

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6615054号
(P6615054)

(45) 発行日 令和1年12月4日(2019.12.4)

(24) 登録日 令和1年11月15日(2019.11.15)

(51) Int.Cl.

G06K 9/72 (2006.01)

F 1

G 06 K 9/72

Z

請求項の数 20 (全 19 頁)

(21) 出願番号 特願2016-123798 (P2016-123798)
 (22) 出願日 平成28年6月22日 (2016.6.22)
 (65) 公開番号 特開2017-21792 (P2017-21792A)
 (43) 公開日 平成29年1月26日 (2017.1.26)
 審査請求日 令和1年6月24日 (2019.6.24)
 (31) 優先権主張番号 14/794,479
 (32) 優先日 平成27年7月8日 (2015.7.8)
 (33) 優先権主張国・地域又は機関
米国(US)

早期審査対象出願

(73) 特許権者 518359845
 コンデュエント ビジネス サービシーズ
 エルエルシー
 アメリカ合衆国 O 7 9 3 2 ニュージャ
 ージー州 フローラム パーク キャンパ
 ス ドライブ 1 0 0 スイート 2 0 0
 (74) 代理人 100079049
 弁理士 中島 淳
 (74) 代理人 100084995
 弁理士 加藤 和詳
 (72) 発明者 アルバート・ゴード・ソルデヴィラ
 フランス共和国 グルノーブル 3 8 0 0
 O リュ・サン・ジャック 2 6

最終頁に続く

(54) 【発明の名称】 辞書を使わない、マッチングベースの単語画像認識

(57) 【特許請求の範囲】

【請求項 1】

画像の中の英数字文字を認識する方法であって、
 単語転写長の範囲内のあらゆる文字位置においてアルファベットのすべてのアルファベット文字の表現を計算することと、
 埋め込みアルファベット文字を作り出すためにマトリックスを使用して共通のベクトル部分空間に前記アルファベット文字の前記表現を埋め込むこと、

埋め込み英数字文字画像を作り出すために前記マトリックスを使用して前記共通のベクトル部分空間に英数字単語の受け取った画像を埋め込むことであって、前記埋め込み英数字文字画像は文字位置における画像文字を表すこと、

前記埋め込み英数字文字画像の各前記文字位置における前記埋め込みアルファベット文字のひとつひとつの有向非巡回グラフ(DAG)の効用値を計算することと、
 前記単語転写長の範囲内で、各前記文字位置における各々の前記埋め込みアルファベット文字の前記効用値に基づいて前記文字位置のひとつひとつに対して前記DAGに最高の効用の経路を見いだすことにより最高の効用の転写アルファベット文字を計算することと、

前記受け取った画像内で認識された英数字単語を作り出すために各前記文字位置に対して前記最高の効用の転写アルファベット文字を割り当てることと、

を含む、方法。

【請求項 2】

前記DAGはノードを有し、

10

20

各前記ノードは前記文字位置の1つにおいて前記埋め込みアルファベット文字の1つを表し、

各前記文字位置は前記埋め込みアルファベット文字の各々の異なる1つに対して異なるノードを含み、

前記ノード間のあらゆるエッジは各前記ノードを選択することの前記効用値およびコスト値を包含し、

前記DAG上の前記最高の効用の経路を前記見いだすことは各前記ノードを選択することとの前記効用値および前記コスト値に基づいて開始する、

請求項1に記載の方法。

【請求項3】

10

前記DAG上の前記最高の効用の経路を前記見いだすことは、前記DAGの各々の異なる経路の前記効用値および前記コスト値を累積することに基づいて開始する、請求項2に記載の方法。

【請求項4】

前記効用値および前記コスト値を前記累積することは、各々の前記経路に対して、

経路に対する合計効用値を作り出すために前記経路のノード間のすべてのエッジの効用値を合計することと、

前記経路に対する合計コスト値を作り出すために前記経路のすべての前記エッジのコストを合計することと、

前記合計コスト値の標準コスト値を算出することと、

20

前記合計効用値を前記標準コスト値で割ることと、

によって前記DAGの各々の経路のスコアを計算することを含む、請求項3に記載の方法。

【請求項5】

前記最高の効用の経路を前記見いだすことは、

全ての経路が調査される全数探索手法を使用することと、

最も有望な経路だけが調査される近似探索手法を使用することと、

のうちの1つを含む、請求項2に記載の方法。

【請求項6】

30

前記効用値を計算することは、前記埋め込みアルファベット文字の各々に前記画像文字の各々を転写することの効用を表す効用テンソルを作り出すために、前記埋め込みアルファベット文字画像の前記受け取った画像をテンソルで処理することを含む、請求項1に記載の方法。

【請求項7】

前記埋め込みアルファベット文字をテンソルに格納することをさらに含む、請求項1に記載の方法。

【請求項8】

画像の中の英数字文字を認識する方法であって、

画像プロセッサを使用して初期処理を自動的に行うことと、

前記初期処理の後に、前記初期処理を繰り返すことなく自動的に後続の処理を繰り返しを行うことと、

40

を含み、

前記初期処理は、

単語転写長の範囲内のあらゆる文字位置においてアルファベットのすべてのアルファベット文字の表現を計算することと、

埋め込みアルファベット文字を作り出すために、マトリックスを使用して共通のベクトル部分空間に前記アルファベット文字の前記表現を埋め込むことと、

前記埋め込みアルファベット文字を保存することと、

を含み、

前記後続の処理は、

50

前記画像プロセッサに作動的に接続された画像センサデバイスから英数字単語の電子画像を、前記画像プロセッサに、受け取ることと、

前記画像プロセッサを使用して、埋め込み英数字文字画像を作り出すために前記マトリックスを使用して前記共通のベクトル部分空間に前記電子画像を埋め込むことであって、

前記埋め込み英数字文字画像は文字位置における画像文字を表すことと、

前記画像プロセッサを使用して、前記埋め込み英数字文字画像の各前記文字位置における前記埋め込みアルファベット文字のひとつひとつの有向非巡回グラフ（DAG）の効用値を計算することと、

前記画像プロセッサを使用して、前記単語転写長の範囲内で、各前記文字位置における各々の前記埋め込みアルファベット文字の前記効用値に基づいて前記文字位置のひとつひとつに対して前記DAGに最高の効用の経路を見いだすことにより最高の効用の転写アルファベット文字を計算することと、

前記画像プロセッサを使用して、前記電子画像内で認識された英数字単語を作り出すために各前記文字位置に対して前記最高の効用の転写アルファベット文字を割り当て、前記認識された英数字単語の文字長を決定することと、

前記画像プロセッサから前記認識された英数字単語を出力することと、
を含む、
方法。

【請求項 9】

前記DAGはノードを有し、
各前記ノードは前記文字位置の1つにおいて前記埋め込みアルファベット文字の1つを表し、

各前記文字位置は前記埋め込みアルファベット文字の各々の異なる1つに対して異なるノードを含み、

前記ノード間のあらゆるエッジは各前記ノードを選択することの前記効用値およびコスト値を包含し、

前記DAG上の前記最高の効用の経路を前記見いだすことは各前記ノードを選択することの前記効用値および前記コスト値に基づいて開始する、

請求項8に記載の方法。

【請求項 10】

前記DAG上の前記最高の効用の経路を前記見いだすことは、前記DAGの各々の異なる経路の前記効用値および前記コスト値を累積することに基づいて開始する、請求項9に記載の方法。

【請求項 11】

前記効用値および前記コスト値を前記累積することは、各々の前記経路に対して、経路に対する合計効用値を作り出すために前記経路のノード間のすべてのエッジの効用値を合計することと、

前記経路に対する合計コスト値を作り出すために前記経路のすべての前記エッジのコストを合計することと、

前記合計コスト値の標準コスト値を算出することと、

前記合計効用値を前記標準コスト値で割ることと、

によって前記DAGの各々の経路のスコアを計算することを含む、請求項10に記載の方法。

【請求項 12】

前記最高の効用の経路を前記見いだすことは、

全ての経路が調査される全数探索手法を使用することと、

最も有望な経路だけが調査される近似探索手法を使用することと、

のうちの1つを含む、請求項9に記載の方法。

【請求項 13】

前記効用値を計算することは、前記埋め込みアルファベット文字の各々に前記画像文字

10

20

30

40

50

の各々を転写することの効用を表す効用テンソルを作り出すために、前記埋め込みアルファベット文字画像の前記電子画像をテンソルで処理することを含む、請求項 8 に記載の方法。

【請求項 1 4】

前記埋め込みアルファベット文字をテンソルに格納することをさらに含む、請求項 8 に記載の方法。

【請求項 1 5】

画像センサデバイスと、

前記画像センサデバイスに作動的に接続された画像プロセッサデバイスと、
を備えるシステムであって、

前記画像プロセッサデバイスは、

単語転写長の範囲内のあらゆる文字位置においてアルファベットのすべてのアルファベット文字の表現を計算することと、

埋め込みアルファベット文字を作り出すためにマトリックスを使用して共通のベクトル部分空間に前記アルファベット文字の前記表現を埋め込むことと、

前記埋め込みアルファベット文字を保存することと、

を含む初期処理を自動的に行い、

前記初期処理の後に、前記画像プロセッサデバイスは前記初期処理を繰り返すことなく後続の処理を繰り返し自動的に行い、前記後続の処理は、

前記画像センサデバイスから英数字単語の電子画像を受け取ることと、

埋め込み英数字文字画像を作り出すために前記マトリックスを使用して前記共通のベクトル部分空間に前記電子画像を埋め込むことであって、前記埋め込み英数字文字画像は文字位置における画像文字を表すことと、

前記埋め込み英数字文字画像の各前記文字位置における前記埋め込みアルファベット文字のひとつひとつの有向非巡回グラフ (DAG) の効用値を計算することと、

前記単語転写長の範囲内で、各前記文字位置における各々の前記埋め込みアルファベット文字の前記効用値に基づいて前記文字位置のひとつひとつに対して前記 DAG に最高の効用の経路を見いだすことにより最高の効用の転写アルファベット文字を計算することと、

前記電子画像内で認識された英数字単語を作り出すために各前記文字位置に対して前記最高の効用の転写アルファベット文字を割り当て、前記認識された英数字単語の文字長を決定することと、

を含み、

前記画像プロセッサデバイスは前記認識された英数字単語を自動的に出力する、
システム。

【請求項 1 6】

前記 DAG はノードを有し、

各前記ノードは前記文字位置の 1 つにおいて前記埋め込みアルファベット文字の 1 つを表し、

各前記文字位置は前記埋め込みアルファベット文字の各々の異なる 1 つに対して異なるノードを含み、

前記ノード間のあらゆるエッジは各前記ノードを選択することの前記効用値およびコスト値を包含し、

前記 DAG 上の前記最高の効用の経路を前記見いだすことは各前記ノードを選択するとの前記効用値および前記コスト値に基づいて開始する、

請求項 1 5 に記載のシステム。

【請求項 1 7】

前記 DAG 上の前記最高の効用の経路を前記見いだすことは前記 DAG の各々の異なる経路の前記効用値および前記コスト値を累積することに基づいて開始する、

請求項 1 6 に記載のシステム。

10

20

30

40

50

【請求項 18】

前記効用値および前記コスト値を前記累積することは、各々の前記経路に対して、経路に対する合計効用値を作り出すために前記経路のノード間のすべてのエッジの効用値を合計することと、

前記経路に対する合計コスト値を作り出すために前記経路のすべての前記エッジのコストを合計することと、

前記合計コスト値の標準コスト値を算出することと、

前記合計効用値を前記標準コスト値で割ることと、

によって前記DAGの各々の経路のスコアを計算することを含む、請求項17に記載のシステム。

10

【請求項 19】

前記最高の効用の経路を前記見いだすこととは、

全ての経路が調査される全数探索手法を使用することと、

最も有望な経路だけが調査される近似探索手法を使用することと、

のうちの1つを含む、請求項16に記載のシステム。

【請求項 20】

前記効用値を計算することは、前記埋め込みアルファベット文字の各々に前記画像文字の各々を転写することの効用を表す効用テンソルを作り出すために、前記埋め込みアルファベット文字画像の前記電子画像をテンソルで処理することを含む、請求項15に記載のシステム。

20

【発明の詳細な説明】**【技術分野】****【0001】**

本明細書のシステムおよび方法は、一般に自動化された認識システムに関し、より具体的には光学式文字認識(OCR)および類似のシステムに関する。

【背景技術】**【0002】**

単語画像およびテキスト文字列を、関連類似度測定基準を用いて共通の空間に表すことは、ユーザが単語画像の認識を検索問題として投げかけることを可能にし、単語画像を所与として、システムはすべての可能な転写(すなわち、辞書)を順位付けし、画像単語に最も類似したものを見出された転写として使用することができる。これは、システムが可能な有効な転写のセットに関して先進的な知識を有し、かつこのセットが小さい場合、非常に有用である。例えば、システムは、道路画像の中の店に関連したキーワードを見いだすことに、またはナンバープレートを認識することに専念があり、小さい有効にされたセットの中に正しい転写があることを知っている場合がある。

30

【発明の概要】**【発明が解決しようとする課題】****【0003】**

あいにく、可能な転写の空間が非常に莫大である、制約がないナンバープレート認識などの、他の重要な対象領域において、すべての可能な転写を符号化し順位付けすることは実行可能ではない。例証となる実施例として、

40

【0004】

$$36^8 \approx 3 * 10^{12}$$

【0005】

、3兆個の、数字およびラテンアルファベットの26文字だけを使用した8文字の可能なナンバープレートがある。192次元のテキスト特徴を使用してすべてのこれらのナンバープレート転写を表すことは、ほとんど2000テラバイトを必要とすることになり、ただ1つの画像とすべてのナンバープレート転写との間の類似度を計算することはおよそ32時間を要することになる。したがって、大域的特徴で単語画像を記述する利益を利用

50

するが、潜在的に無限の辞書でなお認識を行うことができる解決策が求められる。

【課題を解決するための手段】

【0006】

本明細書の例示的な方法は画像中の英数字文字を認識する。このような方法は、ある一定の単語転写長の範囲内のあらゆる文字位置においてアルファベットのあらゆる文字の個別の表現を計算（または事前計算）する。

【0007】

それから、これらの方法は、テンソルに格納され得る「埋め込みアルファベット文字」を作り出すために、（マトリックスを使用して）共通のベクトル部分空間に各々のアルファベット文字の個別の表現を埋め込む。これらの方法は、「埋め込み英数字文字画像」を作り出すために、（マトリックスを使用して）共通のベクトル部分空間に英数字単語の受け取った画像を同様に埋め込む。埋め込み英数字文字画像は、「文字位置」における「画像文字」を表す、またはこれらを暗黙的に符号化する。

【0008】

このような方法は、埋め込み英数字文字画像に対して、文字位置のひとつひとつにおける埋め込みアルファベット文字のひとつひとつの効用値を計算する。効用値を計算するプロセスは、各埋め込みアルファベット文字に各画像文字を転写することの効用を表す効用テンソルを作り出すために、受け取った電子画像をテンソルで処理することを含み得る。

【0009】

同様に、これらの方法は、（単語転写長の範囲内の）各々の文字位置における各々の埋め込みアルファベット文字の効用値に基づいて、各文字位置に対する最良の転写アルファベット文字を計算する。最良の転写アルファベット文字を計算するとき、これらの方法は有向非巡回グラフ（DAG）上の最良の経路を見いだす。DAGは、各ノードが文字位置の1つにおいて埋め込みアルファベット文字の1つを表すノードを有し、各文字位置は、各々の異なる埋め込みアルファベット文字に対して異なるノードを含む。ノード間のあらゆるエッジは、各ノードを選択することの効用値およびコスト値を包含する。したがって、本明細書の方法は、各ノードを選択することの効用値およびコスト値に基づいてDAG上の最良の経路を見いだす。これは、すべての経路が調査される全数探索手法を使用して、または最も有望な経路だけが調査される近似探索手法を使用して行うことができる。

【0010】

換言すれば、方法は、DAGの異なる経路の各々のエッジの効用値およびコスト値を累積することによって、DAG上の最良の経路を見いだす。効用値およびコスト値を累積するプロセスは、各々の前記経路に対して、その経路に対する合計効用値を作り出すために所定の経路のノード間のすべてのエッジの効用値を合計することと、経路に対する合計コスト値を作り出すためにその経路のすべてのエッジのコストを合計することと、合計コスト値の標準コスト値を算出することと、合計効用値を標準コスト値で割ることによって、DAGの各々の経路のスコアを算出することを伴う。

【0011】

それからこのような方法は、受け取った画像内の認識された英数字単語を作り出し認識された英数字単語の文字長を決定するために、各文字位置に対して最良の転写アルファベット文字を割り当てる。

【0012】

本明細書のシステムは、画像センサデバイス、および画像センサデバイスに作動的に接続された画像プロセッサデバイス（他の構成要素に加えて）同様に含む。画像プロセッサデバイスは、単語転写長の範囲内のあらゆる文字位置においてアルファベットのすべてのアルファベット文字の個別の表現を計算することと、埋め込みアルファベット文字を作り出すためにマトリックスを使用して共通のベクトル部分空間にアルファベット文字の個別の表現を埋め込むことと、埋め込みアルファベット文字を保存することによって初期処理を自動的に行う。

【0013】

10

20

30

40

50

初期処理の後に、画像プロセッサデバイスは、初期処理を繰り返すことなく、後続の処理を繰り返し自動的に行う。この後続の処理は、画像センサデバイスから英数字単語の電子画像を受け取ることと、埋め込み英数字文字画像を作り出すためにマトリックスを使用して共通のベクトル部分空間に電子画像を埋め込むことを含む。埋め込み英数字文字画像は、文字位置における画像文字を表す、またはこれらを符号化する。

【0014】

この後続の処理は、画像プロセッサデバイスを使用して埋め込み英数字文字画像に対して各文字位置における埋め込みアルファベット文字のひとつひとつの効用値を計算することと、単語転写長の範囲内で、各文字位置における各埋め込みアルファベット文字の効用値に基づいて文字位置のひとつひとつに対して最高の効用の転写アルファベット文字を計算することとを含む。後続の処理の間に、画像プロセッサデバイスは、電子画像内で認識された英数字単語を作り出すために各文字位置に対して最高の効用の転写アルファベット文字を割り当て、認識された英数字単語の文字長を決定する。画像プロセッサデバイスは認識された英数字単語を出力する。

【0015】

これらのおよび他の特徴は、以下の詳細な説明で、説明され、または明らかになる。

【0016】

さまざまな例示的なシステムおよび方法は、添付の図面を参照して、下に詳細に説明される。

【図面の簡単な説明】

【0017】

【図1】図1は、本明細書の方法およびデバイスによって使用される有向非巡回グラフ(DAG)である。

【図2】図2は、本明細書のさまざまな方法の流れ図である。

【図3】図3は、本明細書のシステムを例示する概略図である。

【図4】図4は、本明細書のデバイスを例示する概略図である。

【図5】図5は、本明細書のデバイスを例示する概略図である。

【発明を実施するための形態】

【0018】

上述のように、単語認識を行うために辞書を必要とする多くの大域的画像マッチングフレームワークがある。一般に、これらの手法は、最初に、d次元の共通の実部分空間に、画像および文字列をそれぞれ埋め込む、2つの埋め込み関数

【0019】

【数1】

$$\phi : \mathcal{I} \rightarrow \mathbb{R}^d$$

【0020】

および

【0021】

【数2】

$$\psi : \mathcal{S} \rightarrow \mathbb{R}^d$$

【0022】

を定義する。これらの関数は、ラベル付けされた訓練データを使用して学習され、それから画像xとテキスト文字列yとの間の類似度が埋め込み間のドット積として定義される、すなわち、

$$s(x, y) = (x)^T (y) \quad (1)$$

である。

【0023】

10

20

30

40

50

辞書

【0024】

【数3】

\mathcal{Y}

【0025】

を考えると、辞書のすべての単語を順位付けして最大の類似度を有するものを使用することによって、単語画像 x の最適な転写を見いだすことができる。すなわち、

【0026】

【数4】

10

$$\hat{y} = \operatorname{argmax}_{y \in \mathcal{Y}} s(x, y) = \operatorname{argmax}_{y \in \mathcal{Y}} \phi(x)^T \psi(y) \quad (2)$$

【0027】

である。

【0028】

最初に、オフラインで辞書中のすべての文字列の埋め込みを計算し、それから、オンラインで辞書のすべてにわたって探索を行い転写する単語を与えることを必要とするので、これは辞書

【0029】

20

【数5】

\mathcal{Y}

【0030】

のサイズが非常に大きいとき問題を呈する。辞書が大きいときこれは非実用的になる場合があり、莫大な辞書で（例えばナンバープレート認識のために）、直接これを適用することは実行不可能である。辞書が完全でないとき、辞書順位付け手法は同様に問題になる。すなわち、単語および完全に順位付けされた辞書を所与として、一番上の結果が（辞書が転写を包含するので）正しいか（辞書が転写を包含しないので）間違っているかを決めて、辞書にない可能性がある単語を認識しようとなることになる。

30

【0031】

本明細書のシステムおよび方法によって提供される解決策において、問題は順位付け方法の代わりに最適化方法を使用することによって取り組まる。画像を所与として、これらのシステムおよび方法は、この類似度を最大にする、転写長および転写の個別の文字を決定する。単語が包含し得るすべての可能な文字を包含するアルファベット（例えば、この場合サイズ $| \mathcal{A} | = 36$ の $\mathcal{A} = \{ A, \dots, Z, 0, \dots, 9 \}$ ）を考慮する場合、 c_i で、

【0032】

【数6】

$$\hat{y} = \operatorname{argmax}_{l, c_1, \dots, c_l} s(x, \{c_1, \dots, c_l\}) = \operatorname{argmax}_{l, c_1, \dots, c_l} \phi(x)^T \psi(\{c_1, \dots, c_l\}) \quad (3)$$

40

【0033】

である。式の最適化を可能にするこの中の特徴は、関数 s が、 ϕ によって作り出される最終的な表現の中で単語の各々の文字が有する（非線形の）影響力を示すために分解されることである。実際には、本明細書のシステムおよび方法は、式の類似度を 2 つの量の比率として書き直すことができるることを示す。分子は、単語中のすべての文字に対して、 x の i 番目の文字を c_i として転写することの個別の効用を符号化し、認識するべき単語に依存する。分母は、単語中のすべての文字に対して、単語の i 番目の文字を c_i として転写するために発生したコストを符号化し、認識するべき単語に依存しないので、それは

50

すべての可能な値に対してオフラインで容易に事前計算することができる。いったん認識するべき単語が提供されると、分子は速やかにオンラインで事前計算することができ、類似度の評価を非常に効率的にすることができます。この手法は、中間結果を記憶することによって、明示的に (y) を計算することなく画像とテキストとの間の類似度を速やかに計算することを可能にする。しかしながら、探索空間はまだ劇的に大きく、最も可能性が高い転写を見いだすことはまだ実行不可能である。本明細書のシステムおよび方法は、分枝限定技術に類似した、探索アルゴリズムの近似バージョンを提供する。この実施例の場合、探索空間の最も有望な部分だけが調査されて、膨大なスピードアップをもたらし、そしてこれはすべての個別の文字の効用を活用することによって達成することができる。この手法は、正確な解を保証しないけれども、この近似による精度の損失は無視できることを示す。したがって、本明細書のシステムおよび方法は、大域的画像表現手法を使用して辞書なしで単語を認識する能力を提供する。

【0034】

さらに本明細書のシステムおよび方法は、同様に、速度およびメモリ効率に役立つ。辞書が既知であるときでさえ、大きな辞書に対して、提案された類似度を使用することは、辞書のひとつひとつの単語を符号化し、認識するべきあらゆる新しい単語に対して辞書のすべてを通して探索することをユーザに要求する、式(1)の単純な手法を直接使用するより効率的である。

【0035】

システムおよび方法は同様に「モデルイントロスペクション」を可能にする。訓練時に単語画像の最も可能性が高い転写を作り出すことによって、モデルがどこでテスト辞書によって限定されずに失敗しているか気づくことができる。これは、基礎となっているモデルが認識目的用であるかのように良くはないという事実を強調し、そしてこれらの辞書なしの転写は、ハード・ネガティブ・マイニング、新しい訓練単語の統合など、モデルを改善するために潜在的に使用され得る。

【0036】

本開示は、用いられた 関数を解析することによって開始する。この関数の目標は、辞書編集上の類似度が保たれるように、ベクトル空間にテキスト文字列を埋め込むことである（すなわち、レーベンシュタイン距離などの単語距離を使用してより接近している単語はベクトル空間で接近しているべきである）。このタスクのために、他の大域的画像フレームワークは、26字のラテン文字プラス10個の数字で構成されるアルファベットを使用して単語の文字のヒストグラムを本質的に計算する、文字の空間ピラミッド(SPOC)を提供する。他の方法は、同じ埋め込みを計算する、文字のピラミッドヒストグラム(PHOC)を最初に提供し得る。これらの表現は、「聞く(1 i s t e n)」および「静かな(s i l e n t)」などの単語は同じ埋め込みを有することになるから、しかしながら判別可能ではない。これに対処するために、ヒストグラムはマルチレベルで計算される。例えば、レベル2において、2つの別個のヒストグラムを作り出して連結することになる。単語の前半の文字のヒストグラムを符号化するものおよび単語の後半の文字のヒストグラムを符号化するものである。レベル3は3つのヒストグラムを作り出す、などとなる。文字の一部だけが領域に現れる（例えば、3文字の単語を2つの領域に分割する）場合、文字は両方のヒストグラムに柔軟に割り当てされる。

【0037】

しかしながら、若干の微妙な相違にもかかわらず、SPOCおよびPHOCは、本質的に同等である。

【0038】

【数7】

$$h : \mathcal{S} \rightarrow \mathbb{R}^D$$

【0039】

で示されるこの実施例において、この埋め込み関数は所定の数のレベルを使用する。レ

10

20

30

40

50

ペル 2 から ペル 8 を使用する場合、これは $(2 + 3 + 4 + 5 + 6 + 7 + 8) * 36 = 1260$ 次元をもたらす。1つの興味深い特性は、テキスト文字列の埋め込みは付加的であり、(それらの位置を保っている) 個別の文字の埋め込みの合計に分解され得ることである。すなわち、記述法をわずかに乱用すると、

【0040】

【数8】

$$h(y) = \sum_{i=1}^{|y|} h(y_i, i, |y|) \quad (4)$$

10

【0041】

であり、式中 y_i は y の i 番目の文字であり、 $|y|$ は y の長さである。その絶対位置を与えた文字の相対的位置付けは単語の長さに依存するから、 h を計算するために単語の長さが必要であることに留意されたい(例えば、長さ 4 の単語の 3 番目の文字は単語の後半にあるが、長さ 8 の単語の 3 番目の文字は単語の前半にある)。 h を導入後に、本明細書のシステムおよび方法は今、 h の関数として ψ を定義することができる。他の大域的画像マッチングソリューションにおいて、 ψ は単に l_2 正規化された h 、すなわち、

【0042】

【数9】

$$\psi(y) = \frac{h(y)}{\|h(y)\|}$$

20

【0043】

である。

【0044】

他の大域的画像マッチングソリューションにおいて、 h は同様に別の部分空間に投影され、

【0045】

【数10】

$$\psi(y) = \frac{W^T \frac{h(y)}{\|h(y)\|}}{\|W^T \frac{h(y)}{\|h(y)\|}\|} = \frac{W^T h(y)}{\|W^T h(y)\|}$$

30

【0046】

である。式(1)を思い出すと、数式は今、

【0047】

【数11】

$$s(x, y) = \phi(x)^T \psi(y) = \frac{\phi(x)^T W^T h(y)}{\|W^T h(y)\|} = \frac{\sum_{i=1}^{|y|} \phi(x)^T W^T h(y_i, i, |y|)}{\|\sum_{i=1}^{|y|} W^T h(y_i, i, |y|)\|} \quad (5)$$

40

【0048】

として書き直すことができる。

【0049】

最も興味深いことに、ユーザが単語の最大長を制約する(例えば、 L 文字以下の単語だけ認識するようにする)場合、本明細書のシステムおよび方法は、 W でそれらを投影する h のすべての可能な値を容易に事前計算し、サイズ $|y| \times L \times L \times d$ のテンソルにそれらを格納することができ、1と $|y|$ との間のすべての i 、1と L との間のすべての j 、および1と L との間のすべての l に対して、

$$T(i, j, l) = W^T h(i, j, l) \quad (6)$$

である。システムおよび方法が $L = 20$ 文字の最大長および 192 の次元数 d を想定す

50

る場合、これは、格納するために 2 M b のメモリだけを必要とし、ただ 1 度だけ計算する必要がある、 $36 \times 20 \times 20 \times 192$ のテンソル T をもたらす。さらに、認識するべき画像 x を考えると、本明細書のシステムおよび方法はその埋め込み (x) を事前計算しそれをテンソル T と掛け算することができ、単語の i 番目の文字を単語が長さ l であると想定する 中のシンボルのいずれかとして転写することの効用を表す、サイズ | | × L × L の効用テンソル U_x をもたらす。1 と | | との間のすべての i、1 と 1 との間のすべての j、および 1 と L との間のすべての l に対して、

$$U_x(i, j, l) = (x)^T T(i, j, l) \quad (7)$$

である。この場合、式はさらに、

【0050】

【数12】

$$s(x, y) = \frac{\sum_{i=1}^{|y|} \phi(x)^T W^T h(y_i, i, |y|)}{\|\sum_{i=1}^{|y|} W^T h(y_i, i, |y|)\|} = \frac{\sum_{i=1}^{|y|} U_x(y_i, i, |y|)}{\|\sum_{i=1}^{|y|} T(y_i, i, |y|)\|} \quad (8)$$

【0051】

として書き直すことができる。

【0052】

式 (3) を思い出すと、システムおよび方法は、転写の最も可能性が高い文字および長さを見いだす問題の中に、画像 x の最も可能性が高い転写を見いだす問題に対する解決策を提供し、それを式 (8) の類似度と組み合わせることにより、システムおよび方法は、

【0053】

【数13】

$$\hat{y} = \underset{l, c_1, \dots, c_l}{\operatorname{argmax}} s(x, \{c_1, \dots, c_l\}) = \underset{l, c_1, \dots, c_l}{\operatorname{argmax}} \frac{\sum_{i=1}^l U_x(c_i, i, l)}{\|\sum_{i=1}^l T(c_i, i, l)\|} \quad (9)$$

【0054】

を取得する。

【0055】

簡単のために、1 L のあらゆる値に対して独立して式 (3) を解き、それから最も大きい類似度を有する解をとることを考慮してもよい。転写長 l を考えると、この問題は有向巡回グラフ (DAG) の最良の経路を見いだすこととして見ることができ、グラフの各々のノードは単語中の所定の位置でアルファベットの所定の文字を表し、あらゆるエッジは選ばれたノードを選択することの効用およびコストを包含する。図 1 を参照されたい。この最良の経路は、深さ優先のやり方であらゆる位置においてあらゆる文字の影響力を漏れなく評価し、効率について各々のノードにおける中間の効用およびコストを累積することによって、再帰的に見いだすことができる。この方法は、所定の長さまですべての可能な単語を調査し、したがって、より効率的な方法にもかかわらず、明示的に辞書単語を作り出し、でそれらを符号化し、式 (1) および (2) を使用して取得するであろうものとまったく同じ解を取得する。

【0056】

ただ最良だけでなく N 個の最も良い解を作り出すために、これを同様に一般化することができる。これは、容量 N の最大ヒープを使用し、新しい解のスコアが最大ヒープによって現在保持される最悪の解のスコアより大きいときはいつでも現在の解およびそのスコアをプッシュすることによって、ほとんど余分のコストなしで達成することができる。1つより多い解を取得することは、より強力な記述で最良の解を再度順位付けするのに有用である場合がある。例えば、バイグラムを同様に符号化するテキスト埋め込みを使用することを考慮してもよい。バイグラムは、連続した文字を交換することに起因する誤りに役立つ場合があるが、式 (4) で記述された付加的な特性はもう保持されないことになるので

10

20

30

40

50

残念ながら提案されたフレームワークに容易に統合することができない。しかしながら、結果は、さらに結果を改善するために再順位付け段階で容易に使用され得る。

【0057】

前の段落で説明されたアルゴリズムは、ユーザが正確かつ効率的なやり方で、長さ 1 の単語の最も可能性が高い転写を見いだすことを可能にする。しかしながら、探索空間の指數関数的な性質を考えると、単語の潜在的な長さが増加するとき、5 または 6 より大きい 1 の値に対してこのアルゴリズムを適用することは困難である。本明細書で説明されたシステムおよび方法は、最適化プロセスを劇的にスピードアップする近似探索を提供する。この中の特徴は、図 1 の D A G を調査するとき、本明細書のシステムおよび方法は若干のノードが最良の経路によって使用されないのであろうことを速やかに見いだすことができ、それでそれらをグラフから除去することである。直感的に、これは各々のレベルにおいて調査される文字の数を減少させることを暗示する。枝刈戦略を適用することによって、本明細書のシステムおよび方法は、精度を保つと同時に探索空間を大幅に減少させることができる。探索空間を減少させる 1 つの明らかな手法は、分枝限定戦略を適用することである。しかしながら、式(9)の分母で導入された非線形性は、最適解が除去されないことを保証すると同時にグラフで効果的にノードを枝刈りするために厳しい上限を見いだすことを困難にする。本明細書のシステムおよび方法は、探索空間を減少させるために高速のヒューリスティックを提供する。10

【0058】

本明細書のシステムおよび方法は、単語長を考えると、 $|T(i, j, l)|$ の値が限定された分散を有し、そのため効用 U_x が最終的なスコアの重みの大部分を担うことに気づく。したがって、ヒューリスティックとして、所定のレベルにおいて、この開示は、 U_x に従って K 個の最も見込みがある文字ノードだけを調査する選択をした。ここで K は、探索空間減少とヒューリスティックの精度との間のトレードオフを制御する。K の大きい値は、見いだされた解はまだ最適であることを確実にするが、探索空間の減少は限定されることになる。K の小さい値は最適解を捨てる場合があるが、探索空間を著しく減少させることになる。20

【0059】

前の定式化は辞書なしの認識の場合に焦点を合わせる。しかしながら、手法は、辞書についての若干の情報が前もって知られている場合に対処するように拡張することができる。本明細書のシステムおよび方法は 2 つのシナリオに焦点を合わせることができる。第 1 のシナリオにおいて、辞書は前もって完全に知られており、既知の、適度に大きい辞書に対してすべての辞書単語の埋め込みを事前計算することおよび順位付けを行うことはまだ可能とすることができますが、非実用的である。本明細書のシステムおよび方法は、既知の辞書を取り扱うために提案された手法を修正することができる。この中の特徴は、辞書を考えると、本明細書のシステムおよび方法は、最初にテキスト単語を接頭辞木またはトライ木に符号化し、そして次に、認識時に、辞書の中の単語をもたらす文字だけを各々のレベルで調査するためにこの情報を使用することができるということである。直接的な手法に対するこれの利点は、明示的に辞書を符号化する必要がないということの他に、あらゆる接頭辞に対する増分の効用およびコストが記憶され、したがって、それらがあらゆる辞書単語に対して暗黙的に再計算されないことになるということであろう。30

【0060】

第 2 のシナリオは、単語のパターンが既知であるということである。認識する単語のタイプの高度な知識を有する場合もある。例えば、ナンバープレートの場合、次のパターン、3 つの数字 - 3 つの欧字 - 2 つの数字、を満たす転写を探索することができる。再び、これは、対応するノードだけを展開することによって、非常に容易に対処され得る。この実施例について、本明細書のシステムおよび方法は、最初の 3 つのレベルに対して数字だけを、次の 3 つに対して欧字だけを、そして最後の 2 つに対して再び数字だけを展開することになる。これは探索空間を著しく減少させ、したがって正しい転写を見いだすために必要とされる時間を著しく減少させる。40

【 0 0 6 1 】

式(9)の最適化問題は、長さ1を考えると、開始から終了ノードまでの有向非巡回グラフ(DAG)上の最良の経路を見いだすこととして図1で見ることができる。グラフ中のあらゆる黒エッジは、それに関連付けられた効用値およびコストベクトルを有し、所定のレベルにおいて所定の文字を選択することの効用およびコストを表し、単語画像およびテキスト埋め込みに依存する。グレーエッジは同期目的のためであり、ゼロの効用およびコストを有する。完全な経路のスコアは、経路が使用するエッジのコストの合計のノルムで割った効用の合計として測定される。単純に最良の経路を取得するために、本明細書のシステムおよび方法は、すべての可能な組合せを漏れなく探索する。効率目的で、効用およびコストの現在の合計を各々のノードにおいて累積することが可能である。最良の経路はそれから、深さ優先のやり方でノードを調査することによって再帰的に見いだすことができる。アルゴリズム(1)を参照されたい。10

【 0 0 6 2 】

図2は、画像中の英数字文字を認識する本明細書の例示的な方法を例示するフローチャートである。英数字文字を認識する前に、これらの方法は、項目150～項目154で初期処理(例えば、前処理)を行う。具体的には、項目150で、これらの方法は、ある一定の単語転写長の範囲内のあらゆる文字位置においてアルファベットのあらゆる文字(例えば、A～Zおよび0～9、ならびに潜在的に同等な句読点およびヌル文字、など)の個別の表現を計算(または事前計算)する。ヌル文字は、空白スペースまたはアルファベットの範囲内でない文字を表す。それから、項目152で、これらの方法は、「埋め込みアルファベット文字」を作り出すために、(マトリックスを使用して)共通のベクトル部分空間に各々のアルファベット文字の個別の表現を埋め込む。これらの方法はそれから、項目154で、例えば、テンソルに、埋め込みアルファベット文字を保存する。20

【 0 0 6 3 】

(その後繰り返される必要がない)初期処理を行った後で、これらの方法は、項目154で格納された埋め込みアルファベット文字を使用して、その後受け取った画像について(項目156～項目164で)後続の処理を繰り返し行う。具体的には、項目156で、これらの方法は、画像プロセッサに何らかの方法で作動的に(直接的に、または、間接的に、を意味する)接続された画像センサデバイス(例えば、カメラ、スキャン、光センサ、など)から取得した電子画像を(上述の画像プロセッサまたは別の画像プロセッサの中に)受け取る。換言すれば、カメラまたはスキャナは英数字単語の電子(例えば、デジタル)画像を取得することができ、その画像は(例えば、ネットワークを通して、またはデバイス内で内部的に)画像プロセッサに転送することができる。30

【 0 0 6 4 】

それから、項目158で、これらの方法は、「埋め込み英数字文字画像」を作り出すために、(同じマトリックスを使用して)同じ共通のベクトル部分空間に英数字単語の受け取った電子画像を埋め込む。これは、初期処理を行った画像プロセッサ、または格納された埋め込みアルファベット文字にアクセスを有する任意の他の画像プロセッサを用いて行うことができる。専門用語を簡単にするために、埋め込み英数字文字画像は、「文字位置」における「画像文字」を表す、またはこれらを符号化すると言ふことができる。40

【 0 0 6 5 】

このような方法はそれから、(項目160で)、埋め込み英数字文字画像に対して、文字位置のひとつひとつにおける、事前計算された埋め込みアルファベット文字のひとつひとつの効用値を計算するために、画像プロセッサを使用する。換言すれば、アルファベット(A～Zおよび0～9、ならびに潜在的に同等な句読点およびヌル文字、など)の各々の英数字文字の埋め込み表現の効用値は、画像のあらゆる可能な文字位置に対して計算される。例えば、項目160で効用値を計算するプロセスは、各埋め込みアルファベット文字に各画像文字を転写することの効用を表す効用テンソルを作り出すために、受け取った画像をテンソルで処理することを含み得る。

【 0 0 6 6 】

10

20

30

40

50

次に、項目 162 で、これらの方法は、（単語転写長の範囲内の）各々の文字位置における各々の埋め込みアルファベット文字の効用値に基づいて、各文字位置に対して「最高の効用」の転写アルファベット文字（または句読点もしくはヌル文字）を計算する。項目 162 で最高の効用の転写アルファベット文字を計算するとき、これらの方法は有向非巡回グラフ（DAG）上の最高の効用の経路を見いだす。DAG は、各ノードが文字位置の 1 つにおいて埋め込みアルファベット文字の 1 つを表す、ノードを有し、各文字位置は、各々の異なる埋め込みアルファベット文字に対して異なるノードを含む。ノード間のあらゆるエッジは、各ノードを選択することの効用値およびコスト値を包含する。したがって、項目 162 で、本明細書の方法は、各ノードを選択することの効用値およびコスト値に基づいて DAG 上の最高の効用の経路を見いだす。これは、すべての経路が調査される全探索手法を使用して、または最も有望な経路だけが調査される近似探索手法を使用して行うことができる。10

【0067】

換言すれば、項目 162 で、方法は、DAG の異なる経路の各々のエッジの効用値およびコスト値を累積することによって、DAG 上の最高の効用の経路を見いだす。項目 162 で効用値およびコスト値を累積するプロセスは、各々の前記経路に対して、その経路に対する合計効用値を作り出すために所定の経路のノード間のすべてのエッジの効用値を合計することと、経路に対する合計コスト値を作り出すためにその経路のすべてのエッジのコストを合計することと、合計コスト値の標準コスト値を算出することと、合計効用値を標準コスト値で割ることによって、DAG の各々の経路のスコアを算出することを伴う。20

【0068】

それから、項目 164 で、このような方法は、受け取った画像内の認識された英数字単語を作り出し認識された英数字単語の文字長を決定するために、各文字位置に対して最高の効用の転写アルファベット文字を割り当てる。項目 164 で、最高の効用スコアを有するアルファベット文字（または句読点もしくはヌル文字）は、所定の文字位置に対して正しく認識された文字であると決定される。同様に、最高の効用スコアを有するヌル文字は連続的な文字の連鎖の中の区切りを示し、そしてそれは連続的な文字の異なる単語間の境界を決定し、対応して項目 164 で連続的な文字の各認識された英数字単語の単語長を決定する。認識された英数字単語は、項目 166 でプロセッサから出力され、処理は、項目 156 で文字認識のために次の画像を受け取るように戻る。30

【0069】

本明細書で説明されたハードウェアは、解決策がより速く、（すなわち、計算を行うためのコンピュータの利用を通して）達成されるのを許すためのメカニズムとして単に機能するよりむしろ、前述の方法が行われるのを許すこと重要な役割を果たす。

【0070】

当業者によって理解されるように、本明細書で説明されたプロセスは、人だけ（またはペンとメモ帳で働いている人）によって行うことはできず、その代わりに、このようなプロセスは（デジタル画像データを処理し、このようなデータを評価する必要がある速度を考慮するときに）、ただ機械によって行うことができる。具体的には、電子画像データを受け取る、デジタル画像データを処理するなどのプロセスは、異なる専用機械の利用を必要とし、このようなデータは電子情報の形であるから、人はこのようなデータにアクセスすることはできないことになる。40

【0071】

さらに、電子データの処理はプロセスの各々のステップを伴うから、このような機械だけのプロセスは、単なる「解決後の活動」ではない。デジタル画像は、画像プロセッサによって処理されるために導線を通じてまたはワイヤレスネットワークを通じて移動しなければならないから、同様に、データ処理 / 伝送は、本明細書の方法によって行われるプロセスと一体であって、単なる解決後の活動ではない。換言すれば、方法は機械なしで行うことができない（そして人だけによって行うことができない）から、これらのさまざま50

機械は本明細書の方法と一体である。

【0072】

さらに、本明細書の方法は、多くの高度に複雑な技術的問題を解決する。例えば、上述のように、光学式文字認識プロセスは辞書に結び付けられている。本明細書の方法は、辞書を提供する必要性を回避することによってこの技術的問題を解決し、その代わりに各々の文字位置は、既知の単語長なしで、アルファベット全体に対して評価され、それは辞書の必要性を回避する。このことは処理を単純化し、それによって画像プロセッサが維持しなければならない電子記憶装置の量を減少させ、画像プロセッサに同じ量の時間でより多くのタスクを完遂させるように処理リソースを解放する（画像プロセッサの処理速度を改善する）。このような利益を付与することによって、本明細書の方法は、必要とされるハードウェアの量および複雑度を減少させ、それによって提供者が今日経験する実質的な技術的問題を解決する。10

【0073】

図3に示したように、本明細書の例示的なシステムおよび方法は、さまざまな異なる物理的位置206に位置しているさまざまなコンピュータ化デバイス200、204を含む。コンピュータ化デバイス200、204はサーバ、光デバイス、パーソナルコンピュータなどを含み、局所的なまたは広域の（有線またはワイヤレス）ネットワーク202経由で（互いに作動的に接続され）通信している。

【0074】

図4は、本明細書のシステムおよび方法で使用することができ、例えば、サーバ、パーソナルコンピュータ、携帯用計算機器などを備え得る、コンピュータ化デバイス200を例示する。コンピュータ化デバイス200は、コントローラ／有形プロセッサ216、ならびに、有形プロセッサ216におよびコンピュータ化デバイス200の外部のコンピュータ化ネットワーク202に作動的に接続された通信ポート（入力／出力）214を含む。同様に、コンピュータ化デバイス200は、グラフィカル・ユーザ・インターフェース（G U I）アセンブリ212などの、少なくとも1つの付属機能要素を含み得る。ユーザは、グラフィカル・ユーザ・インターフェースまたは制御パネル212からメッセージ、命令、およびメニューオプションを受け取り、グラフィカル・ユーザ・インターフェースまたは制御パネル212を通して命令を入力することができる。20

【0075】

入力／出力デバイス214は、コンピュータ化デバイス200への通信およびコンピュータ化デバイス200からの通信に使用され、（現在周知であるか将来開発されるかにかかわらず、任意の形の）有線デバイスまたはワイヤレスデバイスを備える。有形プロセッサ216は、コンピュータ化デバイスのさまざまなるまいを制御する。（光学的、磁気的、コンデンサベースなどとすることができます、一時的信号と異なる）非一時的な、有形の、コンピュータ記憶媒体デバイス210は、有形プロセッサ216によって読み取り可能であり、コンピュータ化デバイスが、本明細書で説明したものなどの、そのさまざまな機能を行うことを可能にするために、有形プロセッサ216が実行する命令を格納する。したがって、図4に示したように、本体ハウジングは、電源機構218によって交流（A C）源220から供給される電力で動作する、1つまたは複数の機能要素を有する。電源機構218は、普通の電力変換ユニット、蓄電素子（例えば、バッテリなど）、などを備え得る。30

【0076】

図5は、本明細書のシステムおよび方法で使用することができ、例えば、プリンタ、コピー機、多機能機、多機能デバイス（M F D）などを備え得る、印刷装置204であるコンピュータ化デバイスを例示する。印刷装置204は、多くの上述の構成要素および（画像データ処理に専門化しているので汎用コンピュータと異なる）専門化した画像プロセッサ224に作動的に接続された少なくとも1つのマーキング装置（印刷エンジン）240、シート供給230からマーキング装置240に連続的媒体または媒体のシートを供給するように配置した媒体経路236、などを含む。印刷エンジン240からさまざまなマー40
50

キングを受けた後で、媒体のシートは任意選択で、さまざまな印刷したシートの折り畳み、ホチキス留め、ソートなどをすることができる、フィニッシャ 234 に渡ることができる。同様に、印刷装置 204 は、(電源機構 218 を通して) 外部電源 220 から供給された電力で同様に動作する(スキャナ / ドキュメントハンドラ 232 (自動原稿送り装置 (ADF))、などの) 少なくとも 1 つの付属機能要素を含む。

【0077】

1つまたは複数の印刷エンジン 240 は、現在周知であるか将来開発されるか、および、例えば、感光体ベルトもしくは中間体転写ベルトを使用するデバイス、または印刷媒体に直接印刷するデバイス(例えば、インクジェットプリンタ、リボンベースのコンタクトプリンタ、など)を含み得るかどうかにかかわらず、連続的な媒体または媒体のシートにマーキング材料(トナー、インクなど)を塗布する任意のマーキング装置を例示するよう意図される。

【0078】

項目 242 は、品物の画素値を取得することが可能な任意のデバイスとすることができる、別の画像キャプチャデバイス(例えば、カメラ)を例示する。本明細書で論じた画像入力デバイスのセットは、例えば、デジタル・ドキュメント・デバイス、コンピュータシステム、メモリおよび記憶装置、ソースデバイスから画素値を取得できるサーバおよびクライアントデバイスなどのネットワーク化プラットフォーム、ならびに画像キャプチャデバイスなどの多種多様なデバイスを網羅するように意図される。本明細書の画像キャプチャデバイスのセットは、スキャナ、カメラ、写真撮影機器、ファクシミリ装置、写真複製機器、デジタル印刷機、乾式複写装置、などを含む。スキャナは、光学的に画像、印刷媒体などを走査する 1 つの画像キャプチャデバイスであって、走査した画像をデジタル化されたフォーマットに変換する。一般的な走査デバイスは、専門化した受像体がプラテンの下を動きプラテンの上に置かれた媒体を走査する、当該技術分野で一般に知られた、フラットベッドスキャナの変形を含む。近代的なデジタルスキャナは通常、画像感知受容体として電荷結合素子(CCD)または密着型イメージセンサ(CIS)を含む。走査デバイスは走査された画像データの信号を作り出す。このようなデジタル信号は、色値、輝度、および走査画像内のそれらの位置についての情報を包含する。

【0079】

したがって、上に示したように、本明細書のシステムは、画像センサデバイス 232 / 242、および画像センサデバイス 232 / 242 に作動的に接続された画像プロセッサデバイス 216 / 224 を(他の構成要素に加えて) 同様に含む。画像プロセッサデバイス 216 / 224 は、単語転写長の範囲内のあらゆる文字位置においてアルファベットのすべてのアルファベット文字の個別の表現を計算することと、埋め込みアルファベット文字を作り出すためにマトリックスを使用して共通のベクトル部分空間にアルファベット文字の個別の表現を埋め込むことと、埋め込みアルファベット文字を保存することによって初期処理を自動的に行う。

【0080】

初期処理の後に、画像プロセッサデバイス 216 / 224 は、初期処理を繰り返すことなく、後続の処理を繰り返し自動的に行う。この後続の処理は、画像センサデバイス 232 / 242 から英数字単語の電子画像を受け取ることと、埋め込み英数字文字画像を作り出すためにマトリックスを使用して共通のベクトル部分空間に電子画像を埋め込むことを含む。埋め込み英数字文字画像は、文字位置における画像文字を表す、またはこれらを符号化する。

【0081】

この後続の処理は、画像プロセッサデバイス 216 / 224 を使用して埋め込み英数字文字画像に対して各文字位置における埋め込みアルファベット文字のひとつひとつの効用値を計算することと、単語転写長の範囲内で、各文字位置における各々の埋め込みアルファベット文字の効用値に基づいて文字位置のひとつひとつに対して最高の効用の転写アルファベット文字を計算することとを含む。後続の処理の間に、画像プロセッサデバイス 2

10

20

30

40

50

16 / 224 は、電子画像内で認識された英数字単語を作り出すために各文字位置に対して最高の効用の転写アルファベット文字を割り当て、認識された英数字単語の文字長を決定する。画像プロセッサデバイス 216 / 224 は認識された英数字単語を出力する。

【0082】

若干の例示的な構成が添付の図面で例示されるとしても、当業者は、図面が単純化された略図であって、以下に提示された特許請求の範囲が、例示されない（または潜在的に多くのより小さい）、しかしこのようなデバイスおよびシステムで一般に利用される、さらに多くの特徴を網羅することを理解するであろう。したがって、出願者は、以下に提示された特許請求の範囲が添付の図面によって限定されることを意図するものではなく、その代わりに添付の図面は、特許請求された特徴を実施することができるいくつかの方法を例示するために単に提供される。10

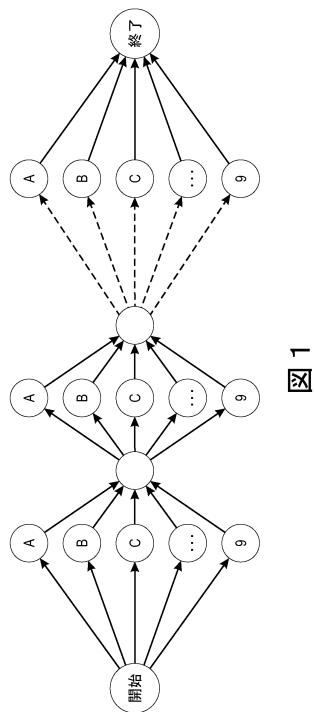
【0083】

多くのコンピュータ化デバイスが上に論じられる。チップベースの中央処理装置（CPU）、（グラフィック・ユーザ・インターフェース（GUI）、メモリ、比較器、有形プロセッサ、などを含む）入力／出力デバイスを含むコンピュータ化デバイスは周知であり、アメリカ合衆国テキサス州ラウンドロックのデルコンピュータ、アメリカ合衆国カリフォルニア州クバチーノのアップルコンピュータ社などの製造業者によって生産される容易に入手可能なデバイスである。このようなコンピュータ化デバイスは一般に、入力／出力デバイス、電源機構、有形プロセッサ、電子記憶メモリ、配線、などを含み、その詳細は、読者が本明細書で説明されるシステムおよび方法の顕著な実施態様に焦点を合わせることを可能にするために、本明細書から省略される。同様に、プリンタ、コピー機、スキャナおよび他の類似の周辺機器は、アメリカ合衆国コネチカット州ノーウォークのゼロックス株式会社から入手可能であり、このようなデバイスの詳細は、簡潔さと読者焦点の目的のため、本明細書では論じない。20

【0084】

本明細書で使用される用語「プリンタ」または「印刷装置」は、任意の目的のために印刷出力機能を行う、デジタルコピー機、製本機、ファクシミリ装置、多機能機、などの任意の装置を網羅する。プリンタ、印刷エンジンなどの詳細は周知であり、本開示の焦点を提示された顕著な特徴に合わせておくために、本明細書では詳細に説明しない。本明細書のシステムおよび方法は、カラー、モノクロで印刷する、またはカラーもしくはモノクロ画像データを扱うシステムおよび方法を網羅し得る。すべての前述のシステムおよび方法は、具体的には静電写真および／または乾式複写の機械および／またはプロセスに適用できる。同様に、用語「埋め込む」は、サブグループであるグループなどの、別のインスタンス内に包含される若干の数学的な構造のインスタンスを意味すると理解される。さらに、用語「自動化」または「自動的に」は、いったんプロセスが（機械またはユーザによって）開始されると、1つまたは複数の機械が任意のユーザからのさらなる入力なしでプロセスを行うことを意味する。30

【図1】



【図2】

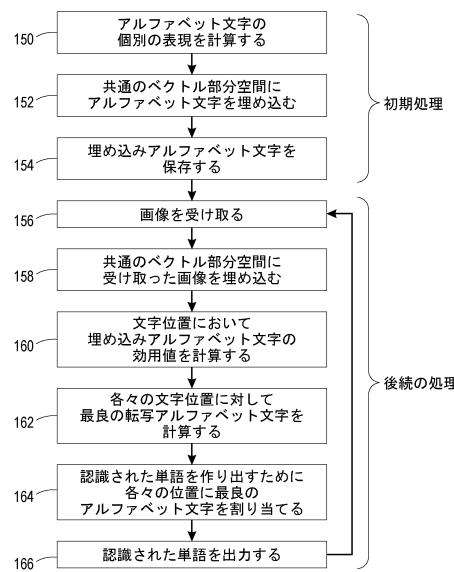


図2

【図3】

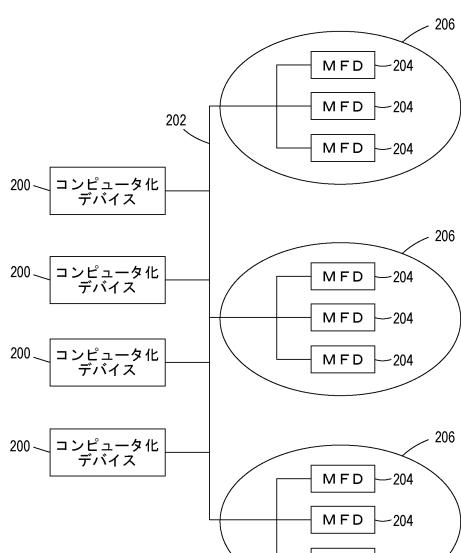


図3

【図4】

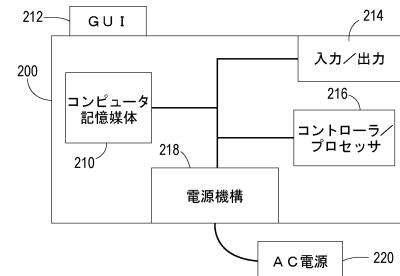


図4

【図5】

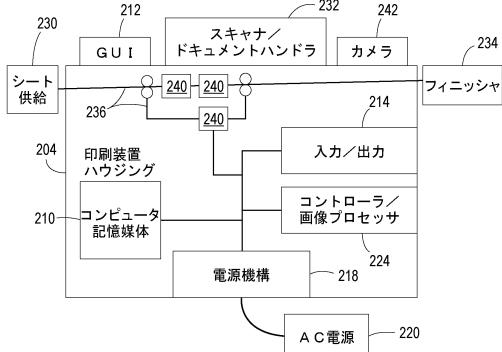


図5

フロントページの続き

(72)発明者 ジョン・アルマーザン
フランス共和国 グルノーブル 38000 リュ・サン・ジャック 26

審査官 千葉 久博

(56)参考文献 特開2015-121708(JP,A)
特開2014-232533(JP,A)
特開2012-98905(JP,A)
米国特許出願公開第2014/0219563(US,A1)
Jon Almazan, 外3名, "Word Spotting and Recognition with Embedded Attributes", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 米国, IEEE, 2014年7月17日, 第36巻, 第12号, p.2552-2566

(58)調査した分野(Int.Cl., DB名)

G 06 K 9 / 72