしないが、無線波通信および赤外線データ通信中に生成されるような無線波、音波、および/または光波を含めて)波の形態を取り得る。

20

## 【 0 0 3 7 】

物理的および/または有形のコンピュータ可読媒体の一般的な形態として、たとえば、フロッピー(登録商標)ディスク、フレキシブルディスク、ハードディスク、磁気テープ、または他の任意の磁気媒体、CD-ROM、他の光媒体、パンチカード、紙テープ、何らかのパターンで穴がけられている他の物理的媒体、RAM、PROM、EPROM、FLASH-EPROM、他のメモリチップもしくはカートリッジ、これ以降説明するような搬送波、またはコンピュータが命令および/またはコードを読み出すことができる他の媒体が挙げられる。

30

## 【 0 0 3 8 】

様々な形態のコンピュータ可読媒体は、実行のためにプロセッサ110に1つまたは複数の命令の1つまたは複数のシーケンスを搬送することに関与し得る。単に例として、命令は、最初に、リモートコンピュータの磁気ディスクおよび/または光ディスク上で搬送され得る。リモートコンピュータは、命令をその動的メモリ内にロードし、コンピュータデバイス100によって受信されかつ/または実行される伝送媒体上の信号として、命令を送信し得る。電磁信号、音響信号、光信号などの形態であり得るこれらの信号は、すべて、本発明の様々な実施形態による、命令が符号化され得る搬送波の例である。

40

## 【 0 0 3 9 】

通信サブシステム130(および/またはそれらの構成要素)は、一般に、信号を受信し、次いでバス105は作業メモリ135に信号(および/またはその信号によって搬送されるデータ、命令など)を搬送し、作業メモリ135から、プロセッサ110は命令を検索して実行し得る。作業メモリ135によって受信された命令は、場合によっては、プロセッサ110による実行の前または後のいずれかに、非一時的記憶デバイス125に記憶され得る。

## 【 0 0 4 0 】

上記で論じた方法、システム、およびデバイスは、例である。様々な実施形態は、様々な手順または構成要素を、適宜、省略し、置換し、または加えることができる。たとえば

50

、代替構成では、説明する本方法は、説明する順序とは異なる順序で実施されてもよく、ならびに/または、様々なステージが加えられ、省略され、および/もしくは組み合わせられてもよい。また、いくつかの実施形態に関して説明する特徴が、様々な他の実施形態と組み合わせられてもよい。実施形態の様々な態様および要素を同様に組み合わせることができる。また、技術が発展し、したがって、要素の多くは、本開示の範囲を特定の例に限定しない例である。

#### 【0041】

実施形態の完全な理解を与えるために、説明に具体的な詳細が与えられる。しかしながら、実施形態は、これらの具体的な詳細なしに実践することができる。たとえば、実施形態を不明瞭にすることを避けるために、よく知られている回路、プロセス、アルゴリズム、構造、および技法は、不要な詳細なしに示してきた。この説明は、例示的な実施形態のみを提供し、本発明の範囲、適用可能性、または構成を限定しない。むしろ、これらの実施形態の上述の説明は、本発明の実施形態を実装することを可能にする説明を当業者に提供することになる。本発明の趣旨および範囲から逸脱することなく、要素の機能および構成に様々な変更を行うことができる。

#### 【0042】

また、いくつかの実施形態は、フロー図またはブロック図として図示されているプロセスとして説明する。各構成では動作を逐次プロセスとして記述し得るが、動作の多くを並行して実施するか、あるいは同時に実施することができる。さらに、動作の順番は並び替えられてもよい。プロセスは、図に含まれていない追加のステップを有することができる。さらに、本方法の実施形態は、ハードウェア、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語、またはそれらの任意の組合せによって実装され得る。ソフトウェア、ファームウェア、ミドルウェア、またはマイクロコードで実装されるとき、関連するタスクを実施するプログラムコードまたはコードセグメントは、記憶媒体などのコンピュータ可読媒体に記憶され得る。プロセッサは関連するタスクを実施し得る。

#### 【0043】

いくつかの例示的な構成について本明細書に説明してきたが、様々な変更、代替構成、および均等物は、本開示の趣旨から逸脱することなく使用され得る。たとえば、上記の要素は、単により大きいシステムの構成要素であり得、他のルールは、本発明の適用よりも優先するか、そうでなければ本発明の適用を変更することができる。また、要素が考慮される前、間、または後に、いくつかのステップを行うことができる。したがって、本明細書の説明は、本開示の範囲を制限しない。

#### 【0044】

本明細書で説明するように、画像にあるグラフィカルオブジェクトには、限定はしないが、1つまたは複数のシンボル、英数字、単語、標識、数、テキストまたはジェスチャが含まれ得る。たとえば、画像中のグラフィカルオブジェクトは、任意の言語における単語または文であってよい。単語は、いくつかの言語におけるシンボルとして表現されることがある。同様に、1つの言語または文化における文は、単に、別の言語または文化におけるシンボルを使用して表現されることがある。別の例では、グラフィカルオブジェクトは、STOP、YIELDなどの道路標識であってよい。異なるグラフィカルオブジェクトは、同じ意味を表現するために異なる地方、言語、または文化において使用されることがある。

#### 【0045】

図2は、経時的に画像を処理するための、本発明の例示的な実施形態を示すフロー図である。図1を参照して上で論じたコンピュータデバイス100は、図2に記載する本発明の実施形態を実施するために使われるモバイルデバイスの構成要素のうちいくつかを表し得る。図11で論じるモジュールおよび構成要素は、図1で論じたコンピューティングデバイス100の構成要素として実装することができ、図2で論じる本発明の実施形態を実施する際に使うことができる。後で論じるモジュールおよび構成要素は、ソフトウェア、ハードウェア、ファームウェアまたはそれらのどの組合せを使っても実施することができる。一態様

では、本明細書でOCRキャッシングモードと呼ぶモードにおいて、コンピュータデバイス100に結合されたカメラは、時間をかけて画像を継続的に入手する。

【0046】

図2において、複数の画像が1(202)で表され、各1202は、異なる受信画像を表す。一実施形態では、画像キャプチャモジュール1102は、コンピューティングデバイス100に結合された1つまたは複数のカメラ150を使って画像を入手または受信することができる。本発明の態様において、画像が受信されると、OCRエンジン206(図11の1104)は、各画像についてのOCR結果210を生じる画像を処理して、複数の画像の各々の中のデータを認識することができる。OCRエンジン206は、ハードウェア、ソフトウェアまたはファームウェアで実装することができる。ソフトウェア/ファームウェアで実装されるOCR用の命令は、作業メモリ135に一時的に記憶し、コンピューティングデバイス100の1つまたは複数のプロセッサ110上で実行することができる。いくつかの実施形態では、OCRエンジン206によるOCR処理の一部または全部は、ネットワークを介してコンピューティングデバイス100に接続されたリモートサーバ(図示せず)によって実施することができる。いくつかの実施形態では、記憶された結果の画像品質は、時間とともに改善される。たとえば、一実施形態では、グレアは、OCRの実施に加え、またはOCRの実施によらずに、画像から除去することができる。OCRキャッシングモードでは、時間をかけてスナップショットモードを使って入手された画像は、画像をOCRキャッシュに知的および自動的に記憶するのに、本明細書に記載する技法を用いることもできる。OCRキャッシュは、記憶デバイス125、作業メモリ135、ハードウェアレジスタ(図示せず)もしくはバッファを使って、またはそれらのどの組合せでも実装することができる。

【0047】

OCR結果は、画像の特性を記述する、画像についての追加データからなり得る。一実施形態では、シーン記述子生成器モジュール1106は、画像についてのシーン記述子として、画像についてのOCR結果を生成し、OCRキャッシュに一時的に記憶する。シーン記述子モジュール1106がシーン記述子を生成した後、照合モジュール1108など、コンピューティングデバイス100の構成要素は、照合プロセスを使うことによって、同じシーンに関連付けられた画像を検出することができる。一実施形態では、照合モジュールは、照合プロセス中に、画像に関連付けられたシーン記述子を比較することができる。図2において、OCR結果212および214からの認識データはシーンAに関連付けられ、OCR結果216はシーンBに関連付けられ、OCR結果218および222はシーンCに関連付けられ、OCR結果220はシーンDに関連付けられる。

【0048】

212~222のOCR結果および関連画像すべてをOCRキャッシュ224に記憶するのではなく、OCRキャッシングシステムは、各シーンについて、ただ1つの精製画像およびOCR結果を知的および自動的に記憶してもよい。選択モジュール1110など、コンピューティングデバイス100の構成要素は、シーンにある複数の画像から、OCRキャッシュ224に記憶すべき画像を選択することができる。選択モジュール1110によって画像を選択するプロセスは、選択プロセスと呼ばれ得る。たとえば、図2において、シーンAを表すOCR結果212および214に対応する画像は、A'(228)としてOCRキャッシュ224に記憶することができる。最終的な合成または選択画像およびシーン記述子は、212および214に対応する画像を使って、OCRキャッシュ224に記憶される。一態様では、選択モジュールは、選択プロセスの一部として、各画像に関連付けられたエネルギーを算出する。画像に関連付けられたエネルギーは、画像に関連付けられた明瞭度および解像度を表し得る。2つの画像(212および214に対応する)のうち、より高いエネルギーをもつ画像および関連シーン記述子は、A'(228)としてOCRキャッシュ224に記憶することができる。OCRキャッシュ224は、図1を参照して上述したように、非一時的作業メモリ135を使って実装することができる。

【0049】

別の実装形態では、統合モジュール1114は、212と214との間の異なる画像品質およびOCR結果を使って、キー画像に欠けている画像データを追加するとともに、最終的なA'(228)

を合成することができる。最終結果を合成するプロセスは、統合プロセスと呼ばれ得る。一例では、シーンにある画像のうち1つ(たとえば、212に対応する)は、画像の領域に関連付けられたグレアを有し得る。統合プロセスは、第1の領域および第2の領域がシーンAの同じ領域を表す、212に対応する画像から、グレアのある第1の領域を識別し、214に対応する画像から、グレアのない第2の領域を識別し、212に対応する画像と、214に対応する画像にある、グレアのない第2の領域とを使って、グレアを低減するように最終画像を合成することによって、グレアを低減するのを容易にし得る。

#### 【0050】

同様に、218および222に対応する画像は、同じシーンCに関連付けられているものと認識され、最終画像およびシーン記述子C'(232)を形成するように合成される。本明細書に記載する画像の合成は、OCRキャッシュ内の画像または画像についてのシーン記述子もしくは画像の表現を含む任意の関連データの更新または置き換えを含み得るが、それに限定されない。したがって、いくつかの実施形態では、画像の合成または複数の画像の合成は、第1の画像に関連付けられたデータを、第1の画像の外見または視覚表現を変更も修正もせずに、たとえば第2の画像にあるデータで(または複数の画像で)更新または修正することを含み得る。216に対応する画像は、シーンBを表す唯一の画像である。216に対応する画像が、シーンBを表す唯一の画像であると、照合モジュール1108が判断すると、挿入モジュール1116は、216に対応する画像を、OCRキャッシュ224内のキャッシュロケーション230に記憶してよい。同様に、220に対応する画像は、シーンDを表す唯一の画像であり、OCRキャッシュ224内でキャッシュロケーション234に記憶することができる。

#### 【0051】

受信画像は、一時バッファ、OCRキャッシュ224または他の適切な記憶媒体に記憶された同じシーンに関連付けられた他の画像を有するシーンに関連付けられ得る。最近受信した画像が、同じシーンについての、OCRキャッシュ224に記憶された画像およびOCR結果と一致する場合、OCRキャッシュ224内の情報は、最も最近受信した画像の結果を統合した後、そのシーンについて、統合モジュール1114によって更新することができる。

#### 【0052】

一態様では、画像は、受信されると直ちに処理される。別の実施形態では、入手された画像は、一時的に記憶され、何回かに分けて一緒に処理され、OCRキャッシュ更新をより少なくすることができる。さらに別の実施形態では、スライディングウィンドウを使って、特定の時間ウィンドウ内の画像を一緒に処理することができる。処理は、モバイルデバイス、たとえばコンピューティングデバイス100において実施することができ、または情報は、処理のためにリモートサーバまたはプロセッサに送信されてよい。たとえば、一実装形態では、コンピューティングデバイス100は、複数の画像を受信し、画像または画像に関連した情報を、OCR、照合プロセス、選択プロセス、および/または統合プロセスを実施するためにリモートサーバに送信することができる。

#### 【0053】

この知的および自動OCRキャッシュ更新は、同じシーンまたは複数のシーンに関連付けられた情報すべてにユーザがアクセスできるようにするので、ユーザにとって有利である。さらに、長い時間にわたる同じシーンについての複数の画像の間に、画像情報とのオーバーラップがあるので、本明細書に記載する技法は、シーンに関連付けられた最良の情報または新たに取得された情報を記憶すればよく、したがって、時間とともに結果をさらに精製する。本明細書に記載する技法に関連付けられた他の利点は、OCR結果を記憶するために必要とされるメモリ空間の削減、およびOCRキャッシュに対するヒット数の削減を含んでよく、それにより全体的システムにおける性能がより速くなる。

#### 【0054】

例として、ユーザが、レストランで席につき、スナップショットモードまたはプレビューモードを使って画像を取得するための従来の手段を使って、ディナーメニューに対してOCRを実施している場合、ユーザは、複数の画像を別個に取得し、それらの画像の各々に関連付けられた情報を連続して処理する必要がある。ユーザが、ディナーメニューが

ら遠く離れた画像を取得するカメラを持っている場合、その画像は、画像に対してOCRを適切に実施するのに十分な解像度をもたない可能性がある。一方、ユーザが、画像を取得するカメラをあまりにも近くにもってきた場合、画像は、関心エリア全体を含めることができない。ただし、自動および知的OCRキャッシュ更新を用いるOCRキャッシングモードでの本発明の態様を使って、本明細書に記載する技法により、ユーザは、ユーザにとっての関心エリアを含む単一の画像を、適正に焦点を合わせ、良好な解像度でメニューから生成することができる。ある事例では、ユーザは、メニューからテキストを選択し、別の言語に翻訳し、または所有もしくはオンラインサーチエンジンを使って探索を実施することを選ぶことができる。いくつかの実施形態では、ユーザは、アプリケーションまたはプログラム、たとえばテキストメッセージ通信もしくはeメールプログラムまたはワードプロセッサに入れるべきテキストを、メニューから選択することができる。

10

## 【0055】

図3Aは、本開示の1つまたは複数の例示的態様による、画像の処理のための本発明の非限定的な例示的实施形態を示すフロー図を示す。1つまたは複数の態様によれば、本明細書で説明されている方法および/または方法ステップのどれか、および/またはすべては、たとえば、コンピューティングデバイス100および/または図1でより詳しく説明されているデバイスなどの、コンピューティングデバイス100によって、および/またはコンピューティングデバイス100内に実装され得る。一実施形態では、図3Aに関して以下で説明されている方法ステップのうちの1つまたは複数の、プロセッサ110または別のプロセッサなどのデバイス100のプロセッサによって実装される。図11および図12で論じるモジュールおよび構成要素は、コンピューティングデバイス100の構成要素として実装することもでき、図3Aで論じる本発明の実施形態を実施する際に使うことができる。それに加えて、または代替的に、本明細書で説明されている方法および/またはステップのどれか、および/またはすべては、メモリ135、ストレージ125または別のコンピュータ可読媒体などのコンピュータ可読媒体上に記憶されているコンピュータ可読命令などのコンピュータ可読命令で実装され得る。

20

## 【0056】

図3Aにおいて、図1を参照して上で論じたコンピューティングデバイス100ならびに図11および図12で論じるモジュールおよび構成要素は、本発明の実施形態を実施するために使われるデバイスのモジュールおよび構成要素のいくつかを表し得る。ステップ302で、コンピューティングデバイス100に結合されたカメラ150は、複数の画像を受信または入手する。カメラ150は、図1に記載した入力デバイス115のうち1つを表し得る。一実装形態では、図11にある画像キャプチャモジュール1102は、画像を入手するために使うことができる。複数の画像は、いくつかの実施形態では順次受信され得る。他の実施形態では、たとえば、複数のカメラ150またはステレオカメラがコンピューティングデバイス100に結合されているとき、画像は同時に受信され得る。ステップ304で、受信画像および画像に関連付けられたデータは、互いと、ならびにOCRキャッシュ224に記憶されたOCR結果および画像とも比較されて、複数の画像が同じシーンに関連付けられているかどうか判断される。一実装形態では、OCRキャッシュは、図1に記載した作業メモリ135から動作する。

30

## 【0057】

ブロック306で、同じシーンに関連付けられた複数の画像が検出された場合、これらの画像は、統合プロセスを通して合成画像を形成するように統合または合成され得る。一態様では、選択プロセスが統合プロセスに先行する。選択プロセス中、シーンを表すキー画像を、シーンを表す複数の画像から選択することができ、複数の画像のうちの1つまたは複数の、コンピューティングデバイス100に記憶されてよい。キー画像は、各画像に関連付けられたエネルギーを算出することによって判断することができる。一実装形態では、各画像に関連付けられたエネルギーは、OCRエンジンを使って認識された単語の数と、各画像についてのOCR結果に関連付けられた信頼度とに正比例する。一実施形態では、キー画像は、OCRキャッシュに記憶されている画像であってよい。他の実装形態では、画像をOCRキャッシュに記憶する前に、キー画像は、同じシーンにある様々な画像からの情報を使

40

50

って最終画像を合成することによって強調または補強され得る。たとえば、キー画像中に表示される単語が判読可能でも検出可能でもない場合、同じシーンにある他の画像からの情報を使って、画像および画像に関連付けられたデータを強調すればよい。別の例では、キー画像の領域からのグレアは、キー画像と同じ領域に対するグレアをもたない同じシーンにある別の画像からの情報を使うことによって除去することができる。上述したように、グレアは、OCRを行わない場合でも、画像から低減または除去することができる。様々な画像にある情報は、OCRキャッシュまたは別のキャッシュ、たとえばOCR情報を含んでも含まなくてもよい画像キャッシュに記憶することができる最終画像を合成する際に使うことができる。たとえば、画像キャッシュは、グレアを低減された、最終画像を記憶するのに使うことができる。そのような画像キャッシュは、OCR情報を省いてもよく、いくつかの実施形態ではOCR情報を含んでもよい。

10

#### 【0058】

ブロック308で、コンピューティングデバイス100によって最終画像が合成されると、その画像は、OCRキャッシュに記憶することができる。いくつかの事例では、OCRキャッシュ中の既存の画像およびOCR結果は、更新され、または置き換えられる。たとえば、受信または入手された画像が、OCRキャッシュに記憶されているものと同じだが、OCRキャッシュに記憶されている画像よりも良好な解像度および明瞭度をもつシーンに関連付けられている場合、OCRキャッシュ中の画像およびOCR結果は、新たに受信された画像に関連付けられた情報を反映するように更新され、または置き換えられてよい。本明細書に記載する画像の合成は、OCRキャッシュ内の画像または画像についてのシーン記述子もしくは画像の表現の部分を含む任意の関連データの更新または置き換えを含み得るが、それに限定されない。

20

#### 【0059】

図3Aに例示されている特定のステップは、本発明の一実施形態による、動作モードを切り替える特定の方法を構成することは理解されるであろう。ステップの他の順序も、代替的な実施形態に従って実施され得る。たとえば、本発明の代替的な実施形態は、異なる順序で、上で概説されたステップを実施することができる。説明のために、ユーザは、第3の動作モードから第1の動作モードに、第4のモードから第2のモードに、またはそれらの間の任意の組合せで変化させることを選択することができる。さらに、図3Aに例示されている個別のステップは、個別のステップに適している通りに様々なシーケンスで実施され得る複数のサブステップを含み得る。さらに、追加のステップが、具体的な用途に応じて、追加または削除され得る。当業者であれば、プロセスの多くの変更形態、修正形態、および代替的形態を認識し、理解するであろう。

30

#### 【0060】

図3Bは、本開示の1つまたは複数の例示的態様による、画像の処理のための本発明の別の非限定的な例示的实施形態を示すフロー図を示す。1つまたは複数の態様によれば、本明細書で説明されている方法および/または方法ステップのどれか、および/またはすべては、たとえば、コンピューティングデバイス100および/または図1でより詳しく説明されているデバイスなどの、コンピューティングデバイス100によって、および/またはコンピューティングデバイス100内に実装され得る。一実施形態では、図3Bに関して以下で説明されている方法ステップのうちの1つまたは複数は、プロセッサ110または別のプロセッサなどのデバイス100のプロセッサによって実装される。図11および図12で論じるモジュールおよび構成要素は、コンピューティングデバイス100の構成要素として実装することもでき、図3Bで論じる本発明の実施形態を実施する際に使うことができる。それに加えて、または代替的に、本明細書で説明されている方法および/またはステップのどれか、および/またはすべては、メモリ135、ストレージ125または別のコンピュータ可読媒体などのコンピュータ可読媒体上に記憶されているコンピュータ可読命令などのコンピュータ可読命令で実装され得る。

40

#### 【0061】

図3Bにおいて、図1を参照して上で論じたコンピューティングデバイス100ならびに図11

50



および図12で論じるモジュールおよび構成要素は、本発明の実施形態を実施するために使われるデバイスのモジュールおよび構成要素のいくつかを表し得る。ステップ310で、コンピューティングデバイス100に結合されたカメラ150は、複数の画像を受信または入手する。カメラ150は、図1に記載した入力デバイス115のうち1つを表し得る。一実施形態では、図11にある画像キャプチャモジュール1102は、画像を入手するために使うことができる。複数の画像は、いくつかの実施形態では順次受信され得る。他の実施形態では、たとえば、複数のカメラ150またはステレオカメラがコンピューティングデバイス100に結合されているとき、画像は同時に受信され得る。

【0062】

ステップ312で、OCRエンジン1104または/およびシーン記述子モジュール1106など、コンピューティングデバイス100の構成要素は、複数の画像の各々の中のデータを認識することができ、複数の画像の各々の中のデータを認識することは、複数の画像に対して光学式文字認識を実施することを含む。データを認識する際にOCRエンジン1104およびシーン記述子1106によって実施されるステップについては、図4および図11においてより詳しく記載する。

【0063】

ステップ314で、照合モジュール1108など、コンピューティングデバイス100の構成要素は、認識されたデータに少なくとも部分的に基づいて、複数の画像のうち少なくとも2つの画像が、第1のシーンにある情報を含むと判断し得る。照合モジュール1108によって実施されるステップについては、図5、図6、図7、図8および図11においてより詳しく記載する。

【0064】

ステップ316で、統合モジュール1114など、コンピューティングデバイス100の構成要素は、少なくとも2つの画像のうち少なくとも1つの画像と、少なくとも1つの画像に関連付けられた認識データとを使って、第1のシーンを表す最終画像を記憶することができる。一実施形態において、同じシーンに関連付けられた複数の画像が検出された場合、これらの画像は、統合プロセスを通して合成画像を形成するように統合または合成され得る。統合モジュール1114によって実施されるステップについては、図8、図9、図10および図11においてより詳しく記載する。

【0065】

図3Bに例示されている特定のステップは、本発明の一実施形態による、動作モードを切り替える特定の方法を構成することは理解されるであろう。ステップの他の順序も、代替的な実施形態に従って実施され得る。たとえば、本発明の代替的な実施形態は、異なる順序で、上で概説されたステップを実施することができる。説明のために、ユーザは、第3の動作モードから第1の動作モードに、第4のモードから第2のモードに、またはそれらの間の任意の組合せで変化させることを選択することができる。さらに、図3Bに例示されている個別のステップは、個別のステップに適している通りに様々なシーケンスで実施され得る複数のサブステップを含み得る。さらに、追加のステップが、具体的な用途に応じて、追加または削除され得る。当業者であれば、プロセスの多くの変更形態、修正形態、および代替的な形態を認識し、理解するであろう。

【0066】

コンピューティングデバイス100によって受信された画像は、シーン記述子を生成するように処理され得る。シーン記述子を生成する構造およびプロセスについては、図4、および図13においてより詳しく記載する。受信画像についてシーン記述子が生成されると、受信画像は、他の画像と照合または比較されて、コンピューティングデバイス100に記憶された画像用のシーンのうち少なくとも1つに属すかどうか判断される。図5および図6は、照合プロセスについてより詳しく記載する。一致が見つかった場合、同じシーンに属する複数の画像から、各画像に関連付けられたエネルギーを使って、キー画像が選択される。図7および図8は、キー画像用の選択プロセスについてより詳しく記載する。キー画像が選択されると、画像は、更新されたシーン記述子をもつ新規合成画像を形成するように統合

10

20

30

40

50

されてよく、OCRキャッシュは更新されてよい。統合プロセスは、図9および図10においてより詳しく説明される。図11は、本発明の実施形態の異なるモジュールおよび構成要素、ならびにそれらの構成要素の間の関連フローを記載する。図12は、追加追跡構成要素をもつ図11をさらに記載する。図14および図15は、画像をキャッシュに記憶するとき、画像からグレアを低減または除去するための本発明の例示的な実施形態を記載する。図16A、図16B、図16Cおよび図16Dは、例示的なコンピューティングデバイス100用の、本明細書に記載する技法を使う例示的なOCRキャッシュ更新シナリオを記載する。

【0067】

図4は、本発明の例示的な実施形態による本の表紙および関連場面記述子の代表的画像を示すブロック図を示す。本発明の態様において、画像キャプチャモジュール1102は、1つまたは複数の画像を受信し、OCRエンジン1104は、1つまたは複数の画像に対してOCRを実施する。シーン記述子モジュール1106は、画像を処理することができ、画像に関連付けられた特性を表す、画像についてのシーン記述子が生成され得る。シーン記述子は、照合モジュールによって、受信または記憶された画像のいずれかが同じシーンに属すかどうか判断する際の照合プロセス中に様々な画像を比較する際に使うことができる。

【0068】

シーン記述子は、グラフィカルオブジェクトを表す画像の部分によって定義され得る。グラフィカルオブジェクトには、限定はしないが、シンボル、文字、単語、テキスト、標識、および数が含まれ得る。図4に示すように、一実装形態では、シーン記述子は、テキスト色、背景色および1つまたは複数の単語の周りの境界ボックスなど、テキストについてのデータをカタログ化する。コンピューティングデバイス100によって検出されたグラフィカルオブジェクトについての特性は、単項特性および二項特性を含み得る。単項特性は、グラフィカルオブジェクトを備える画像の部分のグラフィカルオブジェクト内容、グラフィカルオブジェクト色、前景色、背景色および寸法を含み得るが、それに限定されない。グラフィカルオブジェクトの二項特性は、互いおよび画像に関して、グラフィカルオブジェクトの幾何学的関係を含み得る。本発明のいくつかの態様では、二項特性は、互いに画像中のグラフィカルオブジェクトの異なる幾何学的関係を記述する類似性行列を使って表すことができる。

【0069】

図4のブロック402に戻って参照すると、シーン記述子モジュール1106は、単語「LEARNING from DATA」、「Concepts, Theory, and Methods」、および「SECOND EDITION」を含む画像の部分を識別することができる。図4の「DATA」406などの各単語は、シーン記述子内の別個の記述エンティティとして扱うことができる。また、単語は、「SECOND EDITION」412などの集合的記述エンティティを形成するようにグループ化することができる。

【0070】

図4において、テキスト「DATA」を含む、ブロック406によって表される部分は、シーン記述子404中で「単語3」によって表される別個の記述エンティティ408として記述される。単語記述エンティティ408についての単項特性は、テキスト/内容(すなわち、DATA)、テキスト色(すなわち、白)、背景色(すなわち、黒)ならびに境界ボックスまたは/および寸法を表すフィールドを有するデータ構造により表すことができる。シーン記述子モジュール1106は、各単語についての情報を編成し、データ構造、クラスオブジェクトまたは他のどの適した手段にも記憶することができる。同様に、画像についてのシーン記述子に関連付けられた単語を表す構造は、リンクリスト、キュー、または記憶デバイス125もしくはメモリ135に記憶された情報を取り出し、索引付けし、かつ/もしくは更新するための、他のどの適した手段を使っても編成することができる。

【0071】

単項特性に加え、各単語についての二項特性も、シーン記述モジュール1106によって、画像(図示せず)中の各単語について記憶することができる。二項特性は、互いに関しての単語の幾何学的関係を含むが、それに限定されない。図4において、単語「LEARNING」についての二項特性は、他の単語および画像自体に対する単語の空間関係を含み得る。たと

えば、「LEARNING」は、空間的には左および上からは第1の単語であり、単語「from」を右に、および「Concepts」を下に有する。類似文字をもつ類似単語が異なるシーン中で使われて異なる文を生成する場合、異なるシーンについての二項特性を比較することが有利であり得る。一例では、単語の二項特性を比較することにより、コンピューティングデバイスは、互いに対する単語の相対的な位置を使って、一致をより効率的に見つけることができるようになり得る。

#### 【0072】

図5は、コンピューティングデバイス100によって受信された複数の画像と様々なシーンとの間の照合プロセスを示す図を示す。図1を参照して上で論じたコンピューティングデバイス100は、図5に記載する本発明の実施形態を実施するために使われるデバイスの構成要素のうちいくつかを表し得る。図11および図12で論じるモジュールおよび構成要素、たとえば照合モジュール1108は、コンピューティングデバイス100の構成要素として実装することもでき、図5で論じる本発明の実施形態を実施する際に使うことができる。画像502は、コンピューティングデバイス100に結合されたカメラ150を使って受信された画像を表す。一実施形態では、画像キャプチャモジュール1102は、画像をさらなる処理のためにキャプチャし、画像データをOCRエンジン1104に渡す。OCRエンジン1104は、画像502を処理して、画像に関連付けられたデータなどのOCR結果を生成する。図4で論じたように、シーン記述子が生成され、画像502に関連付けられる。画像504および画像506は、画像502に先立って入手された画像であり、関連シーン記述子を有する。コンピューティングデバイス100の構成要素は、画像504および506を、一時バッファまたはOCRキャッシュのいずれかに記憶することができる。一実装形態では、一時バッファまたはOCRキャッシュ1112は、図1に記載した作業メモリ135のセクションを使うことができる。

#### 【0073】

照合モジュール1108など、コンピューティングデバイス100の構成要素は、画像502用のシーン記述子についての単項および/または二項特性を、画像504および506用のシーン記述子についての単項および/または二項特性と比較することができる。たとえば、画像502は、「LEARNING」、「from」および「DATA」についての単語記述エンティティを有する。ただし、画像506は、「Learning」についての単語記述エンティティを有するだけである。画像502にある「LEARNING」と画像506にある「Learning」との間に肯定的単項一致があった場合であっても、これら2つの単語についての二項特性は一致しない。画像506中で、「Learning」は左からの第1の単語でなく、画像502でのように、右にfromをもたない。したがって、本発明の実施形態は、画像502および画像506は同じシーンに属しないと結論づけることができる。

#### 【0074】

対照的に、画像502および画像504についてのシーン記述子に関連付けられた単項特性および二項特性は、肯定的一致をなし得る。両方の画像は、異なる解像度となるように係数によってスケーリングされた画像中の各単語についての同じ単項特性を有する。また、単語の間の幾何学的関係は、両方の画像において同じである。画像502と画像504との間の唯一の大きな違いは、画像504が画像の一部として「SECOND EDITION」をもたないことである。本発明の態様は、同じシーンにある画像の間のそのような違いを減じるための機構を実装することができる。

#### 【0075】

一実装形態では、照合モジュール1108など、コンピューティングデバイス100の構成要素は、シーン記述子についての任意の2つの所与の特性の間の各類似および違いに重みを割り当てることができる。割当て重みは、類似に対しては加算され、違いについては減算されてよい。閾値は、生成されても所定であってもよい。2つの画像についてのシーン記述子中の類似と違いの加重和が閾値を上回る場合、照合モジュール1108は、2つの所与の画像を同じシーンに関連付けることができる。2つの画像についてのシーン記述子中の類似と違いの加重和が閾値を下回る場合、画像は、異なるシーンを表すものと見なされ得る。

## 【 0 0 7 6 】

図6は、入手画像を記憶画像についてのシーンに関連付けるための非限定的な例示的照合プロセスを示すフロー図を示す。ステップ602で、照合モジュール1108など、コンピューティングデバイス100の構成要素は、受信画像に対してOCRを実施する。図1を参照して上で論じたコンピュータデバイス100は、図6に記載する本発明の実施形態を実施するために使われるモバイルデバイスの構成要素のうちいくつかを表し得る。画像は、コンピューティングデバイス100に結合されたカメラ150を使って入手される。ステップ604で、たとえばシーン記述子生成器1106を使って、受信画像についてシーン記述子が生成され得る。図4を参照して記載したように、テキストなどのグラフィカルオブジェクトを備える画像の部分が、認識され、シーン記述子中でさらに特徴を述べられ得る。各グラフィカルオブジェクトについて、単項特性と二項特性の両方が定義され得る。単項特性は、グラフィカルオブジェクトを備える画像の部分のグラフィカルオブジェクト内容、グラフィカルオブジェクト色、前景色、背景色および寸法を含み得るが、それに限定されない。グラフィカルオブジェクトの二項特性は、互いおよび画像に関して、グラフィカルオブジェクトの幾何学的関係を含み得る。各グラフィカルオブジェクトについての定義された特性は、データ構造、クラスオブジェクト中に、または他のどの適切な手段を使っても記憶することができる。

10

## 【 0 0 7 7 】

ステップ610で、受信画像のシーン記述子は、事前に入手されるとともにコンピューティングデバイス100に記憶およびバッファリングされた画像のシーン記述子と、たとえば照合モジュール1108によって比較される。事前入手画像およびその関連シーン記述子は、一時バッファ、OCRキャッシュまたはストレージ125、メモリ135もしくはハードウェアレジスタを使う他のどの適した記憶媒体にも記憶することができる。シーン記述子を比較する間(ステップ610)、ステップ606で、受信画像についてのシーン記述子の単項特性が、記憶された画像に関連付けられたシーン記述子の単項特性と比較され得る。ステップ608で、受信画像についてのシーン記述子の二項特性が、記憶された画像に関連付けられたシーン記述子の二項特性と比較され得る。

20

## 【 0 0 7 8 】

一実装形態では、シーン記述子の任意の2つの所与の特性の間の各類似は、関連付けられた、割当て重みを有し得る。割当て重みは、類似に対しては反映され、違いについては減算されてよい。閾値は、生成されても所定であってもよい。ステップ612で、2つの画像についてのシーン記述子中の類似の加重和が閾値を上回る場合、コンピューティングデバイス100の構成要素は、たとえば選択モジュール1110および/または統合モジュール1114を使って、画像を同じシーンに関連付けることができる(ブロック616)。一方、2つの画像についてのシーン記述子中の類似の加重和が閾値を下回る場合、これらの画像は、異なるシーンを表すものと見なされ得る(ブロック614)。方法600に記載する部分フローは、一時メモリ、OCRキャッシュ1112または他の任意の適した記憶媒体にある、事前記憶画像およびそれらの関連シーン記述子すべてのシーン記述子と比較するように、反復的または再帰的に繰り返してよい。

30

## 【 0 0 7 9 】

図6に示す具体的なステップは、本発明の一実施形態による、動作のモード間を切り替える特定の方法を提供することを諒解されたい。ステップの他の順序も、代替的な実施形態に従って実施され得る。たとえば、本発明の代替的な実施形態は、異なる順序で、上で概説されたステップを実施することができる。説明のために、ユーザは、第3の動作モードから第1の動作モードに、第4のモードから第2のモードに、またはそれらの間の任意の組合せで変化させることを選択することができる。さらに、図6に例示されている個別のステップは、個別のステップに適している通りに様々なシーケンスで実施され得る複数のサブステップを含み得る。さらに、追加のステップが、具体的な用途に応じて、追加または削除され得る。当業者であれば、プロセスの多くの変更形態、修正形態、および代替的形態を認識し、理解するであろう。

40

50

## 【0080】

図7は、同じシーンに属す複数の画像からの、キー画像としての画像の選択プロセスを示すブロック図を示す。図4、図5および図6に記載したように、コンピューティングデバイス100によって画像が受信された後、コンピューティングデバイス100は、画像についてのシーン記述子を生成し、画像を、事前入手および記憶された他の画像と照合することができる。一実装形態では、コンピューティングデバイスは、同じシーンに属す複数の画像からキー画像を選択する。複数の画像からキー画像を選択するプロセスは、選択プロセスと呼ぶことができ、図11および図12にある選択モジュール1110など、コンピューティングデバイス100の構成要素によって実施することができる。図7を参照すると、グループIは、第1のシーンに関連付けられた3つの画像を有し、グループIIは、第2のシーンに関連付けられた3つの画像を有する。最終的なキー画像は、選択され、グループIおよびグループIIに対して最右列に表示される。キー画像は、シーンを表す、OCRキャッシュ1112に記憶されている候補画像であり得る。いくつかの事例では、キー画像またはキー画像に関連した情報は、キー画像を記憶する前に、他の画像からの情報を使って強調され得る。たとえば、キー画像以外の1つまたは複数の画像中で認識される単語が、キー画像に関連付けられたデータに加えられ得る。各画像グループからキー画像を見つける際に、以下に記載する機構を遂行すればよい。

10

## 【0081】

キー画像は、同じシーンに関連付けられた他の画像よりも良好な解像度および明瞭度を有する画像を表し得る。一態様では、コンピューティングデバイス100の構成要素は、各画像に関連付けられたエネルギーを算出することができる。エネルギー値は、画像についてのシーン記述子の一部として記憶することができる。最も高いエネルギーをもつ画像が、同じシーンに関連付けられた複数の画像から、キー画像として選択され得る。一実装形態では、コンピューティングデバイス100(たとえば、選択モジュール1110を使う)は、画像にある、認識された単語の数と、OCRからの単語の信頼平均値とを使ってエネルギー値を判断することができる。

20

## 【0082】

図8は、同じシーンにある複数の画像からの、キー画像としての画像の選択プロセスのための非限定的な例示的方法を示すフロー図を示す。1つまたは複数の態様によれば、本明細書で説明されている方法および/または方法ステップのどれか、および/またはすべては、たとえば、コンピューティングデバイス100および/または図1でより詳しく説明されているデバイスなどの、コンピューティングデバイス100によって、および/またはコンピューティングデバイス100内に実装され得る。一実施形態では、図8に関して以下で説明されている方法ステップのうちの1つまたは複数の、プロセッサ110または別のプロセッサなどのデバイス100のプロセッサによって実装される。選択モジュール1110など、図11および図12で論じるモジュールおよび構成要素は、コンピューティングデバイス100の構成要素として実装することもでき、図8で論じる本発明の実施形態を実施する際に使うことができる。それに加えて、または代替的に、本明細書で説明されている方法および/またはステップのどれか、および/またはすべては、メモリ135、ストレージ125または別のコンピュータ可読媒体などのコンピュータ可読媒体上に記憶されているコンピュータ可読命令などのコンピュータ可読命令で実装され得る。

30

40

## 【0083】

ステップ802で、図5および図6で論じたように、1つまたは複数(M個)の画像が、同じシーンに属すものとして検出される。いくつかの実装形態では、OCRキャッシュ1112は、任意の所与のシーンに関連付けられた、ただ1つの画像を有する。OCRキャッシュ1112に記憶されているものと同じシーンに関連付けられたどの入手画像も、直ちに処理され、OCRキャッシュ1112に統合され得る。新たに入手された画像が、キー画像として選択され、または付加情報を有する場合、OCR画像は更新されてよい。そのような実装形態において、選択プロセスは、2つの画像、すなわち、OCRキャッシュに記憶された画像と、カメラによって最近入手された画像との間でのみ行われればよい。

50

## 【 0 0 8 4 】

ただし、本発明の他の実装形態では、キー画像を見つけるために、複数の画像が同時に比較されてよい。たとえば、一実装形態では、複数の画像が一定期間にわたって入手され、同時に処理され得るウィンドウプロトコルが実装されてよい。ユーザは、同じシーンの上で短期間に複数回モバイルデバイスを滑らせる場合があるので、入手され、一時的に記憶されている複数の画像のうち、通常は複数の画像が同じシーンに属し得る。したがって、同じシーンに関連付けられた任意の所与の時間ウィンドウからの1つまたは複数の画像が、互いまたは同じシーンに関連付けられた、OCRキャッシュ1112に記憶された画像と比較され得る。

## 【 0 0 8 5 】

10

Mは、任意の所与の時点における特定のシーンに関連付けられた画像の数を表す。各画像に関連付けられたエネルギーは、反復的または再帰的に算出することができる。反復の回数は、Xで表される。例示的な一実装形態では、Xは1に初期化され得る。Xは、現在処理されている画像も表す。一態様では、画像のエネルギーは、画像のシーン記述子に記憶される。

## 【 0 0 8 6 】

ステップ804で、OCRを使って、画像X中の単語の数(N)が認識される。ステップ806で、OCRからの単語の信頼平均値(C)が導出される。ステップ808で、認識された単語に関連付けられた信頼度が、スケール係数( )によってスケーリングされる。一実施形態では、スケール係数( )は、経験的に決定され得る。ステップ810で、画像Xについてのエネルギーは、以下の式に従って算出される。

20

$$EX=N+C$$

## 【 0 0 8 7 】

ステップ812で、画像について導出されたエネルギーは、任意選択で、画像のシーン記述子に記憶することができる。ステップ814で、 $X=M$ である場合、最も大きいエネルギー値をもつ画像が、同じシーンに関連付けられた複数の画像からキー画像として選択される(ステップ816)。XがMに等しくない場合、Xは増分され(ステップ818)、次の画像Xが、方法800によりフェッチされ、処理されてよく、ステップ804で再開する。

## 【 0 0 8 8 】

別の実施形態では、画像についてのエネルギーの反復/再帰的算出は使われない。画像が入手された後、または画像に関連付けられたエネルギーに影響を与え得る付加情報で画像が更新された場合、画像のエネルギーを算出することができる。一実装形態では、画像のエネルギーは、画像についてのシーン記述子に記憶される。

30

## 【 0 0 8 9 】

図8に示す具体的なステップは、本発明の一実施形態による、動作のモード間を切り替える特定の方法を提供することを諒解されたい。ステップの他の順序も、代替的な実施形態に従って実施され得る。たとえば、本発明の代替的な実施形態は、異なる順序で、上で概説されたステップを実施することができる。説明のために、ユーザは、第3の動作モードから第1の動作モードに、第4のモードから第2のモードに、またはそれらの間の任意の組合せで変化させることを選択することができる。さらに、図8に例示されている個別のステップは、個別のステップに適している通りに様々なシーケンスで実施され得る複数のサブステップを含み得る。さらに、追加のステップが、具体的な用途に応じて、追加または削除され得る。当業者であれば、プロセスの多くの変更形態、修正形態、および代替的形態を認識し、理解するであろう。

40

## 【 0 0 9 0 】

図9は、同じシーンにある複数の画像からの、OCRキャッシュ1112に記憶される最終画像としての画像の合成を示す図を示す。一実施形態では、画像902および904は、OCRエンジン1104、次いで、それぞれ、画像に関連付けられたグラフィカルオブジェクトを識別し、シーン記述子を生成するためのシーン記述子生成器1106を通してルーティングされる。照合モジュール1108は、画像902および904を同じシーンに関連付ける。一実施形態では、照

50

合プロセスの後、画像およびそれらの関連シーン記述子は、選択プロセスを通して、キー画像を判断するための選択モジュール1110によってルーティングされる。例示目的で、画像902が、仮定として、選択プロセスからキー画像として選択される。

#### 【0091】

選択プロセスの後、画像は、統合プロセスのための統合モジュール1114を通してルーティングされてよく、統合プロセスにおいて、2つの画像は、単一の合成画像906に統合され得る。図9において、画像902が、OCRキャッシュに保存され得る画像として選択される。ただし、画像904は、画像902中では示されない追加の一意の情報を有し得る。画像904にとって一意の付加情報は、画像902と合成されて、最終画像906を形成し得る。一実施形態では、画像906は、画像902および画像904から新たに生成することができ、その結果すべてのOCR結果が画像906に示され得る。

10

#### 【0092】

シーン記述子は、更新することもできる。たとえば、画像906についてのシーン記述子は、902についてのシーン記述子用に定義された特性を有する。ただし、追加の一意の単語が画像904から追加されるので、画像906についてのシーン記述子は、画像904にある「Methods」について定義された特性を継承することができる。

#### 【0093】

図10は、同じ場面にある複数の画像のうちの1つの画像を合成するための非限定的な例示的方法を示すフロー図である。1つまたは複数の態様によれば、本明細書で説明されている方法および/または方法ステップのどれか、および/またはすべては、たとえば、コンピューティングデバイス100および/または図1でより詳しく説明されているデバイスなどの、コンピューティングデバイス100によって、および/またはコンピューティングデバイス100内に実装され得る。一実施形態では、図10に関して以下で説明されている方法ステップのうちの1つまたは複数は、プロセッサ110または別のプロセッサなどのデバイス100のプロセッサによって実装される。統合モジュール1114など、図11および図12で論じるモジュールおよび構成要素は、コンピューティングデバイス100の構成要素として実装することもでき、図8で論じる本発明の実施形態を実施する際に使うことができる。それに加えて、または代替的に、本明細書で説明されている方法および/またはステップのどれか、および/またはすべては、メモリ135、ストレージ125または別のコンピュータ可読媒体などのコンピュータ可読媒体上に記憶されているコンピュータ可読命令などのコンピュータ可読命令で実装され得る。

20

30

#### 【0094】

本明細書に記載する画像の合成は、OCRキャッシュ内の画像または画像についてのシーン記述子もしくは画像の表現を含む任意の関連データの更新または置き換えを含み得るが、それに限定されない。たとえば、OCRキャッシュに記憶された、レストランにあるディナーメニューの画像は、カメラがディナーメニューをプレビューモードまたはスナップショットモードでホバリングし、より多くの画像を受信すると、受信画像からの付加情報で補強され得る。ステップ1002で、図7および図8で論じたように、選択モジュール1110を使って、同じシーンに属す複数の画像から、キー画像が選択される。ステップ1004で、本発明の構成要素、たとえばシーン記述子生成器1106によって識別される、すでにキー画像の一部ではないどの一意のグラフィカルオブジェクトも、キー画像に追加され得る。いくつかの事例では、受信画像は、OCRキャッシュにすでに記憶されていないシーンの領域などの付加情報を有し得る。逆もまた同様に、受信画像は、OCRキャッシュ1112中の記憶された画像に含まれる情報すべてを有しているわけではない場合がある。ある事例では、キー画像が選択されると、すでにキー画像の一部ではない他の画像にあるどのグラフィカルオブジェクトも、画像に追加される。

40

#### 【0095】

ステップ1006で、任意選択で、合成画像についてのエネルギーは、画像に追加されるどの新規情報も説明するように更新されてもよい。一実施形態では、コンピューティングデバイス100は、図8の810で論じたように、合成画像についてのエネルギーを算出する。ス

50

テップ1008で、シーン記述子も、付加情報で更新され得る。たとえば、グラフィカルオブジェクトに関連付けられた追加画像データおよびシーン記述子データが、合成画像を形成するキー画像についてのデータと統合されてから、合成画像をOCRキャッシュ1112に記憶すればよい。上で論じた同じプロセスは、複数の受信画像およびOCRキャッシュ1112に記憶された画像など、どの2つ以上の画像についても同時に起こり得る。

【0096】

図10に示す具体的なステップは、本発明の一実施形態による、動作のモード間を切り替える特定の方法を提供することを諒解されたい。ステップの他の順序も、代替的な実施形態に従って実施され得る。たとえば、本発明の代替的な実施形態は、異なる順序で、上で概説されたステップを実施することができる。説明のために、ユーザは、第3の動作モードから第1の動作モードに、第4のモードから第2のモードに、またはそれらの間の任意の組合せで変化させることを選択することができる。さらに、図10に例示されている個別のステップは、個別のステップに適している通りに様々なシーケンスで実施され得る複数のサブステップを含み得る。さらに、追加のステップが、具体的な用途に応じて、追加または削除され得る。当業者であれば、プロセスの多くの変更形態、修正形態、および代替的形態を認識し、理解するであろう。

【0097】

図11は、本発明の実施形態によって提供される方法を実施するための例示的な構成要素およびモジュールを示すブロック図である。図1を参照して上で論じたコンピュータデバイス100は、図11に記載する本発明の実施形態を実施するために使われるモバイルデバイスの構成要素のうちいくつかを表し得る。図11で論じた構成要素およびモジュールは、ハードウェア、ソフトウェア、ファームウェア、またはそれらの任意の組合せで実装され得る。

【0098】

一実施形態では、コンピューティングデバイス100は、自動および知的OCRキャッシュを使うOCRキャッシングモードに切り替えられ得る。OCRキャッシュ1112は、図1に記載した作業メモリ135の複数の部分を使用することができる。一実施形態では、カメラキーボードなどのプログラムまたはアプリケーションがオンであるとき、自動および知的OCRキャッシュ更新を用いるプレビューモードがトリガされ、カメラキーボードがオフにされると、プレビューモードがオフにされる。

【0099】

画像キャプチャモジュール1102は、1つまたは複数の画像を入手し、キャプチャし、または受信する。一実施形態では、画像キャプチャモジュール1102は、コンピューティングデバイスに結合された1つまたは複数のカメラ150を使って、1つまたは複数の画像を入手する。OCRエンジン1104は、(たとえば、プロセッサ110を使って)画像を処理し、画像に埋め込まれたグラフィカルオブジェクトを備える画像の部分の認識する。グラフィカルオブジェクトには、限定はしないが、シンボル、文字、単語、標識、および数が含まれ得る。さらに、OCRエンジンは、画像を処理し、各画像についてのシーン記述子への投入を始めることができる。

【0100】

シーン記述子生成器1106は、OCRキャッシュシステム1124の一部として実装することができ、シーン記述子を生成し、シーン記述子に投入を行うことができる。シーン記述子は、画像に関連付けられたグラフィカルオブジェクトの特性を記述する。シーン記述子は、複数のグラフィカルオブジェクトを表す画像の部分によって定義され得る。グラフィカルオブジェクトについての特性は、単項特性および二項特性を含み得る。単項特性は、グラフィカルオブジェクト内容、グラフィカルオブジェクト色、前景色、背景色、OCR結果に関連付けられた信頼度、ならびにグラフィカルオブジェクトを備える画像の部分の寸法および境界を含み得るが、それに限定されない。グラフィカルオブジェクトの二項特性は、互いおよび画像に関して、グラフィカルオブジェクトの幾何学的関係を含み得る。本発明のいくつかの態様では、二項特性は、互いに画像中のグラフィカルオブジェクトの異なる



幾何学的関係を記述する類似性行列を使って表すことができる。図4および図6は、シーン記述子生成器1106によって実施されるステップのいくつかをより詳しく論じている。

【0101】

照合モジュール1108は、1つまたは複数の受信画像のシーン記述子と、OCRキャッシュ1112にある記憶画像を照合することができる。任意の2つの画像のシーン記述子の間の一致は、これら2つの画像が同じシーンに関連付けられていることを示す。画像の間のシーン記述子における類似および違いは、同じシーンに合理的に属するように互いと十分近接して位置合せされている画像を見つけるのに使用することができる。図5および図6は、一致を見つけるための例示的な実施形態をより詳しく記載している。

【0102】

照合モジュール1108が、画像キャプチャモジュール1102からの受信画像と、OCRキャッシュ1112に記憶されている画像との間の一致を見つけなかった場合、画像は、OCRキャッシュ1112に記憶されるために挿入モジュール1116に渡される。OCRキャッシュ1112に記憶されている新規画像は、コンピューティングデバイス100によって入手された新規シーンを表す。

【0103】

照合モジュール1108が、受信画像と、OCRキャッシュ1112にある画像との間に一致を見つけた場合、一致画像(または画像へのポインタ)は、キー画像を選択するための選択モジュール1110にフォワードされる。ブロック1114で、2つの画像からのキー画像が、受信画像およびOCRキャッシュ1112からの一致画像に関連付けられたエネルギーを使う選択プロセスを使って判断される。一実施形態では、エネルギーは、画像についてのシーン記述子に記憶される。別の実施形態では、各画像に関連付けられたエネルギーは、選択モジュール1110において導出される。一実装形態では、画像のエネルギーは、画像について認識される単語の数、および単語の数に関連付けられた信頼度に正比例するように定義される。以下の式は、エネルギーを導出するための実装を表し得る。

$$E=N+C$$

Nは、画像中で認識される単語の数であり、Cは、OCRからの単語の信頼平均値であり、はスケール係数である。一態様では、は、経験的データを使って導出される。画像が受信された後、または画像に関連付けられたエネルギーに影響を与え得る付加情報で画像が更新された場合、画像のエネルギーを算出することができる。

【0104】

統合モジュール1114は、シーン記述子および非キー画像からの他の画像データを、キー画像と統合して、合成画像を形成することができる。本明細書に記載する画像の合成は、OCRキャッシュ内の画像または画像についてのシーン記述子もしくは画像の表現を含む任意の関連データの更新または置き換えを含み得るが、それに限定されない。ある事例では、OCRキャッシュ1112に記憶された画像および受信画像は、同一のグラフィカルオブジェクトについての2つの画像の間で同じであるシーン記述子情報をもつオーバーラップ領域を有し得る。ただし、いくつかの事例では、受信画像は、OCRキャッシュにすでに記憶されていないシーンの領域などの付加情報を有し得る。逆もまた同様に、受信画像は、OCRキャッシュ1112中の記憶された画像に含まれる情報すべてを有しているわけではない場合がある。ある事例では、キー画像が選択されると、すでにキー画像の一部ではない他の画像にあるどのグラフィカルオブジェクトも、画像に追加される。図10は、統合モジュール1114によって実施されるステップのうちいくつかについてより詳しく論じる。

【0105】

一例では、OCRキャッシュに記憶された、レストランにあるディナーメニューの画像は、カメラがディナーメニューをプレビューモードまたはスナップショットモードでホバリングし、より多くの画像を受信すると、受信画像からの付加情報で補強され得る。別の例では、統合モジュールは、OCRキャッシュ1112などのキャッシュに合成画像を記憶する前に、画像からグレアを除去または低減することができる。統合プロセスは、第1の領域および第2の領域がシーンの同じ領域を表す、非キー画像からグレアのある第1の領域を識別

10

20

30

40

50

し、非キー画像からグレアのない第2の領域を識別し、非キー画像と、非キー画像にある、グレアのない第2の領域とを使って、グレアを低減するために最終画像を合成することによって、グレアを低減するのを容易にし得る。

【0106】

一実施形態では、自動および知的OCRキャッシュ更新を用いるOCRキャッシングモードは、カメラキーボードなどのプログラムまたはアプリケーションがオンにされると動作可能であり、このモードは、カメラキーボードがオフにされ、または他のやり方で停止またはサスペンドされると、非アクティブ化される。いくつかの実施形態では、カメラキーボードにより、カメラによってキャプチャされた情報、たとえば画像中で認識され、またはシーンのキー画像に関連付けられたテキストが、アプリケーションに入力される。さらに、コンピューティングデバイス100に結合されたデバイスディスプレイは、OCRキャッシングモード(いくつかの実施形態では、自動および知的OCRキャッシュ更新を用いる)が動作可能である(すなわち、カメラキーボードがオンにされ、またはプログラムもしくはアプリケーションが再開され、もしくは他のやり方で動作可能である)とき、OCRキャッシュ1112の内容を継続的に表示することもできる。また、カメラキーボードがストップされると、OCRキャッシュ1112用のキャッシュデータは、メモリからデバイスストレージ1120に保存されてよい。デバイスストレージ1120は、図1に記載した記憶デバイス125のうち1つであってよい。カメラキーボードが開始されると、OCRキャッシュ1112用のキャッシュデータが、デバイスストレージ1120からメモリにロードされ得る。一実装形態では、OCRキャッシュ1112は、画像のキューを使って実装される。ただし、リンクリスト、クラスオブジェクト、またはいくつかの画像もしくはある程度の量のデータを含む他のどの適した手段が、OCRキャッシュ1112を実装するために使われてもよい。

【0107】

図12は、本発明の実施形態によって提供される方法を実施するための例示的な構成要素を示す別のブロック図を示す。図11に記載した構成要素に加え、図12は、画像キャプチャモジュール1102からOCRキャッシングモードでシーンを継続的に追跡し、ブロック1108で照合論理に入力を供給する追跡モジュール1222を有する。シーンの追跡は、シーンに関連付けられたより良好または大きい画像を生成するように同じシーンの上でユーザがモバイルデバイスを継続的に滑らせているとき、照合論理を支援する際に特に有用である。このシナリオでは、照合論理の複雑さおよび処理時間および電力は、入手されたあらゆる新規画像に対して大幅に削減される。一実施形態では、グラフィカルオブジェクト(およびグラフィカルオブジェクトをフレーミングするのに、グラフィカルオブジェクトの周囲で使われる関連ボックス)の移動は、照合問題を解消する際に使われ得る。図11にあるディナーメニュー例に戻って参照すると、追跡モジュールは、コンピューティングデバイス100がディナーメニューの上をホバリングするとき、画像を絶えず追跡し、コンピューティングデバイス100が、ディナーメニューの単一の統合画像を構築するのを助けることができる。いくつかの実施形態では、モジュール1104、1106、1108、1110、1112、1114、1116、1124、および1222のうち1つまたは複数は、デバイス100の1つまたは複数のプロセッサまたは処理コア内で、たとえばプロセッサ110内で実装される。

【0108】

図13Aおよび図13Bは、テキストを含む画像についての例示的な場面記述子を示す。図13Aは、シーン記述子生成器1106によって生成されるシーン記述子である。シーン記述子は、画像解像度、色の濃度および他のそのような画像特性などの画像データを含む。さらに、シーン記述子は、テキスト内容、関心領域(ROI)、OCRを使った単語の適正な認識に関連付けられた信頼度、推定テキスト色および背景色など、各テキスト単語に関連付けられた特性を含み得る。

【0109】

図13Bは、更新されたシーン記述子を示す。シーン記述子についてのエネルギー値は、各画像についてのエネルギーが算出または更新されると、選択モジュール1110および統合モジュール1114によって更新され得る。同様に、新規テキストが、統合モジュール1114に

においてOCRキャッシュ1112用のキー画像に追加された場合、新規テキスト用の追加エントリが追加され得る(例として、図13BのTextN+1)。

【0110】

図14は、画像からグレアを除去または低減するための本発明の実施形態を示す図である。屋内設定においてカメラ150を使って取得された画像は、反射グレアによって影響されることがある。たとえば、ディナーメニューは、ディナー用照明によってメニューから反射するグレアを有し得る。反射グレアは、画像のキャプチャ、検出および認識のための、光学式文字認識(OCR)など、カメラベースのアプリケーションの性能を低下させ得る。反射グレアは、関心オブジェクトとグレア光源との間の大幅な輝度比によって引き起こされ得る。関心オブジェクト、グレア光源およびカメラの間の角度などの係数は、グレア体験に大幅な影響を与え得る。

10

【0111】

本発明の実施形態は、2つ以上の位置にある同じ関心対象の画像を使って反射グレアを検出および低減するための技法を提供する。画像は、2つ以上の位置の間でカメラ150を動かすことによって、またはステレオカメラなど、複数のカメラ150を使うことによって取得することができる。2つ以上の位置から画像を取得することにより、グレア光源に対するカメラと、関心オブジェクトとの間の角度を変更することができる。わずかに異なる観点からの2つの取得された画像は、同じ関心オブジェクト用の画像の異なる部分におけるグレアを表すことができ、グレアが低減された画像の合成を可能にする。

【0112】

20

図14は、本発明の例示的な実施形態を表すブロック図である。図14において、デバイス(デバイス1402)に結合されたステレオカメラ150を有するステレオカメラコンピューティングデバイスが、同じ関心オブジェクトの2つ以上の画像を取得するために使われる。一実装形態では、画像は同時に取得される。別の実施形態では、画像は、異なる時間に異なる観点から、単一のカメラを使って次々に取得され得る。図14に見られるように、左ビュー(画像1404)を使って取得された画像および右ビュー(画像1406)を使って取得された画像は両方とも、反射グレア(1410および1412)を有する。ただし、2つのカメラがグレア光源および関心オブジェクトに対して画像をキャプチャする角度は異なるので、グレアは、画像の異なる部分に存在する。図14に示すように、本発明の実施形態は、左ビュー画像および右ビュー画像を使って強調画像を合成する。一実施形態では、2つの画像の、1つの強調画像への合成は、統合モジュール1114によって実施することができる。コンピューティングデバイス100が、照合モジュール1108において2つの一致する画像を識別すると、統合モジュールは、図15において後でさらに論じるように、最終画像1408を生成するためにグレアを除去または低減することによって画像を強調してから、最終画像をOCRキャッシュ1112に記憶すればよい。

30

【0113】

図15は、本開示の1つまたは複数の例示的態様による、画像中の反射グレアを低減または除去するための本発明の実施形態を示すフロー図である。1つまたは複数の態様によれば、本明細書で説明されている方法および/または方法ステップのどれか、および/またはすべては、たとえば、コンピューティングデバイス100および/または図1でより詳しく説明されているデバイスなどの、コンピューティングデバイス100によって、および/またはコンピューティングデバイス100内に実装され得る。一実施形態では、図15に関して以下で説明されている方法ステップのうちの1つまたは複数は、プロセッサ110または別のプロセッサなどのデバイス100のプロセッサによって実装される。シーン記述子生成器1106、照合モジュール1108、および/または統合モジュール1114など、図11および図12で論じたモジュールおよび構成要素は、コンピューティングデバイス100の構成要素として実装することもでき、図16で論じる本発明の実施形態を実施する際に使うことができる。それに加えて、または代替的に、本明細書で説明されている方法および/またはステップのどれか、および/またはすべては、メモリ135、ストレージ125または別のコンピュータ可読媒体などのコンピュータ可読媒体上に記憶されているコンピュータ可読命令などのコンピュ

40

50

ータ可読命令で実装され得る。

【0114】

一実施形態では、ステレオカメラ150は、2つの異なる位置にある、関心オブジェクトの2つの画像を取得するのに使うことができる。別の実施形態では、2つの画像は、2つの異なる位置から次々に取得される。ステップ1502で、左ビューからの第1の画像が取得され、ブロック1504で、右ビューからの第2の画像が取得される。ブロック1502および1504で実施されるステップは、同時にまたは次々と起こり得る。別の実施形態では、カメラ150からの受信画像およびOCRキャッシュ1112からの画像が、OCRキャッシュ1112に記憶された画像からグレアを除去するために使われてもよく、OCRキャッシュ1112からの2つの画像が使われてもよい。ステップ1506で、画像の両方からの特徴を抽出することができ、2つの画像の間の幾何学的関係を導出することができる。一実施形態では、OCRエンジン1104およびシーン記述子生成器1106は、図4で前述したように、画像中の様々なグラフィカルオブジェクトと、互いのオブジェクトとの関係とを識別することができる。別の実施形態では、Harrisコーナー、fastコーナーなどのような従来の特徴抽出アルゴリズムを使って、特徴が抽出される。グレア領域中の特徴は、推定し、なくすことができる。照合モジュール1108は、照合プロセスを使って、画像と、グラフィカルオブジェクトなど、画像の特徴とを照合することができる。別の実施形態では、特徴は、従来の特徴記述子(SIFT、SURFなど)およびRANSAC(RANdom SAmple Consensus)などの照合方式を使って照合され得る。特徴の間の対応を見つけるための探索空間を制約する際に、ベースライン制約システムが使用され得る。ブロック1508で、画像は、その幾何学的関係に基づいて位置合せされ得る。さらに、画像のグレア領域が検出される。一実施形態では、グレア領域は、2つの画像のオーバーラップ領域の間の違いを検出することによって判断される。一実施形態では、コンピューティングデバイス100は、同じシーンにある第1の画像および第2の画像の複数の特徴における違いを検出することによって、画像のうち1つの、ある領域中のグレアを識別する。たとえば、いくつかの実施形態では、領域中の削減された数の特徴が、グレアを識別するのに使われ得る。

【0115】

最後に、ブロック1510で、たとえば統合モジュール1114を用いる統合プロセスを使って、グレア効果を低減するために、グレアが比較的少ない、2つの画像にあるオーバーラップ領域を使って、2つの画像を合成することによって強調画像が生成され得る。一実施形態では、本発明の実施形態は、第1のシーンにある2つ以上の画像のうちの画像のうち1つから、グレアのある第1の領域を識別し、第1のシーンにある2つ以上の画像のうちの第2の画像から、グレアのない第2の領域を識別し、ここで第1の領域および第2の領域は、第1のシーンの同じ領域を表す。第1の画像および第2の画像を使って、グレアが低減された、またはグレアのない最終画像が合成され得る。2つの画像の間の情報パッチの選択、より良好な認識のための前景および背景のセグメント化など、異なる技法が、画像を合成する際に使われ得る。グレアが低減または除去された合成画像は、OCRキャッシュ1112などのキャッシュに記憶することができる。

【0116】

図15に示す具体的なステップは、本発明の一実施形態による、動作のモード間を切り替える特定の方法を提供することを諒解されたい。ステップの他の順序も、代替的な実施形態に従って実施され得る。たとえば、本発明の代替的な実施形態は、異なる順序で、上で概説されたステップを実施することができる。説明のために、ユーザは、第3の動作モードから第1の動作モードに、第4のモードから第2のモードに、またはそれらの間の任意の組合せで変化させることを選択することができる。さらに、図15に例示されている個別のステップは、個別のステップに適している通りに様々なシーケンスで実施され得る複数のサブステップを含み得る。さらに、追加のステップが、具体的な用途に応じて、追加または削除され得る。当業者であれば、プロセスの多くの変更形態、修正形態、および代替的形態を認識し、理解するであろう。

【0117】

図16は、自動および知的OCRキャッシュ更新を用いるOCRキャッシングモードで動作するモバイルデバイスなど、コンピューティングデバイス100の簡略化表現を使った、ユーザインターフェースのための例示的な実装形態を示す。たとえば、ユーザインターフェースは、出力デバイス120によって表示されてもよい。図16A、図16B、図16C、および図16Dは、連続した時点におけるコンピューティングデバイス100を表す。コンピューティングデバイス100の上部分に位置するエリア1604は、コンピューティングデバイス100に結合されたカメラのレンズを通してユーザに見える視野を表す。図面の下部に向かって示されている細い帯状部は、OCRキャッシュ1112を表す。一態様では、OCRキャッシュ1112中の画像の表現は、モバイルデバイスの下部に向かって示される細い帯状部中のサムネイル(1606、1608および1610)としてユーザに見える。一実施形態では、OCR結果は、画像ではなく帯状部に示される。別の実施形態では、画像とOCR結果の両方が示される。ユーザは、OCRキャッシュ画像を選択し、テキストを選択するとともに編集、コピーおよびペーストなどの共通テキスト編集機能を実施することによって画像中のテキストを操作することができる。例示的な使用ケースでは、ユーザは、たとえば上述したユーザインターフェースを表示するタッチスクリーン上の画像またはOCR結果にタッチすることによって、OCRキャッシュからテキストを選択することができる。細い帯状部にある画像を選択すると、その画像はサイズが大きくなり、たとえば、OCR結果がより正確に選択され得るように、エリア1604に表示され得る。いくつかの実施形態では、選択は、細い帯状部から直接行われてもよい。ユーザによって、たとえば、ユーザの手、指、目、頭、またはユーザによって制御される他のアイテムで実施することができる、ユーザのタッチまたはタッチなしジェスチャに基づいて識別または選択された情報は、翻訳され、探索を実施するために提出され、または、たとえばテキストボックスに入力されることによって、他のやり方でプログラムまたはアプリケーションに入力され得る。いくつかの実施形態では、画像またはOCR結果の数は、コンピューティングデバイス100のスクリーンに表示可能であるよりも大きい場合がある。そのような一実施形態において、ユーザは、画像もしくは結果をスクロールし、または他のやり方で、たとえば帯状部を前に出し、もしくは引っ込めるためにタッチスクリーンに指を打ちつけることによって、可視画像もしくは結果をトグルすることができる。

#### 【0118】

図16Aを参照すると、シーン1は、カメラ150の視野を表し得る。カメラ150は、シーン1を表す画像をキャプチャすることができ、図16Aに関連付けられた時点においてOCRキャッシュ1112が空なので、本発明の実施形態に従って、画像をシーン記述子とともにOCRキャッシュに記憶する。図16Bにおいて、シーン2は、カメラの視野内にある。カメラ150は、シーン2を表す画像をキャプチャすることができる。本明細書で論じる技法を用いて、デバイス100の構成要素は、照合プロセスを実施し、OCRキャッシュからシーン1についてキャプチャされた画像を、シーン2についてキャプチャされた画像と比較する。これらの画像は異なるシーンに属するので、一致は見つからず、OCRキャッシュは、シーン2についてキャプチャされた画像で更新される。

#### 【0119】

図16Cにおいて、シーン1はやはり、カメラ150の視野内にある。シーン1を表す画像は、カメラ150によって入手される。今度は、デバイスの構成要素が、受信画像についてのシーン記述子と、OCRキャッシュに記憶されている画像のシーン記述子との間の比較を実施すると、一致が見つかる。シーン記述子にある、画像両方の算出エネルギーを使って、2つの画像からキー画像を選択することができる。統合プロセスの一部として、非キー画像中で見つかったどの一意の単語も、画像に追加されてよく、シーン記述子は更新されてよい。さらに、シーン記述子中のどの単語または結果も、より高い信頼度をもつ他の単語または結果で更新され、または置き換えられてよい。統合プロセスは、図9および図10においてより詳しく説明している。一実施形態では、さらに、図14および図15で論じたように、キー画像からグレアも低減または除去することができる。図16Cに示すように、シーン1用のOCRキャッシュ1112は、新規画像およびシーン記述子で更新される。図16Bと同様、図

16Dは、新規シーンが視野内にある。シーン3を表す画像がキャプチャされ、OCRキャッシュ1112は新規画像で更新される。いくつかの実施形態では、キー画像は、たとえば図16A～図16Dに示すように、シーンが識別された順序で、細い帯状部に表示される。いくつかの実施形態では、キー画像の順序は、どのシーンが最も最近検出されたかに基づいて変わる。たとえば、最も最近検出されたシーンを表すキー画像が常に、細い帯状部の左端側に示され得る。さらにまたは代わりに、キー画像を表示する他の順序または技法が使われてもよい。

#### 【 0 1 2 0 】

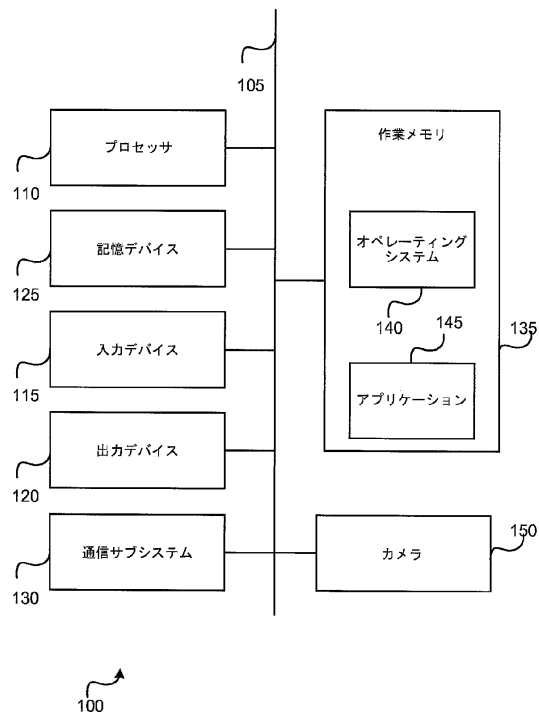
本明細書に記載したキャッシュ更新は、同じシーンまたは複数のシーンに関連付けられた情報すべてにユーザがアクセスできるようにするので、ユーザにとって有利であり得る。さらに、長い時間にわたる同じシーンについての複数の画像の間に、画像情報とのオーバーラップがあるので、本明細書に記載する技法は、シーンに関連付けられた最良の情報または新たに取得された情報を記憶すればよく、したがって、時間とともに結果をさらに精製する。本明細書に記載する技法に関連付けられた他の利点は、OCR結果を記憶するために必要とされるメモリ空間の削減、およびOCRキャッシュに対するヒット数の削減を含んでよく、それにより全体的システムにおける性能がより速くなる。

#### 【 符号の説明 】

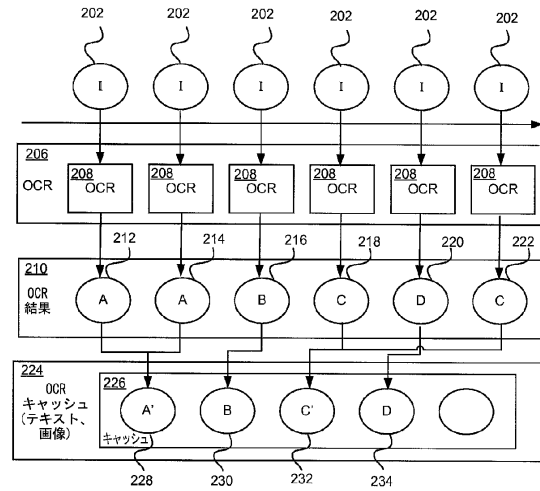
#### 【 0 1 2 1 】

100	コンピュータデバイス	
105	バス	20
110	プロセッサ	
115	入力デバイス	
120	出力デバイス	
125	記憶デバイス	
130	通信サブシステム	
135	作業メモリ	
140	オペレーティングシステム	
145	アプリケーションプログラム	
150	カメラ	
206	OCRエンジン	30
224	OCRキャッシュ	
230	キャッシュロケーション	
234	キャッシュロケーション	
1102	画像キャプチャモジュール	
1104	OCRエンジン	
1106	シーン記述子モジュール、シーン記述子生成器	
1108	照合モジュール	
1110	選択モジュール	
1112	OCRキャッシュ	
1114	統合モジュール	40
1116	挿入モジュール	
1118	デバイスディスプレイ	
1120	デバイスストレージ	
1124	OCRキャッシュシステム	
1222	追跡モジュール	
1402	デバイス	

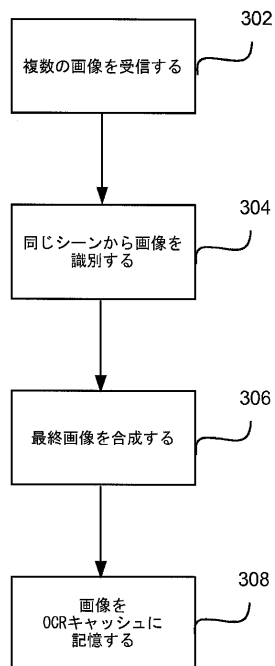
【図 1】



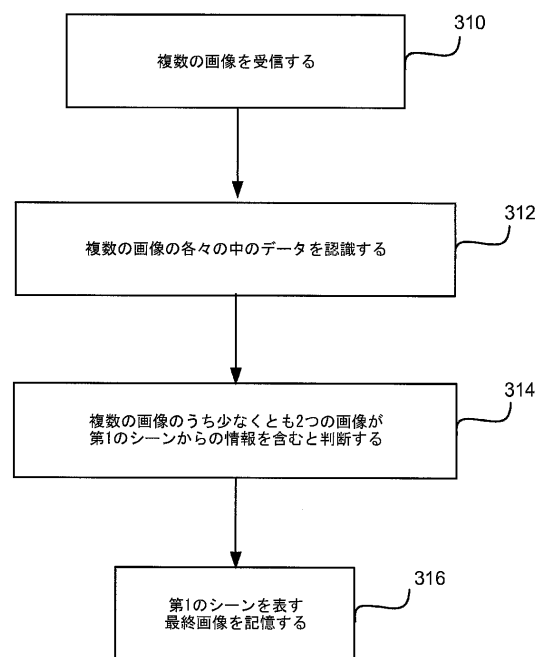
【図 2】



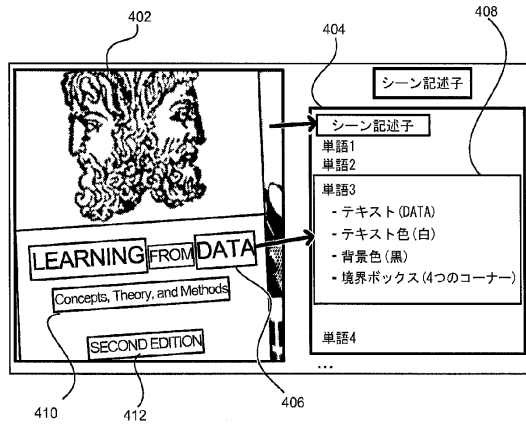
【図 3 A】



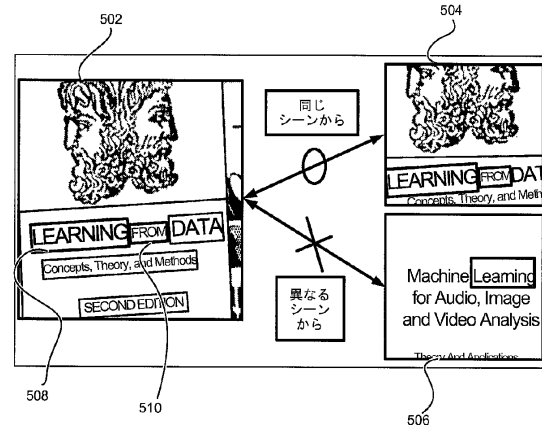
【図 3 B】



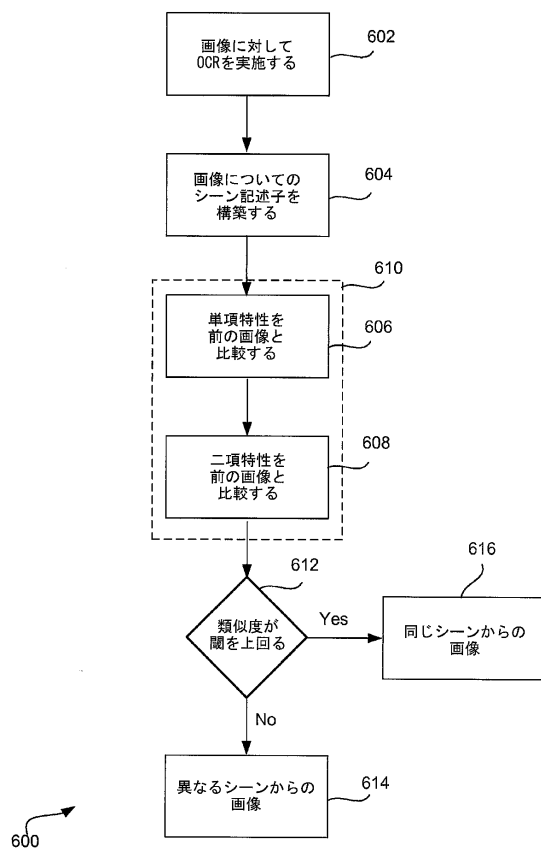
【図 4】



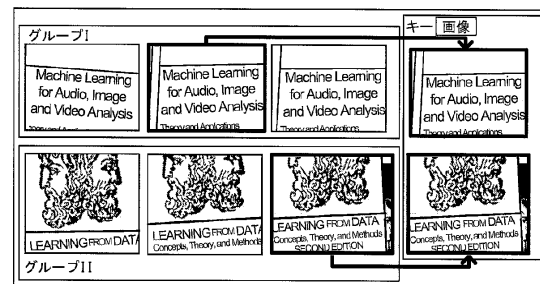
【図 5】



【図 6】

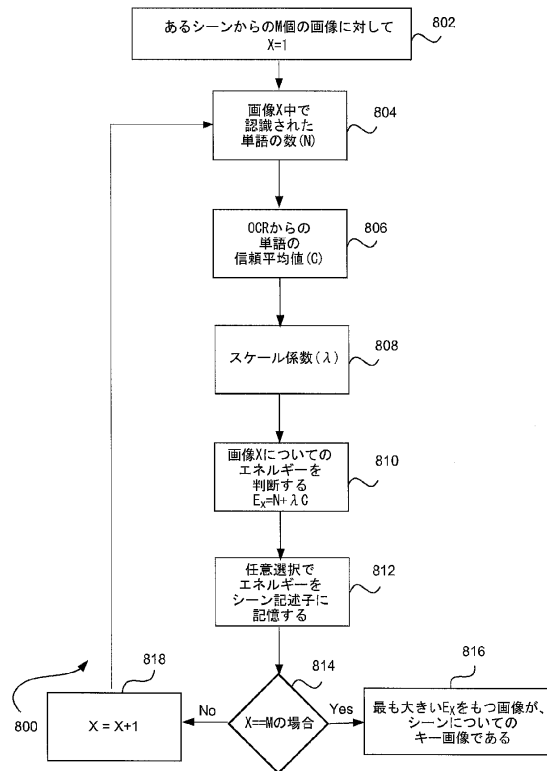


【図 7】





【図 8】



【図 9】

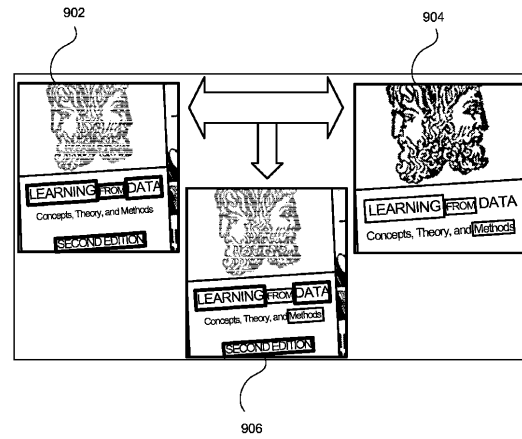
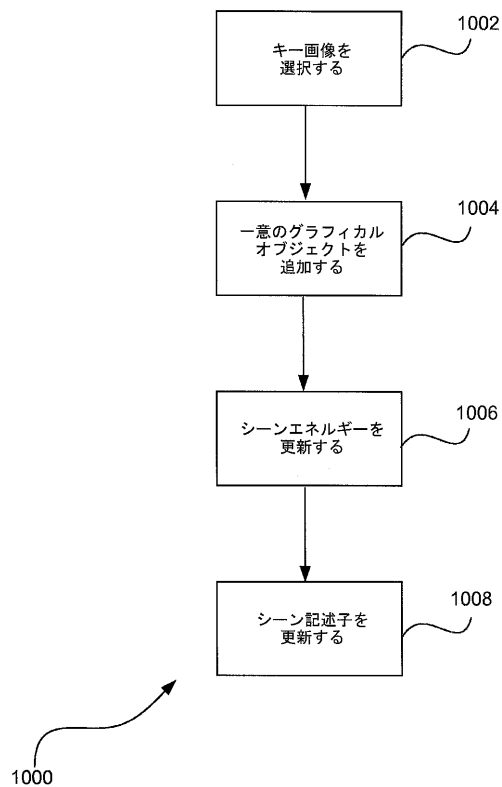
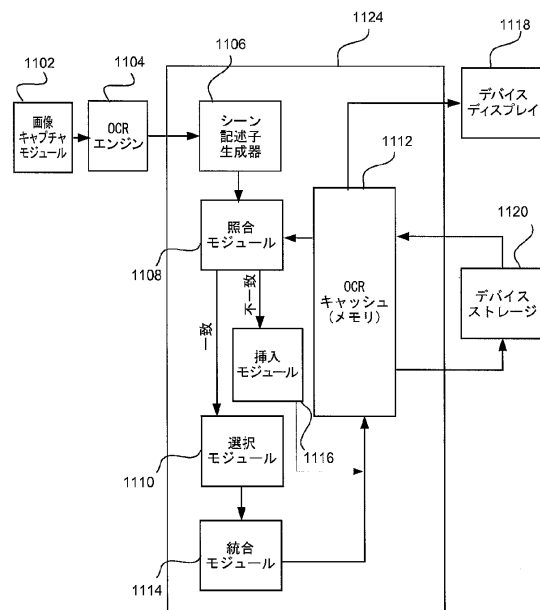


FIG. 9

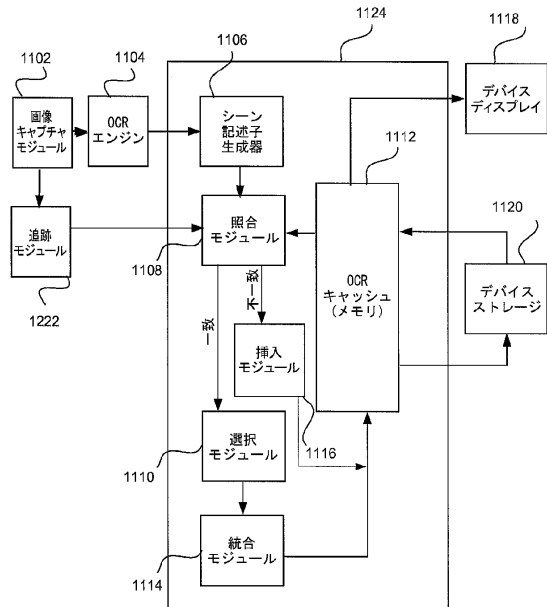
【図 10】



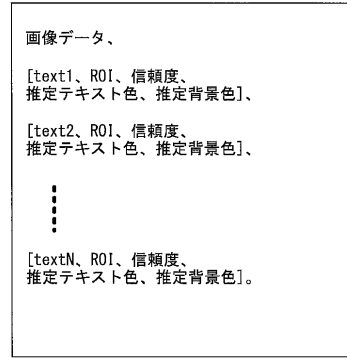
【図 11】



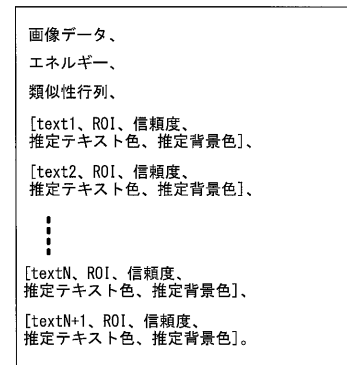
【 図 1 2 】



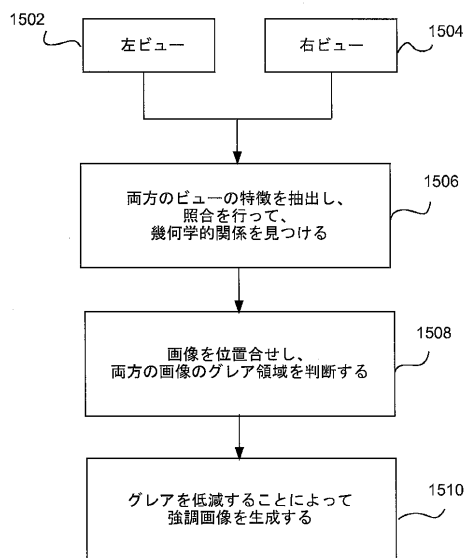
【 図 1 3 A 】



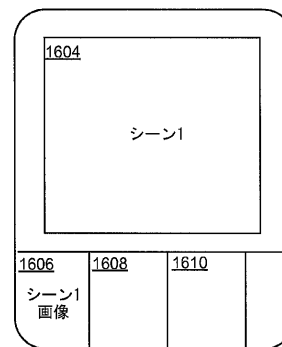
【 図 1 3 B 】



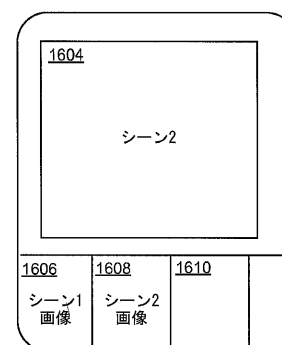
【 図 1 5 】



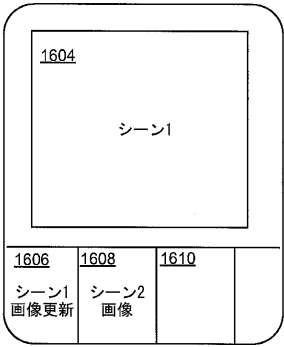
【 図 1 6 A 】



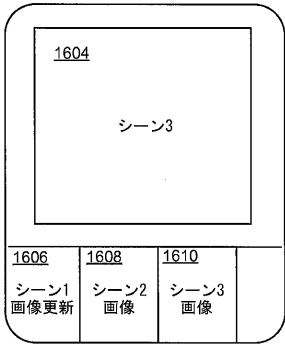
【 図 1 6 B 】



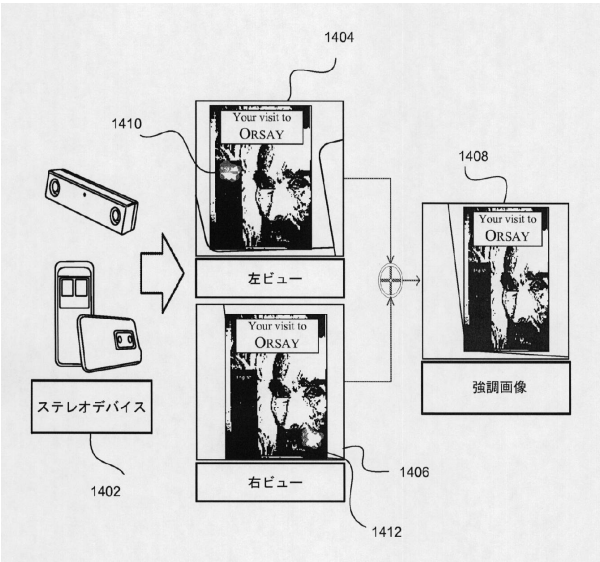
【図 16 C】



【図 16 D】



【図 14】



## フロントページの続き

- (72)発明者 ヒュン・イル・コー  
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
- (72)発明者 ドック・フン・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

審査官 真木 健彦

- (56)参考文献 特開２００３－０６０８９４（ＪＰ，Ａ）  
特開２０００－２７８５１４（ＪＰ，Ａ）  
再公表特許第２００６／０１１２５５（ＪＰ，Ａ１）  
特開２００７－０７９６７５（ＪＰ，Ａ）  
特開２００７－３３４５５８（ＪＰ，Ａ）  
米国特許出願公開第２００６／００１７８１０（ＵＳ，Ａ１）  
竹内 俊一，投影特徴を利用した階層的画像モザイク化手法による絵画等静止画像の高精細ディ  
ジタル化，電子情報通信学会論文誌 D - I I Vol.J83-D-II No.4，日本，社団法人電子情報通  
信学会，２０００年 ４月２５日，Vol.J83-D-II No.4，P.1090-1099，ISSN 0915-1923

## (58)調査した分野(Int.Cl.，ＤＢ名)

G 0 6 T	7 / 0 0	
G 0 6 T	1 / 0 0	
G 0 6 T	3 / 0 0	- 3 / 6 0
G 0 6 T	5 / 5 0	
G 0 6 K	9 / 0 0	- 9 / 0 3
G 0 6 K	9 / 4 6	- 9 / 5 2
G 0 6 K	9 / 6 2	- 9 / 8 2