

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3576570号  
(P3576570)

(45) 発行日 平成16年10月13日(2004.10.13)

(24) 登録日 平成16年7月16日(2004.7.16)

(51) Int. Cl.<sup>7</sup>

F I

G06K 9/46

G06K 9/46

A

G06K 9/36

G06K 9/36

請求項の数 3 (全 29 頁)

<p>(21) 出願番号 特願平4-325993                  (22) 出願日 平成4年11月11日(1992.11.11)                  (65) 公開番号 特開平5-282495                  (43) 公開日 平成5年10月29日(1993.10.29)                      審査請求日 平成11年11月11日(1999.11.11)                      審査番号 不服2001-14335(P2001-14335/J1)                      審査請求日 平成13年8月13日(2001.8.13)                  (31) 優先権主張番号 795169                  (32) 優先日 平成3年11月19日(1991.11.19)                  (33) 優先権主張国 米国(US)</p>	<p>(73) 特許権者 596170170                  ゼロックス コーポレイション                  XEROX CORPORATION                  アメリカ合衆国 コネチカット州 スタン                  フォード、ロング・リッジ・ロード 80                  O                  (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 ダニエル ビー、ハッテンロッカー                  アメリカ合衆国 14850 ニューヨー                  ク州 イシカ コムストック ロード 3                  14</p>
--	--

最終頁に続く

(54) 【発明の名称】 比較方法

(57) 【特許請求の範囲】

【請求項1】

複数の記号ストリングの等価性を判定する方法であり、この記号シンボルは記号シンボル間の相似性の相対計測を判断するために、画像を定義するデータ内で、隣接する行中の隣接する単語同士から鮮明に分離されたプロピファイされた単語である単語対象物を形成し、

画像を定義するデータの第1の離散記号ストリングを検出し、第1の記号ストリングによって表される単語対象物を分離するステップと、

第1の記号ストリングのバウンディングボックスと黒画素との距離である、該第1の記号ストリングの形状の第1の輪郭信号を示す値を推定するステップと、所定の基準線から第1の記号ストリングまたはバウンディングボックスへ向かう第1の中央輪郭信号を示す値を推定するステップと、

第1の記号ストリングのエックス・ハイト、アセンダー、及び、デセnderの何れかの大きさである特徴的大きさを測定するステップと、

画像を定義するデータの第2の離散記号ストリングを検出し、第2の記号ストリングによって表される単語対象物を分離するステップと、

第2の記号ストリングのバウンディングボックスと黒画素との距離である、該第2の記号ストリングの形状の第2の輪郭信号を示す値を推定するステップと、第2の記号ストリングのエックス・ハイト、アセンダー、及び、デセnderの何れかの大きさである特徴的大きさを測定するステップと、

所定の基準線から第2の記号ストリングまたはバウンディングボックスへ向かう第2の中央輪郭信号を示す値を推定するステップと、

第2の記号ストリングの特徴的な大きさで、第1の記号ストリングの特徴的な大きさを割ることにより、縮小拡大率を計算するステップと、

第1の輪郭信号および第1の中央輪郭信号に対する次の比較のため、新しい第2の輪郭信号および第2の中央輪郭信号を作るために前記縮小拡大率に従って、二次元において第2の輪郭信号および第2の中央輪郭信号を縮小又は拡大するステップと、

前記単語対象物の単語の上部、下部、及び、中央部である信号が定義される範囲にわたって、第1及び第2の輪郭信号間、および、第1および第2の中央輪郭信号間の差である差信号を示す値を求めるステップと、

第1及び第2ストリングの間の相対相似の差計測表示に到達するために、前記定義された範囲の部分にわたって、前記差信号を評価するステップと、

を含む、複数の記号ストリングの等価性を判定する方法。

#### 【請求項2】

差計測に対して、第1と第2の記号ストリングが等しいかどうか判断するステップを含む、請求項1に記載の複数の記号ストリングの等価性を判定する方法。

#### 【請求項3】

第1と第2の記号シンボルが等しいかどうか判断するステップにおいて、

所定の閾値と差計測を比較するステップと、

差計測が閾値よりも小さいならば、第1及び第2の輪郭信号が、同一の記号シンボルを表す語形から生成されたと判定するステップと、

を含む、請求項2に記載の複数の記号ストリングの等価性を判定する方法。

#### 【発明の詳細な説明】

##### 【0001】

##### 【産業上の利用分野】

本発明は、画像データの配列において形状で表されたテキストまたは文字列を、前記列を構成する文字または複数の文字を個々に検出及び/または確認する必要なしに認識する方法に関する。

##### 【0002】

##### 【従来の技術】

電子的に符号化されたドキュメント（電子ドキュメント）のテキストは、互いに異なる2つのフォーマットのいずれかであることが多い。第一のフォーマットでは、テキストはビットマップフォーマットであり、この場合、テキストは画像データまたは画素の配列によってのみ定義され、本質的に同様に表された隣接する画像との区別はできない。このフォーマットでは、通常、本文の内容にのみ基づくコンピュータ処理にかけるとはできない。以下、文字コードフォーマットと呼ばれる第二のフォーマットでは、テキストが文字コード列（例、アスキーコード）として表される。文字コードフォーマットでは、テキストの画像またはビットマップは利用できない。

##### 【0003】

光学文字認識（OCR）処理を利用したビットマップから文字コードフォーマットへの変換では、膨大な時間および処理労力が必要なる。文字の各ビットマップはその近隣のものや解析されたその外観から区別しなければならず、さらに決定工程において、所定の集合の文字における個別の文字として識別される必要がある。たとえば、米国特許第4,864,628号では、文字画像を一周するデータの読み取り方法について開示する。米国特許第4,326,190号では、英数文字を読み取る文字特徴検出システムについて開示する。米国特許第4,813,078号では同様の処理を用いた文字認識装置が開示され、さらに米国特許第4,833,721号でも同様のシステムについて開示されている。

##### 【0004】

デジタル化された画像内の文字を識別する他の方法や装置について開示しているその他の引例としては、米国特許第3,755,780号、米国特許第3,899,771号、米

10

20

30

40

50

国特許第4,817,166号、さらに米国特許第4,566,128号がある。さらに、米国特許第4,956,869号は、複数の線から成る画像内の形状の輪郭座標を用意するためのさらに効率のよい輪郭のトレース方法について示している。

【0005】

【発明が解決しようとする課題】

原稿を走査することによって電子ドキュメントが作成された場合でも、その再生画像の画質および雑音はビットマップの実際の外観の不確定性によるものである。さらに、質の悪いオリジナルドキュメント(原稿)、走査ミス、または画像のデジタル化表現に影響を与える同様の要素によって劣化したビットマップ外観が生じる場合もある。したがって、文字識別で用いられる決定工程は本質的に文字識別に関する不確定性を有する。ここで特に問題となるのはテキスト中の文字が不鮮明になったり(ぼやけたり)または重なってしまう点である。ほとんどの文字識別工程では、文字が連結された画素の独立集合であるという仮定から出発する。入力画像の質のためにこの仮定が誤っていた場合、文字識別も失敗する。米国特許第4,926,490号では、歪んだドキュメントの文字が認識された場合のドキュメントの文字認識方法及びその装置について開示している。米国特許第4,558,461号は、ドキュメントの垂直パッチを解析することで歪んだテキストを調節するテキスト行限定システムについて開示している。米国特許第3,295,105号では、文字認識装置における文字正規化のための走査制御装置について開示する。米国特許第4,918,740号は、光学文字認識システムで利用するための処理手段について開示する。米国特許第4,809,344号では文字認識に必要なデータを得るための文字認識の事前処理について開示する。

10

20

【0006】

OCR法の信頼性を向上させるため、米国特許第4,010,445号に開示されるような辞書単語確認法がこれまで用いられてきた。論文「F6365日本製ドキュメント読み取り機」(富士通科学技術ジャーナル、26号、3巻、224頁乃至233頁、1990年10月刊)には、ブロック抽出段階、歪み調節、ブロック分割、隣接文字セグメント化、行抽出、辞書検査を併用するパターン整合および比較による文字認識を利用した文字読み取り機が示される。テキスト列に対して、その読み取り方位に平行なテキストを走査する3個で一組のスキャンが用いられ、各スキャンが黒から白への推移に関する情報を引き出す米国特許第2,905,927号に示されるように、単語または文字列それぞれ自身を形成する文字集合を識別するのが好ましい。米国特許第4,155,072号は、文字の前エッジおよび後エッジを表示する値の集合を生成するために動作可能な同様の装置を示す。

30

【0007】

プリントまたはタイプされた本文画像を処理するOCRシステムのほかに、多数の引例で電子表示に変換された手書きテキストの認識を扱っている。

【0008】

署名認識でもすでに考慮された認識の基本単位として単語全体を選択することが、米国特許第3,133,266号により示されており、これは今までのとおり文字画像の特性を識別する連続特徴識別方法に基づく。

40

【0009】

米国特許第4,499,499号は、単語のビットマップ表現をビットマップ表現辞書と比較する画像圧縮方法について示す。ビットマップをコンピュータで情報として処理可能な文字へ符号化するOCR法も、画像走査用のビットマップ法も、テキスト操作またはテキスト処理のあらゆる目的に対して完全に満足できる方法というわけではなかった。また、米国特許第4,949,281号は、画像物体を生成および再生する多項式の利用法を開示する。ただし、前記物体は輪郭座標中に基準輪郭の形状であらかじめ定められている。

【0010】

単語が比較的単純な方法で表現可能であれば、既知信号と未知信号とを比較するある種の

50

信号処理技術が利用できる。米国特許第4,400,828号では、入力された単語が基準単語の集合から認識される音声単語認識装置について開示する。正規化および線形タイムワーブ装置が開示される。入力単語および基準単語の集合は電氣的に一致を確定するために処理される。米国特許第4,977,603号では、多重相似法を用いたパターン認識装置が開示される。さらにアーキン、チュウ、フッテンロシエル、ケデム、ミッシェル共著「多角形状の比較のための効率的計算可能計量」(「離散数学に関する第1回年次ACM-SIAM[アメリカ計算機学会/工業・応用数学界]シンポジウム議事録」129頁乃至137頁、1990年1月刊)では、形状マッチングが計量可能であることが示される。

【0011】

【課題を解決するための手段】

本発明の一態様によれば、画像データ内に単語対象物を形成する複数の記号ストリングの同値を確定する方法が提供される。この方法は離散した記号ストリングを検出し、これによってストリングで表現される単語対象物を分離し、記号ストリングの形状を表す輪郭信号を導出する。検出および導出を繰り返すことで第二単語対象物の形状が分離され、第二輪郭信号で表現される。続いて、この方法は、2つの輪郭信号の差を表す差信号を両信号が定義される範囲にわたって確定するステップと、最終的に記号ストリングまたは単語対象物同士の相似を示す差計測(寸法)となる差信号の数値を求める。

本発明の別の態様によれば、複数の記号ストリングの等価性を判定する方法が提供される。この記号シンボルは記号シンボル間の相似性の相対計測を判断するために、画像を定義するデータ内で単語対象物を形成する。この方法は、画像を定義するデータの第1の離散記号ストリングを検出し、第1の記号ストリングによって表される単語対象物を分離するステップと、第1の記号ストリングの形状の第1の輪郭信号代表値を推定するステップと、第1の記号ストリングの特徴的な大きさを測定するステップと、画像を定義するデータの第2の離散記号ストリングを検出し、第2の記号ストリングによって表される単語対象物を分離するステップと、第2の記号ストリングの形状の第2の輪郭信号代表値を推定するステップと、第2の記号ストリングの特徴的な大きさを測定するステップと、第2の記号ストリングの特徴的な大きさを、第1の記号ストリングの特徴的な大きさを割ることにより、縮小拡大率を計算するステップと、第1の輪郭信号に対する次の比較のため、新しい第2の輪郭信号を作るために前記縮小拡大率に従って、二次元において第2の輪郭信号を縮小又は拡大するステップと、第1及び第2の両信号が定義されている範囲にわたって、第1及び第2の輪郭信号間の差である差信号代表値を求めるステップと、第1及び第2ストリングの間の相対相似の差計測表示に到達するために、前記定義された範囲の部分にわたって、前記差信号を評価するステップと、を含む。

上記の複数の記号ストリングの等価性を判定する方法において、差計測に対して、第1と第2の記号ストリングが等しいかどうか判断するステップを含む、方法も提供される。

さらに、上記の複数の記号ストリングの等価性を判定する方法において、所定の閾値と差計測を比較するステップと、差計測が閾値よりも小さいならば、第1及び第2の輪郭信号が、同一の記号シンボルを表す語形から生成されたと判定するステップと、を含む、方法も提供される。

本発明の別の態様では、複数のストリングの記号の等価性を判定する装置が提供される。この記号ストリングは、記号ストリング間で相対計測の相似を判定するために、画像を定義するデータ内の単語対象物から形成される。この装置は、画像を定義するデータ内で、第1及び第2の離散記号ストリングを検出し、ストリングによって表される単語対象物に離散記号ストリングを分離する分離手段と、第1の記号ストリングの形状の第1の輪郭信号代表値と第2の記号ストリングの形状の第2の輪郭信号代表値とを推定する手段を備える。上記推定する手段は、記号ストリングの全ての記号に共通する大きさを決定し、前記決定手段は、第1の共通測定値及び第2の共通測定値の両方を生成するために第1及び第2の記号ストリングの両方において作用する手段と、第2の共通測定値に対する第1

10

20

30

40

50

の共通測定値の比に比例して、輪郭信号の一つを縮小又は拡大する手段と、定義された範囲にわたって第1の輪郭信号の大きさと第2の輪郭信号の大きさの間の差の差信号代表値を求める手段と、第1及び第2のストリング間で相対相似の差計測値に到達するために、定義された範囲の一部にわたって差信号の評価を行う演算処理手段と、を含む。

本発明の別の態様では、記号の第1のストリングを比較する方法が提供される。第1のストリングは既知のストリングの記号とともに画像を定義するデータの中で単語対象物を形成し、記号の既知のストリングは第1のストリング及び既知のストリング間の相似の相対計測を求めるために単語対象物を形成する。この方法は、第1の画像を定義するデータの第1の離散記号ストリングを検出し、記号の第1のストリングによって表される単語対象物を分離するステップと、記号の第1のストリング形状の第1の輪郭信号代表値を推定するステップと、既知の記号のストリングの形状の第2の輪郭信号代表値を推定するステップと、第1と第2の信号が定義された領域にわたって、第1と第2の輪郭信号間の差の差信号代表値を求めるステップとを有する。上記差信号代表値を求めるステップは、第1の輪郭によって形成される円弧に対して、重心を検出するステップと、第2の輪郭によって形成される円弧に対して、重心を検出するステップと、これらの重心間の相対変位を決定するために、第1と第2の輪郭に対する重心を揃えるステップと、輪郭の極値で、輪郭の重複しない範囲において定義される唯一の輪郭を有する範囲を判定するステップと、輪郭の重複しない範囲において、この範囲内で定義された輪郭信号と所定の定数の間の差としての差信号を定義するステップと、両輪郭が定義された重複する範囲において、この範囲内で定義された輪郭信号間の差としての差信号を定義するステップと、第1と第2のストリング間の相対相似の差測定値に到達するために、定義された範囲の一部にわたって、差信号を評価するステップと、を含む。

本発明の別の態様では、2つの印刷されたテキスト・ドキュメントの両方に現れる共通な単語の数によって定義された関連性を決定する方法が提供される。前記ドキュメントは画像データの配列として表される。この方法は、ストリング内の分離している隣接記号の距離と、分離している隣接記号ストリングの距離との差に基づいて、第1のドキュメントを表す画像データの第1の配列の中で複数の記号ストリングに対する語形輪郭を生成するステップと、分離しているストリング内の隣接記号の距離と、分離している隣接記号ストリングの距離の差に基づいて、第2のドキュメントを表す画像データの第2の配列において複数の記号ストリングに対して語形輪郭を生成するステップと、比較される語形輪郭の全ての対に対して差計測を生成するために、第2の配列の語形輪郭の全てと第1の配列の語形輪郭の全てを比較するステップと、所定の閾値の元で、差計測を有する語形対の数を決定するステップと、前記数に基づいて、2つのドキュメントに含まれる単語の割合が相似であることを判定し、2つのドキュメントの関係の表示を生成するステップと、を含む。

#### 【0012】

本発明は、単語およびテキストストリングの基本特性を利用すると同時にOCR方法に固有の問題を解消することを目的とする。特に、画像導出またはイメージング工程に固有の信号対雑音比は文字に対して比較的小さいが、より大きな文字列と比較すれば比較的大きい。さらに、逐語間隔は逐字間隔よりも大きくなる傾向があり、このため、個々の文字の識別と比較した場合、文字列(ストリング)の分離および識別が向上される。OCR方法も、アセンダー、デセンダー、カーブ等の間違いやすい文字の部分の識別を含む正しい識別の準備として、文字の態様に関するいくつかの正しい決定を必要とする傾向がある。単語の境界を識別するためにはまず画像内のテキスト行の特性を確定しなければならない。さらに、本発明によれば単語形状を表す1次信号を導出し、導出された信号と他の単語形状信号とを比較することができる。このため、比較がなされるまで単語に関する仮定はなされず、これによって連続的な誤った比較および決定を生じさせる可能性のある無効文字による仮定の影響を解消する。

#### 【0013】

コンピュータ処理されたテキストの潜在的利用法について調べると、少なくともいくつか

10

20

30

40

50

のケースでは単語の各文字を引き出すことが必ずしも処理上の必要条件でないことが明らかとなった。OCR方法で文字を誤って確定する確率は比較的小さいかもしれないが、全単語にわたるとこの確率は増加的に累加される。このため、単語を探索したりあるいは認識する前に単語を文字コード列へ変換するためにOCRを用いると、相当なエラーが生じる場合がある。本発明では、人間がテキスト句を読んだりざっと目を通す場合に用いるのと似たやり方で連続単語認識することを可能とするため画像データの単語レベルセグメント化を利用する。OCR方法はビットマップから表示文字コードへ変換するが、これによってビットマップの情報内容を失うことになる。通常、文字コードから原ビットマップを得るために工程を逆にすることはできない。しかしながら、本発明によって開示されるように形状に基づく単語の識別では、さらに多くのビットマップ情報をさらに認識工程に保存させることとなり、これによって1次元信号からビットマップの適当な再構成が可能となる。したがって、ビットマップ情報の重要な部分は、選択されたテキストまたは文字列の形状を表すために利用される1次元信号によって保存される。

【0014】

【実施例】

これより図面を参照するが、これらは本発明の好ましい実施例を示すためのものであり、これに限定されるものではない。図1は、本発明が有利に利用できるさまざまな状況を対象とする一般化された画像処理システムを示す。通常、ソース画像は、スキャナ、ファクシミリ装置、または記憶システムでもよいソース画像導出システム2から導出することができる。ソース画像はコンピュータ処理装置4へ進む。コンピュータ処理装置4はここで開示される新規の装置だけでなくすでに公知のいかなる装置であってもよい。ユーザインタフェース6で入力されたコマンドに対して、処理装置4は出力装置8で出力を生成する。出力装置8はプリンタ、ディスプレイ、ファクシミリ装置またはその他の記憶装置であってもよい。本質的には、図1の上側で図示されるように、入力ドキュメント(文書)はシステム内へ方向づけられ、出力ドキュメントはシステムより引き出される。

【0015】

図2は、単語の形状によって単語を導出し、定義し、比較するため本発明を具体化したシステムを示す。当然のことながら、このシステムの各要素は多数の装置であってもよいし、または単一装置内で作動するプログラムであってもよい。まず、入力ビットマップ10のソースは不確定であり、本発明の範囲ではないが、ビットマップは最初にセグメント化システム12へ方向づけられる。セグメント化システム12において単語、文字列(ストリング)、または理解に関するその他の複合文字単位が導出される。まず、画像ビットマップは画像中のテキストの方向の角度を確定する歪み(スキュー)検出器14を通過する。画像の方向に関する情報および画像自身を用いて、テキスト底線プロセッサ16では、テキストの頂上線および底線が確定され、これによって画像中のテキストの行の上方境界および下方境界が識別される。中央フィルタ18では「プロビファイ(ぼかし、塊化)」とよばれる機能が実行され、これは画像上で作動し、行中の各単語群を単一単位として扱うことができる。ここで用いられるような「単語」、「記号列」、「文字列」という言葉は、ともに意味理解の単一単位を形成する連結された英数文字または句読点要素の集合、さらに広義には、印または記号の集合を示す。このような理解の単一単位は、前記単位を構成する要素、印、記号を分離するものより広い字間あけにより分離される場合に画像中で特徴づけられる。プロビファイをかけた画像に対して、空白行(白線)の集合がブロック20で加えられ、テキストの隣接する行を鮮明に分離する。空白行はプロセッサ16によって提供される底線確定に基づく。この情報、すなわち、隣接する行中の隣接する単語同士から鮮明に分離されたプロビファイされた単語を用いて、ブロック22で単語についてのバウンディングボックス(境界付けボックス)が定義され、これにより単語を識別するとともに囲む。

【0016】

その後、単語形状信号コンピュータ24が、原(元の)画像および前記バウンディングボックス確定に基づき画像中の個々の単語を表す単語形状信号を導出する。単語形状辞書2

10

20

30

40

50

8 から既知の単語を表現する単語形状信号とまだ未確認の単語形状信号とを比較する場合、この情報が単語形状コンパレータ 26 で利用可能である。他の実施例では単語形状コンパレータ 26 は画像 10 から確定された 2 個またはそれ以上の単語形状を比較するために利用することができる。さらに重要なことは、単語形状コンパレータ 26 が認識されていない文字列から既知の単語形状への単語形状比較に限定されていない点である。単純化された状況では、コンパレータ 26 は単なる一単語形状をもう一つの単語形状と比較する装置にすぎず、これによって 2 つの形状の間の相似度を相対的に表示することができる。

#### 【0017】

一般的に単語形状認識工程の段階について説明したが、このようなシステムの数学的根拠について考えてみる。例えば、画像データ  $i(x, y)$  を考えてみると、これは通常の場合ではビットマップの形状の一系列の画像データである。文字集合はおそらく上述のように多数の方法の一つで識別され、画像データ配列の部分集合内に選択された文字列を囲む境界を定義する。 $i(x, y)$  から、囲まれた境界内で検出された  $i(x, y)$  のエッジを表すエッジ信号  $e(x, y)$  が導出される。 $i(x, y)$  へ追加データを加えることにより前記エッジ信号が拡張され、その結果、 $e(x, y)$  は囲まれた境界内の独立変数に対してその全領域にわたって定義された連続信号  $e'(x, y)$  となる。各 1 次元信号  $g'(t)$  である  $e'(x, y)$  からは 1 個、2 個またはそれ以上の信号を導出することができる。ただし、 $g$  は基準フレーム従属パラメータである独立変数  $t$  の関数である。

10

#### 【0018】

単語形状導出工程に関するこの数学的説明から、他にも単語形状信号導出方法が存在することがわかる。可能性のある他の実施例は、極座標等の他の座標図表を用いた 1 次元信号の確立する手法である。もう一つの可能性は信号  $g(t)$  の生成である。ただし、 $g(t)$  は各輪郭点からそれに続く輪郭点への方向を示し、 $t$  は点番号を表す。

20

#### 【0019】

次に単語形状計算の計算方法について説明する。画像の方向、行間、単語群の字間あけがいったん決まると、各単語はバウンディングボックスによって囲むことができる。さらに文字列画像を通過して伸長する基準線が作成される。この基準線はエックスハイト ( $X - height$ ) の 3 分の 2 からエックスハイトの 3 分の 1 までの限定的な厚さを有するブロックでもよく、実際には幅がゼロであってもよい。画像の分析において、基準線からテキスト輪郭またはバウンディングボックスの上縁部までの距離が基準線に垂直な方向に計測される。同様に、基準線から下方バウンディングボックス縁部またはテキスト輪郭までの距離が単語の下方部分に沿って計測される。導出された値の集合は計算上、長手方向に対する水平軸に沿った位置によって表せるため、この信号は 1 次元信号と考えることができる。これらの値のいずれか、または両方を用いて単語形状を記述することができる。さらに、あまり望ましいことではないかもしれないが、本発明の範囲内では、バウンディングボックスの頂上またはバウンディングボックスの底部から引かれ、単語または基準線に最初に接する垂直線の距離を必要に応じて計測することができる。

30

#### 【0020】

ここで重要なのは、1 次元信号の導出に用いられる数学的工程が本質的に可逆的であるということである。基準線が限定的な厚さであるために画像から取り出された場合、その画像部分は識別不能であるものの、基準線の幅がゼロであれば、この情報はそのまま残るものとする。

40

#### 【0021】

認識辞書または単語形状の参照テーブルを、上述の工程を用いることで鮮明に作成することができる。走査された単語を情報のソース (源) として利用するか、またはより「完璧な」辞書を目指してコンピュータ生成単語を実際に用いることで、前記工程が影響される。

#### 【0022】

単語形状比較のための一般的な方法および装置について定義し、この方法の基礎となる数学的根拠も説明してきたが、これより各段階についてさらに詳細に説明する。本発明の工

50

程をさらに詳述するため、図3では公有ソースからとられた数行のテキストのサンプル画像が示される。図3はテキストページの画像がどのように表示されるかを示し、図4はこのページを走査した画像を示すが、これは公知のOCR方法で問題を引き起こすビットマップ画像を拡大して示すものである。例えば、テキスト画像の1行目の単語50「practitioner」を見ると、数個の文字が互いに連続している（くっついている）ように見える。また、画像の囲み52で囲まれた右下の部分では雑音がある。囲み54で囲まれた単語「practitioner's」を見ると、句読点と文字の連続（くっつき）がさらにはっきりとしている。

#### 【0023】

画像のテキスト行の方向を確定する一般的な方法では、任意に選択された少数のエッジ画素（少なくとも白色画素一つと隣接する黒画素として定義される）を調べ、各エッジ画素に対して行数を考慮する。サンプルとしては56a、56b、56cがあり、前記画素からそれぞれ指定された範囲の角度で離間された角度増分で伸びている。エッジ画素は全画像画素の集合から関数  $RandamEdgePixel$ （ランダムエッジピクセル）によって任意に選ばれる。図5（線56a、56b、56c参照）、図6（線58a、58b、58c参照）、図7（線60a、60b、60c参照）は、画像内のテキストの角度方向を正確に確定するため上述の技術が表示エッジ画素に対して適用されるますます狭められていく一連の角度範囲を表す。エッジ画素を見つけ、線を定義した後、歪み（スキュー）検出器14は各行の経路をトレースし、線とぶつかる連続する黒画素の列（ストリング）の長さを画素内に確定する画像境界に到達すると、個々の列（ストリング）の長さを合計するとともにその合計を求められた個別の列の総数で除すことによって黒画素列の平均的な長さが計算される。この演算はすべての線に対して実行され、これによって選択されたエッジ画素から伸びる各線に対する黒画素列の平均的な長さを求める。これらの長さは図8に曲線Aとして描かれ、約0ラジアンおよび3.14ラジアンで最小となることを示す。曲線Aは、一連の角度を設けられエッジ画素から伸びる線の各々に関する合計/平均化関数をグラフに表したものであり、その範囲は0から2ラジアンまでである。いったん第一の最小値が決まると、第二の最小値が第一の最小値からほぼラジアンで存在するかどうかを確定することで最小値の確認（サンプルでは、ほぼ0ラジアン）が行われる。第二最小値の存在が確認されると（サンプルでは、ほぼ3.14またはラジアン）、大まかな歪み（スキュー）角度が識別される。続いて、テキストの歪み角度をさらに密着的に確定する必要がある。これは、任意に選ばれたエッジ画素から伸びる多くの線を用いて達成できる。ただし、これらの線は最小角度増分だけ異なっており、さらに角度範囲は前記の大まかな歪み角度が中心とされる。しかしながら、所定の長さの線に沿って含まれる黒画素の総数を解析することで微細な歪み角度を確定することができる。さらに詳細に言えば、単位距離に関する画素数は図8の曲線Bとして描かれ、前記の微細な歪み角度は曲線の最大値を識別することで確定される。いいかえれば、単位線長あたり最も高濃度の黒画素が存在する曲線の点は、画像中のテキスト行の角度をより正確に表す。曲線Bで示されるように、これによってほぼ0ラジアンの微細な歪み角度となる。ただし、線はその長手方向に沿って最多数の黒画素と交差するため、確定に必要な方向づけのもっとも近接した角度を表す。

#### 【0024】

または、微細な角度確定について記述される手続きを多重反復することで歪み角度を確定することができる。図5乃至7に示されるように、所望の歪み角度精度が達成されるまで各反復もますます狭められる角度範囲をカバーする線を利用することとなる。図5乃至7により図示される実現方法では、それぞれ選ばれたエッジ画素を中心として180個の個別の角度を用い、一連の3回反復により所望の精度が達成される。

#### 【0025】

次の工程段階では、図9および図10に示されるように、テキスト底線プロセッサ16がテキストの各行の特徴線、すなわち上方頂上線および下方底線を識別する。テキスト底線プロセッサ16により実行される工程段階は、図11および図12に詳細に図示される。

10

20

30

40

50

画像に沿って左手に示される図9のヒストグラムは画像分析で線を調べることで導出され、すでに確定された歪み角度によって定義される場合、画像の歪み方向に平行に方向づけられる。画像を走るこれらの平行線は、各線と交差する黒画素数を確定するために用いられる。テキストを通過する線に沿って切り取られる黒画素は多数存在するが、テキスト行間の間隔を貫通する線に沿って切り取られる黒画素は存在しない。

#### 【0026】

さらに詳細には、関数ベースラインズがまず「主」線の座標を求める(ブロック142)。この主線は画像の中心を通過して構成され、前記関数へ送られた歪み角度により確定されるようなテキスト行に垂直である(ブロック140)。次に、ラインエンジン手続き144が実行される。ただし、主線に沿って一端から他端へ進むことにより、主線の一連の点では固定された距離だけ主線から外方へ向けて伸びる垂直支線が構成される(ブロック146)。前記支線に沿って黒の垂直エッジ画素の数がカウントされ(ブロック148)、線が交差する黒画素の数がカウントされるとともに(ブロック150)、さらに対向する対になった線に対して合計される(ブロック152)。ブロック148でカウントされたように黒の垂直エッジ画素は、上方または下方の近隣画素位置で白色画素と隣接した黒画素として定義される。ラインエンジンLine Engine( )手続き144は、決定ブロック154により確定されるように主線に沿ったすべての点および関連する支線が処理され終わるまで繰り返される。

#### 【0027】

続いて、すべての支線のカウン트가解析されることで、黒垂直エッジ画素対黒画素の最も高い比率を有する支線の対を確定することができる。通常、最も高いパーセンテージを有するこれらの線は、テキスト行を形成する文字の上方および下方エッジに沿って通過する線に対応する。図10の拡大図に示されるように、高い垂直エッジ画素比を有するこれらの支線(位置82)と、比率の低い支線(位置84)の間には明確な区別が存在する。フィルタマスクを適用し、マスク内の最大ピークを比較することで、線82等のテキスト頂上線および底線を表すこれらの線の識別が可能となる。ステップ150のヒストグラム動作に対して追加試験を行うこともできる。この追加試験、すなわちブル試験(boollean test)は、線の解析中に検出された黒画素は最小であったと保証するために用いられる。この試験は、垂直エッジ画素比が高いため小さな雑音または画像人工物が底線として認識されることはないということを保証する。

#### 【0028】

他の方法としては、支線に沿って並ぶ黒画素の総数を利用することで底線の位置を確定することができる。支線に沿ってカウントされた黒画素の数を表すヒストグラム曲線BLを用いると、どの支線が最多数の黒画素との交点を有するかを確定することができる。最大値の閾を適用することによって、各テキスト行に対して上方および下方の特性線の対を確定することができる。このため、ヒストグラム曲線BLの起伏部分はテキストの特性線を構成し、さらに前記閾は、介在される最小値を囲む局所化された最大値を明確に識別するために用いられることとなり、これによって、さらに処理するために用いられる底線位置を識別することができる。さらに重要なことは、ブロック162で図示されるこの手法を用いることで、BLヒストグラム曲線の勾配に基づき底線対の上方および下方の底線を識別することができる点である。重要なのは、ヒストグラム情報がステップ150ですでに収集されている場合、識別段階に関連する追加処理はほとんど存在しないことである。いったん予備特性線または底線対が識別されると(ブロック162)、確認段階(ブロック164)が実行されることで前記底線対が最小距離よりもさらに離間されていることを確認する。前記最小距離は画像中のすべての線対に対して平均的な線対分離を計算することで得られたものである。確認の後、有効な底線情報は出力ブロック166により格納され、後で空白行追加ブロック20とセグメント化ブロック22において利用される。

#### 【0029】

これらの底線確定方法の重要な利点は、これらの方法が行間の雑音または無関係なマークに対してほとんど反応しないという点である。図13はサンプルテキストの例文画像上で

10

20

30

40

50

底線を確定した結果を示す。底線対、すなわち底線  $B_n$  および頂上線  $B'_n$  がそれぞれ画像上に配置され、テキストの際立つ部分が生じる画像におけるそれらの部分を示す。文字アセンダーstrookのいくつかの部分は底線より外側であるものの、残りの工程に支障はきたさない。もちろん、閾値が小さいほど、システムはアセンディングstrookをより多く捕らえることができる。

#### 【0030】

図14および図15とともに図2を参照すれば、次の工程段階は単語群分離段階である。フィルタ18は画像のコピーに適用され、その結果のフィルタ画像は単語を互いに区別できるしみとする傾向がある。各領域に対して小さな窓を有するフィルタが適用され、部分的に黒いこれらの領域を黒として表現する。図14に示されるように、プロビファイ（プロブ化）関数はまず、マスクサイズおよび角度、を決めるマスク変数を初期設定し（ブロック180）、さらに上方走査線を処理することでデータ配列を初期設定する（ブロック182）。画像の中でマスク窓を順に移動させることで中央フィルタリングが達成される。さらに、窓に現れる黒画素の数が閾値を越える場合はいつでも、窓が配置される中心となる目標画素が黒に設定される。図15はフィルタ工程のいくつかのサンプルを示し、画像の一部に対してマスク窓（ウィンドウ）200が配置されている。例えば、ほぼ20%の閾であるとともに21個の画素を有する通常の矩形のマスクがテキストに対してすでに確定された歪み角度にほぼ等しい角度で配置された場合、窓200内のフィルタリングの結果、画素204は黒へ設定される。同様にして、最初に文字表示「r」と「o」の間の文字間内に存在する窓206は画素208を黒へ設定させる。他方、単語群の間の領域に存在する窓210内には、画素212を黒へ設定させるだけの十分な数の黒画素が現れていない。マスク窓200のサイズ、形状および方向は最適化され、単一単語に共通な文字間の充填物を最大化する一方でテキスト行間の充填物を減らす。

#### 【0031】

図16に示されるように、中央フィルタリングの結果では、単語の文字間の比較的小さな間隔あけは通常、取るに足らず、さらに黒画素で充填される。単語は画素の単一連続集合となる。すなわち、単一単語において完全に文字を分離させる空白間隔が存在しない。しかしながら、記号列間または単語間の比較的大きな間隔あけは、黒となるにはフィルタの能力を超えるほど大きな空間であり、したがって、隣接する記号列を区別するよう作用する。図13および図16によれば、サンプルテキストの最初の2単語「A」と「practitioner」はこの工程の呼び名でいえば「プロビファイ（プロブ化）」され、このため、例えば「practitioner」の「p」はもはやその単語の「r」から分離されない（図4と比較されたい）。つまり、文字のプロビファイングまたは不鮮明さにもかかわらず、「A」および「practitioner」は連結された記号または単語の別個のしみのままである。

#### 【0032】

図2によれば、この工程の付属物として空白行追加20は図16のプロビファイされた画像に対して一連の白色画素行を重ね、確実にテキストの行が隣接するテキスト行から分離されたままとなる（すなわち、フィルタをかけたテキスト行はまったく重複しない）。図16および図17によれば、囲み領域258および258'はアセンダーとデセンダーが組合わさり2つの単語の行間がマージされたことを示す。図16の領域258に図示されたテキスト行の重複は、プロビファイされたまたはフィルタをかけた画像に空白行を重ねることで確実に解消できる。その結果が図17に図示される空白行（白線）を重ねる演算は、関数DrawMiddleLines（ドロミドルライン）で実行される場合、図18に図示される工程によって実行される。通常、空白行WLは隣接する底線と頂上線対の間のほぼ中央の画像に追加され、これによって確実にプロビファイするクロステキスト行がなくなる。図17は空白行を図16のプロビファイされた画像に追加した結果を示す。

#### 【0033】

図18によれば、空白行追加ブロック20はステップ280で変数を初期設定することで

10

20

30

40

50

開始し、続いて第一テキスト行の底線情報から頂上線の位置を読み込む。頂上線情報は破棄され(ブロック282)、次の底線および頂上線の位置が記憶スタックまたはリストからポップされる(ブロック284および286)。画像に関しては、この底線-頂上線の対がそれぞれ隣接するテキスト行の底部および頂上部を表す。次に、ステップ288では前記対の中心に存在する点が配置され、画像の中心から外方向に引かれる空白行の支点を提供する。図2の歪み検出器14により確定される歪み角度を用いて空白行の終点がステップ290で計算される。空白行はステップ292でプロビファイされた画像の上に引かれるかまたは重ねられ、さらに試験ブロック294で制御されるようにすべてのテキスト行が効果的に分離され終わるまでこの工程は続く。

**【0034】**

図2によれば、プロビファイまたは中央フィルタリングの結果として、プロビファイ工程で形成されたそれぞれ連結された画素の集合を中心とするバウンディングボックスの位置を確定することができる。バウンディングボックスは重ねられた空白行の間のテキスト行内に存在するこれらの接続された構成要素または単語を中心としてのみ配置される。テキスト行の方向および画像座標系に対向するようにテキスト行に直交する方向に接続された画素の各群の先端点を識別することでバウンディングボックスはテキスト行の方向に配置される。この演算は関数 `Find Borders` (ファインドボーダー) によって実行される。通常、関数 `ファインドボーダー` は画像内のすべての画素を進み、接続された文字のバウンディングボックスを求め(ペイント構成要素)、ボックスの長さおよび幅だけでなく各ボックスの上方の左隅の座標を確定する。

**【0035】**

ファインドボーダー手続きを詳細に示す図19および20によれば、セグメント化工程22は、ステップ300でフィルタをかけた画像の周囲に完全に空白ボーダーを配置することにより開始される。これによって画像画素の配列の境界線を越えて流れることを避ける。次に、画素  $x$  および行カウンタ  $y$  はそれぞれ、ボーダー内の第一画素位置へ初期設定される。手続き `Read pixel` (リードピクセル) を呼び出し(ブロック304)、ブロック306で画素カラー(黒または白)が返されテストされる。画素が白であれば、もはや処理の必要はなく処理はブロック322へ続く。白でなければ、手続き `Paint Component` (ペイントコンポーネント) が呼び出され、まず待ち行列に黒画素の位置を格納する(ブロック308)。続いて、ブロック310および312ではそれぞれ、画像のコピーにおいて画素が白へ設定され、連結された画素または構成要素を囲むボックスの境界線が更新される。次に、隣接する黒画素を白へ設定し(ブロック314)、黒画素の位置が待ち行列の最後へ追加される(ブロック316)。ブロック318では、待ち行列ポインタがテストされ待ち行列が空かどうかを確定する。空でなければ、待ち行列の次の画素が取り出され(ブロック320)、処理はブロック312へ続く。逆に待ち行列が空であれば、連結された黒画素すべてが白へ設定され、ボックス境界線は連結された構成要素を包囲するボックスを表すこととなる。さらに、単語セグメントを包囲するボックスの境界線が確認されるとともに、テキスト行の歪みに対して方向づけられた直交座標系に調節することができる(ブロック322)。

**【0036】**

ループ処理はブロック324へ続き、画素カウンタ  $x$  を検査することで走査線の最後へ達したかどうかを確定する。達していなければ、ブロック304で処理を続ける前にブロック326でカウンタを増分する。走査線の終わりまで達していれば、画素カウンタ  $x$  がリセットされるとともに走査線カウンタ  $y$  はブロック328で増分される。さらに、ブロック330は走査線カウンタ  $y$  の値を検査し、画像全体が処理されたかどうかを確定する。処理されていれば、処理は終了である。そうでなければ処理は新たな走査線の第一画素を処理するためブロック304へ続く。

**【0037】**

したがって図21に示されるように、単語「`practitioner`」に対しては連結された文字画像の先端部分がバウンディングボックスを決める。いったんバウンディング

10

20

30

40

50

ボックスが決まると、この段階でこれからの考察の中から雑音マークを除去することができる。雑音マークが確定されるのは、1) バウンディングボックスコーナが画像画素の配列の外部にある場合、2) ボックスがこの配列における複数のテキスト行にわたる、またはボックスがテキスト行から完全にはみ出す場合、3) ボックスが縦横寸法のいずれかまたは両方において標準値より小さく、したがって破棄される場合、である。雑音マーク70および72等は考察される単語には含まれない。

#### 【0038】

図2によれば、単語形状コンピュータ24で、単語の画像または少なくともその一部分を表す信号は隣接する信号からは孤立し、その信号が導出される。導かれたこの信号は単語形状輪郭と呼ばれる。図22(A)に図示されるように、この関数はまず各バウンディングボックスの頂上に沿って移動し、各画素位置からはじめてボックスの頂上に沿って、黒画素またはボックスの底部のいずれかに到達するまでページ方向に対して下向きに走査する。ボックスの頂上と黒画素またはボックス底部との間の距離dの集合の記録が維持される。距離dの集合はボックスの長手方向にわたって蓄積され、単語形状の頂上の生輪郭を構成する。続いて、図22(B)に図示されるように、順にボックスの底部を走査して移動し、上方向に最初の黒画素またはバウンディングボックスの頂上を探査することにより、図22(A)に示される同一単語に対して底部の生輪郭が同様に生成される。図23は図3のテキストサンプルに対して作成された輪郭位置の画像である。重要なのは、ほとんどの部分でその輪郭のみで句の単語を認識することが比較的易しい図23の情報内容である。

#### 【0039】

図24によれば、この工程でも利用可能なフィルタをかけた画像に対して実際の画像上で作動するのが好ましいブロック100では、1またはそれ以上の基準線が各単語により作成される。一実施例では限定された厚さまたは厚さゼロであってもよい削除バーが単語によって構成され、これはエックスハイトのほぼ3分の2で上限または基準線を有し、さらにエックスハイトのほぼ3分の1で下限を有するのが好ましい。計算102では、上方または下方バウンディングボックスの間の距離dに対して、単語または基準線のより近い方へ組の計測が引き出される。この計算は画像の分析で行われる。図25(A)によればこの計算が図式的に示され、この基準線によってこの段階から最終的に引き出されることとなる信号が単語の長手方向にわたる各サンプリング位置で定義可能となることが理解できる。好ましい実施例では、この計算は実際にはすでに収集されたデータから生成され、さらに調節されることにより距離dを図示されるような上方または下方削除バーのいずれかで限定する。図示される実施例では、必要条件はないものの、バウンディングボックスの上方線から上方基準線に対して計測がなされている。したがって、例えば基準線から上方または下方バウンディングラインまたは文字に向けて計測してもよい。図25(B)は、計測の集合がブロック104から出力された信号を形成するためにどのように用いられるかさらに詳細に示す。輪郭は基準線に対して距離d'として表される。基準線に対して距離を計算することで、単語形状輪郭を通常のエックスハイトに対してスケールアップすることができ、これによって形状に関するあらゆる比較が容易となる。図25(C)および25(D)は、d'値の集合がグラフ状に示され、1次元信号を形成するというを示す。

#### 【0040】

英単語の外見によって引き出された情報の研究では、ほとんどの場合、単語の画像のほぼ頂上3分の1のみを見ることで単語が識別されているということが判明している。いいかえれば、単語の上部がその識別に必要な情報を大量にもっている。残りのケースの重要部分では、単語の画像の上部3分の1のみでは識別不能な単語は、識別労力が単語画像の下方3分の1により運ばれる情報を含む場合、識別不能となる。比較的小さなクラスの単語は識別を行う前に単語の中央3分の1に関する情報を必要とする。したがって、必要に応じて、単語形状の順序づけされた検査において、まず上方単語形状信号または輪郭を引き出し、第2に下方単語形状または輪郭を引き出し、第3に単語形状信号中央輪郭を(基

10

20

30

40

50

準線から単語またはバウンディングボックスへ向けて)引き出すという段階的工程が利用できることが理解できる。図25(A)、(B)、(C)、(D)のサンプルでは、単語「from」が上部のみから完全に識別可能である。図26(A)、(B)、(C)、(D)の例において、単語「red」は上部からでは完全に識別することはできない。というのは、「red」は単語「rod」や「rad」と混乱しやすいためである。文字「a」の下部があれば「red」と「rad」を区別することはできるが、文字「o」の下部が「rod」と「red」を区別できるかどうか疑わしい。しかしながら、「red」「rad」「rod」の中央部は完全に別個のものである。

【0041】

図2によれば、次に実行される段階は単語形状コンパレータ26での比較である。一実施例では、この比較は実際にはいくつかの小さな段階であり、そのそれぞれについて説明する。図27によれば、通常、既知の単語ともう一つの未知の文字列との2個の単語形状信号が比較され、それらが類似しているかどうか調べる。この場合、信号Rは単語「red」の上方輪郭であり、信号Fは単語「from」の上方輪郭である。実際には、文字フォントの通常の違いや再生方法、走査された画質が与えられても完全に同一であると期待できる信号は相対的に極めて少ない。しかしながら、比較される単語形状信号は互いにスケールすることができるため、同一のエクスハイトを有することができる。これは比較される単語形状輪郭の対のエクスハイトを確定することにより達成される。いったんエクスハイトが決まると、エクスハイトの比率は、輪郭の一方に適用される倍率を確定するために利用される。エクスハイトはフォントに対する特性計測であるので、水平方向および垂直方向の両方における倍率を確定するために利用できる。または、異なる長さのため異なって計測された部分に正規化および重量係数を強制せずに形状信号を比較することができる。さらに、信号の振幅または高さが正規化されることにより、単語形状比較に与えるフォントサイズの影響をさらに低減することができる。

【0042】

アセンダー/デセンドーの正規化演算を詳細に図示する図28(A)、(B)、(C)によれば、形状信号はそれぞれアセンダーハイトとデセンドーハイトとテキスト文字のエクスハイトとの共通関係に基づき正規化される。図示されるように、一見類似したフォントサイズまたは適当にスケールされたフォントサイズでプリントされた文字の実際のアセンダーハイトはわずかに異なる場合がある。これは小型字体または大型字体の結果として生じるもので、同一文字であっても24ポイントフォント等の同一サイズのフォントのハイト(高さ)には多様性があることを示す。図示されるように、図28(A)の距離d1は2個の文字「h」におけるアセンダーハイトの差を示す。同様に、距離d2は図28(B)の文字「f」の高さの差を示す。図28(C)に図示されるように、通常、文字はアセンダー部390、エクスハイト部392、デセンドー部394の3つのセクションに分けることができる。さらに、これらのセクションの相対的な高さをそれぞれa、b、cで示す。このエクスハイト上に存在する輪郭の領域は以下のようにスケールされる。

【数1】

$$f(t) = \frac{1.5}{a+c} \cdot f(t)$$

同様に、デセンドーは次の式によってスケールされる。

【数2】

10

20

30

40

$$f(t) = \frac{1.5}{a+b} \cdot f(t)$$

ただし、両者の場合、分子で用いられた値(1.5)はアセンダーまたはデセnderハイトとエクソハイトとの関係を計測することにより求められたものである。これにはテキスト列の部分を表さない輪郭部分を削除する演算も含まれる。これらの領域は図21で図示されたバウディングボックスの端部に存在する。例えば、図21で単語「practitioner」を囲むボックスは実際の単語画像を超えて伸びていることがわかる。図25(A)、(B)、(C)、(D)で単語「from」の端部でさらに図示されるように、輪郭は有用な情報をもっているわけではない。これらの領域を輪郭形状から除去することで、比較演算に生じるエラーを抑える。

#### 【0043】

正規化演算に続いて、標準信号処理段階を用いることで比較されている2つの信号の類似性または非類似性を確定することができる。または、次の式を用いることができる。

#### 【数3】

$$\Delta_{\text{string}} = \sqrt{\int_0^1 (f(x) - g'(x))^2 dx}$$

ただし、string(ストリング)は2つの信号の差であり、 $f(x)$ は既知信号であり、 $g'(x)$ は未知の信号である。

#### 【0044】

単純に確定する場合、差を調べ、もし差がゼロに近ければ2つの信号にはほとんど相違がないことを示すことになる。しかしながら、差が大きければ大きいほど、その単語が比較されている単語と別個のものであるという可能性が高くなる。

#### 【0045】

重要なのは、上述の実施例が各単語に対する上方および下方輪郭を用いて単語形状輪郭を互いに比較するという点である。これは特定の決定方法を実現したものであって、本発明を頂上輪郭と底部輪郭のみを用いた比較に限定するものではない。実際、上方輪郭内だけでも下方輪郭を比較する必要性を相当減少させ、これによって処理労力が相当軽減できるだけの十分な情報が含まれている場合がある。

#### 【0046】

この単純化された比較方法の段階は、最初に述べたように図29に図示される。まずステップ410では第一単語形状に対する輪郭がメモリから取り出され、さらに第二単語形状がステップ412で取り出される。次に、ステップ414で上方および下方輪郭により定義される単語形状の重心が確定され位置合わせされる。この段階の目的は単語輪郭の中心を位置合わせすることにより、比較中の2組の輪郭のあらゆる相対的変位に対して原因となりうる輪郭の差を低減することにある。前記重心を確定するには円弧の重心を確定する標準的な方法による。両組の輪郭対に対していったん確定すると、前記対の間の相対的な変位が決まり(ステップ416)、輪郭は輪郭同士の差を計算する前に変位する。輪郭の変位は、単語形状境界線の確立と図2のブロック24の単語形状の計算とに関連するあらゆるエラーを低減するために必要である。ステップ418は変位した輪郭対の重複範囲の外部に存在する領域を処理し、非重複領域におけるゼロ振幅信号に対する差を確定する。これは、輪郭の非重複端部での上方輪郭および下方輪郭の平方値を合計することで達成される。続いて、輪郭の重複領域が比較される(ステップ420)。この領域の差は、上方曲線と下方曲線との差の二乗の合計として確定される。さらにステップ418および42

10

20

30

40

50

0 から返された値を加算することで、変位した輪郭により定義される全範囲にわたる差の合計を確定する。この値は比較中の2個の単語形状に対する輪郭対の間の相似の相対表示として用いることができる。

【0047】

重心比較法 (center-of-gravity comparison method) に対する別の方法は、タイムワーピングとして公知の信号処理方法を利用する。これについては論文「分離された単語認識のための動的タイムワーブアルゴリズムにおける性能分担」(メーヤーズ、ラビナー、ローゼンバーグ共著、「音響・音声・信号処理に関する電気電子学会会報」、ASSP-28巻、6号、1980年12月刊)、および文献「タイムワーブ、記号列編集および高分子：列比較の理論と実際」(サンコフ、クラスカル共著、マサチューセッツ州リーディング、アジソン・ウエスレー出版社、1983年刊)の第1章から第4章までに開示されており、この文献では最良の照合ができるまで輪郭に沿って点を圧縮および拡大させるために利用することができる。さらに比較中の輪郭同士の差の量と輪郭を一致させるために必要な伸縮とに基づきスコアが引き出される。このスコアは比較中の2個の信号同士の照合の相対表示を提供する。

10

【0048】

動的ワーピング法の一般的な段階を示す図30によれば、この方法は、第一輪郭の各点とそれが比較されている輪郭の点との距離を記録するため、差の配列または行列の利用に依存している。図示されるように、この工程は比較法にも適用できるすべての計測に類似している。まず、さまざまな比較関数の演算を動的に制御するためにデータ構造が利用できるようにコードが編成される。構造DiffDescriptor(ディフディスクリプタ)は、比較を制御するために用いられる他の因数だけでなく、輪郭に対して適用される寸法を定義する変数を有する。これらの因数には比較前の輪郭長の正規化、上方輪郭および下方輪郭のための分離された比較、ワーピング経路を方向づけるcenterWeight(センターウェイト)因数、ワーブ経路を制限する帯域幅、底部輪郭比較に対して重みづけを重くまたは軽くした頂上輪郭比較を可能にするtopToBottom(トップ-ボトム)比、未知の輪郭が既知のまたはモデル単語形状輪郭と比較されている場合、輪郭差の重みづけを選択的に制御するhillToValley(ヒル-バレー)比がある。

20

【0049】

通常、各計測によって比較技術が実現されるが、それぞれの計測は特定の型の動的比較に対して最適化される。例えば、勾配限定動的ワーブ法では一方より大きな非単位的な中心重みおよびトップ-ボトム重みを有する。選択の第一レベルにより、比較、無条件ワーブ、単純なワーブされない比較に対して勾配条件付きワーピング関数を利用することが可能となる。それぞれのワーブ比較法には、頂上輪郭および底部輪郭が独立的にワーブされる分離比較関数と、頂上輪郭および底部輪郭の両者に対してワーブが同時に適用される平行比較関数とがある。

30

【0050】

通常の実施例では、動的ワーピング工程はまず、比較中に生成された距離値を保持するとともに一方の単語形状輪郭を他方に対してワーピングさせることを保持する経路/距離配列に対して空間を割り当てる(ステップ450)。通常、この配列は比較中の輪郭に対して可能性のある差計測をすべて識別するために用いられる。配列要素に差を充填した後に、すべての要素に対する差を走査経路内に蓄積する一方、前記配列は一コーナから斜めに対向するコーナへ向けて走査する。これまで説明した実施例は、差が確定した場合、蓄積された差を反復的に充填することによりこの工程を多少最適化する。本質的に、これによっていったん輪郭の間の相対差がすべて計算されてしまうと前記配列を走査する冗長な段階をなくすことができる。さらに変更すると最終的に大きな配列が除去され、配列の「走査」中に距離値を格納し、取り出し、更新するために操作される一対の線形配列で置換される。

40

【0051】

50

空間を割り当てた後、すべてのワーピング計測により用いられる工程が、輪郭の累加的な差を確定するため事前に配列内に格納されたデータを利用する反復工程である場合、配列のボーダー領域は初期設定されなければならない。ステップ452では、配列ボーダーが初期設定される。配列の第一行を初期設定するには、第一輪郭上の第一点と第二輪郭上の各点との差の二乗を確定する必要がある。ボーダーの初期設定に続き、列インデックス値11および行インデックス値12はそれぞれ1にリセットされ、輪郭に沿って個々のボーダーでない点処理し始める。

#### 【0052】

ステップ458乃至464まで、輪郭の処理が進む。第二輪郭に沿った各点と第一輪郭の点との間の距離における差が計算される。さらにこの差または距離が計算されると、すでに確定された差と合計する。さらに、すでに確定された差のいくつかはそれぞれ異なって重みづけすることができる。例えば、一実施例では配列斜線に沿った差の重みはセンターウェイト重みづけ因数により修正することができる。図のように、まず、上方輪郭および下方輪郭にわたって第一輪郭上の点と第二輪郭上の点との差の二乗の合計として距離が計算される。ただし、頂上輪郭差はトップ-ボトム変数によって重みづけされる。この距離は連続反復で水平差、垂直差、斜め差を確定するために用いられる。これらの値をそれぞれ確定するため、現在の距離値が下配列位置、左配列位置および左下配列位置における以前の値に加算される。この左下配列位置は既に説明したようなセンターウェイト因数により重みづけされる斜め位置である。配列位置502ですでに確定された値 $x$ と次の配列位置との位置関係を図示する図31(A)によれば、全体差を蓄積するためその後の位置の差に値 $x$ を加算することができる。配列位置504に対する差を計算する場合、位置502の値は下値として利用されることになる。同様に、位置506の値を計算する場合、位置502の値はセンターに重みづけされた左下または斜め値として利用される。3個の異なる値をステップ458、460、462で計算した後、ステップ464で3個の値のうち最小のものを選択し、ステップ466で現在の配列位置内へ挿入する。

#### 【0053】

続いて、図30に図示される工程では11で表される第一輪郭上の点と12で表される第二輪郭上の点との差を確定する。決定ステップ468は輪郭の端部または幅を調べることにより第二輪郭に沿った点の反復処理を制御する。限定幅または帯域幅の場合、比較中の輪郭領域のひとつまたは両方の領域が、輪郭が定義される領域の部分集合に制限される。限定が達成されていなければ、12の値がステップ470で増分された後に次の点の処理がステップ458で行われる。同様に、決定ステップ472が第一輪郭に沿った各点の処理を制御するとともにステップ474で増分する。すべての点が互いに処理され終わると、配列が走査され終わったことを示すステップ472の肯定応答により明らかのように、最も斜めの位置である配列(11, 12)において相対差のスコアが最上のスコアとなる。続いて、ステップ476で確定された値は比較中の輪郭同士の動的にワーブされた差を示すものとして返される。

#### 【0054】

コード化の実現により、大きな2次元配列を必要に応じて更新される一対の線形配列に減少させることで上述のワーピング工程の実行を最適化することができた。この修正により、ワーブ比較値に対する最小差または最良のスコアが1次元配列の最終位置で求められる。さらに、最終差を正規化することにより、比較中の2組の輪郭の間に長さの差を生じさせることができる。最後に、このような値を閾または同様にして得られた差の集合と比較し、単語が一致したといえるほど輪郭同士が近接しているかどうか確定することができる。または一連の単語形状比較から最良の一致を確定することができる。

#### 【0055】

他の実施例では、上述の動的タイムワーピング工程を変形し、差配列に含まれる差を周期基底上の閾値と比較することができる。比較において、比較中の輪郭が互いに一致しないと確定できるだけの十分な差が存在した場合にはこの工程を中断することができ、これによって貴重な処理時間が節約できる。さらに、単語形状コンパレータ26の連続的な演算

10

20

30

40

50

は単語形状コンピュータ24からの連続的な出力とともに実行され、これによってキーボードを探している場合、テキスト画像の平行処理が可能となる。

【0056】

これまで動的ワーピング比較計測に関する基本的な実現方法について説明してきたが、本発明の数多くの可能な実施例を示すため、他の動的ワーピング比較法と上述の制御因数の適用との違いを簡単に説明する。まず、すでに述べた動的ワーピング法も、配列を走査する場合、条件づけされているワーピング経路の勾配で実現することができる。この計測法は図31(B)でさらに詳細に図示される。ただし、配列位置512の値Xは図示される3個の連続する配列位置にのみ加算することができる。例えば、位置514に対してd211として考えられる場合、Xは配列位置514へ加算できる。図中の変数に対する命名法は次の通りである。d211は下に2行、左に1列の配列位置を示す。d111は左斜め下方の配列位置を、d112は現在の配列位置から1列下の左へ2行進んだ配列位置を示す。同様にして、配列位置516の累加的な差を計算するためにXをd112値として加算することができる。

10

【0057】

図31(A)および31(B)の比較から明らかなように、勾配条件付きワーピング計測は、累加的な差の生成中に利用できるワーピング経路を局所的に限定する。このような条件づけを行うのは、ワーピング工程が比較中の2つの輪郭の一方の大きな領域を除去または圧縮することを、このような圧縮に対してあまり「コスト」をかけずに避けるためである。

20

【0058】

平行ワーピング処理に対してすでに説明したこの方法も、一度に一对のみ2個の単語形状の上方輪郭等の輪郭について実現することができる。この関数は、上述の非勾配条件付きの方法および勾配条件付きの方法の両方で分離照合計測を実現する。通常、これらの計測では一对の単語形状の頂上輪郭または底部輪郭同士の差を別個に計算する。この計測に対して示される通常の実現から、これらの計測が通常は順に利用され、まず頂上輪郭に対してワーピングされた差を確定し、それに底部輪郭比較からワーピングされた差を加算し、その結果、単語形状の総計差が得られることがわかる。

【0059】

「それぞれの部分に関して」段階化されたやり方で記述されたこの比較方法を実行することにより、さらに処理上の利点が引き出される。詳細にいえば、段階化された比較にはまず単語を識別するかまたは少なくとも可能性のある他の集合を狭めるために比較中の単語の上方輪郭を利用する必要がある。第二に完全な識別を提供するため下方輪郭比較を用いる必要がある。単語形状比較演算26に対するこのような手法が、図2に図示されるような既知の単語形状の辞書28と比較することによって未知の単語形状を識別するために要する処理時間を相当に低減することとなる。段階化された比較法の重要な点は、輪郭に作用する頂上ワーピングおよび底部ワーピングが相対的に同値でなければならないという条件である。この必要条件は、上方曲線および下方曲線が共通の単語に対して関係をもっており、この関係がワーピング解析中に維持されなければ、比較の精度が落ちるという事実から生じる。

30

【0060】

または、最良の一致を達成する上方曲線および下方曲線に作用する相対的ワーピングを蓄積するのに適した関数を加えて動的ワーピング技術を上記のように適用することができる。例えば、既知の強調されていない単語形状を未知の単語形状と比較した場合、下方曲線に対して上方曲線へ作用するワーピングにおける変位は強調単語を示すことができる。しかしながら、ワーピングされた領域の長さは頂上ワーピングおよび底部ワーピングに対する長さと同じままである。このような技術は、大き目の字体のテキスト内の重要な単語が時折強調されている場合の識別に有用である。

40

【0061】

bandwidth(帯域幅)因数はすでに説明した制御因数の一つである。実現された場合、この帯域幅因数は、ワーピング信号がその内部で条件づけされている信号帯域の相

50

対幅を制御する。詳細に説明すれば、配列を走査するワーブ経路が条件づけされた配列対角線を中心とする領域を定義することにより帯域幅の限定が実現できる。大きな値を帯域幅の外部であるこれらの領域へ割り当て、前記経路が条件を超えることがほとんどないようにすることによってこの条件が実現される。

#### 【0062】

トップ - ボトム比は、すでに簡単に説明した他の因数である。これが適用されると、この変数は頂上輪郭ワーピング工程のために確定された差を重みづけするために用いられる。したがって、2以上の数字を用いると、上方輪郭差を下方輪郭差よりも重く重みづけすることになる。非常に大きな数字を用いれば下方輪郭差を効果的かつ完全に除去することとなる。さらに、値がゼロであれば上方輪郭差を完全に除去することとなる。上方輪郭はたいてい下方輪郭よりも単語に関する情報を多くっており、この因数は通常、上方輪郭をその情報内容に比例して重みづけできることから重要であると考えられている。

10

#### 【0063】

ヒル - バレ - 比は、単語形状輪郭の既知のまたはモデル集合が未知の画像から単語形状輪郭の集合と比較中である状況で通常適用される変数である。このオプションを利用する場合、輪郭のモデル集合は比較計測関数として通用する。輪郭上の点の差を確定する場合、比較関数は通常、関数スクエアディフレンスと呼び出し、二乗された差の合計を確定する。スクエアディフレンスは、モデル輪郭の値が比較中の輪郭よりも小さいと確定したときは必ず、ヒル - バレ - 比を二乗された差に適用する。1を上回るヒル - バレ - 値を適用した結果、モデル輪郭が目的輪郭よりも小さい場合の差の相対的な「コスト」は、モデル輪郭が目的輪郭より大きい場合の同一差よりも小さくなる。この型の重みづけの基本は、モデル輪郭と比較する場合、モデル輪郭未満の輪郭位置により明らかであるように、この比較が、走査中または同様のデジタル化演算の間に充填されそうにない領域よりも軽い重みで「充填」される必要がある目的輪郭の領域を扱うということである。例えば、アセンダーが文字体とぶつかる領域は走査中に充填されやすく、これによって目的輪郭がこれらの領域の段階的な輪郭となる。しかし、モデル輪郭はこれらの領域で定義されたピークまたは谷を有する確率が最も高い。このため、たとえ文字が同一であったとしてもモデルの輪郭値は目的の輪郭値よりも小さくなる。したがって、ヒル - バレー変数はこれらの領域にわたって計算された差に対する影響を最小化しようとする。

20

#### 【0064】

重要なのは、上述の計測および制御因数により前記比較計測が多数の順列で行うことができる点である。しかしながら、これらの計測がもつ柔軟性は比較工程の適応性を向上させようとするため、コンピュータ生成文字フォント列から生成されたモデル輪郭のように情報が特定の単語形状輪郭について既知であった場合、前記計測において比較をより強固にするその情報に依存することができる。

30

#### 【0065】

好ましい実施例にしたがって本発明を説明してきたが、ソフトウェア構成はコンピュータシステムで実現可能となるように設計され、デジタル信号データの処理に関して上述の演算を達成するため、コンピュータシステムにはあらかじめ定義された命令を実行する一またはそれ以上のマイクロプロセッサまたは演算処理装置が用いられている。さらに本発明は、ここで開示された演算を実現するよう設計された特定のハードウェアを利用することで達成される。さらに本発明は、テキスト画像に関して説明した。しかしながら、本発明は非テキスト画像部分を有する画像であっても適用可能である。また、図面とともに本明細書を読み理解する中で修正点が生じることは明らかである。この実施例は一例であって、請求項で請求される内容から当業者によりさまざまな代替案、修正、変更、改良がなされうるものとする。

40

#### 【図面の簡単な説明】

【図1】本発明で使用される画像処理システムの概要システム図を示す。

【図2】新規な単語形状認識システムの一実施例を構成するシステム構成要素の配列のブロックシステム図を示す。

50

【図3】新規の工程が実行される例文テキストの画像サンプルを示す。

【図4】例文テキストの走査画像のコピーである。

【図5】処理を進める前に画像サンプルにおいて例文テキストを方向づける角度を確定するために用いられる工程を画像で図示したものである。

【図6】処理を進める前に画像サンプルにおいて例文テキストを方向づける角度を確定するために用いられる工程を画像で図示したものである。

【図7】処理を進める前に画像サンプルにおいて例文テキストを方向づける角度を確定するために用いられる工程を画像で図示したものである。

【図8】処理を進める前に画像サンプルにおいて例文テキストを方向づける角度を確定するために用いられる、例文テキストから得られたレスポンスのグラフを示す。

10

【図9】画像内のテキストの底線を確定するため例文テキストのサンプル画像を調べるグラフの導出および利用法を示す。

【図10】画像内のテキストの底線を確定するため例文テキストのサンプル画像を調べるグラフの導出および利用法を示すものであって、図9の部分的拡大グラフのみを示す。

【図11】図9に示される底線を確定するために実行される手順を示すフローチャートである。

【図12】図9に示される底線を確定するために実行される手順を示すフローチャートである。

【図13】図9乃至10に図示されるデータから導出した後、画像上に底線が示された例文テキストの走査画像を示す。

20

【図14】図3の画像に対して中央フィルタを適用した場合に用いられる各段階を示すフローチャートである。

【図15】前記中央フィルタの適用を示す図3の画像の部分的な拡大画像表示である。

【図16】ここではプロピファイニングとしてよばれる工程で文字列を連結画素の単一集合として表示する傾向のある例文テキストの走査画像に対して中央フィルタを適用した後の結果画像を示す。

【図17】白色画素の行が不鮮明な画像へ加えられることにより文字列の隣接する行から文字列の行を鮮明に描写することができる前記工程における連続段階を示す。

【図18】図17の白線を加えるために必要な段階を示すフローチャートである。

【図19】図16の不鮮明画像による画像データをセグメント化するために行われる手順を示すフローチャートである。

30

【図20】図16の不鮮明画像による画像データをセグメント化するために行われる手順を示すフローチャートである。

【図21】バウディングボックスが各文字列を含む画像画素の部分集合をただ一つ識別するように各単語群の周囲に配置された例文テキストを示す。

【図22】(A)及び(B)は例文テキストのサンプル画像中に現れる例文単語「from」を用いて単一独立変数信号の導出を図示する。

【図23】図22に図示された導出工程により形成される輪郭の結果を図示する。

【図24】単語形状信号の導出に関連する各段階を示す。

【図25】例文単語「from」を用いて単一独立変数信号の導出を示す。

40

【図26】例文テキストのサンプル画像中に現れない例文単語「red」を用いて単一独立変数信号の導出を示す。

【図27】信号正規化法を用いて単語「red」および「from」に対して導出された信号の単純な比較を示す。

【図28】フォント高さにおける矛盾を詳細に図示し、さらにこのような矛盾の正規化方法を図示したものである。

【図29】単語形状輪郭の間の相対差を確定する一方法に対して用いられる段階を詳細に示すフローチャートである。

【図30】単語形状輪郭の間の相対差を確定する第二方法の段階を詳細に示すフローチャートである。

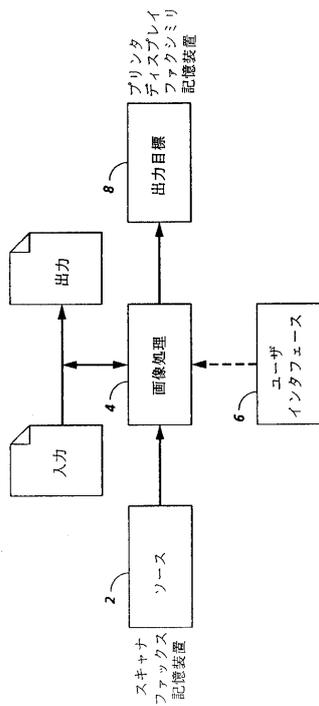
50

【図3 1】非勾配条件付き比較および勾配条件付き比較の両方に対して計算されるとともに一配列に格納される相対差値間の関係図である。

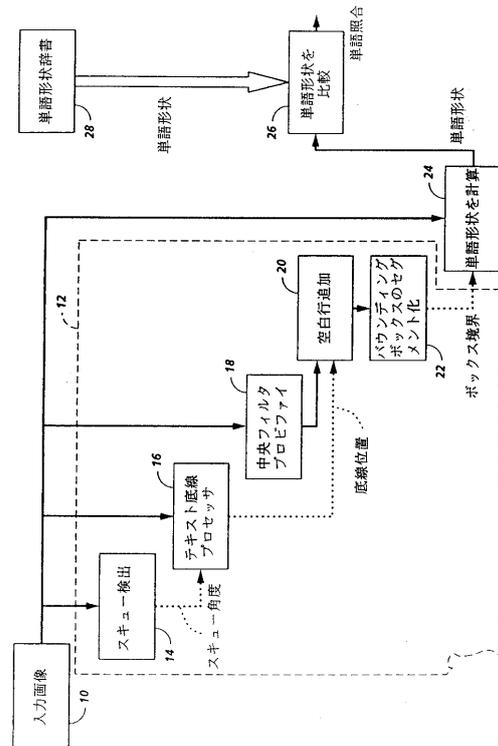
【符号の説明】

- 2 ソース画像導出システム
- 4 コンピュータ処理装置
- 6 ユーザインタフェース
- 8 出力装置
- 10 入力画像
- 14 スキュー検出器
- 16 テキスト底線プロセッサ
- 18 中央フィルタ
- 24 単語形状信号コンピュータ
- 26 単語形状コンパレータ

【図 1】



【図 2】



【 3 】

A practitioner may not solicit professional employment from a prospective client with whom the practitioner has no family or prior professional relationship, by mail, in-person or otherwise, when a significant motive for the practitioner's doing so is the practitioner's pecuniary gain under circumstances evidencing undue influence, in-

【 4 】

A practitioner may not solicit professional employment from a prospective client with whom the practitioner has no family or prior professional relationship, by mail, in-person or otherwise, when a significant motive for the practitioner's doing so is the practitioner's pecuniary gain under circumstances evidencing undue influence, in-

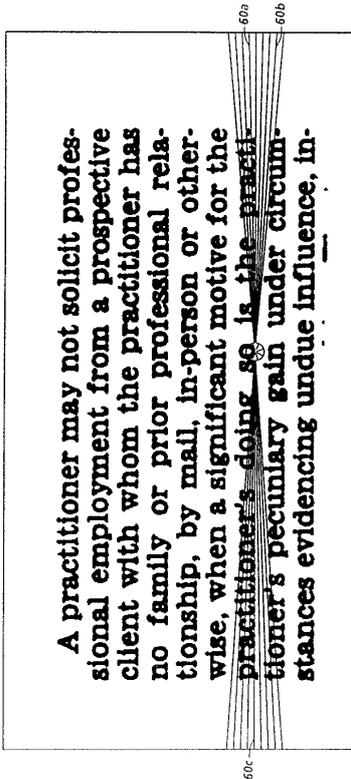
【 5 】

A practitioner may not solicit professional employment from a prospective client with whom the practitioner has no family or prior professional relationship, by mail, in-person or otherwise, when a significant motive for the practitioner's doing so is the practitioner's pecuniary gain under circumstances evidencing undue influence, in-

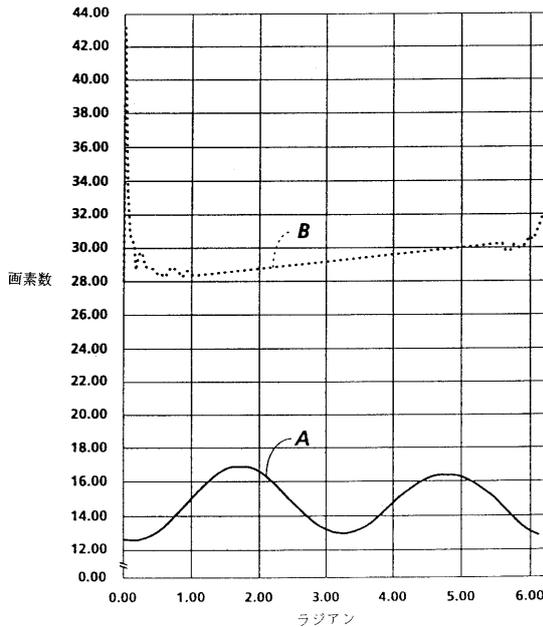
【 6 】

A practitioner may not solicit professional employment from a prospective client with whom the practitioner has no family or prior professional relationship, by mail, in-person or otherwise, when a significant motive for the practitioner's doing so is the practitioner's pecuniary gain under circumstances evidencing undue influence, in-

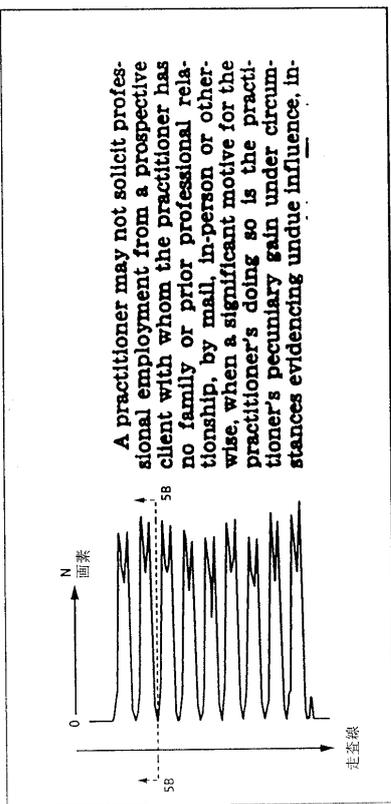
【 図 7 】



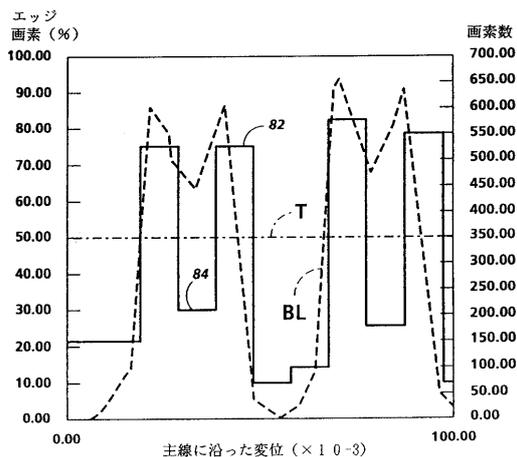
【 図 8 】



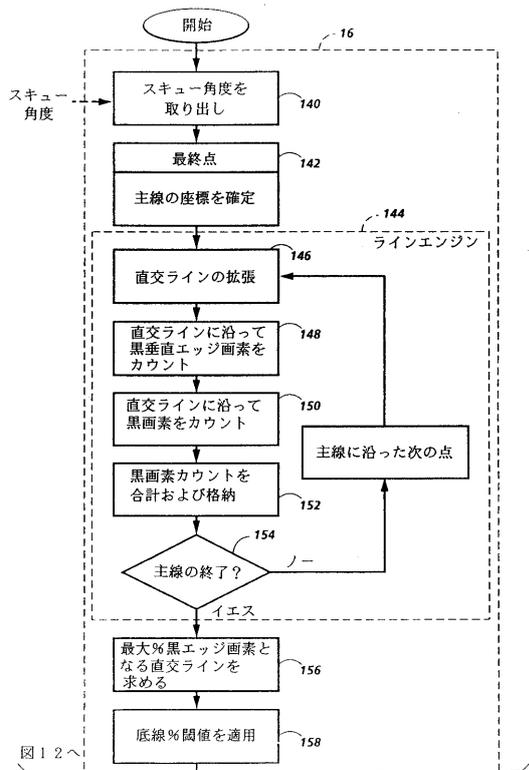
【 図 9 】



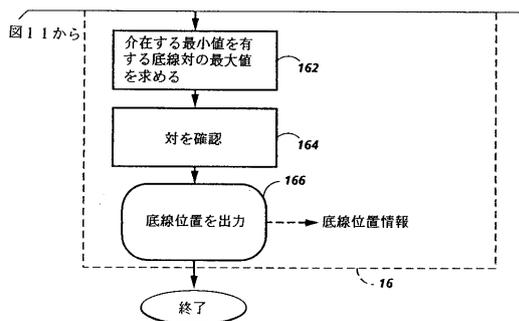
【 図 10 】



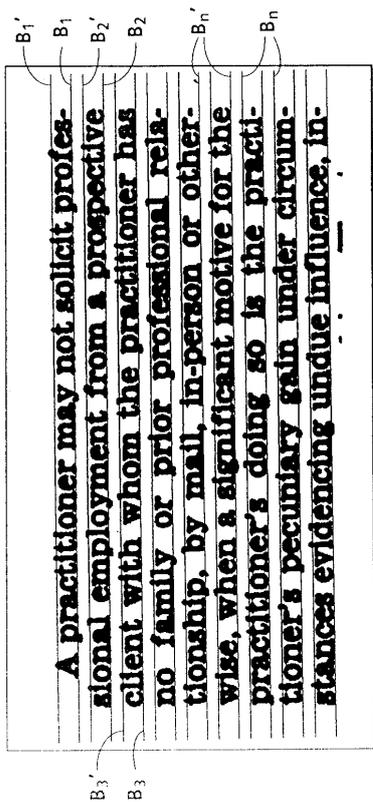
【 図 1 1 】



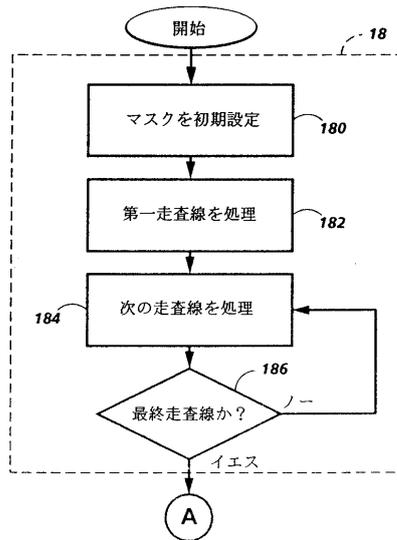
【 図 1 2 】



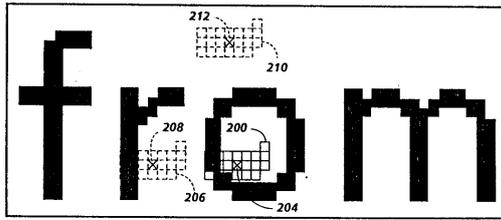
【 図 1 3 】



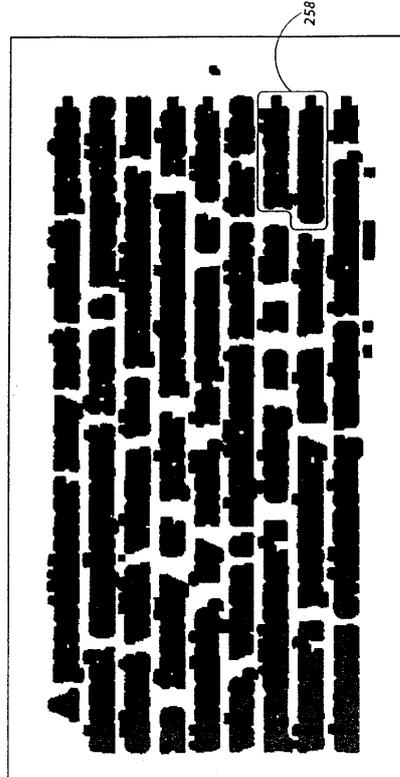
【 図 1 4 】



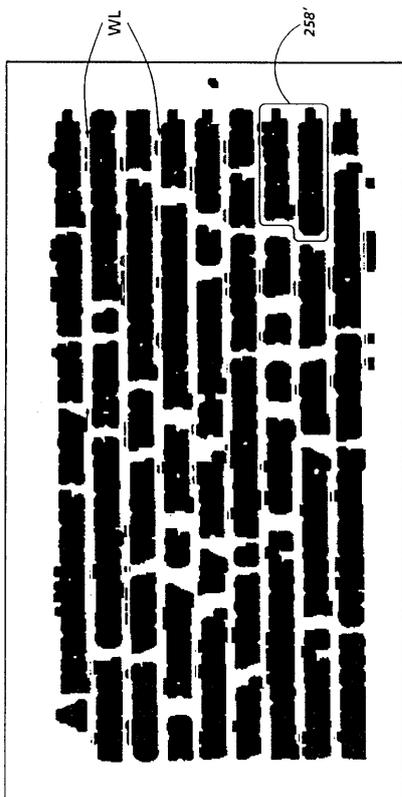
【 図 15 】



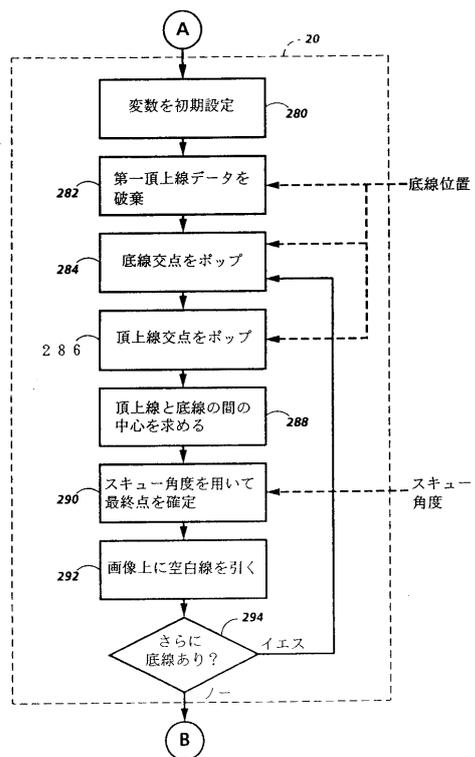
【 図 16 】



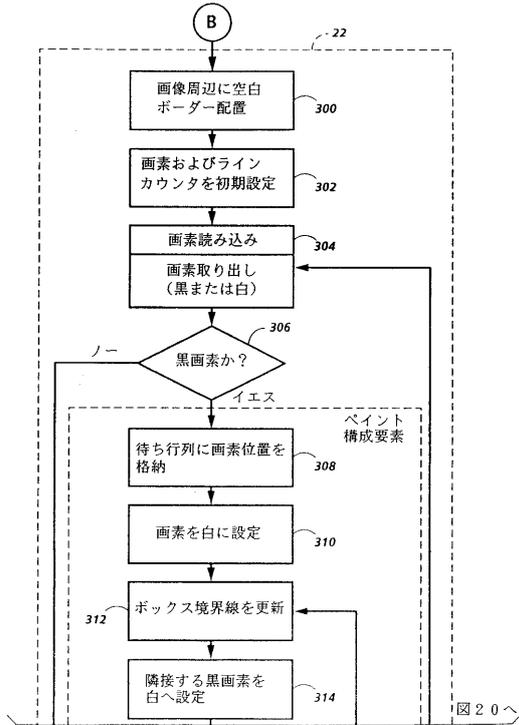
【 図 17 】



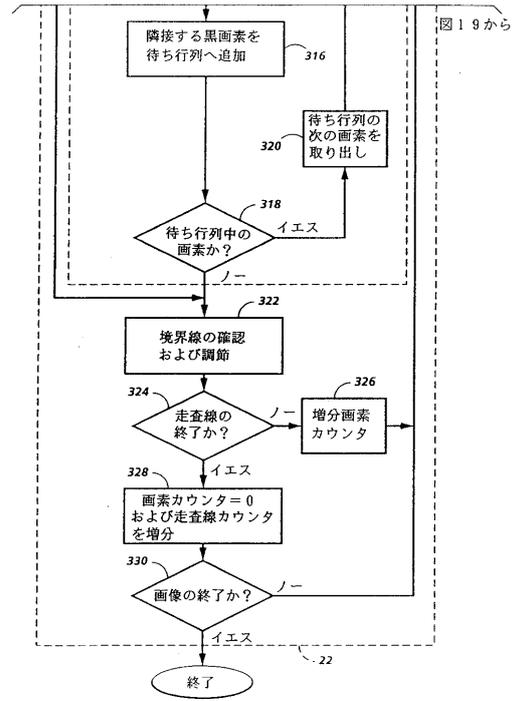
【 図 18 】



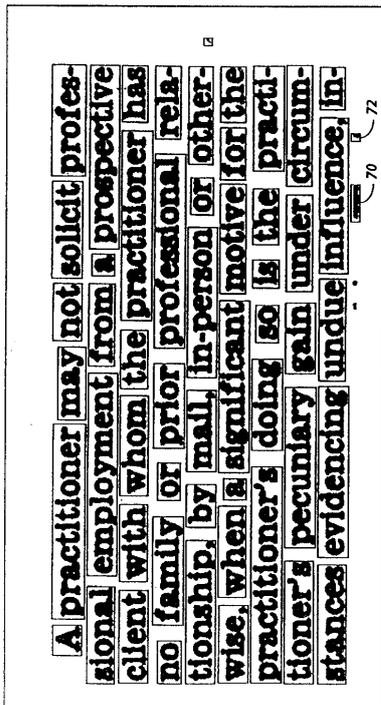
【 図 1 9 】



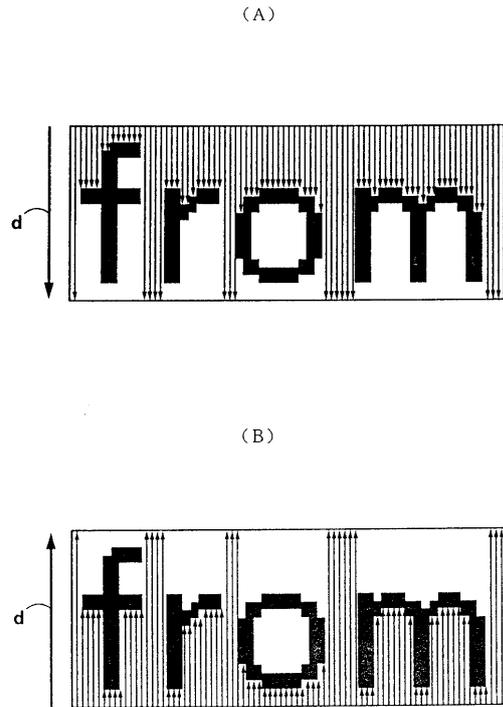
【 図 2 0 】



【 図 2 1 】



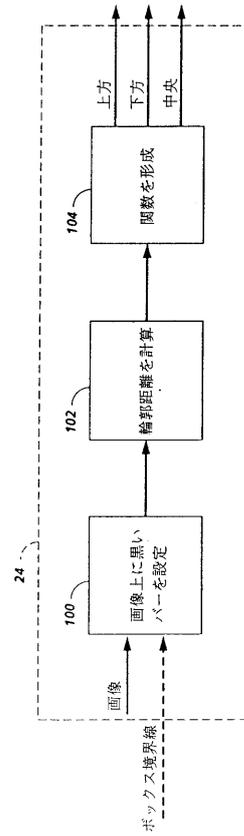
【 図 2 2 】



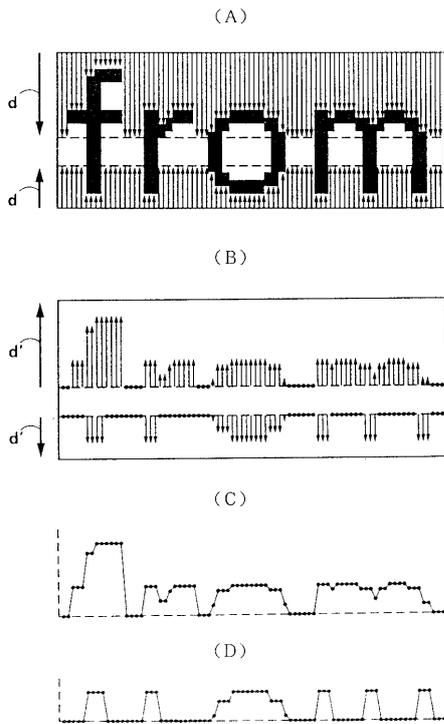
【 図 2 3 】

A practitioner may not solicit profes-  
sional employment from a prospective  
client with whom the practitioner has  
no family or prior professional rela-  
tionship, by mail, in person or other-  
wise, when a significant motive for the  
practitioner's going so is the practi-  
tioner's pecuniary gain under circum-  
stances of having undue influence, in-

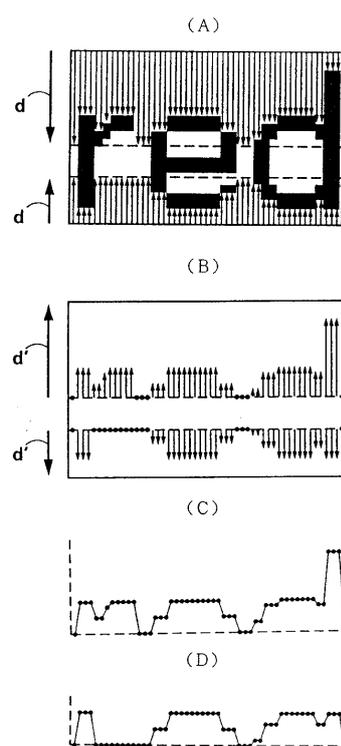
【 図 2 4 】



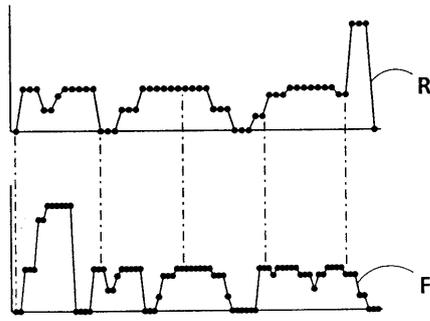
【 図 2 5 】



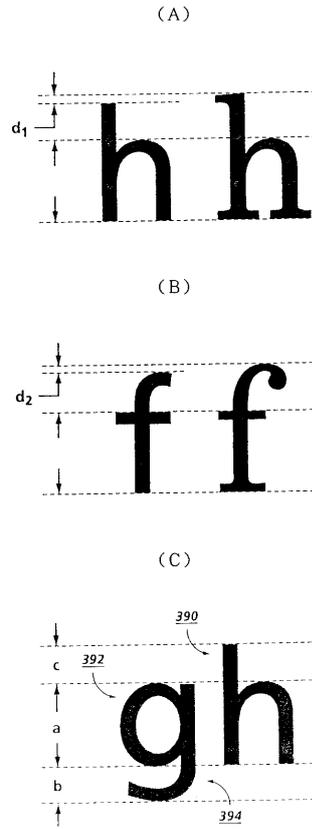
【 図 2 6 】



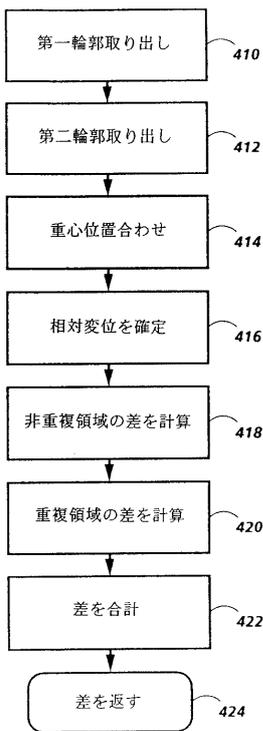
【図 27】



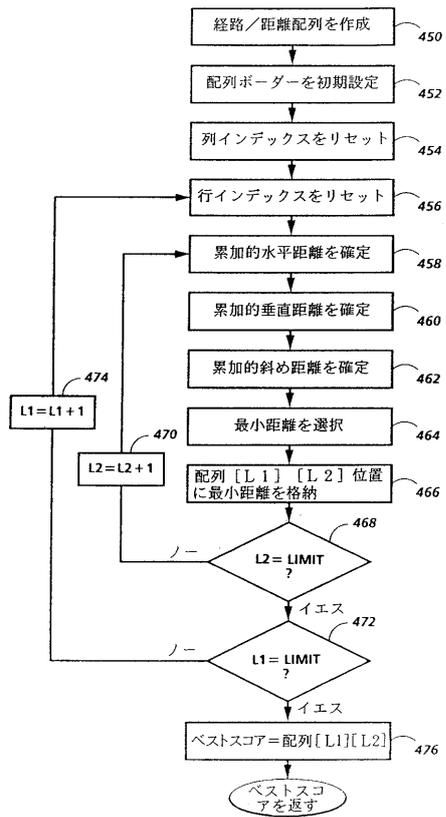
【図 28】



【図 29】

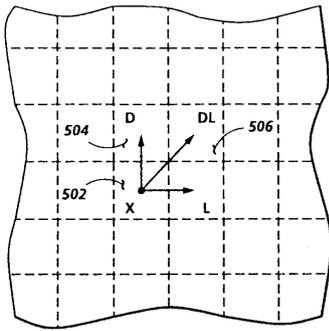


【図 30】

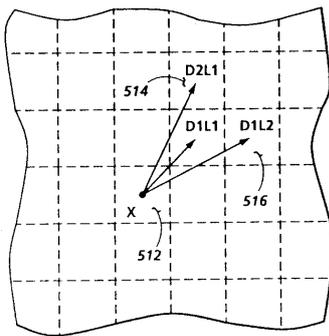


【 図 3 1 】

(A)



(B)



---

フロントページの続き

(72)発明者 マイケル ジェイ . ホブクロフト  
アメリカ合衆国 14850 ニューヨーク州 イシカ ロウエル プレイス 10

合議体

審判長 小川 謙

審判官 井上 信一

審判官 深沢 正志

(56)参考文献 特開平1 - 265378 (JP, A)  
特開昭64 - 36387 (JP, A)  
特開昭58 - 201184 (JP, A)  
特開昭62 - 127988 (JP, A)