

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4987960号  
(P4987960)

(45) 発行日 平成24年8月1日(2012.8.1)

(24) 登録日 平成24年5月11日(2012.5.11)

(51) Int.Cl. F I  
**G06K 9/48 (2006.01)** G O 6 K 9/48  
**G06T 11/20 (2006.01)** G O 6 T 11/20 I I O A

請求項の数 17 (全 16 頁)

(21) 出願番号	特願2009-503161 (P2009-503161)	(73) 特許権者	507241665
(86) (22) 出願日	平成19年3月22日 (2007.3.22)		アマゾン・テクノロジーズ・インコーポレイテッド
(65) 公表番号	特表2009-531788 (P2009-531788A)		アメリカ合衆国、ネバダ・89451、インクライン・ビレッジ、インクライン・ウェイ・920、スイート・シー
(43) 公表日	平成21年9月3日 (2009.9.3)	(74) 代理人	100062007
(86) 国際出願番号	PCT/US2007/064616		弁理士 川口 義雄
(87) 国際公開番号	W02007/121029	(74) 代理人	100114188
(87) 国際公開日	平成19年10月25日 (2007.10.25)		弁理士 小野 誠
審査請求日	平成20年11月25日 (2008.11.25)	(74) 代理人	100140523
(31) 優先権主張番号	11/392, 213		弁理士 渡邊 千尋
(32) 優先日	平成18年3月29日 (2006.3.29)	(74) 代理人	100119253
(33) 優先権主張国	米国 (US)		弁理士 金山 賢教

最終頁に続く

(54) 【発明の名称】 レンダリングのためにトークンベースファイルへの文字列を含むデジタル画像の変換

(57) 【特許請求の範囲】

【請求項 1】

電子画像を処理するためにコンピュータで実施される方法であって、  
 書籍の走査された画像に対応する電子画像における1つまたは複数のトークンを、接続構成要素解析を用いて識別することと、  
 トークンの形状にしたがって、識別されたトークンを複数のトークングループに分類することと、  
 トークングループのそれぞれに関して、トークングループに含まれるトークンの形状を数学的に表現するベクトル化トークンを作成することと、  
 出力媒体上でレンダリングするように、画像を表すベクトル化トークンの集合を生成することと、  
ベクトル化トークンによって表されるトークンの位置を各ベクトル化トークンに関連付けることとを含む、方法。

【請求項 2】

少なくとも2つのトークンが、同一の英数字に対応する、請求項1に記載の方法。

【請求項 3】

出力媒体が、ハンドヘルド型媒体デバイスである、請求項1に記載の方法。

【請求項 4】

出力媒体が、印刷媒体である、請求項1に記載の方法。

【請求項 5】

類似の形状を有するトークンをグループに分類し、それによって、複数のトークングループを形成することと、

各トークングループにおいてトークンの形状を形態構造的に表現する代表的なトークンを識別し、それによって複数の代表的なトークンを識別することと、

各代表的なトークンをベクトル化して、代表的なトークンの形状を数学的に表現するベクトル化トークンを作成することとをさらに含み、

各トークングループにおいてトークンの形状を形態構造的に表現する代表的なトークンを識別することが、各トークングループにおいてトークンの形状を結合して、トークングループにおいてトークンの組み合わせである代表的なトークンを作成することを含む、請求項 1 に記載の方法。

10

【請求項 6】

類似の形状を有するトークンをグループに分類することが、類似の形状および類似のサイズの両方を有するトークンをグループに分類することを含む、請求項 5 に記載の方法。

【請求項 7】

各トークングループにおいてトークンの形状を結合することが、トークンの形状を平均化して、平均化したトークンを作成することを含む、請求項 6 に記載の方法。

【請求項 8】

電子画像においてトークンを識別することが、接続構成要素解析を用いることを含む、請求項 5 に記載の方法。

【請求項 9】

電子画像においてトークンを識別することが、接続構成要素が所定の閾値サイズより大きいように求められると、その最短経路に沿って接続構成要素を分離し、それによって 2 つのトークンに識別することを含む、請求項 8 に記載の方法。

20

【請求項 10】

電子画像においてトークンを識別することが、ピクセルがピクセル強度、エッジマグニチュードおよび方向に基づくエッジウェイトを用いてグラフとして表されるエッジグラフ解析を用いることを含む、請求項 5 に記載の方法。

【請求項 11】

類似の形状を有するトークンをグループに分類することが、

各トークンに関する重心を計算することと、

重心の値に基づいてトークンを整列することとを含む、請求項 5 に記載の方法。

30

【請求項 12】

各トークンに関する重心を計算することが、

各トークンを 2 つ以上の部分に分割することと、

部分のそれぞれに関して重心を計算することとをさらに含み、請求項 11 に記載の方法

【請求項 13】

類似の形状を有するトークンをグループに分類することが、トークンを比較するために誤差解析技術を用いることを含む、請求項 5 に記載の方法。

【請求項 14】

類似の形状を有するトークンをグループに分類することが、トークンを比較するために二乗平均平方根 (RMS) 誤差解析を用いることを含む、請求項 13 に記載の方法。

40

【請求項 15】

各代表的なトークンをベクトル化してベクトル化トークンを作成することが、ベジエ曲線において代表的なトークンの形状を数学的に表現することを含む、請求項 5 に記載の方法。

【請求項 16】

各ベクトル化トークンをベクトル化トークンによって表されるトークンの位置に関連付けることが、位置に関連付けてベクトル化トークンに対するポイントを格納することを含む、請求項 5 に記載の方法。

50

## 【請求項 17】

コンピュータによって実行されたときに、前記コンピュータに、請求項 1 から 16 のいずれかに記載の方法のステップを実行させるためのコンピュータプログラム。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明は、デジタル画像の処理に関し、さらに詳細にはその中に文字列を有するコンテンツの画像の処理に関する。

## 【背景技術】

## 【0002】

コンピュータおよびコンピュータに基づくネットワークの使用が拡大し続けているため、コンテンツプロバイダは、電子形態におけるさらに多くのコンテンツを調製して配信している。このコンテンツは、書籍、雑誌、新聞、ニュースレター、マニュアル、ガイド、参考文献、記事、レポート、ドキュメントなどの印刷における存在する従来の媒体のほか、前述のコンテンツがデジタル形態で存在し、走査デバイスを用いて印刷形態からデジタル形態に変換される電子媒体を含む。特に、インターネットは、コンテンツの画像のダウンロードおよび表示を通じてデジタルコンテンツのより広い公開を容易にしてきた。データ伝送速度が増大するにつれて、コンテンツのページのますます多くの画像が、オンラインで利用可能になっている。ページ画像は、読者が印刷物のようにコンテンツのページを見ることを可能にする。

## 【0003】

コンテンツのデジタル画像を提供する大きな魅力にもかかわらず、多くのコンテンツプロバイダは、コンテンツの画像を生成して格納する場合、特に、画像における文字列を認識する精度が重要である場合に、問題点に直面する。例えば、ユーザがコンピュータスクリーン上で書籍または雑誌からページ画像を読むこと、または後に読むために印刷することを可能にするために、画像は、可読な文字列を提供するほど十分に明瞭でなければならない。現在、画像は、デジタル文字認識を含む光学文字認識（OCR）などの種々の文字認識技術を用いてコンピュータ読み出し可能なデータに変換される。光学文字認識の精度は一般に高いが、一部のページ画像は、OCR処理後であっても、種々のアーチファクトのために簡単に読み出すことが可能でない。手動補正は可能であるが、誤識別された文字の手動補正または欠けている文字の挿入のコストは、特に大量のページを走査する場合には、きわめて高い。

## 【0004】

デジタルコンテンツプロバイダが直面する別の問題点は、コンテンツの画像を格納するコストである。格納コストを削減するために、コンテンツプロバイダは、画像を格納するために用いられるファイルのサイズを最小限にすることを望む。デジタル画像は、種々の解像度で表されてもよく、水平方向および垂直方向の両方向における画像のピクセルの数によって通常表される。通常、常にではないが、より高い解像度の画像は、より大きなファイルサイズを有し、格納のために大量のメモリを必要とする。書籍、雑誌などの大容量の媒体を捕捉して格納するために要する画像の数を考慮すると、コンテンツの画像の格納コストは、著しく増大する可能性がある。画像のサイズおよび解像度の削減は、画像を格納するための要件を削減することが多いが、低解像度の画像は最終的には、画像、特にその中に含まれる任意の文字列が表示される場合に、読者が認識することが困難な事態に達する。文字列を備えたページ画像を提供したいコンテンツプロバイダは、表示された文字列が可読であるように、画像が十分に高い解像度でレンダリングすることができることを確保しなければならない。コンテンツプロバイダが直面するさらに別の問題点は、スケラブル、すなわち、例えば、画像における文字列の最小品質および可読性を確保すると同時に、比較的高い解像度で種々のサイズのディスプレイ上にレンダリングするために容易に拡大縮小し得るページ画像を提供することである。

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0005】

必要とされるものは、レンダリング時に、ページ画像中の文字列が、十分に高い解像度で可読され、さらに、格納のために、過剰な量のメモリ空間を必要とすることなく、さらにスケーラブルであるように、文字列を含むスキャンインされたページ画像を確実に処理するための方法およびシステムである。

## 【課題を解決するための手段】

## 【0006】

従来技術における上述の問題および他の欠点に対処するために、本発明は、スキャンインされたコンテンツのページを「トークンベース」ファイルに変換するために、コンピュータで実施され得る方法を提供する。本明細書で用いられるとき、トークンは、グラフィカルユニットを指し、単独の文字または記号を表してもよく、または表さなくてもよい。スキャンインされたページ画像から、数々のトークンが分離される。次に、類似の形状のトークンが、共にグループ化され、それらの形状が結合され、結合トークンを形成する。結合トークンは、グループに含まれるトークンのすべてを形態構造的に表している。結合トークンはさらに、ベクトル化トークンに変換される。ベクトル化トークンは、結合トークンの数学的表現であり、明瞭な曲線で結合トークンの形状を表すことができる。コンテンツのスキャンインされたページ、例えば、1冊の書籍に関して、複数のベクトル化トークンが、このような方式で作成され、それぞれのベクトル化トークンは、同様に形成されたトークンのグループを表す。その後で、グループを形成する（元の、未処理の）トークンのそれぞれの位置は、トークンのグループを表すベクトル化トークンに関連付けられる。例えば、各トークンの位置は、ページ番号およびトークンが現れる各ページ内の位置のX-Y座標によって画定されてもよく、位置は、ポイントによって対応するベクトル化トークンに対して関連付けられる。したがって、レンダリング時に、ベクトル化トークンは、元のトークンとは対照的に、この位置で表示され、それによってベクトル化トークンのみからなるページ画像を作成する。ベクトル化トークンがトークン形状の数学的表現であるために、高解像度をはじめとする任意の解像度でレンダリングされ得、表示されたときに整って可読であるように見え得る。さらに、同様に形成されたトークンの複数の位置が、ポイントを用いて代表的なベクトル化トークンに単に関連付けられるため、ページ画像に関する格納要件を最小限に抑えることができる。

## 【0007】

本発明の一実施形態によれば、コンピュータで実施される方法は、文字列を含む電子画像をトークンベースファイルに変換するために設けられる。この方法は、一般に5つのステップを含む。第一に、種々のトークン（すなわち、グラフィカルユニット）は、電子画像に識別される。第二に、類似の形状を有する識別されたトークンは、共にグループ化され、トークングループを形成する。したがって、複数のトークングループが形成され、各トークングループは、類似の形状を有する1つまたは複数のトークンを含む。第三に、各トークングループにおいて、グループに含まれるトークンの形状を形態構造的に表現する代表的なトークンが、生成される（または求められる）。例えば、代表的なトークンは、トークングループにおけるトークンの形状を結合（例えば、平均化）することによって生成されてもよい。第四に、各代表的な（例えば、結合された）トークンは、ベクトル化トークンに変換される。ベクトル化トークンは、代表的なトークンの形状の数学的表現である。したがって、この時点で、複数のベクトル化トークンが、形成される。各ベクトル化トークンは、代表的なトークンの形状を数学的に表しており、今度は、1つのトークングループに分類される1つまたは複数のトークンの形状を形態構造的に表現する。第五に、ベクトル化トークンのそれぞれは、ベクトル化トークンによって表されるトークンの位置に関連付けられ、それによって、トークンベースファイルを形成する。言い換えれば、グループを形成するトークンのそれぞれの位置は、トークンのグループを表すベクトル化トークンに関連付けられる。したがって、レンダリング時に、ベクトル化トークンは、元のトークンとは対照的に、この位置で表示され、それによって、ベクトル化トークンに

10

20

30

40

50

基づく非常に鋭いトークン画像のみからなるページ画像を作成する。

【0008】

本発明の一態様によれば、トークンを分離するステップは、接続構成要素（または「フラッドフィル」）解析を用いることによって実現される。本発明の別の態様によれば、類似の形状を有するトークンのグループ化ステップは、各トークンに関する重心を計算し、重心を用いてトークンを整列し、例えば、2つのトークン間の二乗平均平方根誤差を計算することによって、1対のトークン間の「距離」を計算し、互いと所定の距離内のトークンをグループ化することによって実現される。本発明のさらなる態様によれば、ベクトル化トークンを形成するために各代表的なトークン（例えば、結合トークン）をベクトル化するステップは、ベジエスプラインなどの数学的表現を用いて、ラスタベクトル変換方法に基づいて実現される。

10

【0009】

本発明のさらなる実施形態によれば、電子画像をトークンベースファイルに変換するためのシステムが、設けられる。このシステムは一般に、2つの構成要素、すなわち、ページ画像などの文字列を含む電子画像を格納するためのページ画像データベースと、ページ画像データベースと通信状態にある計算デバイスとを含む。計算デバイスは、文字列を含む電子画像を処理して、その中のトークンを識別し、識別されたトークンを複数のトークングループに分類するように動作する。計算デバイスはさらに、トークングループのそれぞれに関して、トークングループに含まれるトークンの形状を数学的に表現するベクトル化トークンを作成し、各ベクトル化トークンがベクトル化トークンによって表されるトークンの位置に関連付けられるトークンベースファイルを生成するように動作する。

20

【0010】

本発明のさらに別の実施形態によれば、その上に符号化された命令を有するコンピュータによるアクセス可能な媒体が、トークンベースファイルを作成するために設けられる。命令は、計算装置によって実行されると、計算デバイスに（1）その中に文字列を有する画像を処理して、その中にトークンを識別させ、（2）識別されたトークンをそれらの形状にしたがって複数のトークングループに分類させ、（3）トークングループのそれぞれに関して、トークングループに含まれるトークンの形状を数学的に表現するベクトル化トークンを形成させ、（4）ベクトル化トークンによって表されたトークンをベクトル化トークンに置き換えさせる。

30

【0011】

本発明の前述の態様および付随する利点の多くは、添付図面と共に解釈すれば、以下の詳細な説明を参照することによってよりよく理解されるように、さらに容易に認識されるであろう。

【発明を実施するための最良の形態】

【0012】

本発明は、走査された文字列を含む電子（デジタル）画像を例えば、過度の量の格納空間を必要とすることなく、高解像度のレンダリングに適したトークンベースファイルに変換するための命令を有する方法、システムおよびコンピュータによるアクセス可能な媒体に関する。説明のための実施形態において、トークンベースファイルのレンダリングは、デジタルディスプレイおよび印刷媒体などの種々の出力媒体上で行われ得る。

40

【0013】

以下の詳細な説明は、本発明の例示的な実施を提供する。特定のシステム構成および流れ図が示されているが、提供された実施例は、開示された正確な形態に対して包括的なものではなく、本発明を限定するものではないことを理解すべきである。コンピュータおよびデジタル画像化の分野における当業者は、他の構成要素またはステップあるいは構成要素またはステップの組み合わせによって置換可能であり、さらに本発明の利益および利点を依然として達成し得る本明細書に記載される構成要素および処理ステップを認識するであろう。また、以下の説明は、従来コンピュータ構成要素によって実行され得る論理および動作に関して主に提供されることも理解されるべきである。これらのコンピュータ

50

構成要素は、単独の位置でグループ化されてもよく、広範囲に分散されてもよく、一般に、コンピュータプロセッサ、メモリ格納デバイス、ディスプレイデバイス、入力デバイスなどを含む。コンピュータ構成要素が分散される環境において、コンピュータ構成要素は、通信リンクを介して互いにアクセス可能である。

【0014】

以下の説明において、多数の特定の詳細項目は、本発明の全体的な理解を提供するために記載される。しかし、本発明は、これらの詳細項目の一部またはすべてがなくても実行され得ることは当業者には明白であろう。他の場合において、よく知られている処理ステップは、本発明を必要以上に曖昧にしないようにするために、詳細に記載されていない。

【0015】

本発明の実施形態を記載するための文脈を提供するために、図1は、本発明を実施するために用いられ得る計算システム10の機能ブロック図を示している。コンピュータの当業者は、図1に示される計算システム10などのシステムを構成するために用いられ得る市販の構成要素の広範囲の選択を認識している。計算システム10は、ネットワークインターフェイス14、入力/出力インターフェイス16およびメモリ19などの種々の計算要素と通信状態にあるプロセッサ12を有する計算デバイス11を含む。ネットワークインターフェイス14は、計算デバイス11がデータ、制御信号、データ要求および他の情報をコンピュータネットワーク15(LAN、WAN、インターネットなど)と通信することを可能にする。例えば、計算デバイス11は、ネットワークインターフェイス14を介してコンピュータネットワーク15に接続されるページ画像データベース17から、書籍、雑誌などのページ画像を含むファイルを受信してもよい。トークンベースファイルデータベース18は、コンピュータネットワーク15に接続されてもよく、計算デバイス11によって生成されたトークンベースファイルは、格納のためにネットワークインターフェイス14を介して送信される。当業者は、コンピュータネットワーク15がインターネット、ローカルエリアネットワークまたはワイドエリアネットワークであってもよく、インターネット、ローカルエリアネットワークまたはワイドエリアネットワークは、関連ドキュメントおよび関連付けられるファイル、スクリプトおよびデータベースを格納するためのサーバ、あるいは音声ファイルまたは映像ファイル、ドキュメント、スクリプト、データベースなどにアクセスするセットトップボックスまたは他の情報家電を含む放送通信ネットワークに接続する。

【0016】

入力/出力インターフェイス16は、計算デバイス11が種々の局所的な入力および出力デバイスと通信可能にする。入力/出力インターフェイス16と通信状態にある入力デバイス20は、スキャナ、走査ペン、デジタルカメラ、ビデオカメラ、コピー機、キーボード、マウス、外部メモリ、ディスクドライブなどの計算デバイス11に入力信号を提供する計算要素を含んでもよい。例えば、スキャナおよびカメラを含む入力デバイスは、文字列を含むページ画像などの電子画像を計算デバイス11に提供するために用いられてもよく、計算デバイス11は、本発明によれば、これらの電子画像をトークンベースファイルに変換する。

【0017】

入力/出力インターフェイス16と通信状態にある出力デバイス22は、コンピュータディスプレイ(例えば、CRTスクリーンまたはLCDスクリーン)、テレビ、プリンタ、ファクシミリ機、コピー機などの通常の出力デバイスを含んでもよい。本発明に関して、出力デバイス22は、オペレータが、それらの精度および可読性を手動で確認するために、トークンベースファイル画像を表示するために用いてもよい。

【0018】

プロセッサ12は、メモリ19などのメモリに格納されるコンピュータプログラム命令によって動作するように構成される。プログラム命令はまた、プログラムされるデジタル信号プロセッサなどのハードウェアフォーマットにおいて具体化されてもよい。メモリ19は一般に、RAM、ROMおよび/または固定記憶装置を含む。メモリ19は、本発

10

20

30

40

50

明によれば、処理、伝送および表示のために文字列のデジタル画像を格納するように構成されてもよい。メモリ19は、計算デバイス11の一般的な動作を制御するためのオペレーティングシステム23を格納する。オペレーティングシステム23は、Microsoft（登録商標）オペレーティングシステム、UNIX（登録商標）オペレーティングシステムまたはLinux（登録商標）オペレーティングシステムなどの汎用オペレーティングシステムであってもよい。メモリ19は、その中に文字列を含むデジタル画像を解析するためのプログラムコードおよびデータから構成される光学文字認識（OCR）アプリケーション24をさらに格納してもよい。当業者は、画像において文字列を解析および認識することができる種々のアルゴリズムおよび技術があることを認識されよう。しかし、本発明の目的のためには、アルゴリズムおよび技術が実際に、種々のOCRルーチンによって達成されるような個別の文字または記号をそれ自体として実際に認識するか、それらの意味を解釈することは必要ではない。市販のOCRソフトウェアの例としては、ScanSoft, Inc.によるOmniPage Pro（商標）およびSmartLink CorporationによるFineReader（商標）が挙げられる。メモリ19はさらに、トークンベースファイルジェネレータアプリケーション25を格納する。トークンベースファイルジェネレータアプリケーション25は、トークンベースファイルを生成するために、ネットワークインターフェイス14、入力/出力インターフェイス16などを介して受信された文字列を含む電子画像を処理するためのプログラムコードおよびデータを含む。トークンベースファイルは次に、トークンベースファイルデータベース18に送信されて格納されてもよい。

#### 【0019】

図2は、本発明の一実施形態によれば、文字列を含む1つまたは複数の電子画像をトークンベースファイルに変換するために、トークンベースファイルジェネレータアプリケーション25によって実施される例示的な方法30の流れ図である。本発明の文脈において、「文字列」という用語は、電子画像（例えば、ページ画像）において情報を表すために用いられ得る字、文字、記号、数字、数式、グラフィックス、線図、表の縁などのすべての形態を含むものと理解すべきである。この方法30は、計算デバイス11が、文字列を含む電子画像（例えば、ページ画像）を受信するブロック31から始まる。例えば、ページ画像データベース17（図1）の中に前にスキャンインされたようなページ画像が、検索されてもよく、またはページ画像は、適切なスキャナ入力デバイス20（図1）を用いてスキャンインされてもよい。受信された画像は、300dpi（ドット/インチ）などの比較的low解像度からなってもよい。受信されたようなページ画像のフォーマットは、可変であってもよく、JPEG、TIFF、GIFまたはBMPファイルなどの非文字列がアクセス可能なフォーマットで表されたページ画像のコンテンツ、またはAdobe Portable Document File（PDF）などの文字列のアクセス可能なフォーマットで表されたページ画像のコンテンツを含むことができる。ページ画像には、ページ画像が受信されるフォーマットに関係なく、トークンベースファイル生成処理方法30を施す前に、標準的なOCRまたはOCRに似た前処理技術、例えば、コントラスト調整、デスクューイング、デスクベリングおよび/またはページ回転補正などを施してもよい。

#### 【0020】

ブロック32で、その中に文字列を有するページの受信画像において、トークンは、識別される。トークンは、グラフィカルユニットを指し、単独の文字または記号を表してもよく、または表さなくてもよい。むしろ、トークンは、純粋に図形という意味で十分に別個であるように識別されるユニットであり、それにより単独のユニットを形成する。本発明の種々の例示的な実施形態において、電子画像におけるトークンに関する探索は、通常は白色である背景領域の中で生じる。トークンは、ピクセルの色が背景色から十分に逸脱する場所ならどこでも推測される。次に、当分野ではよく知られている接続構成要素解析（またはフラッドフィル解析）が、そのすべての隣接する（または接続される）ピクセルを求めるために最初のピクセルに施されてもよい。この実施例において、単独の接続構成

10

20

30

40

50

要素に関連付けられるピクセルはすべて、トークンと識別される。しかし、別の実施例において、2つ以上の接続構成要素が、トークンと識別されてもよい。

【0021】

図3を参照すると、参照符号42、文字列構成要素「every day」を含む電子画像が、接続構成要素技術に基づいて解析され、「e」、「v」、「e」、「r」、「y」、「d」、「a」および「y」を個別のユニット、すなわち、トークンとして識別する。さらに、これらのトークンのそれぞれは、図示されているように、境界ボックス内に境界を定められ得る。接続構成要素解析および境界ボックスの発見は、当分野ではよく知られているように、メモリ19(図1)に格納される適切なOCRまたはOCRに似たソフトウェアプログラムを用いて実行されてもよい。

10

【0022】

別の実施形態において、他の計算ジオメトリ構成技術を用いて、電子画像におけるトークンを識別してもよい。例えば、電子画像内部のピクセルは、ピクセル強度およびエッジマグニチュードおよび方向に基づいてエッジウェイトを有するグラフとして表されてもよい。接続決定は、ピクセルの2つの集合間の最短経路を決定することによって行われ得る。ピクセルの集合が十分に接続される場合には、単独のトークンを共に形成するものとして識別されてもよい。

【0023】

場合によっては、2つの字(または文字)は、図4Aに示される「ra」の場合のように、互いに接触して単独の接続構成要素を形成してもよい。この場合には、元の単語「raw」は、接続構成要素解析に基づいて、2つのトークン「ra」50および「w」51に分離される。本発明の方法では、個別のトークンとして各字を識別することは重要でないため、格納要件を低減するために、トークンタイプの数削減を目的のために、これを実行することが望ましい場合がある。したがって、2つ以上の字または記号の組み合わせであってもよい接続構成要素を分離するためにさらなる処理が行われてもよい。例えば、図4Aの場合には、トークン「ra」50に関する境界ボックスは、このトークンに関して単独の字または記号を表すためには、その水平寸法において大きすぎる可能性があることが決定されてもよい。一般に、その境界ボックスが垂直寸法より長い水平寸法を有するトークンは、2つ以上の字または記号を潜在的に表している疑いがある可能性がある。そのように決定される場合には、疑われたトークンは、適切なOCRまたはOCRに似たソフトウェアプログラム(例えば、迷路アルゴリズム)を用いてさらに解析し、トークンを2つの部分に切り離すために、一方の側(例えば、上側)から他方の側(例えば、下側)への最短経路52を識別してもよい。OCRまたはOCRに似たソフトウェアはまた、上記の実施例において、接触している可能性がある隣接する字(すなわち、上記の実施例では「a」)から分離しやすくするために、「r」が所与の位置にあることを認識する際に有用である可能性がある。さらに別に、場合によりトークンを2つの部分に切り離す経路は、ピクセル強度およびエッジマグニチュードおよび方向に基づいて、エッジウェイトを有するグラフとしてピクセルを表すことによって、計算され得る。次に、最短経路52は、トークンの対向する側にある2つの点の間(例えば、上エッジの中心と下エッジの中心との間)で求められ得る。図4Aの実施例において、最短経路52は、トークン「ra」50を2つのトークン「r」および「a」に分けるために求められる。その後で、分離された「r」トークンおよび「a」トークンの精度は、「r」トークンおよび「a」トークンをそれぞれ既にトークンとして明確に識別された他の「r」トークンおよび「a」トークンと比較することによって確認されてもよい。

20

30

40

【0024】

一実施形態において、2つの字を1つのトークンに接続する(例えば、上述のトークン「ra」50)方が、字を2つのトークンに分離する(例えば、「H」を2つの垂直部分に分離する)よりよい。前者は、メモリ要件をわずかに増大するが、後者は、ページにおいて誤って現れる不正確な(または誤って分割された)トークンを作成する。例えば、字「H」が、2つの垂直部分に「誤って分割される」場合には、その間に間隙を有する2つ

50

の垂直部分として現れる。したがって、不明確なトークン（すなわち、2つ以上の字または記号を含む疑いのあるトークン）は、すべて明確なトークンが識別された後のみ、複数のトークンへの可能性の或る分離に関して評価されてもよい。例えば、結果として生じる分離点、いくつかの明確に識別されたトークンに適合する場合に限り、任意の不明確なトークンが複数のトークンに分離されてもよい。特定の実施例として、図4Aにおけるトークン「r a」50は、結果として生じる分離された「r」トークンおよび「a」トークンが明確に定義されたトークン「r」および「a」にそれぞれ適合する場合に限り、「r」トークンおよび「a」トークンに分離されてもよい。結果として生じる分離されたトークンのそれぞれは、明確に定義されたトークンと厳密に適合することができない場合には、不明確なトークンは、複数のトークンに分離される必要はない。

10

## 【0025】

図2に戻って参照すると、ブロック33で、分離されたトークンが互いと比較され、類似の形状を有するトークンが共にグループ化される。例えば、図3において、ステップ43、文字列「every day」からの2つの「e」トークンは、「e」バケットに類似の形状を有するものとして共にグループ化され、同一の文字列から2つの「y」トークンは、「y」バケットに類似の形状を有するものとして共にグループ化される。グループ化は、トークンの形態構造学的特性に基づいて、実行されることに留意されたい。したがって、例えば、通常の「e」およびボールド体で記載された「e」は、十分に異なる形状を有するものとして良好に処理され、2つの異なるバケットでグループ化されてもよい。結果として生じるトークンベースファイルがレンダリング時に、元の電子画像の全体的な

20

## 【0026】

種々のパターンマッチング方法または形状マッチング方法は、それらの形状にしたがって、トークンを分類するために用いられてもよい。本発明の種々の例示的な実施形態において、「重心」が、各トークンに関して計算され、互いと比較することができるようにトークンを整列するために用いられる。本明細書で用いられるとき、グレースケール画像におけるピクセルの「質量」は、背景色（通常は、純粋な白色）からのその偏差として定義される。グレースケール画像が質点の格子、各ピクセルに関する1つの質点として処理される場合には、画像の「重心」は、画像の代表的な点と見なされ得る。色画像がこれも色における背景と共に用いられる場合には、「質量」および「重心」は依然として、任意の適切な変換方法を用いて、色画像をグレースケール画像に最初に変換することによって、同様に計算され得る。各トークン画像に関して計算された重心は次に、それぞれの重心値にしたがって、トークン画像を整列するために用いられてもよい。

30

## 【0027】

説明のための実施形態において、一旦、トークンが整列されると、整列されたトークンは、トークンが十分に類似であるかどうかを決定するために比較され得る。各（グレースケール）画像における各ピクセルは、0.0が白色を表し、1.0が黒色を表すように正規化されてもよい。その後で、1対の画像に関して、トークン画像間の形状における類似性を確認するために、画像間の「距離」が、計算される。そのような距離を計算するために、種々の方法が可能である。一実施形態において、二乗平均平方根（RMS）誤差に関して、距離を計算することができる。具体的に言えば、各一致する対のピクセル（すなわち、重心に対して同一の位置にあるピクセル）に関して、2つのトークン画像を比較するために、グレースケール（色）値における差の平方を計算し、すべての一致するピクセルに関して平方値の和をとり、一致するピクセルの数によってその和を除算し、最後にその平方根をとってRMS誤差値を生成することができる。0.0のRMS誤差値は、2つの画像が同一であることを意味するのに対し、任意のより大きなRMS誤差値は、2つの画像が互いから一層著しく異なっていることを示す。一実施形態において、2つのトークン

40

50

画像は、RMS誤差値が0.10などの所定の閾値以下である場合には、同一のトークングループに属するように、互いに対して形状において同一または十分に類似であるものと考えてもよい。

#### 【0028】

図4Bは、境界ボックス53におけるトークン「e」を示しており、重心が点「x」で求められる。それらの重心値に基づいて、トークンの整列および/または比較のための種々の別の方法が、可能である。例えば、図4Bをさらに参照すると、境界ボックスは、重心の点53'を通る水平線および垂直線を用いて、図示されているように、4つの部分54、55、56および57などの複数の部分に分割してもよい。重心値は、示されているように、4つの部分に関してそれぞれ4つの点「x」で求められてもよい。4つの重心値は、原点として用いられる重心の点53'に対する(x, y)座標として表されてもよい。次に、問題のトークンがどのトークングループに属する可能性があるかどうかを概ね決定するために、4つの重心値はそれぞれ、(例えば、4つの重心値の2つの集合間の平均平方差をとることによって)トークングループからの別のトークンの対応する重心値と比較されてもよい。特に、トークンを比較するために多数のトークングループが存在する場合には、この実施例の場合のように、4つの重心値のみの比較は、予備マッチング処理を著しくスピードアップする。しかし、この方法に基づいてマッチングが求められる場合には、真のマッチングは、上述したRMS誤差に基づく方法などの包括的な比較試験を用いて確認されてもよい。

10

#### 【0029】

4つの重心値の使用に対してさらにまたはあるいは、上述のように、種々の他の方法が、トークンをトークンが属し得る候補トークングループに予備的に分類するために用いられてもよい。例えば、OCRまたはOCRに似た処理は、検出された実際の文字などの字の情報およびフォント、近似フォントサイズ、字がボールド体、イタリック体またはアンダーライン付きなどであるかどうかなどの種々のフォーマットの詳細項目を得るように行われてもよい。2つのトークンが同一のOCR文字および同一サイズを有するように検出される場合には、2つのトークンは、互いに形状において類似であると予備的に決定されてもよい。しかし、既に述べたように、マッチングがこの方法に基づいて求められたとしても、真のマッチングはさらに、RMS誤差に基づく方法などのさらに包括的な比較方法を用いて確認されてもよい。

20

30

#### 【0030】

上述したRMS誤差解析を用いなくても済む種々のトークンの形状を推定して比較するための他の包括的なグラフィカル解析技術もまた、当業者には明白であるように、本発明によれば採用され得る。例えば、黒度解析の和を用いて、種々のトークンの形状を比較してもよい。別の実施例は、クロスエントロピー方法である。2つのトークンAおよびBが与えられると、Aに対するBのクロスエントロピーが、ガイドとしてのAに関するトークン画像における情報を用いてBに関するトークン画像を圧縮することによって計算される。次に、トークン画像Bに関する最終的な圧縮ファイルにおけるビットの数が、取得される。同様に、Bに対するAのクロスエントロピーが、Bに関するトークン画像における情報を用いてAに関するトークン画像を圧縮し、トークン画像Aに関する最終的な圧縮ファイルにおけるビットの数を取得することによって、計算され得る。次に、Bに対するAのクロスエントロピーとAに対するBのクロスエントロピーとの間の最大値が取得され、2つのトークン画像の間の「距離」(すなわち、形状における近似性)の尺度として用いられる。

40

#### 【0031】

ブロック33を参照し続けると、説明のための実施形態において、同様に形成されたトークンは、共に分類され、トークングループを形成することができる。図4Cは、種々に形成されたトークンをさまざまなトークングループに分類する際に使用するのに適した1つの技術を概略的に示している。本質的に、探索木または分類木は、分類処理をスピードアップするように構築されてもよい。コンピュータに基づく探索および分類の当分野にお

50

いてよく知られているように、探索木または分類木は、各状態（ノード）が、状態の新たな集合（子ノード）に生じてもよく、これらのそれぞれが今度はそれ自身の後を継ぐ状態（孫ノード）を生じてもよい、などといった分岐構造を含む。それらの形状（例えば、一定の参照画像に対する RMS 誤差値に基づく）によってトークンに関してそのような木を構築し、根ノードから子ノードに、孫ノードになど木を横断することによって、コンピューターは、新たなトークンを種々のトークングループ（葉ノードを形成する）に迅速に分類してもよい。

#### 【 0 0 3 2 】

図 4 C の実施例において、「 e 」、「 a 」および「 b 」に関するトークングループが、形成されている。さらに、（例えば、「 e 」トークン画像と「 a 」トークン画像との間の RMS 誤差値に関して）「 e 」および「 a 」に関するトークングループの間の差は、0 . 3 であることが求められ、「 e 」および「 b 」に関するトークングループの間の差は、0 . 4 であることが求められた。この実施例において、「 e 」トークングループは、基準点として用いられる。分類されることになっている次のトークン「 ? 」と「 e 」トークングループとの間の RMS 誤差値が、「 ? 」として計算される。  $x$  が 0 . 3 未満である場合には、「 e 」トークングループから 0 . 3 未満だけ異なる既存のトークングループはないため、「 ? 」は、まだ作成されていない新たなトークングループに分類され得る。同様に、  $x$  が 0 . 4 より大きい場合には、「 e 」トークングループから 0 . 4 より大きく異なる既存のトークングループはないため、「 ? 」は、まだ作成されていない新たなトークングループに分類され得る。したがって、  $0 . 3 < x < 0 . 4$  である場合のみ、「 ? 」がこれらのトークングループのいずれかに属するか、またはまだ作成されていない新たなトークングループにおそらく属するかを確認するために、「 ? 」をそれぞれ「 a 」および「 b 」に関するトークングループと比較することが必要である。比較は、例えば、「 ? 」トークン画像と「 e 」トークン画像との間の RMS 誤差値または「 ? 」トークン画像と「 a 」トークン画像との間の RMS 誤差値を計算することによって、実行され得る。適切な探索木または分類木を用いて、種々に形成されたトークンは、それらの形状にしたがってトークングループに迅速に分類されてもよい。

#### 【 0 0 3 3 】

一旦、ページ画像、例えば、書籍において求められたトークンのすべてが、種々のトークングループに分類されると、各グループにおけるトークンには、任意の適切な画像処理または前処理を施されてもよい。具体的に言えば、図 2 に戻って参照すると、ブロック 3 4 で、任意に、種々のデジタル画像処理フィルタが、各グループに分類されるトークンに適用され、例えば、トークンのアウトラインを平滑化し、明白なアーチファクトを除去することなどを行ってもよい。これに関して種々のフィルタは、当分野では知られており、メモリ 1 9（図 1）に格納される市販の OCR または OCR に似たソフトウェアプログラムの一部であってもよい。

#### 【 0 0 3 4 】

ブロック 3 5 で、各トークングループに関して、トークングループにおいて分類されるトークンのすべてを形態構造的に表現する代表的なトークンが求められる。例えば、トークングループにおけるすべてのトークンは、結合トークンを得るために結合されてもよい。例えば、平均化、中央値をとるなどの当業者には明白であるようなトークンを、またはさらに具体的に言えば、トークン形状を結合するための種々の方法が、可能である。本発明の種々の例示的な実施形態において、代表的なトークンは、平均化されたトークン（図 3）として求められる。平均化は、すべてのトークン画像の重心の点を整列し、あらゆるピクセルの位置の平均をとることによって（例えば、各一致するピクセルをとり、各ピクセルに関する平均色（グレースケール）値を計算し、色（グレースケール）値のすべての和をとり、トークン画像の数によってその和を除算することによって）実行され得る。さらに、内挿が、サブピクセルレベルの平均値を得るために行われてもよい。平均化されたトークンが、トークングループにおけるトークン画像のすべての平均化から作成される場合には、元のトークン画像において提示され得る種々の不完全さまたはアーチファクト

10

20

30

40

50

は、ぼやけが増大するという潜在的な代償を払うことになるが、平均化されて（または最小化されて）略平滑化されたエッジを作成する。

【0035】

一部の実施形態において、結合トークンを作成するために、トークングループに含まれるトークンのすべてが、結合（例えば、平均化）される必要はないことに留意すべきである。例えば、1000個を超えるトークンなどトークングループにおける大量のトークンがある場合には、平均化されたトークン画像の品質が、数百のトークンの後で認識可能なほど向上しないため、すべてのトークンを平均化する必要がない可能性もある。そのような場合には、わずか100個程度の「最も近い」トークン画像が取得されて平均化され、平均化されたトークンを生成してもよい。

10

【0036】

ブロック36で、トークングループにおいてトークンのすべての形状を形態構造的に表現するが、ある程度のぼやけを有する代表的な（例えば、結合または平均化された）トークンが、代表的なトークンの数学的表現であるベクトル化トークンに変換される。本明細書で用いられるとき、「ベクトル化」という用語は、代表的なトークンの形状を最もよく表すアウトラインを求め、（任意の囲まれた部分を満たすために、適切なフィル命令と共に用いる）数式においてそのアウトラインを表す処理を指す。ビットマップをベクトル図形に変換するための任意の適切なラスターベクトル変換ソフトウェアは、本発明によれば、代表的なトークンをベクトル化するために用いられてもよい。説明のための実施形態において、ブロック36でベクトル化する前に、コントラスト調整、デスクューイング、

20

【0037】

本発明の一部の例示的な実施形態において、代表的なトークンに関するアウトラインは、トークン領域の解析に基づいて求められる。具体的に言えば、各代表的なトークンは、2つ以上の領域に分割される。例えば、字「e」は、3つの領域、すなわち、背景、「e」を表す実線部分、および「e」の上部分における半円形状の孔を有する。アウトラインは、任意の隣接する領域間の境界の集合として求められ得る。例えば、「e」のアウトラインは、実線部分と半円形状の孔との間の別の境界と組み合わせた背景と実線部分「e」との間の境界として求められ得る。同様に、字「i」は、3つの領域、すなわち、背景および2つの実線部分を有し、そのアウトラインは、第2の（下部の）実線部分と背景との間の別の境界と組み合わせた第1の（上部の）実線部分と背景との間の境界として求められ得る。

30

【0038】

さらに、代表的なトークンのアウトラインを数学的に表現するための種々の方法が可能である。例えば、ベジエ円弧/曲線表現技術は、コンピュータフォントのレンダリングの当分野ではよく知られており、制御点および制御点の各隣接対との間の曲線（直線を含む）に基づく任意のアウトラインを表すために用いられてもよい。さらに図3を参照すると、46で、例えば、ベクトル化トークン「e」は、9つの端点1から9と、各隣接対の端点の間の9つの曲線1-2、2-3、3-4、4-5、5-6、6-7、7-1および8-9を表す数式とによって表される。各隣接対の端点はまた、ベジエ曲線の外観（または「曲線性」）を制御するために用いられる2つの他の制御点を有する。本発明の一実施形態において、代表的なトークンの各曲線を定義するために用いられるベジエ曲線の数は、代表的なトークン（またはさらに具体的に言えば代表的なトークンによって表されるトークン）がどれほど頻繁にドキュメント中に現れるかに応じて変化してもよい。例えば、一部のトークンは、ドキュメント中に数千回生じ、他のトークンは、数回程度しか生じない可能性がある。頻繁に生じるトークンを定義するためにより多くのベジエ曲線を使用することを可能にすることによって、ドキュメントにおけるトークンの大部分に関して画像品質を向上すると同時に、あまり頻繁に生じないトークンの良好な圧縮を依然として達成することができる。

40

50

## 【0039】

Bスプラインまたはエルミート技術などの他の数学的表現技術もまた、当業者には明白であるように、ベクトル化トークンを定義するために用いられてもよい。ベクトル化トークンは、形状の数学的表現であるために、任意の解像度、例えば、2400 dpi または 19200 dpi もの比較的高解像度でレンダリングされ得る。また、ベクトル化トークンは、そのメモリ空間に関して、表す元のトークンのいずれと比較しても、著しく圧縮される。例えば、本発明の種々の例示的な実施形態において、単独のベクトル化トークンを表すために180バイト程度をとってもよい。

## 【0040】

したがって、多くのベクトル化トークンが定義され、それぞれのベクトル化トークンは類似の形状を有するトークンのグループを表してもよい。例えば、200ページの書籍からのページ画像は、2,000個を超えるベクトル化トークンを作成するために処理され、それぞれが、同様に形成されたトークンのグループを表すようにしてもよい。書籍において最初に識別されたトークンのすべては今度は、2,000を加えたベクトル化トークンの1つによって表されることに留意されたい。本発明の方法は、特定の文字または一定のフォントタイプとして認識することなく、ベクトル化トークンを定義することにも留意されたい。むしろ、方法は、スキャンインされた書籍などの元のドキュメントにおいて求められるすべてのトークンの形態構造的な特徴の解析に基づいて、画像として純粋にベクトル化トークンを定義する。スキャンインされたドキュメントを処理するために、この画像に基づく手法は、トークンベースファイルを作成するためのキーの1つであり、印刷した元のドキュメントと同じルックアンドフィールを維持すると同時に、高解像度でレンダリングされ得る。

## 【0041】

図2を参照すると、ブロック37で、トークンベースファイルは、ブロック36において前に定義されたベクトル化トークンに基づいて作成される。このファイルにおいて、各ベクトル化トークンは、トークン番号に割り当てられ、トークングループを形成するトークンのそれぞれの位置は、トークンのグループを表すベクトル化トークン（またはさらに具体的に言えば、そのトークン番号）に関連付けられる。例えば、各トークンの位置は、ページ番号およびトークンが現れる各ページ内の位置のX-Y座標によって定義されてもよく、位置は、ポインタによって対応するベクトル化トークンに対して関連付けられる。したがって、レンダリング時に、ベクトル化トークンは、元のトークンとは対照的に、この位置で表示され、それによってベクトル化トークンのみからなるページ画像を作成する。ベクトル化トークンがトークン形状の数学的表現であるために、高解像度をはじめとする任意の解像度でレンダリングされ得、表示されたときに整って可読であるように見える。さらに、同様に形成されたトークンの複数の位置が、ポインタを用いて代表的なベクトル化トークン（小さなメモリサイズを有する）に単に関連付けられるため、これらの位置に関して元のトークンを格納する必要はなく、したがって、ページ画像に関する格納要件を最小限に抑えることができる。例えば、平均で、書籍は、約2MBのメモリサイズを有するトークンベースファイルに変換され得る。さらに別に、各ベクトル化トークン（例えば、180バイト）の小さなメモリサイズに起因して、トークンベースファイルのきわめて高速のレンダリングが可能である。さらに別に、トークンベースファイルは、任意の数の印刷媒体にさらにレンダリングされてもよい。

## 【0042】

本発明の説明のための実施形態について図示して記載してきたが、本発明の趣旨および範囲を逸脱することなく、その中で種々の変更が成され得ることは十分に認識されよう。

## 【図面の簡単な説明】

## 【0043】

【図1】本発明の実施形態を実施するために用いられ得る例示的な計算システムの機能ブロック図である。

【図2】本発明の一実施形態による文字列を含む電子画像をトークンベースファイルに変

10

20

30

40

50

換するための例示的な方法の流れ図である。

【図3】図2に示された文字列を含む電子画像をトークンベースファイルに変換する方法のステップの一部を概略的に示す絵図である。

【図4A】本発明の種々の実施形態による文字列を含む電子画像をトークンベースファイルに変換する方法において、それらの形状にしたがってトークンを識別して分類するために用いられ得る種々の技術を示している。

【図4B】本発明の種々の実施形態による文字列を含む電子画像をトークンベースファイルに変換する方法において、それらの形状に基づきトークンを識別して分類するために用いられ得る種々の技術を示している。

【図4C】本発明の種々の実施形態による文字列を含む電子画像をトークンベースファイルに変換する方法において、それらの形状に基づきトークンを識別して分類するために用いられ得る種々の技術を示している。

10

【図1】

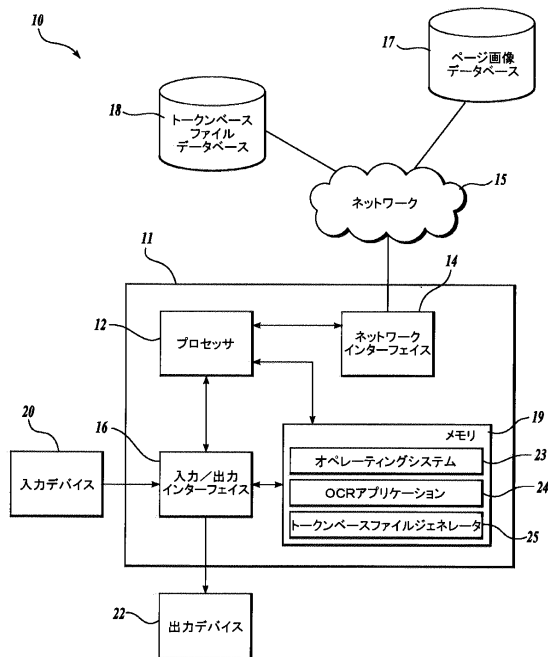


Fig.1.

【図2】

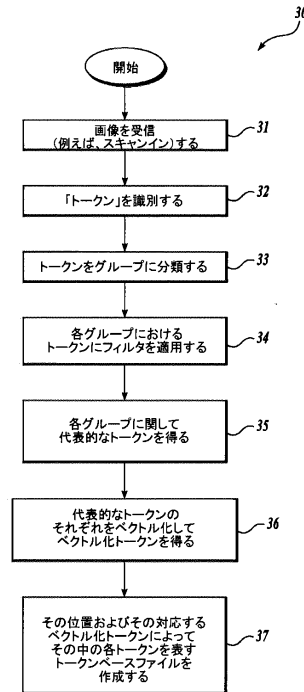


Fig.2.

【図3】

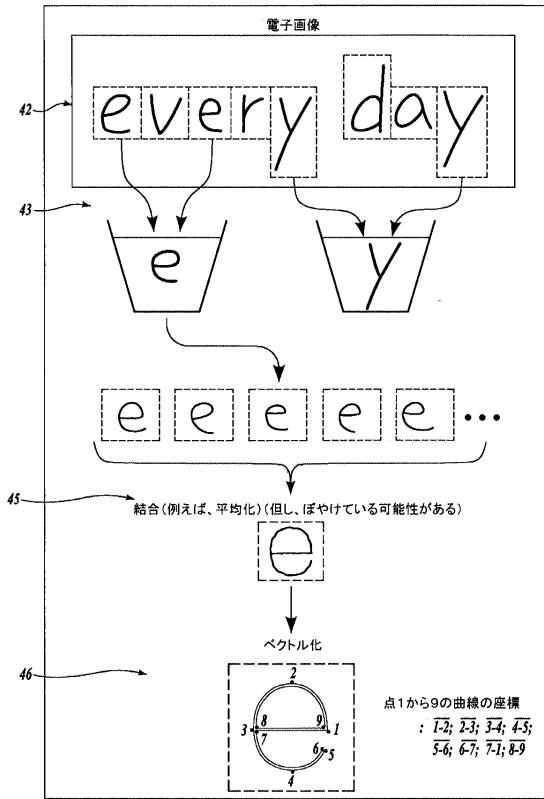


Fig.3.

【図4A】

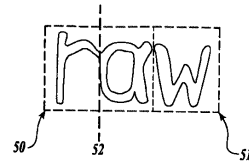


Fig.4A.

【図4B】

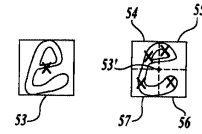


Fig.4B.

【図4C】

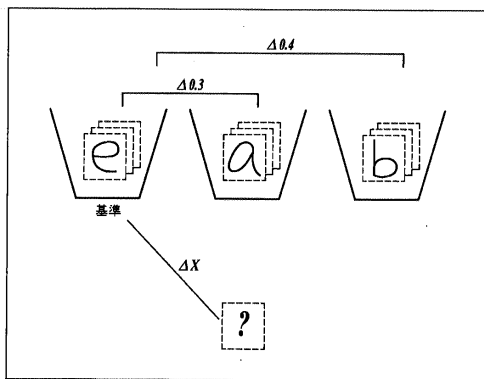


Fig.4C.

## フロントページの続き

- (74)代理人 100103920  
弁理士 大崎 勝真
- (74)代理人 100124855  
弁理士 坪倉 道明
- (72)発明者 コーチ, アダム・ブライアン  
アメリカ合衆国、ワシントン・98144、シアトル、1200 - トウエルブス・アベニュー・サ  
ウス、スイート・1200
- (72)発明者 アカルン, フレデリック・ジーヤ・ラモス  
アメリカ合衆国、ワシントン・98144、シアトル、1200 - トウエルブス・アベニュー・サ  
ウス、スイート・1200
- (72)発明者 グツドウイン, ロバート・エル  
アメリカ合衆国、ワシントン・98144、シアトル、1200 - トウエルブス・アベニュー・サ  
ウス、スイート・1200
- (72)発明者 シヤガム, ヨシユア  
アメリカ合衆国、ワシントン・98144、シアトル、1200 - トウエルブス・アベニュー・サ  
ウス、スイート・1200

審査官 鹿野 博嗣

- (56)参考文献 特開2005-346586(JP, A)  
特開平06-180771(JP, A)  
特開平10-091724(JP, A)  
特開2000-113112(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06K 9/48

G06T 11/20