

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

**特許第3754716号  
(P3754716)**

(45) 発行日 平成18年3月15日(2006.3.15)

(24) 登録日 平成17年12月22日(2005.12.22)

(51) Int. Cl.		F I			
<b>G06K</b>	<b>9/20</b>	<b>(2006.01)</b>	<b>G06K</b>	<b>9/20</b>	<b>340L</b>
<b>G06K</b>	<b>9/32</b>	<b>(2006.01)</b>	<b>G06K</b>	<b>9/32</b>	

請求項の数 14 (全 34 頁)

(21) 出願番号	特願平6-320955	(73) 特許権者	000001007
(22) 出願日	平成6年12月22日(1994.12.22)		キヤノン株式会社
(65) 公開番号	特開平7-234918		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成7年9月5日(1995.9.5)	(74) 代理人	100076428
審査請求日	平成13年12月25日(2001.12.25)		弁理士 大塚 康德
(31) 優先権主張番号	08/171 720	(74) 代理人	100112508
(32) 優先日	平成5年12月22日(1993.12.22)		弁理士 高柳 司郎
(33) 優先権主張国	米国(US)	(74) 代理人	100115071
			弁理士 大塚 康弘
前置審査		(74) 代理人	100116894
			弁理士 木村 秀二
		(72) 発明者	シン・ヤン ワング
			アメリカ合衆国 カリフォルニア州 92
			680, タスティン, リチェン レーン
			2251
			最終頁に続く

(54) 【発明の名称】 文書のテキスト及び／または非テキストブロックの選択方法及び装置

(57) 【特許請求の範囲】

【請求項1】

画像データにおける黒画素の連続成分を検出する連続成分検出工程と、  
前記検出された連続成分を、サイズ及び他の連続成分との相対的位置に基づいて、テキスト又は非テキストに分類する分類工程と、

少なくとも一部が前記分類工程で分類された非テキストの連続成分のエッジに沿っており且つ当該非テキストの連続成分に隣接する領域へ延びる不可視の直線をサーチするサーチ工程と、

前記分類工程で分類されたテキストの連続成分の相対的位置に基づいて、テキストブロックを形成するテキストブロック形成工程と、

前記テキストブロック形成工程で形成されたテキストブロックを、前記非テキストの連続成分の可視線または前記サーチ工程でサーチされた不可視の直線が横切っている場合、当該テキストブロックを当該可視線または不可視の直線で分割する分割工程と、

前記分割工程で分割されたテキストブロックに対して、所定の後処理を実行する後処理工程とを備えることを特徴とするブロック選択方法。

【請求項2】

前記画像データを入力する入力工程を更に備え、該入力工程は、前記画像データが2値画像データではない場合に、当該画像データを2値画像データに変換してから入力することを特徴とする請求項1に記載のブロック選択方法。

【請求項3】

10

20

前記連続成分検出工程で検出された連続成分の階層ツリーを、各連続成分の位置に基づいて形成するツリー形成工程を更に備えることを特徴とする請求項 1 に記載のブロック選択方法。

【請求項 4】

前記後処理工程では、前記テキストブロックにおける文字画像を文字認識することを特徴とする請求項 1 に記載のブロック選択方法。

【請求項 5】

前記後処理工程では、前記テキストブロックをデータ圧縮することを特徴とする請求項 1 に記載のブロック選択方法。

【請求項 6】

前記分類工程で分類された非テキストの連続成分について、少なくとも可視のラインであるか否かに分類する非テキスト連続成分分類工程を更に備えることを特徴とする請求項 1 に記載のブロック選択方法。

【請求項 7】

前記非テキスト連続成分分類工程では、ライン、中間調画像、線画、フレーム、未知のいずれかに分類することを特徴とする請求項 6 に記載のブロック選択方法。

【請求項 8】

画像データにおける黒画素の連続成分を検出する連続成分検出手段と、  
前記検出された連続成分を、サイズ及び他の連続成分との相対的位置に基づいて、テキスト又は非テキストに分類する分類手段と、  
少なくとも一部が前記分類手段で分類された非テキストの連続成分のエッジに沿っており且つ当該非テキストの連続成分に隣接する領域へ延びる不可視の直線をサーチするサーチ手段と、

前記分類手段で分類されたテキストの連続成分の相対的位置に基づいて、テキストブロックを形成するテキストブロック形成手段と、

前記テキストブロック形成手段で形成されたテキストブロックを、前記非テキストの連続成分の可視線または前記サーチ手段でサーチされた不可視の直線が横切っている場合、当該テキストブロックを当該可視線または不可視の直線で分割する分割手段と、

前記分割手段で分割されたテキストブロックに対して、所定の後処理を実行する後処理手段とを備えることを特徴とするブロック選択装置。

【請求項 9】

前記画像データを入力する入力手段を更に備え、該入力手段は、前記画像データが 2 値画像データではない場合に、当該画像データを 2 値画像データに変換してから入力することを特徴とする請求項 8 に記載のブロック選択装置。

【請求項 10】

前記連続成分検出手段で検出された連続成分の階層ツリーを、各連続成分の位置に基づいて形成するツリー形成手段を更に備えることを特徴とする請求項 8 に記載のブロック選択装置。

【請求項 11】

前記後処理手段では、前記テキストブロックにおける文字画像を文字認識することを特徴とする請求項 8 に記載のブロック選択装置。

【請求項 12】

前記後処理手段では、前記テキストブロックをデータ圧縮することを特徴とする請求項 8 に記載のブロック選択装置。

【請求項 13】

前記分類手段で分類された非テキストの連続成分について、少なくとも可視のラインであるか否かに分類する非テキスト連続成分分類手段を更に備えることを特徴とする請求項 8 に記載のブロック選択装置。

【請求項 14】

前記非テキスト連続成分分類手段では、ライン、中間調画像、線画、フレーム、未知の

10

20

30

40

50

いずれかに分類することを特徴とする請求項 1 3 に記載のブロック選択装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】

本発明は、例えば文字認識に先立って行なわれる文字ブロックの処理のための方法及び装置に関する。さらに詳しくは、認識に先立ち、画像データのブロックが画像データの文字に基づいて分類され選択される、文字ブロック処理方法及び装置に関するものである。例えば、画像データがテキスト画像データ（水平及び／または垂直）か、それとも中間調（またはグレースケール）画像、線画、表、垂直または水平線、フレーム等の非テキスト画像データのいずれであるかに基づいて、画像データブロックが選択されて分類される。

10

【0002】

【従来の技術】

近年、テキストデータの画像を分析して、テキストデータ内の個々の文字を認識したり、認識した文字に対応する、コンピュータが読み取れる文字コードファイルを作成することが可能になった。そうしたファイルは、ワードプロセッシングや、データ圧縮、あるいはデータ処理プログラムで操作することができる。そうしたシステムは、以下「文字認識システム」と称されるが、リタイプもしくはテキストデータの再入力が必要なくなるため有益である。例えば、ファクシミリ送信されたり複写機やマイクロフィルムで再現された文書に対して文字認識を行なって、その文書の文字や数字の文字コード（例えばASCIIコード）を含むコンピュータファイルを作成し、リタイプや文書の再入力を必要とせず

20

【0003】

文字認識される文書は、しばしば多くの異なるタイプの画像データを含んでおり、それらの全てが認識できるわけではない。例えば、現在、テキスト画像データの認識が可能だとすると、一方で、非テキスト画像データの認識は非常に困難である。通常、文字認識対象の文書はテキスト画像データブロックと、中間調画像、線画、線等の非テキスト画像データブロックとを含んでいる。さらに、文書は表や、枠組があったり、または枠組みは無いがテーブル状に配置されたデータも含んでいる場合もある。従って、文字認識処理の前に、文書内の個々のブロックをブロックの画像データのタイプに基づいて分類し、テキスト

30

【0004】

図1は代表的な文書の1ページを示している。図1において、文書ページ101は2欄形式に構成されている。ページには、タイトルに適切な大きいサイズのフォントのテキスト情報を含むタイトルブロック102と、テキストデータのラインを含むテキストブロック104と、非テキストのグラフィック画像を含むグラフィックブロック105と、表のテキストまたは数字情報を含む表ブロック106と、グラフィックまたは表情報のブロックの説明文（caption）である小サイズのテキストデータを含むキャプションブロック107とが含まれている。各情報のブロックはそこに含まれる情報のタイプに基づいて分類され、その分類に応じて分割される。

40

【0005】

【発明が解決しようとする課題】

予め、画像データのテキストタイプのブロックを検出するために、画像データにおける黒画素を水平及び垂直に展開して、水平及び垂直に隣接する単数または複数の白画素に展開することで画素画像データをにじませる（スミアー技術）ことが考えられた。こうしたスミアー技術は、スミアーパラメータを適切に選択することができる様に、テキストタイプ画像データの文字の予備知識（例えばフォントサイズ）に依存しているため、不十分である。さらに、スミアーパラメータの小さな変化で選択結果に大きな変化が生じる。スミアー技術はまた、必ず原稿文書の内部構造を保持できるわけではない。例えば、スミアーにより2欄形式の原稿が1欄に変わってしまうことがある。このような状況は、テキストデ

50

ータが格納されている順序を狂わせ、原稿テキストの正確な再構成を不可能にしてしまうことから、不都合である。さらに、スミア技術では、時折テキストタイプデータを非テキストタイプデータ中ににじませ、全体の領域がテキストタイプデータとして誤認識されることが知られている。

#### 【0006】

同出願人による、1992年4月24日出願の米国特許出願第07/873、012号では、格納された文書内の文字ブロックを選択するための他の方法が提案されている。ここでは、格納された文書に対し、最初に所謂「結合成分 (connected components)」のサーチが行なわれる。これは、各画素の周囲の8方向のいずれかに存在する、2つ以上の画素が結合されたものである。次に、テキスト結合成分は非テキスト結合成分から分離されて、非テキスト結合成分は、例えば、表、中間調画像、線画等として分類される。次に、文書の何らかの歪みの方向が検出され、もし歪みが垂直ならば画像は90度回転され、結合成分のサーチが再び行なわれる。歪みの補正の後、テキストの適切なブロック、例えば欄が識別される様に、非テキスト成分のエッジに添った目に見えない白ラインがサーチされる。その後、水平テキスト列とタイトル列が形成され、水平テキスト列はグループ化されて四角形に配置されたテキストブロックになる。そして、後処理が行なわれて、更なる文字認識処理のために、識別されたテキストブロックが準備される。07/873、012号は参照により本願に併合されている。

#### 【0007】

上記のブロック選択技術は水平文書 (例えば英語文書) に適切であるが、1ページが水平及び垂直テキストブロック (2方向) を含んでいる場合にも適用可能である。例えば、日本語の文書は垂直な漢字文字と表や数字記号などの水平な文字との組み合わせを含んでいる。また、ある種の英語文書は特定の情報を強調したり、所望の効果を提供するために、垂直に配置された文字を含んでいる。

#### 【0008】

さらに、スキャンされたページ自体はしばしば歪んでおり、上記のブロック選択技術では、最初に歪みを識別して、テキストブロックを形成する前に画像を回転させることでこの問題に対処している。この技術を実用化するには、処理速度と正確さにおいて2つの実質的な問題となる。さらに詳しくは、上記のブロック選択技術では、ブロック領域は四角形で表され、非テキストブロックの境界もまた記録される。しかし、歪んだ文書の場合、四角形はテキストブロック間の分離を不明瞭にして、実際重なることもある。これにより、ブロック内のデータの誤分類が起こり、文字認識のエラーとなることもある。

#### 【0009】

従って、垂直及び水平テキストブロックが認識され、テキストブロック形成の前に歪んだ文書を回転させる必要の無い、格納文書におけるテキスト及び非テキストブロックを効果的に且つ効率的に選択するための方法及び装置が求められている。これにより、処理時間を節約し、認識精度を向上させる、より柔軟なブロック選択技術が提供される。

#### 【0012】

##### 【課題を解決するための手段】

上記の目的を達成する本発明の一つの態様によるブロック選択方法は、  
画像データにおける黒画素の連続成分を検出する連続成分検出工程と、  
前記検出された連続成分を、サイズ及び他の連続成分との相対的位置に基づいて、テキスト又は非テキストに分類する分類工程と、  
少なくとも一部が前記分類工程で分類された非テキストの連続成分のエッジに沿っており且つ当該非テキストの連続成分に隣接する領域へ延びる不可視の直線をサーチするサーチ工程と、

前記分類工程で分類されたテキストの連続成分の相対的位置に基づいて、テキストブロックを形成するテキストブロック形成工程と、

前記テキストブロック形成工程で形成されたテキストブロックを、前記非テキストの連続成分の可視線または前記サーチ工程でサーチされた不可視の直線が横切っている場合、

10

20

30

40

50

当該テキストブロックを当該可視線または不可視の直線で分割する分割工程と、

前記分割工程で分割されたテキストブロックに対して、所定の後処理を実行する後処理工程とを備える。

【 0 0 1 3 】

また、上記の目的を達成するための本発明の他の態様によるブロック選択装置は以下の構成を備える。即ち、

画像データにおける黒画素の連続成分を検出する連続成分検出手段と、

前記検出された連続成分を、サイズ及び他の連続成分との相対的位置に基づいて、テキスト又は非テキストに分類する分類手段と、

少なくとも一部が前記分類手段で分類された非テキストの連続成分のエッジに沿っており且つ当該非テキストの連続成分に隣接する領域へ延びる不可視の直線をサーチするサーチ手段と、

前記分類手段で分類されたテキストの連続成分の相対的位置に基づいて、テキストブロックを形成するテキストブロック形成手段と、

前記テキストブロック形成手段で形成されたテキストブロックを、前記非テキストの連続成分の可視線または前記サーチ手段でサーチされた不可視の直線が横切っている場合、当該テキストブロックを当該可視線または不可視の直線で分割する分割手段と、

前記分割手段で分割されたテキストブロックに対して、所定の後処理を実行する後処理手段とを備える。

【 0 0 1 7 】

なお、これらの及び他の本発明の特徴及び有利な点は、以下の添付の図面に従う好適な実施例の詳細な説明を参照することにより、さらに容易に理解されるであろう。

【 0 0 1 8 】

【実施例】

以下に添付の図面を参照して本発明の好適な実施例を説明する。

【 0 0 1 9 】

本発明は、複写機、ファクシミリ装置、ビデオまたはスチルビデオカメラ、レーザビームプリンタ等の、文字認識処理が望まれる画像処理または画像再生装置などの様々な装置において実現される。そうした装置では、文字画像を含む画像は、文字画像が文字認識処理される様に処理または再生される。必要であれば、認識された文字画像を標準化された文字セットまたはフォントに換えて、原稿の文字画像ではなく標準化された文字を再送信したり再生したりすることもできる。本発明はまた、汎用コンピュータや、パーソナルコンピュータ、ワードプロセッシングまたはデータ処理機器等のオフィス機器、複数のビジネス機器ツールを1つの統合パッケージにおいて結合した統合オフィスオートメーション機器等においても実現され得るものである。

【 0 0 2 0 】

図2は本発明の代表的な実施例の外観を示す図である。図2において、IBM-PC（商標）またはPC互換性を持つコンピュータ等のコンピュータ10が示されており、コンピュータ10は、マイクロソフトウインドウズ・オペレーティングシステム（商標）等のグラフィカル・ユーザインタフェース・オペレーティングシステムを備えている。コンピュータ10にはカラーモニタ等の表示画面11が設けられている。コンピュータ10は、さらに、白黒、中間調、カラービットマップ画像等のデータファイルを格納したり、コンピュータ10がデータファイルを操作したり、表示画面11を通じてデータをオペレータにファイルの形で提示するためのプログラム・インストラクション・シーケンスを含むアプリケーションプログラムファイルを格納するための、コンピュータディスクドライブ14等のマスク格納装置を備えている。例えば、ディスクドライブ14上のアプリケーションプログラムは、光ディスク上のデータにアクセスするアプリケーションプログラムや、他の情報処理プログラムを含むものである。

【 0 0 2 1 】

テキストデータ入力や、オペレータが表示画面11に表示されたデータの選択や操作を行

10

20

30

40

50

なうためのキーボード 15 がコンピュータ 10 に接続されている。さらに、表示画面 11 上で処理対象物の選択や操作を行なうための、マウス等のポインティングデバイス 16 が設けられている。スキャナ 18 は文書または他の画像を走査して、例えば、それらの文書のビットマップ画像をコンピュータ 10 に提供する。それらの画像はコンピュータ 10 により直ちに使用されるか、あるいはコンピュータ 10 はこれらの画像をディスクドライブ 14 へ格納したり、または光ディスクドライブ 20 を介してディスク 19 等の光ディスクに格納する。引き続きコンピュータ 10 による処理のために、ビットマップ画像データはディスク 19 またはコンピュータディスクドライブ 14 から検索される。電話リンク 21 またはネットワーク 22 を介して、コンピュータ 10 内の不図示のモデムを通じて、ビットマップ画像データ及び他のデータを供給するための他の手段も設けることができる。さらに、取り外し可能なフロッピーディスクドライブを設けることも可能であり、または、不図示のビデオインタフェースを介して、デジタルまたはアナログのビデオ情報をコンピュータ 10 に入力してもよい。

10

#### 【0022】

コンピュータ 10 により処理された情報を入力するために、プリンタ 24 が設けられている。

#### 【0023】

コンピュータオペレータ命令に従って、そしてオペレーティングシステムの制御下で、格納されたアプリケーションプログラムが選択的に作動され、入力されたデータの処理や操作を行なう。例えば、以下に詳細に説明する様に、文字認識プログラムは、オペレータがスキャナ 18 を介して文書画像をスキャン入力し、スキャンされた画像を光ディスク 19 へ格納するために動作する。オペレータはオペレーティングシステム及びアプリケーションプログラムに従って、スキャン入力した文書を光ディスク 19 から検索することができる。

20

#### 【0024】

図 3 は本発明の 1 実施例の詳細なブロック図であり、スキャニング、ファクシミリ、情報送信及び受信、そして情報処理能力を有し、情報処理能力としては選択的にワードプロセッシング及びデータ処理能力を有する統合オフィスオートメーション機器に関するものである。

#### 【0025】

図 3 の装置において、画像はファクシミリ送信や、原稿のスキャン入力や、モデムを介した遠隔受信等によって入力される。本実施例によれば、画像中の文字を認識して、認識された文字のコンピュータテキストファイルを作成し、装置のワードプロセッシング、スプレッドシートプロセッシングまたは他の情報処理能力を使用してテキストファイルを修正することができる。修正されたテキストファイル（または修正を行っていないテキストファイル）は再送信されたり、あるいは、例えばテキストファイルをスピーカまたは通常の音声電話機により音声再生するスピーチ合成技術を用いて出力することができる。

30

#### 【0026】

図 3 では、プログラム可能なマイクロプロセッサ等の中央処理装置 (CPU) 30 がバス 31 と接続されている。さらにバスには、画像を画素単位で画像メモリ (例えば下記の RAM 32) へスキャン入力するスキャナ 18、デジタルデータをアナログ形式で電話線 21a を介して送受信するモデム 33、及び画像を電話線 21b を介して送受信するファクシミリ装置 34 ( 所望により不図示の電話を含む ) が接続されている。電話線 21a、21b は同じ線でもよいし、不図示のネットワーク制御部を通じて統合されていてもよい。バス 31 には、さらに、CPU 30 に実行される単数または複数のコンピュータプログラムを格納する読み取り専用メモリ (ROM) 35、認識処理の間に入力された文字が参照される文字の辞書を格納する文字ディクショナリ 36、入力された画像データ、処理された画像データ、画像構造に関する情報等を格納するランダムアクセスメモリ (RAM) 32、文字認識処理の間に認識された文字の識別が出力される出力装置 37 ( ディスクメモリまたはスピーカ / 音声電話線インタフェースを有するスピーチ合成装置 )、本装置によ

40

50

って処理された画像を表示するプリンタ/表示装置 24、及びオペレータが装置を制御するためのキーボード 15 が接続されている。

【0027】

バス 31 に接続された装置は 1 つの統合オフィスオートメーションツールに収納されているが、これらの装置の幾つかあるいは全てが選択的にスタンドアロン形式で提供され得ることも明らかである。

【0028】

スキャナ 18、モデム 33、ファクシミリ 34 は、画像データを装置へ入力する選択的な形の入力手段を構成している。スキャナ 18 を使用した場合、原稿はライン単位、画素単位でスキャンされて、そして CPU 30 の制御下で、画像データの画素が RAM 32 の画像メモリにビットマップメモリ形式で格納される。モデム 33 を使用した場合、画像データは電話線 21a よりアナログ形式で受信され、モデム 33 によりデジタル画素形式に変換され、RAM 32 の画像メモリに格納される。ファクシミリ 34 を使用した場合、画像データは修正ハフマンランレングス符号化方式等で圧縮または符号化された形で、電話線 21b より受信される。圧縮画像データは、ファクシミリ 34 によって公知の技術によりデジタル画像画素データに伸張 (uncompressed) され、CPU 30 は画像データの画素をビットマップ形式で RAM 32 の画像メモリに格納する。他の入力手段も勿論使用可能である。例えば、画像データは、ディスクメモリ等の大型記憶装置から簡単に検索して獲得することができ、また、ビデオまたはスチルビデオカメラから得ることもできる。

【0029】

ファクシミリ 34 と出力装置 37 は、文字認識された画像データを装置から出力するための選択的な出力手段を構成している。ファクシミリ 34 を使用した場合、本実施例に従って認識処理された文字画像は標準文字セットまたはフォントに変換され、装置から送信することができる。これにより、例えば、文字画像を含む画像の受信、文字画像の文字認識、再送信前に行なう認識された文字の標準文字フォントへの変換が可能となり、劣化した画像品質を向上させる。

【0030】

モデム 33 及び出力装置 37 は、画像データの認識された文字の識別を、例えば ASCII コードで出力したり格納するための選択的な手段である。文字の識別は装置内に (ディスクメモリ内等に) 格納され、または送信のためにモデム 33 を介して遠隔ロケーションへ出力される。ASCII コード等の文字の識別をファクシミリ互換性のある形式へと形式再変換するための手段を設けることも可能で、ファクシミリ 34 を起動することなく、モデム 33 を通じて離れて位置するファクシミリ装置へ送信することができる。

【0031】

プリンタ/表示装置 24 は、文字認識のいずれかのステップのための恒久的な記録を出力及び形成するだけでなく、文字認識処理の過程を監視するための手段である。キーボード 15 は、オペレータが図 3 の装置の操作を制御するためのものである。

【0032】

図 4 は本実施例によるブロック選択技術が有益に使用される 1 方法を理解するのに役立つ、全体的な文字認識処理を示すフローチャートである。図 4 の処理ステップは、プログラム ROM 35 に格納されたコンピュータプログラムに従って CPU 30 により実行される。

【0033】

ステップ S401 では、画素画像データが装置へ入力され RAM 32 へ格納される。画像データは画像を画素単位で表現している。好ましくは、画素データは 2 値画素データ、つまり白黒画像データである。しかし、画像データは、各画素が複数のグレイスケールレベルの内の 1 つで表現される中間調画像データであってもよいし、各画素が、その色を符号化するマルチビットワードで表現されるカラー画像データであってもよい。それらの場合または画素データが 2 値画素データではない他のいずれの場合でも、RAM 32 へ格納する前に、非 2 値画素データを 2 値画素データへ変換するための閾値処理が行なわれる。

10

20

30

40

50

## 【0034】

好ましくは、ステップS401で入力された画素画像データはポートレートタイプの画像、つまり左上隅から右下隅まで読んだ画像である。もし画像がその様に構成されていない場合、例えば、ランドスケープタイプの画像である場合、画素画像データは、ポートレートタイプの画像を表現する様に翻訳される。翻訳は、画像データを翻訳することを命じる、キーボード15を介したオペレータ入力の命令に従って行なわれる。

## 【0035】

ステップS402では、画像データが前処理される。通常、劣化した文字や画像を向上させること等により画像データを補強する様に前処理フィルタが使用される。適切な画像補強技術は、同出願人による、1991年10月4日出願の審査中の米国特許出願第07/771、220号で説明されている。

10

## 【0036】

ステップS402で、必要であれば、認識処理の速度を上げるために、可能な程度に精度とコストを落として、画素画像データの画素数を減らすか、圧縮を行なってもよい。例えば、 $m \times n$ ブロックの画素( $m$ と $n$ は異なる)の画素値を平均して、その $m \times n$ ブロックの平均値の単一の画素に変換することを行ってもよい。

## 【0037】

ステップS403では、画像の各セグメントにおける画像データのタイプを特徴付け、テキスト情報、グラフィックス情報、ライン画像情報、写真情報等の情報のブロックを指定して識別するために、ブロック選択が行なわれる。さらに、ステップS403のブロック選択では、画像の各部分が、後述のステップS412で説明される様な適切な順序に再構成される様に、階層的ツリー構造への画像の配置も行なわれる。例えば、階層的ツリー構造には、2つの欄の画像データが、欄1よりのテキストが欄2よりのテキストへ読み込まれるのを防止する様に、欄1のテキストを欄2のテキストに先行して再構成する情報が含まれている。ステップS403によるブロック選択を以下に詳細に説明する。

20

## 【0038】

最初の情報のブロックがRAM32に格納された画像から選択され、ステップS404では、それがステップS403で決定されたブロック識別に基づいてテキストブロックか否かが決定される。その最初のブロックがテキストブロックではない場合、ステップS405へ進み、次のブロックが選択されて、ステップS404へ戻る。

30

## 【0039】

ステップS404で、ブロックがテキストブロックである場合はステップS406へ進み、テキストブロックに対しライン分割が行なわれる。ライン分割では、テキストブロックにおけるテキストの個々のラインがテキストブロックのテキストの他のラインから分割されて、分割されたラインは、次に詳細に説明される様に順次処理される。

## 【0040】

ステップS407では、あるラインの各文字が、そのラインの他の文字から切り離しあるいは分割され、個々の文字が次に詳細に説明される認識処理ルーティンに提供される。

## 【0041】

ステップS408において、各文字に対して認識処理が文字単位で行なわれ、各文字は、文字ディクショナリ36に格納された標準文字と公知の技術で比較される。文字ディクショナリ36のエントリは、通常単一の文字に対するものであるが、ある文字の連続は切り分けるのが困難であり(例えば「fi」)、また単一の文字も時折不用意に切れるので(例えば「j」)、文字ディクショナリは一字以外の他のエントリも有している。得に、ディクショナリは切り離しが困難な接触文字の組のためのエントリを有している。さらに、ディクショナリはしばしば不意に切れる文字の部分のためのエントリも有している。

40

## 【0042】

比較に基づいて、文字画像に対してある識別が選択されて、RAM32へ格納されるか、または出力装置37へ出力される。また、所望により、識別された文字をプリンタ/表示装置24上で表示してもよい。

50



## 【 0 0 4 3 】

ステップ S 4 0 9 では、テキストブロックに対する処理が完了したかどうかを判定する。処理が完了していない場合は、ステップ S 4 0 6（または、ステップ S 4 0 7 が適切ならばステップ S 4 0 7）へ戻り、更なるライン分割または文字分割処理を行なう。ブロック処理が完了すると、ステップ S 4 1 0 へ進んで、ページに対する処理が完了したかどうかを判定する。ページの処理が完了していない場合は、ステップ S 4 0 5 へ戻り、そこで同じページの次のブロックが処理のために選択される。

## 【 0 0 4 4 】

ページの処理が完了すると、ステップ S 4 1 0 から S 4 1 1 へ進み、後処理が行なわれる。後処理にはコンテキストチェックやスペルチェック等の技術が含まれ、ステップ S 4 0 8 において実行された処理の様に個々の文字に基づくよりはむしろ、それらの文字の識別が生じるコンテキスト（全体的観点）に基づいて、ステップ S 4 0 8 の認識処理で認識された文字の識別結果が修正される。

10

## 【 0 0 4 5 】

ステップ S 4 1 2 では、画像データはブロック選択ステップ S 4 0 3 で規定されたツリー構造に基づいて再構成される。ページ再構成により、画像データは、適切な順序に変換される。例えば、脚注は本文テキストから切り離され、欄は他の欄と混合されるというよりは、むしろその欄の後へと続き、グラフィックまたは線画データはページの認識された文字テキストにおける適切な位置へ挿入される。

## 【 0 0 4 6 】

上述のグラフィック画像または線画像に関連する説明文（caption）はそれらの画像に隣接して挿入される。他のルールも適用可能である。例えば、ページの物理的再構成は必要ではない場合に、ページからテキストを抽出するために再構成を行なう、ということも可能である。

20

## 【 0 0 4 7 】

ステップ S 4 1 3 では、再構成されたページは例えば出力装置 3 7 に格納される。そして、ROM 3 5 に格納された他のアプリケーションプログラムに従い、CPU 3 0 により実行される形で、スプレッドシートやワードプロセッシング等の情報処理が行なわれる。続いて、処理された情報（または必要であれば処理を行なっていない情報）は、様々な手段により、例えばファクシミリ 3 4、モデム 3 5、またはコンピュータテキストファイルを音声再生するスピーチ合成装置による通常の音声電話機により、再送信される。

30

## 【 0 0 4 8 】

図 5 は図 4 のステップ S 4 0 3 におけるブロック選択処理を説明するための全体的なフローチャートである。ただし、ブロック選択処理は文字認識システムとは別に使用される。例えば、画像再生機器において、あるタイプのブロックに用いられる第 1 の画像再生方法を生じ、次のタイプのブロックに用いられる第 2 の画像再生方法を生じるため、またはデータ圧縮方式におけるブロック選択を使用するために、ブロック選択処理を用いることが可能である。

## 【 0 0 4 9 】

ステップ S 5 0 では、ブロック選択処理速度を上げることが必要であれば、画像補強及び/またはデータ縮小を行なうことができる。データ縮小が行なわれると、ブロック選択は縮小された画像に対して行なわれる。しかし、図 4 の文字認識処理（つまりステップ S 4 0 4 ~ S 4 1 3）に影響を与えない様に、ブロック選択処理の終了時点で、選択されたブロックは縮小されていない画素画像データに対して適用される。

40

## 【 0 0 5 0 】

画像データ縮小は  $m \times m$  画素ブロック毎に黒画素の連続性を評価することにより進められる。例えば、 $3 \times 3$  画素のブロックにおいて 2 つの連続する黒画素が存在すれば、その  $3 \times 3$  画素ブロックは単一の黒画素に縮小される。逆に、 $3 \times 3$  画素ブロックに 2 つの連続する黒画素が存在しなければ、その  $3 \times 3$  画素ブロックは単一の白画素に縮小される。好ましくは、 $4 \times 4$  画素ブロックの黒画素の評価が行なわれる。この原理は原画像の連続性

50

を保持するためのものである。

【0051】

ステップS51では、画素画像が分析された、連続成分が検出され、その大きさ及び他の連続成分との相対的位置に基づいて、検出された連続成分が分類される。連続成分は白画素によって完全に囲まれた黒画素のグループである。従って、連続成分は、少なくとも1つの白画素によって他の黒画素のグループから分離された黒画素のグループである。図6を参照して詳細に後述するが、ステップS51では連続成分（各画素を囲む8方向の内いずれかで連続した成分）の検出が行なわれる。ステップS52では、連続成分から得られたサイズ情報及びある統計上の値（下記に詳説する）に基づいて、各連続成分に分類が割り当てられる。詳細は後述するが、各連続成分はテキストユニットまたは非テキストユニットとして分類される。ステップS53において、非テキストユニットは更なる分類に掛けられ、フレームデータか、中間調画像か、線画か、表かまたは他のテーブル状の構造のテキストデータか、垂直線か、水平線か、垂直または水平な斜線であるか、または未知の分類に相当するかどうか、その厚さや黒画素の割合に基づいて判定される。連続成分の組織的データを提供し、ステップS412で前述した様にデジタルデータの再構成を助成するために、各連続成分に対して階層ツリー構造が展開される。

10

【0052】

ステップS54では、非テキスト連続成分のエッジに沿って不可視線（白線）がサーチされる。このような不可視線を使用する理由は、テキスト欄の間の距離が、同じページの大抵の欄の隙間よりも例外的に狭いものがあるためである。こうした不可視線の検出は、後で隣接するテキスト連続成分が実際は同じブロック内にグループ分けされるべきかどうかを判定するのに役立つ。

20

【0053】

ステップS55では、ステップS51で分類することができなかった非テキストユニットが分析されて、大きいフォントサイズのタイトルであるかどうか判定される。タイトルである場合は、ユニットは適切に再指定されて、ツリー構造が更新される。タイトルはステップS412におけるページ再構成に役立つものである。未知の非テキスト及びステップS52、S53で識別されたテキスト成分をグループ化することにより、水平または垂直方向にタイトルラインが形成される。このグループ化の前に、グループ化されると想定される成分の間の距離と、タイトルが水平または垂直である場合に可能なタイトルの長さ

30

【0054】

ステップS56では、後述するが、テキストブロックがテキスト連続成分から形成される。ステップS57では、形成されたテキストブロックを可視線または不可視線が横切っていた場合、テキストブロックは分割される。これらの線を使用する理由は、極端に狭い欄間の隙間のためだけでなく、異なるブロックのテキスト連続成分の間の距離が、斜めの場合に、特に不可視または可視線がそれらを分けている場合に、それらの成分が1グループにするのに十分に短くできるためである。

【0055】

ステップS58では、各テキストブロックの方向が検出される。この様に、本実施例によれば、テキストブロックが形成される前には歪み検出も画像回転も行なわれることはない。こうして、一般のブロック選択アルゴリズムとは異なり、本実施例は予め歪み検出や画像回転を行なうことなく、ページ画像に直接テキストブロックを形成する。これにより、画像回転に費やされる時間が節約され、歪み検出による不正確さを避けることができる。さらに、ある特殊な視覚効果を達成するために、同じページのテキスト領域の幾つかを、特別に斜めに編集することも可能である。こうした場合、歪み検出及び画像回転は、ページ画像を直立スタイルに修正するのに役立つことになる。こうして、本実施例によれば、テキスト及び非テキストブロックは、スキャンされたページの歪みに係わり無くサーチされる。テキストブロックが形成された後、各テキストブロックの歪み方向が個別に識別される。

40

50

## 【0056】

ステップS59では、テキストブロックの方向に基づいて、各テキストブロックのテキストラインが形成される。ステップS60では、テキスト及び非テキストブロック表現をより簡潔で明瞭にするために後処理が行なわれる。例えば、以前のブロック選択アルゴリズムは四角形で表現されたテキストブロックを形成するものであった。しかし、歪んだ文書の場合、そうした四角形ではテキストブロックの分離が不明瞭になる。テキストブロックの四角形領域が重なる場合にテキストブロック間の分離を明白にするために、本実施例では各テキストブロックをより簡潔に囲む不規則な形状の曲線を提供するものである。この曲線による囲みは表示効果を補強するだけでなく、ブロック画像の抽出にも有効である。テキストブロックを囲む不規則形状の曲線は、テキストブロック形成ステップS56でテキストブロックが新たに形成または更新される度に形成することができる。

10

## 【0057】

図6A、図6B、図6C、図6Dは、上述のステップS51～S53の処理を詳細に示すフローチャートである。このフローチャートには、画素画像データの連続成分がどの様にして検出されるか、またそれらの連続成分がどの様にして分類されるかが示されている。図6A～図6Dの各処理ステップは、プログラムROM35に格納されたプログラムステップに従ってCPU30により実行されるものである。

## 【0058】

ステップS601では、輪郭トレーシングにより画素画像データの連続成分が検出される。輪郭トレーシングは図7Aに示す様な画像データをスキャンすることにより行なわれる。スキャンは矢印Aで示される様に画像の右下部から左へと進み、画像の右の境界に到達するまで続く。スキャンは、例えば左上から右下へ向けたり、他の方向へも行うことができる。黒画素があると、隣接する画素が符号71で示される様な放射状のパターンになった順序で調べられ、黒画素の隣接画素も黒画素であるかどうか判定される。放射状パターン71は共通の中心から延びる8つの番号の付いたベクトルを持つので、以下この輪郭トレーシングは「8方向」トレーシングと称される。隣接する黒画素が見つかり、処理は上述の様に画像の外郭のトレースが終わるまで進められる。こうして、図7Bに示す様に、矢印Aの方向のスキャンは文字「Q」の尾の部分に対応する点72を突き止める。隣接画素の捜査は、文字「Q」の外郭がトレースされてしまう様に、放射状パターン71に従って進められる。閉じた輪郭の内側の部分はトレースが行なわれない。

20

30

## 【0059】

1つの連続成分が検出され、その輪郭が8方向トレーシングによりトレースされた後、スキャンは次の黒画素が見つかるまで進められる。こうして、例えば、完全に黒く塗りつぶされた領域を表すオブジェクト74が8方向トレースされる。同様に、単語「non-text」を示す手書き文字の非テキストオブジェクト75が、単語「text」を形成する個々の文字からなるテキストオブジェクト76における個々の単語と同様にトレースされる。図7Aに示すスキャンは画素データの全ての連続成分が検出されてその輪郭が8方向トレースされるまで続けられる。

## 【0060】

そして、ステップS602に進み、各連続成分が四角形化される。詳しくは、各連続成分の周りに最小の可能な四角形が描かれる。こうして、図7Bに示す様に、四角形77がオブジェクト72の周りに、四角形79がオブジェクト74の周りに、四角形80がオブジェクト75の周りに、そして四角形81a、81b、81c、81dがテキストオブジェクト76a、76b、76c、76dの周りに夫々描かれる。

40

## 【0061】

ステップS603では、ツリーの位置が各四角形に割り当てられる。殆どの部分について、ステップS603で得られたツリー構造は画素画像における各オブジェクトに対し直接ツリーの根から進んでいる。これは、連続成分の外郭のみがトレースされ、閉じた輪郭の内側の部分はトレースされないためである。この様に、図7Cに示す様に、連続成分72に相当する四角形77はページの根から直接進んでいる。しかし、非テキストオブジェク

50

ト 7 5 に相当する四角形 8 0 や、テキストオブジェクト 7 6 a、7 6 b に相当する四角形 8 1 a、8 1 b の様な、その四角形が他の連続成分の四角形の中にすっかり収まっている連続成分については、そうした連続成分を囲んでいる成分（この場合成分 7 4）からの末裔として指定されている。さらに、少なくとも 1 つの枝分れ成分を持つ各連続成分については、その成分自身からの「主末裔成分」として指定されている。この様に、成分 7 9 は他の末裔成分 8 0、8 1 a、8 1 b の中に自身 7 9 からの主末裔成分として指定されている。

#### 【 0 0 6 2 】

ステップ S 6 0 4 では、ツリーの最初のレベルにある連続成分が、テキストユニットまたは非テキストユニットに分類される。分類は 2 つのステップで進む。最初のステップでは、連続成分の四角形が所定の大きさの閾値と比較される。連続成分を囲む四角形の高さ及び幅が最大フォントサイズに相当する第 1 の所定閾値より大きい場合、連続成分は非テキストユニットと分類されて「非テキスト」属性がユニットに与えられる。

10

#### 【 0 0 6 3 】

第 2 のステップでは、全てのユニット、つまりまだ非テキストとして分類されていないユニットが、残りの全ての連続成分の集合的なサイズに基づいて適応的に決定された閾値と比較される。詳しくは、非テキストの指定がされていない全ての四角形のうちで通常のテキストサイズの大きい方（1 2 p t）及び中間のサイズ（中間の高さと中間の幅のうち小さい方の四角形が選択される。選択された値はスカラーで乗算され（説明の都合で、1 . 5 として選択されている）、高さ及び幅について適応的に決定された閾値とされる。決定された閾値より大きい全てのユニットは非テキストであると推定され、その様に分類される。一方、適応的に決定された閾値のいずれかより小さく所定のテキスト特性に合うユニットはテキストと推定される。ユニットはこの様に分類され適切な属性が与えられる。これらの分類は図 6 A ~ 図 6 D の残りで述べられる様な精密な分類であり、下記により詳細に説明する。

20

#### 【 0 0 6 4 】

ツリーの第 1 のレベルの各ユニットがテキストまたは非テキストとして分類された後、テキストユニットの主末裔成分を含む全ての末裔成分がテキストユニットとして分類される。非テキストユニットの主末裔成分は非テキストとして分類されるが、非テキストユニットの他の末裔成分はテキストユニットとして分類される。

30

#### 【 0 0 6 5 】

ステップ S 6 0 6 では、最初のユニットが選択される。ステップ S 6 0 6 で、そのユニットがテキストユニットである場合は、ステップ S 6 0 7 へ進んで次のユニットが選択される。非テキストユニットが選択されるまでステップ S 6 0 6 ~ S 6 0 7 が繰り返され、非テキストユニットが選択された時点でステップ S 6 0 8 へ進む。

#### 【 0 0 6 6 】

ステップ S 6 0 8 では、非テキストユニットが調べられ、そこからの末裔が存在するかどうか判定される。例えば、図 7 C に示す様に、非テキストユニット 7 9 は非テキスト主末裔 7 9 とテキスト末裔 8 0、8 1 a、8 1 b を含んでいる。

#### 【 0 0 6 7 】

ステップ S 6 0 8 において何らかの末裔が存在する場合、ステップ S 6 0 9 へ進み、そのユニットがフィルタ処理され、そのユニットが中間調（またはグレースケール）ユニットであるかどうか決定する。中間調フィルタリングでは、ユニットの末裔が調べられ、「ノイズサイズ」ユニットより小さいサイズの末裔の数が決定される。「ノイズサイズ」ユニットは、画像データについて予測された最小フォントサイズよりも小さい高さ及び幅を持つユニットである。ノイズサイズより小さいサイズの末裔の数が末裔の総数の過半数である時、ユニットは中間調画像であると判定される。従って、ステップ S 6 1 0 より S 6 1 1 へ進んで、「中間調」の属性が当該ユニットに与えられる。そしてステップ S 6 0 7 へ戻り、処理のために次のユニットが選択される。

40

#### 【 0 0 6 8 】

50

ステップS 6 0 9 の中間調フィルタリングでユニットが中間調画像ではないと判定されると、ステップS 6 1 0 からS 6 1 3 へ進んで、ユニットからの主末裔が更なる処理のために選択される。そしてステップS 6 1 4 へ戻る。

【 0 0 6 9 】

ステップS 6 0 8 で非テキストユニットに末裔が存在しないと判定されると、あるいはステップS 6 1 3 で主末裔が更なる処理のために選択されると、ステップS 6 1 4 でその主末裔にフレームフィルタリングが行なわれる。フレームフィルタリングは、問題のユニットがフレームであって、ユニットを囲む四角形の幅及びノまたは高さで夫々ほぼ等しい平行な水平線及び平行な垂直線を含んでいるかどうかを判定する様に設計される。特に、連続成分が調べられて、画素の各列について、ユニット内の連続成分の内側の部分をつなぐ最長の距離が決定される。こうして、図 8 A に示す様に、非テキストユニット 8 2 は連続成分 8 3 を含み、その輪郭は 8 4 で示される様に 8 方向トレースされている。列「i」については、連続成分の内部をつなぐ最長距離は輪郭の左の境界 8 5 a から右の境界 8 5 b までの距離  $X_i$  である。一方、列「j」については、連続成分の内部をつなぐ 2 つの距離が存在する。つまり連続成分の境界上の点 8 6 a、8 6 b 間の距離と、点 8 7 a、8 7 b 間の距離である。点 8 6 a、8 6 b 間の距離の方が点 8 7 a、8 7 b 間の距離よりも長い場合、距離  $X_i$  は列 j に関する連続成分の内部をつなぐ最長距離である。

10

【 0 0 7 0 】

非テキストユニット 8 2 における n 個の列の各々について、「x」距離が求められ、非テキストユニットがフレームであるかどうかを判定するために次の不等式がテストされる。

20

【 0 0 7 1 】

【 数 1 】

$$\sum_{k=1}^N \frac{(X_k - W)}{N} < threshold$$

【 0 0 7 2 】

ここで  $X_k$  は k 番目の列に関し連続成分の内部をつなぐ最長の距離であり（上記の通りである）、W は四角形ユニット 8 2 の幅であり、N は列数であり、そしてたとえ画像においてフレームが歪んでいたり斜になっていてもフレームの検出ができる様に、閾値が予め算出されている。1°の歪みまたは傾斜角度を許容するためには、「 $\sin(1^\circ) \times L +$  ステップS 6 0 4 で計算された平均テキスト高さに等しいオフセット」で満足な結果が得られることがわかる。

30

【 0 0 7 3 】

上記の不等式が満足された場合、ユニットはフレームデータと判定され、ステップS 6 1 5 からS 6 1 6 へと進み、「フレーム」属性が当該ユニットに与えられる。

【 0 0 7 4 】

ステップS 6 1 6 の後、フレームデータが表またはテーブル状に組織されたデータを含んでいる可能性を検討する。こうして、ステップS 6 1 7（図 6 B）で、連続成分の内部が調べられて白輪郭が得られる。

40

【 0 0 7 5 】

白輪郭はステップS 6 0 1 で検出された輪郭と同様であるが、黒画素よりも白画素が調べられる。こうして、図 9 A に示す様に、非テキストユニットの内部が、矢印 B の方向に右下から左上に向けてスキャンされる。最初の白画素が見つかったとき、その隣接画素が放射状パターン 9 1 に示される順序で調べられる。ただし、放射状パターン 9 1 は 1 から 4 まで番号が付けられたベクトルを有している。従って、このステップに従った白輪郭トレーシングは以下「4 方向」白輪郭トレーシングと称される。白輪郭トレーシングは、黒画素に囲まれた全ての白輪郭がトレースされるまで 4 方向に行なわれる。例えば、白輪郭トレ

50

ーシングは、96で示される、一般に指定された黒画素の様な、内部の他の黒画素に加えて黒画素セグメント92、93、94、95の内部輪郭を形成している画素に続くものである。白輪郭が見つかる度に、非テキストオブジェクトに囲まれた全ての白輪郭がトレースされるまで、スキヤニングが上述の様に矢印Bの方向に進められる。

【0076】

ステップS618では、非テキストユニットの密度が計算される。密度は、連続成分の黒画素数を数えて、その黒画素数を四角形に囲まれた画素の総数で除算することにより求められる。

【0077】

ステップS619では、非テキストユニット内で見つかった白輪郭の数が調べられる。白輪郭数が4以上の場合は、非テキストユニットは実際は、表またはテーブル状に配置されたテキストブロックの連続である可能性がある。従って、ステップS620（図6D）で、白輪郭フィリング率が決定される。白輪郭フィリング率は、白輪郭が非テキスト画像で囲まれた領域を埋める度合いである。図9Aに示す様に、白輪郭フィリング率は、黒画素が見つかる白空間である100や101の様な領域に加えて、完全に空白の空間である97や99の様な斜線領域を含んでいる。フィリング率が高いと、非テキスト画像が表またはテーブル状に配されたテキストデータのシーケンスである可能性がある。従って、ステップS621でフィリング率が調べられる。フィリング率が高いと、非テキスト画像が表またはテーブル状に配されたテキストデータのシーケンスである可能性がある。この判定の信頼度を増すためには、白輪郭が調べられて、それらが水平及び垂直に広がるグリッド状の構造を形成しているかどうか判定される。詳しくは、ステップS622では、非グリッド配列の白輪郭は、その境界が少なくとも2つの輪郭を横切って水平及び垂直に延びていない場合は再結合される。例えば、図9Aに示す様に、白輪郭99の左の境界102と右の境界103は、白輪郭100の左の境界104と右の境界105に一致する様に垂直に広がっている。従って、これらの白輪郭はグリッド構造に配置されているので、これらの白輪郭は再結合されない。同様に、白輪郭103の上部の境界106と下部の境界107は、白輪郭110の上部の境界108と下部の境界109に一致する様に水平に広がっている。従って、これらの白輪郭はグリッド状構造に配置されているので、これらの白輪郭は再結合されない。

【0078】

図9B～図9Cは、白輪郭が結合される状態を説明するための図である。図9Bは、例えば、ステップS401において前述した様に中間調画像を2値画像に変換する閾値計算を通じて形成される、非テキストユニット111を示している。非テキスト画像111は、白領域114、115、116、117、118、119に加えて黒領域112を含んでいる。推定上、これらの白領域のフィリング率は、ステップS621から再結合ステップS622へ進める様に、十分に高くなっている。最初に、図9Bに示す様に、白輪郭115の上下境界が白輪郭117の上下境界と比較される。これらの上下境界は一致しないので、図9Cに示す様に白輪郭115は白輪郭116と再結合されて、結合された白輪郭116'を生成する。図9Cにおいて、白輪郭117の左右境界が白輪郭118の左右境界と比較される。これらの境界は同じではないため、図9Dに示す様に、白輪郭117と119は単一の白輪郭117'に再結合される。

【0079】

再結合が生じなくなるまで、この処理が水平及び垂直に反復される。

【0080】

こうして、上述した様に、表に関する白輪郭は再結合される可能性が低く、非テーブルの白輪郭、例えば中間調画像または線画は再結合される可能性がより高い。従って、ステップS623で、再結合率が調べられる。再結合率が高い場合、またはステップS621でフィリング率が低い場合、非テキストフレームユニットは斜線、未知の画像(an unknown)、中間調画像または線画である可能性があり、ステップS628へ進んで非テキストユニットは表ではないとの指定がされて、後述のステップS642A（図6B）へ進む。

## 【 0 0 8 1 】

ステップ S 6 2 3 で再結合率が高くない場合は、ステップ S 6 2 4 へ進み、非テキスト画像は「表類」として指定される。ステップ S 6 2 5 では、8 方向に連続成分を検出して分類する様に、新たに指定された表の内部が調べられる。ステップ S 6 2 6 では、新しい内部の連続成分に基づいて、階層構造が更新される。ステップ S 6 2 7 では、内部の連続成分はテキストまたは非テキストとして再分類され、ステップ S 6 0 2 ~ S 6 0 4 で前述した様に、適切な属性が与えられる。ステップ S 6 2 7 の後、表内部のテキスト連続成分が線描画、中間調画、または未知の画であるかがサイズ、密度等に基づいて更に分類される。フレーム内部の非テキスト連続成分及び線描画内部の非テキスト連続成分も、ステップ S 6 0 8 以降に説明した様に、更に分類される。

10

## 【 0 0 8 2 】

ステップ S 6 2 1、S 6 2 3 において、ステップ S 6 2 1 でフィリング率が高くない場合、またはステップ S 6 2 3 で再結合率が高い場合は、非テキストフレームユニットは斜線、未知の画像、中間調画像または線画、つまり表ではない可能性がある。

## 【 0 0 8 3 】

ステップ S 6 1 9 において、白輪郭の数が 4 より少なければ、フレームユニットは表ではないと考えられる。従って、ステップ S 6 4 2 A へ進んで、フレーム及び密度が約 0 . 5 に等しい閾値より小さいかどうか判定される。閾値は、フレーム内部のテキストユニットまたは線画が画素の半分より少ないとの推測に基づいて選択されたものである。フレーム及び密度が閾値より小さい場合、上述のステップ S 6 2 5 ~ S 6 2 7 へ戻り、その後、

20

## 【 0 0 8 4 】

ステップ S 6 4 2 A において、フレーム及び密度が所定の閾値より小さい場合、ステップ S 6 4 2 へ進み、フレームユニットが線画または中間調画像に分類できるかどうか、あるいはフレームは分類できないか（つまり、フレームは「未知の画像」である）どうか、より詳細に後述される様に決定される。

## 【 0 0 8 5 】

ステップ S 6 1 5（図 6 A）に戻って、ステップ S 6 1 4 におけるフレームフィルタリングで非テキストユニット内のフレームが検出されなかった場合、ステップ S 6 3 5（図 6 B）へ進み、非テキストユニットがラインを含んでいるかどうか判定される。ラインはテキスト境界を表す有用な非テキストユニットである。しかし、そうしたラインにより境界が引かれているテキストは、しばしばラインと近接しているため、そのテキストをラインに結びつけることが可能である。従って、テキストが近接した、またはテキストが近接していないラインを検出する様に、ライン検出が指定される。

30

## 【 0 0 8 6 】

ラインを単独で検出するためには、ユニットの長さ方向で非テキストユニットのヒストグラムが計算される。図 8 B に示す様に、あるラインのヒストグラム 8 8 は、ライン幅にほぼ等しい高さの、明らかに均一な分配を示している。ラインの幅は非テキストユニットの幅（W）にほぼ等しい。いずれの差異も、画素画像が形成される時に原稿文書が歪んでいた場合に結果として生じる傾斜角度  $\theta$  によるものである。従って、非テキストユニットがラインを含んでいるかどうかを判定するためには、ヒストグラムにおける各  $cell_k$  が、非テキストユニットの幅 W と比較される。次に示す様に、これらの値の二乗平均の平方根の差が、閾値と比較される。

40

## 【 0 0 8 7 】

## 【 数 2 】

$$\sum_{K=1}^N \frac{(cell_k - W)^2}{N} < threshold$$

50

【 0 0 8 8 】

閾値は、非テキストユニット内のラインの歪みまたは傾斜角度  $s$  を許容するために算出されるものである。  $1^\circ$  の歪みまたは傾斜角度に対しては、

【 0 0 8 9 】

【 数 3 】

$$\sum_{K=1}^N \left[ \frac{k \sin(1^\circ)}{N} \right]^2$$

10

【 0 0 9 0 】

の閾値で満足な結果が得られることがわかる。

【 0 0 9 1 】

上記の式に従って近接物の無いラインが見つからない場合、そのユニットがテキスト近接ラインを含んでいるかどうか判定される。テキスト近接ラインが非テキストユニットに含まれているかどうかを判定するために、テキストユニットが調べられて、あるラインがユニット境界に沿って縦に延びているかどうか判定される。詳しくは、もしあるラインがユニットの縦方向に延びていれば、図 8 C に示す様に、ユニットを囲む四角形の境界がそのラインの非常に近くにあるということになる。従って、四角形の境界内にある最初の複数の黒画素における均一性が、境界からの距離の二乗の数の合計を計算することによって調べられる。このように、図 8 C を参照すると、次の不等式がテストされる。

20

【 0 0 9 2 】

【 数 4 】

$$\sum_{K=1}^N \frac{X_k^2}{N} < threshold$$

【 0 0 9 3 】

二乗の合計が所定の閾値より小さければ、テキストが接着したラインが見つかったことになる。接着物の無いラインに対して得られた閾値と同じ閾値で、満足する結果を得ることができる。

30

【 0 0 9 4 】

ステップ S 6 3 5 でラインが検出されると、ステップ S 6 3 6 から S 6 3 7 A へ進んで、「ライン」の属性が非テキストユニットに与えられる。そしてステップ S 6 0 7 (図 6 A) へ戻り、次のユニットが選択される。

【 0 0 9 5 】

一方、ステップ S 6 3 5 でラインが検出されない場合、ステップ S 6 3 6 から S 6 3 7 へ進んで、非テキストユニットの大きさが調べられる。サイズが所定の閾値より大きくなければ、非テキストユニットの分類は決定できない。閾値は最大フォントサイズ及び文書タイプ(水平、垂直、または両者混合)によって設定される。最大フォントサイズの半分だと満足な結果が得られる。従って、ステップ S 6 3 8 へ進んで「未知」の属性が非テキストユニットに与えられ、ステップ S 6 0 7 (図 6 A) へ戻って、次のユニットが選択される。

40

【 0 0 9 6 】

ステップ S 6 3 7 でサイズが所定閾値より大きければ、ステップ S 6 1 7、S 6 1 8、S 6 1 9 へ進み、前述の様に、非テキストユニットの内部の白輪郭がトレースされ、非テキストユニットの密度が計算され、白輪郭数が調べられる。

【 0 0 9 7 】

50



ステップS 6 1 9で白輪郭数が4より大きくなければ、上述の様にステップS 6 4 2 A、S 6 4 2へ進む。ステップS 6 4 2ではユニットのサイズが計算されて、線画または中間調画像を構成できる程大きいかどうか判定される。このサイズ判定は、黒画素の最大ランレングスに加えて、非テキストユニットの高さ及び幅と文書タイプ、に基づいて行なわれる。例えば、非テキストユニットの高さ及び幅が最大フォントサイズより大きくない場合は、その非テキストユニットは中間調画像または線画になる程大きくはないため、ステップS 6 4 3へ進み、「未知」の属性が与えられる。さらに、非テキストユニットの幅は最大フォントサイズより大きい黒画素の最大ランレングスは最大フォントサイズより大きくない場合も、ステップS 6 4 3へ進み、「未知」の属性が与えられる。そしてステップS 6 0 7 (図6 A)へ戻り、新たなユニットが選択される。

10

#### 【0098】

ステップS 6 4 2で、非テキストユニットが線画または中間調画像になる程大きい場合、ステップS 6 4 3 Aへ進み、非テキストユニットが水平または垂直の傾斜したラインであるかどうか判定される。その後、ステップS 6 4 4へ進み、非テキストユニットが線画か中間調画像であるかが判定される。ユニットが線画として、または中間調画像として分類されるかどうかは、ユニット内の黒画素の平均水平ランレングス、ユニット内の白画素の平均水平ランレングス、白画素対黒画素の比、及び密度に基づいて判定される。一般に、非常に暗い画像は中間調画像と考えられ、明るい画像は線画と考えられる。

#### 【0099】

詳しくは、白画素の平均ランレングスがゼロにほぼ等しく(つまり、暗い部分が優勢または点描画の様な場合)、ステップS 6 1 8で算出される様な密度が、ユニットが白より黒の方が強く(つまり、約1/2に等しい第1の閾値より密度が高い)、そして殆どの列が長い黒ランレングスを含んでいることを示す場合は、フレームユニットは中間調画像であると判定される。密度が第1の閾値よりも高くない場合は、ユニットは線画であると判定される。

20

#### 【0100】

白画素の平均ランレングスがゼロにほぼ等しくはなく、白画素の平均ランレングスが黒画素の平均ランレングスより大きく、そして殆どの列に長い黒ランレングスが含まれていない場合は、フレームユニットは線画であると判定される。しかし、白画素の平均ランレングスが黒画素の平均ランレングスより大きくない(つまり、再び、暗い部分が優勢な画像である)場合、更なるテストが必要である。

30

#### 【0101】

詳しくは、黒画素数が白画素数より大幅に少ない場合(つまり、白画素数で割った黒画素数が、2に等しい第2の閾値より大きい場合)、フレームユニットは中間調のユニットであると判定される。一方、白画素数で割った黒画素数は第2の閾値より大きくないが、ステップS 6 1 8で求めた密度が第1の閾値より大きい場合は、ユニットは中間調画像であると判定される。そうでなければ、ユニットは線画であると判定される。

#### 【0102】

従って、ステップS 6 4 4でユニットが線画であると判定されると、ステップS 6 4 5へ進み、「線画」の属性が与えられ、ステップS 6 4 6で全ての末裔成分が現在の親成分から移動されて、2代前の親の直接の末裔とされる。詳しくは、一度ユニットが線画であると判定されると、文字認識のためにいずれのブロックも線画から選択されることはない。その後、ステップS 6 2 5、S 6 2 6、S 6 2 7へ戻り、それからステップS 6 0 7で次のユニットが選択される。

40

一方、ステップS 6 4 4でユニットが線画ではないと判定されると、ステップS 6 4 7へ進み、「中間調」の属性が与えられ、ステップS 6 4 8で全ての末裔成分が現在の親成分から移動されて、2代前の親の直接の末裔とされる。そして、ステップS 6 0 7へ戻り、次のユニットが選択される。

#### 【0103】

本実施例は、上述のように判定された非テキストユニットに加えて、傾斜した垂直ライン

50

及び水平ラインを、その厚さや完全性に基づいて識別する。

#### 【0104】

ステップS53に2つの付加的機能を取り入れることも可能である。詳しくは、テキスト文字が下線より分離されるような、ライン近接物分離機能の実行が可能である。2番目の機能としては、点線、破線、または装飾された線（直線でも斜線でも）の検出が実行可能である。

#### 【0105】

ライン近接物分離機能は、テキスト（あるいは他のテキストのサイズの近接物）が、何らかの非テキスト成分（直線、水平線、フレーム等）に付随しているように見える所に取り入れることができる。例えば、テキスト文字が下に引かれた水平線に近接しているように見える場所では、テキスト文字がテキスト連続成分として特徴付けられ、ライン画素が非テキスト連続成分として検出される様に、テキスト文字をラインから切り離す必要がある。

10

#### 【0106】

テキストとラインの近接状態を分離する方法は、まずそうした近接物の存在を検出することで実行される。ラインの場合、ラインの片側のエッジはスムーズであると判定され、しかし反対側のエッジはスムーズではないと判定された場合、そのスムーズでない方のエッジに何らかの付着物が存在している。フレームの場合、外郭のエッジのいずれかがスムーズでない場合、そうしたフレームエッジに何らかの付着物が存在していることになる。

#### 【0107】

次に、付着されたライン（またはフレームのエッジ）の2つの端点が算出されるが、これらの2つの端点の位置は非スムーズ側の画像アウトラインに基づいている。そして、「分離線」が算出された2つの端点の間を通過して形成され、付着物は分離線に沿った線から切り離される。付着物はラインから「分離線」に沿って切り離され、付着物とラインはそれぞれ処理されて、付着物がテキスト連続成分か非テキスト連続成分かが判定されるが、好ましい形では、分離された付着物はテキスト連続成分として分類される。

20

#### 【0108】

点線、破線、また装飾された線の存在を検出する方法は、垂直線、水平線、水平斜線、垂直斜線のいずれかの属性を各検出された線に与えることで実行される。そうした非連続線を検出するために、水平及び/または垂直な小さいサイズの（ピリオドの大きさと同様である）テキスト連続成分が収集される。次に、収集された成分は、それらの相対的な距離に基づいて異なるグループに分配される。各グループについては、そのサイズが水平、垂直、または斜線として十分かどうか判定される。十分でなければ、そのグループが点線であり得るかどうかチェックされる。対象のグループがラインとして可能であれば、グループはサイズとグループ間の間隙の明瞭度とに基づいて再編成される。その後、グループは再びチェックされて点線であるかどうか判定される。

30

#### 【0109】

グループの幅が垂直線と同様であって、対象のラインの周辺に垂直なラインが存在し、グループとラインの間隙が明瞭であれば、グループは垂直線と考えられる。そうでない場合は、グループの長さが水平線と同じであるか、周辺に水平なラインが存在し、グループとこのラインの間隙が明瞭であるかどうか判定される。そうである場合には、検出された非連続ラインは水平線であると決定される。

40

#### 【0110】

画素画像における全ての連続成分が検出され、図6（図5のステップS53）で説明した様に分類されると、図10に示されるような修正されたツリーが得られる。図示の様に、ツリーの根は画素画像データのページに相当する。根から出ているのはテキストブロック、内容の不明な非テキストブロック、フレーム、写真、及び線の分類である。フレームから出ているのはテキストブロック、未知の非テキストデータ、写真や線のテキストブロックを含む表である。

#### 【0111】

50

図 5 に戻り、ステップ S 5 3 で全ての非テキスト連続成分が分類されると、ステップ S 5 4 では非テキスト連続成分のエッジに沿った不可視線（白線）がサーチされる。そうした不可視線は（ステップ S 5 2 で求められた不可視線も含わせて）、ステップ S 5 7 で、可視または不可視線がテキストブロックを横切っている場合に、テキストブロックを分離するのに使用される。この手順は後で詳細に説明する。

【 0 1 1 2 】

ステップ S 5 5 で、ステップ S 5 3 で分類された未知の非テキスト連続成分を用いて、水平及び垂直タイトルラインが形成される。これらの未知の非テキスト連続成分は、グループと想定される成分間の距離と、タイトルが水平または垂直である場合に可能なタイトルな長さとはに基づいて、名目上垂直、水平、または傾斜のタイトルの方向を検出するのに使用される。この方法を用いて、各タイトル連続成分が水平及び垂直方向の非常に近接したブロックと比較され、より近いブロックに水平または垂直タイトルが添付される。

10

【 0 1 1 3 】

ステップ S 5 6 では、テキスト連続成分はテキストブロックを形成するのに用いられる。簡潔に述べると、各テキスト連続成分に対し、周辺にある近い水平及び垂直の成分がサーチされて、集成プロセスで、テキストブロックとしてグループ化される。プロセスを図 1 1 に従って詳細に説明する。

【 0 1 1 4 】

図 1 1 において、ステップ S 5 2 で識別されたテキスト連続成分を用いて、ステップ S 5 6 1 で H - G A P 及び V - G A P が計算される。これらはそれぞれ、隣接するテキスト連続成分間の水平及び垂直中央間隙部である。ステップ S 5 6 2 では、全てのテキスト連続成分が、その垂直な位置によりソートされる。ステップ S 5 6 3 では、現在のテキストブロックの形成が開始される。ステップ S 5 6 4 では、ソート順に連続成分が抽出され、値 V - D I S T 及び H - D I S T が計算される。これらの値はそれぞれ、抽出された連続成分と現在のテキストブロックの間の垂直及び水平距離である。利用可能な連続成分が存在しなければ、ステップ S 5 6 3 へ戻り、次のテキストブロックが形成される。

20

【 0 1 1 5 】

ステップ S 5 6 5 では、抽出された連続成分について、V - D I S T が V - G A P より大きいかどうか、または H - D I S T が H - G A P より大きいかが判定される。ステップ S 5 6 5 でいずれかに該当すれば、抽出された連続成分は現在のテキストブロック内には存在しないことになるので、ステップ S 5 6 4 へ戻り、ソート順に次の連続成分が抽出される。一方、ステップ S 5 6 5 でいずれにも該当しない場合は、抽出された連続成分は現在のテキストブロック内に存在することになるので、ステップ S 5 6 6 へ進む。

30

【 0 1 1 6 】

ステップ S 5 6 6 では、抽出された連続成分と、現在のテキストブロックに既に含まれている連続成分とが、所定の条件を満たすかが判定される。好適な実施例では、所定条件は図 1 2 に示される通りである。

【 0 1 1 7 】

図 1 2 において、ステップ S 5 6 6 1 では、抽出された連続成分が現在のテキストブロックに既に含まれている連続成分と比較される。ステップ S 5 6 6 2 では、値 h - d i s t 及び v - d i s t が計算される。これらの値はそれぞれ、抽出された連続成分とステップ S 5 6 6 1 で識別された 1 つの連続成分（テキストブロックに含まれている連続成分）との間の水平及び垂直距離を表している。

40

【 0 1 1 8 】

ステップ S 5 6 6 3 では、v - d i s t が V - G A P より小さいかどうか、そして h - d i s t が H - G A P より小さいかが判定される。ステップ S 5 6 6 3 の条件に該当しない場合は、テキストブロックに含まれている連続成分は抽出された連続成分に十分に近似していないことになるため、ステップ S 5 6 6 1 へ戻って、現在のテキストブロック内の次のテキスト連続成分が抽出された連続成分と比較される。

【 0 1 1 9 】

50

ステップS 5 6 6 3で条件に該当すれば、ステップS 5 6 6 4へ進み、抽出された連続成分がテキストブロックに含まれている連続成分とどのような垂直関係にあるのかが判定される。3つの関係が可能である。つまり、(1)抽出された連続成分が垂直にテキストブロックに含まれている連続成分の上にある、(2)抽出された連続成分は垂直にテキストブロックに既に含まれている連続成分の上ではなく、水平方向にある、(3)抽出された連続成分は垂直にテキストブロックに既に含まれている連続成分の上ではなく、斜め下方向にある、という関係である。条件(1)が満たされる場合、ステップS 5 6 6 5へ進むが、条件(2)または(3)が満たされる場合は、ステップS 5 6 6 7へ進む。

#### 【0120】

ステップS 5 6 6 5では、現在のテキストブロック内に、抽出された連続成分に近い他の連続成分が存在するかどうか判定される。これは、 $h - dist$ が $W - GAP$ より小さいかどうか( $W - GAP$ は $H - GAP$ より僅かに広い)を判定することにより確認される。該当しない場合は、ステップS 5 6 6 1へ戻って、現在のテキストブロック内の次の連続成分を用いて比較を行なう。しかし、ステップS 5 6 6 5において該当する場合は、ステップS 5 6 6 7へ進み、何らかの線描オブジェクト(例えば、ライン、線描画等)が抽出された連続成分とテキストブロックに含まれている連続成分の間に存在するかどうか判定される。そのような線描オブジェクトが存在する場合は、ステップS 5 6 4へ戻り、次の連続成分がパート順に抽出される。しかし、ステップS 5 6 6 7でそうしたオブジェクトが存在しないと判定された場合は、図11のステップS 5 6 7へ戻る。

#### 【0121】

図11に戻り、ステップS 5 6 7では、ちょうど処理されたばかりの、抽出された連続成分が、ステップS 5 2で識別された最後のテキスト連続成分であるかどうか判定される。抽出された連続成分が最後の成分でない場合は、ステップS 5 6 8で、その連続成分は現在のテキストブロックに挿入され、ステップS 5 6 4へ戻って、ソート順に次の連続成分が抽出される。一方、ステップS 5 6 7で最後の連続成分と判定された場合は、その成分は現在のテキストブロックに既に挿入されているので、図5のステップS 5 7へ戻る。

#### 【0122】

図5に戻って、ステップS 5 7では、ステップS 5 6で形成されたテキストブロックが単一のテキストブロックであるかどうか、または形成されたテキストブロックを横切る可視線または不可視線により、複数のテキストブロックに分割されるべきであるかどうか判定される。例えば、図13Aはテキストブロック131を示し、テキストブロック131内には複数の連続成分132が配されている。テキストブロックに隣接するのは非テキストブロック133、134である。非テキストブロックの間にはステップS 5 4で述べた不可視(白)線が存在している。例えば、非テキストブロック133、134が各々写真を含み、ブロックの間に不可視線が引かれている場合は、非テキストブロック133に隣接するテキスト連続成分は非テキストブロック133にのみ関係し、非テキストブロック134に隣接するテキスト連続成分は非テキストブロック134にのみ関係している。ステップS 5 7で形成されたテキストブロック131が単一のテキストブロックである場合は、後に行なわれる文字認識処理で、非テキストブロック133、134を文字として誤認識する可能性がある。

#### 【0123】

図13Bに示すように、不可視線135(または可視線)が先に形成されたテキストブロック131を横切っている場合において、テキストブロックは2つのテキストブロック136、137に分割される。そして、非テキストブロック133、134のそれぞれのテキスト連続成分が適切に処理される。ステップS 5 7の後、ステップS 5 8へ進み、各テキストブロックの方向が検出される。テキストブロックの方向の検出は、テキストブロック内のテキスト連続成分の長さ及び幅、連続成分間の水平及び垂直の間隙、及びそれらの位置に依存して行なわれる。図14A、及び、図14Bは、この処理を詳細に説明したフローチャートである。

#### 【0124】

10

20

30

40

50

図14Aにおいて、まずステップS5801で、ステップS56で形成されたテキストブロック内の全てのテキスト成分について、様々な値が計算される。ただし、幅は水平寸法に相当し、長さは垂直寸法に相当するものである。値MAX-LEN及びMAX-WIDが計算されるが、これらはそれぞれ、テキストブロック内のテキスト連続成分間の最長の長さ及び最大の幅を表している。また値AVG-LEN及びAVG-WIDが計算されるが、これらはそれぞれ、テキストブロック内の全ての連続成分の平均の長さおよび幅を表している。最後に、値STD-LEN及びSTD-WIDが計算されるが、これらはそれぞれ、テキストブロック内の全ての連続成分の長さ及び幅の標準的な偏差を表している。

#### 【0125】

その後、複数の処理ステップが実行されてテキストブロックが水平テキストブロックか垂直テキストブロックかが判定される。ステップS5802では、STD-LENがSTD-WID以下かどうか、MAX-WIDがMAX-LENより大きいかどうか、そしてMAX-WIDが12ptのプリントサイズ、もしくは $1.5 \times \text{MAX-LEN}$ よりも大きいかが判定される(ここで、12pt値は通常のテキストサイズに基づいている。勿論、これより大きいまたは小さいテキストサイズの場合は、好ましくはダイナミックに調整できる異なる値が必要とされる)。これらの関係に該当すると、テキストブロックは水平テキストブロックであると決定される。これらの関係に該当しない場合は、ステップS5803へ進む。

#### 【0126】

ステップS5803では、STD-WIDがSTD-LEN以下であるかどうか、MAX-LENがMAX-WIDより大きいかどうか、そしてMAX-LENが12pt、または $1.5 \times \text{MAX-WID}$ より大きいかが判定される。これらの関係に該当すると、テキストブロックは垂直テキストブロックであると決定される。これらの関係に該当しない場合は、ステップS5804へ進む。

#### 【0127】

ステップS5804では、各連続成分について、値H-SHORT及びV-SHORTが決定される。これらの値はそれぞれ、各連続成分及びその周辺の連続成分からの最短の水平及び垂直の間隙を表している。

#### 【0128】

そしてステップS5805(図14B)へ進み、テキストブロック内の大抵の連続成分について、H-SHORTがV-SHORTより小さいかが判定される。この関係に該当すると、テキストブロックは水平テキストブロックであると決定される。この関係に該当しない場合は、ステップS5806へ進んで、テキストブロック内の大抵の連続成分について、V-SHORTがH-SHORTより小さいかが判定される。この関係に該当すると、テキストブロックは垂直テキストブロックであると決定され、該当しない場合は、ステップS5807へ進む。

#### 【0129】

ステップS5807では、テキストブロック内の連続成分間の水平及び垂直の間隙の数が計算される。その後、ステップS5808では、水平間隙の数が垂直間隙の数より多いかどうか、またテキストブロックの幅が $12\text{pt} \times 2$ より大きいかが判定される。この関係に該当すると、テキストブロックは水平テキストブロックであると決定され、該当しない場合は、ステップS5809へ進む。ステップS5809では垂直間隙数が水平間隙数より多いかどうか、またテキストブロックの長さが $12\text{pt} \times 2$ より大きいかが判定される。この関係に該当すると、テキストブロックは垂直テキストブロックであると決定され、該当しない場合は、ステップS5810へ進む。

#### 【0130】

ステップS5810では、テキストブロックの長さが $12\text{pt} \times 2$ より小さく、全ての連続成分が同一の水平ライン上にあるかが判定される。これに該当する判定であれば、テキストブロックは水平テキストブロックであり、該当しない判定の場合は、ステップS5811へ進む。ステップS5811では、テキストブロックの幅が $12\text{pt} \times 2$ より

10

20

30

40

50

小さく、全ての連続成分が同一の垂直線上にあるかどうか判定される。この関係に該当すると、テキストブロックは垂直テキストブロックであると決定され、該当しない場合は、ステップS5812へ進む。

#### 【0131】

ステップS5812では、テキストブロック内の全ての連続成分が垂直にソートされて、最初の連続成分と最後の連続成分が、重なる垂直領域を有しているかどうか検出される。それらが重なっていれば、テキストブロックは水平テキストブロックであると決定され、重なっていない場合は、ステップS5813へ進む。

ステップS5813では、テキストブロック内の全ての連続成分が水平にソートされて、最初の連続成分と最後の連続成分が、重なる水平領域を有しているかどうか検出される。それらが重なっていれば、テキストブロックは垂直テキストブロックであると決定され、重なっていない場合は、テキストブロックは未知のブロックである。

10

#### 【0132】

識別された全てのテキストブロックの方向が一度決定されると、ステップS59へ進み、ステップS58で識別された方向に基づいて、各テキストブロックに関するテキストラインが形成される。図15はこの処理を詳細に示すフローチャートである。

#### 【0133】

図15は水平テキストブロック内のテキスト連続成分の処理を示している。垂直テキストブロック内の連続成分の処理も同様であるので、詳細な説明は省略する。ステップS591では、水平テキストブロック内のテキスト成分が、その水平位置に従ってソートされる。その後、ステップS592では、ソーティングリストの次の連続成分が抽出され、その連続成分が既に存在しているいずれかのテキストラインに嵌入するかどうか検出される。この判定の詳細な処理は、図16A～図16Eで示されるフローチャートを参照して説明する。

20

#### 【0134】

図16Aでは、ステップS5921で、ステップS592で抽出した連続成分を用いて、値C-LINE1及びC-BLOCK1が計算される。C-LINE1は、抽出された連続成分と現在のテキストラインとの間のパーセントの垂直領域重畳部分を表し、C-BLOCK1は、抽出された連続成分と現在のテキストラインの末尾連続成分との間のパーセントの垂直領域重畳部分を表している。図16AのステップS5921は現在のテキストライン、その末尾の連続成分、抽出された連続成分、及びそれらの垂直領域重畳部分の関係を示している。

30

#### 【0135】

ステップS5922では、C-LINE1がゼロより大きいかが判定される。ゼロより大きくなければ、図16Eに従ってより詳細に説明されるステップS5936へ進む。一方、ステップS5922で、C-LINE1がゼロより大きいと判定されると、3つの値h-dist1、v-dist1、NEXT-CLOSEST-LINEが決定される。h-dist1は、抽出された連続成分と現在のテキストラインの末尾連続成分との間の水平距離である。v-dist1は、抽出された連続成分と現在のテキストラインの末尾連続成分との間の垂直距離である。NEXT-CLOSEST-LINEは、(i)現在のラインの下にあり、(ii)その末尾連続成分が抽出された連続成分に最も近接しており、末尾連続成分と抽出された連続成分の間の距離は所定の距離内である、テキストラインである。そして、図16BのステップS5924へ進む。

40

#### 【0136】

ステップS5924では、値C-LINE2、C-BLOCK2、h-dist2、及びv-dist2が決定される。C-LINE2は、抽出された連続成分とNEXT-CLOSEST-LINEとの間のパーセントの垂直領域重畳部分である。C-BLOCK2は、抽出された連続成分とNEXT-CLOSEST-LINEの末尾連続成分との間のパーセントの垂直領域重畳部分である。h-dist2は、抽出された連続成分とNEXT-CLOSEST-LINEの末尾連続成分との間の水平距離である。v-dist2

50

は、抽出された連続成分とNEXT-CLOSEST-LINEの末尾連続成分との間の垂直距離である。

【0137】

その後、ステップS5925へ進み、C-LINE1が50%未満でC-LINE2が50%以上であり、そして現在のラインが上方向に傾き、NEXT-CLOSEST-LINEが上方向に傾いているかどうか判定される。この関係に該当すると、ステップS5926へ進み、次の水平テキストラインが選択されて、図15のステップS592へ戻る。一方、この関係に該当しない場合は、図16CのステップS5927へ進む。

【0138】

図16Cにおいて、ステップS5927では、8つの関係が存在するかどうか判定される。それらの関係は次の通りである。即ち、

(1) C-LINE2=0%

(2) C-LINE1 50%、C-LINE2 50%、  
C-BLOCK1 0%、C-BLOCK1>C-BLOCK2

(3) h-dist1>0、h-dist2>0、  
(h-dist1<12pt、又はh-dist1<h-dist2)、  
C-BLOCK1>0%、C-BLOCK1>C-BLOCK2

(4) h-dist1>0、h-dist2>0、  
(h-dist1<12pt、又はh-dist1<h-dist2)、  
C-BLOCK1>0%、現在のラインは下方向に傾き、  
NEXT-CLOSEST-LINEも下方向に傾いている

(5) h-dist1>0、h-dist2>0  
(h-dist1<12pt、又はh-dist1<h-dist2)、  
C-BLOCK1=0%、C-BLOCK2=0%、  
v-dist1<h-dist2

(6) h-dist1 0、h-dist1<12pt、  
C-BLOCK1>0%、h-dist2<0

(7) h-dist1<0、h-dist2<0、  
v-dist1>v-dist2

(8) C-BLOCK1>C-BLOCK2

である。

【0139】

そしてステップS5928へ進み、(1)~(8)の関係のいずれかに該当するかどうか判定される。該当するものが無いと、ステップS5929へ進んで、次の水平テキストラインが選択され、図15のステップS592へ戻る。ステップS5928で(1)~(8)の関係のいずれかに該当すると、ステップS5930へ進んで、抽出された連続成分の水平位置が、現在のテキストラインの末尾連続成分と重畳するかどうか判定される。重畳部分が存在すれば、ステップS5931へ進み、重畳部分が存在しなければ、ステップS5932へ進む。

【0140】

図16Dにおいて、ステップS5931では、抽出された連続成分と現在のテキストラインの末尾連続成分を合計した長さが、1文字の長さと同じであるかどうか判定される。合計の長さが1文字の長さと同じである場合は、ステップS5935へ進み、抽出された連続成分は現在のテキストラインに嵌入される。一方、合計の長さが1文字の長さと同じでない場合には、ステップS5931'へ進んで、C-BLOCK1が50%より大きいかどうか、またはv-dist1が1ドットサイズより小さいかどうか判定される。いずれかの関係に該当すると、ステップS5935へ進んで、抽出された連続成分は現在の水平テキストラインへ嵌入される。しかし、ステップS5931'のいずれの関係にも該当しない場合は、ステップS5933へ進み、次の水平テキストラインが選択されて、その後、図15のステップS592へ戻る。

10

20

30

40

50

## 【0141】

ステップS5932では、 $h - dist1$ が $12pt \times 2$ より小さく、( $C - LINE1$ が0%より大きい、または $C - BLOCK1$ が0%より大きい)かどうか、あるいは、 $NEXT - CLOSEST - LINE$ が存在しないかどうか、または $v - dist1$ が $v - dist2$ より小さいかが判定される。これらの条件が満たされると、抽出された連続成分は現在のテキストライン内に当て嵌まると決定され、ステップS5935を介して図15のステップS592へ進む。ステップS5932の關係に該当しない場合は、ステップS5934で次のテキストラインが選択され、再び図15のステップS592へ戻る。

## 【0142】

図16Aへ戻り、ステップS5922で、 $C - LINE1$ が0%より大きくないと判定されると、図16EのステップS5936へ進む。ステップS5936では、ステップS5923、S5924で概略を述べたように、 $NEXT - CLOSEST - LINE$ 、 $h - dist1$ 、 $h - dist2$ 、 $v - dist1$ 、及び $v - dist2$ が決定される。

## 【0143】

ステップS5937では、 $NEXT - CLOSEST - LINE$ が存在し、 $v - dist1$ が $v - dist2$ より大きいかが判定される。該当する場合、ステップS5939へ進み、次のテキストラインが選択され、そして図15のステップS592へ戻る。しかし、ステップS5937の關係が存在しなければ、ステップS5938へ進み、 $h - dist1$ がゼロ未満で $v - dist1$ が1ドットサイズ以下であり、抽出された連続成分と末尾連続成分の合計の長さが1文字の長さと同じであるかが判定される。これらの關係に該当すると、ステップS5941で、抽出された連続成分は現在の水平テキストライン内に当て嵌まると決定され、図15のステップS592へ戻る。ステップS5938の關係に該当しない場合は、ステップS5940へ進み、 $h - dist1$ がゼロ以上で、抽出された連続成分が末尾連続成分の上にあり、抽出された連続成分の垂直位置が末尾連続成分の垂直位置に近いかが判定される。これらの關係に該当すると、ステップS5941へ進み、抽出された連続成分は現在の水平テキストライン内に当て嵌まると判定され、図15のステップS592へ戻る。これらの關係に該当しない場合は、ステップS5942へ進み、次の水平テキストラインが選択され、図15のステップS592へ戻る。

## 【0144】

図5のステップS59において各テキストブロックについて全てのテキストラインが形成されると、ステップS60へ進み、ブロックの後処理が行なわれる。後処理は、文字認識、データ圧縮等の、特定の後続する手順を適用するようにしてもよい。この応用の目的で、後続する文字認識をブロックに対する後処理として行なう場合を説明する。こうした後処理の目的は、テキスト及び非テキストブロック表現をより簡潔にすることである。後処理において、必要であれば、テキストブロックは他のテキストブロックと結合され、非テキストブロックは他の非テキストブロックと結合されて、文字認識等の更なる処理を容易にする。まず、例えば、最小二乗法を用いて、形成されたテキストライン(ステップS59で決定された)に含まれる連続成分に基づいて、そのテキストラインの歪み角度を概算し、全体のページの歪みが計算される。ページの歪みは、ページのテキストラインの歪み角度の平均値となる。

## 【0145】

ページが僅かに歪んでいるだけであると(例えば、図17に示すように、各テキストを囲む四角形または不規則な曲線の間の間隙が明確である)判定された場合は、様々な処理ステップが実行できる。始めに、テキスト(またはタイトル)ブロックと、タイトルブロックが、重畳する四角形または不規則な曲線の囲みが存在するかどうか、それらの方向(orientation)等に基づいて結合できる。次に、ある小さいテキストブロックが実際にテキストブロックであるか、それとも写真等の非テキストブロックの一部であるかを判断するために、ある中間調(またはグレイスケール)画像が四角い領域を占めているかどうか

10

20

30

40

50



判定できる。3番目に、欄の分析を行い、欄情報に基づいて、テキストブロックが結合できる。こうした結合は、基本的に、ブロックが同一の欄内にあるかどうか、お互いに近接しているかどうか、その結合が他のブロックに重なるかどうか、及びそれらのブロックが同一の傾斜を持っているかどうかに基づいて行なわれる。

【0146】

ページの歪みがいくらか大きい場合は、歪みページのケースにおいて上記の3つの処理ステップは抜かされる。そして、小さいテキストブロックがその近接ブロックと結合されるステップが実行される。この結合は、近接ブロックの属性、近接ブロックの近接度、及びテキストブロック自身の特性に基づいて行なわれる。こうした結合の後、結合されたテキストブロックの属性は、結合された近接テキストブロックと一致するように変えられる。

10

【0147】

また、複数のテキストブロックは、複数のタイトルブロックまたは未知の連続成分を1つ含んだ複数ブロックと結合することができる。この結合は、その相対的な位置、ブロックの方向、ブロックの相対的なサイズ等に元づいて行なわれる。

【0148】

最後に、ブロックの後処理として、幾つかの非テキストブロックの結合を行なうこともできる。こうした結合は、非テキストブロックの相対的な位置、ブロックの属性等に基づいて行なわれる。例えば、幾つかの写真ブロックは、それらが同一の写真に属しているということが検出されれば結合される。こうした後処理を、後続の文字認識処理を容易にするために実行することが可能である。

20

【0149】

このように、格納された文書において効果的及び効率的にテキスト及び非テキストブロックを選択する方法及び装置であって、垂直及び水平テキストブロックが認識され、テキストブロックの形成の前に文書の回転を必要としない、方法及び装置を説明した。

【0150】

図面のブロックで概略が示され、または指定された個々の成分は、文字認識の分野では公知のものであり、その特定の構成及び動作は、本実施例の操作または実行の最良の形態において重要なものではない。

【0151】

本発明は好適な実施例について説明されたが、本発明は開示された実施例に限定されるものではないことが理解されるべきである。反対に、発明の精神と範囲内に含まれる、様々な修正や同等の変形が可能であるよう意図されている。特許請求の範囲の記載は、そうした修正及び同等の構成及び機能の全てに渡るように最大の解釈が与えられるものである。

30

【0152】

尚、本発明は、複数の機器から構成されるシステムに適用しても1つの機器からなる装置に適用しても良い。また、本発明はシステム或いは装置に本発明により規定される処理を実行させるプログラムを供給することによって達成される場合にも適用できることはいうまでもない。

【0153】

【発明の効果】

40

以上説明したように、本発明によれば、テキストブロックを非テキストブロックの不可視線、非テキストブロックのエッジに沿った不可視線により分割することにより、より確実にテキストブロックを抽出でき、非テキストブロックを文字とするような誤認識の発生を低減できる。

【0154】

【図面の簡単な説明】

【図1】文字認識される文書の一般的な1ページを示す図である。

【図2】本発明の一実施例に係わる装置の斜視図である。

【図3】本発明の一実施例に係わる装置のブロック図である。

【図4】本実施例のブロック選択の特徴がどの様に文字認識のプロセスに組み込まれてい

50

るかを示す、一般化されたフローチャートである。

【図 5】本発明に係わるブロック選択技術を示す全体的なフローチャートである。

【図 6 A】図 5 のステップ S 5 1 ~ S 5 3 の詳細を示すフローチャートである。

【図 6 B】図 5 のステップ S 5 1 ~ S 5 3 の詳細を示すフローチャートである。

【図 6 C】図 5 のステップ S 5 1 ~ S 5 3 の詳細を示すフローチャートである。

【図 6 D】図 5 のステップ S 5 1 ~ S 5 3 の詳細を示すフローチャートである。

【図 7 A】輪郭トレースを説明するための図である。

【図 7 B】輪郭トレースを説明するための図である。

【図 7 C】輪郭トレースを説明するための図である。

【図 8 A】非テキストユニットに対する分類処理を説明するための図である。

10

【図 8 B】非テキストユニットに対する分類処理を説明するための図である。

【図 8 C】非テキストユニットに対する分類処理を説明するための図である。

【図 9 A】白輪郭処理を説明するための図である。

【図 9 B】白輪郭処理を説明するための図である。

【図 9 C】白輪郭処理を説明するための図である。

【図 9 D】白輪郭処理を説明するための図である。

【図 10】本実施例における、テキスト及び非テキスト文字を決定するためのルートまたはツリー階層構造を概略的に示す図である。

【図 11】図 5 のステップ S 5 6 を示すフローチャートである。

【図 12】図 11 のステップ S 5 6 6 を示すフローチャートである。

20

【図 13 A】図 5 のステップ S 5 7 を概略的に示す図である。

【図 13 B】図 5 のステップ S 5 7 を概略的に示す図である。

【図 14 A】図 5 のステップ S 5 8 を示すフローチャートである。

【図 14 B】図 5 のステップ S 5 8 を示すフローチャートである。

【図 15】図 5 のステップ S 5 9 を示すフローチャートである。

【図 16 A】図 15 のステップ S 5 9 2 を示すフローチャートである。

【図 16 B】図 15 のステップ S 5 9 2 を示すフローチャートである。

【図 16 C】図 15 のステップ S 5 9 2 を示すフローチャートである。

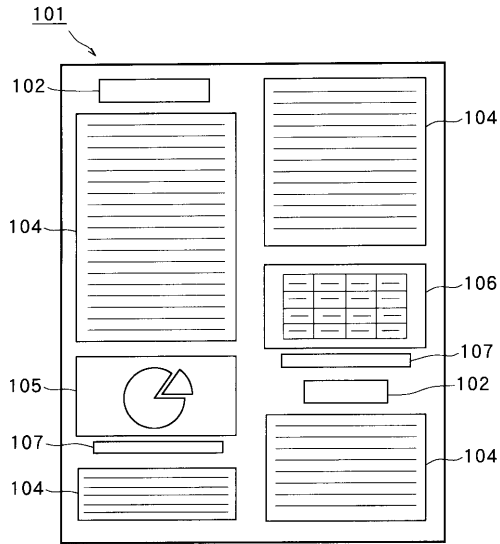
【図 16 D】図 15 のステップ S 5 9 2 を示すフローチャートである。

【図 16 E】図 15 のステップ S 5 9 2 を示すフローチャートである。

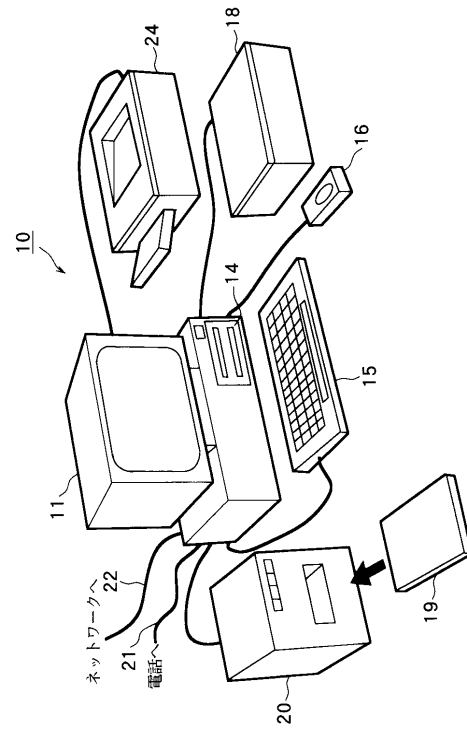
30

【図 17】不規則な形状のテキストブロックを概略的に示す図である。

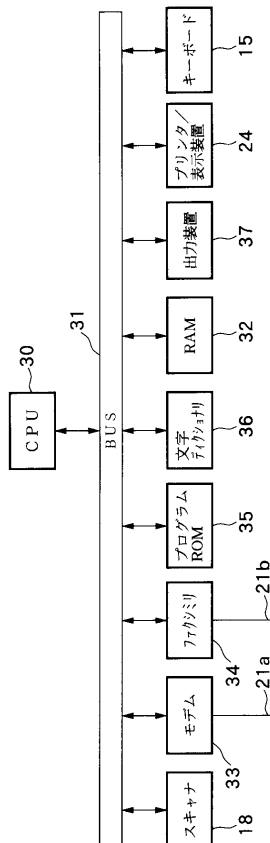
【図 1】



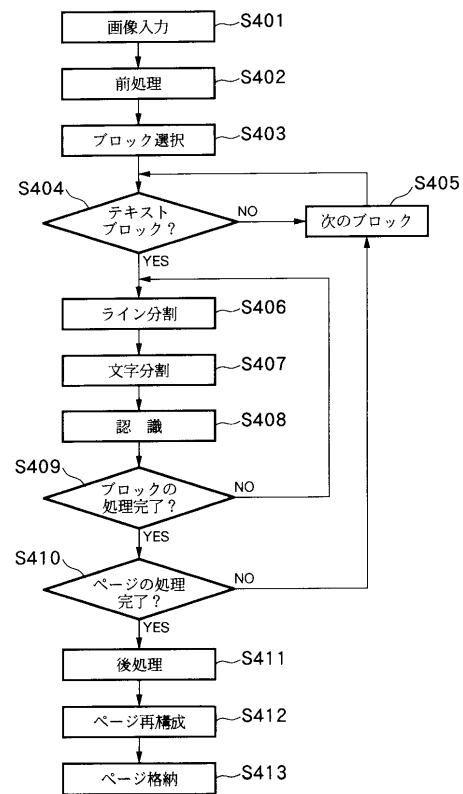
【図 2】



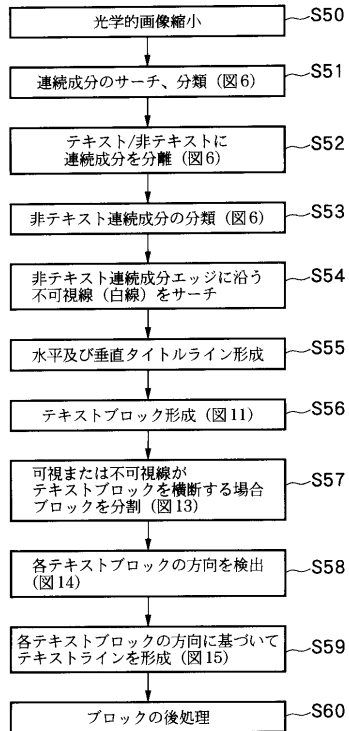
【図 3】



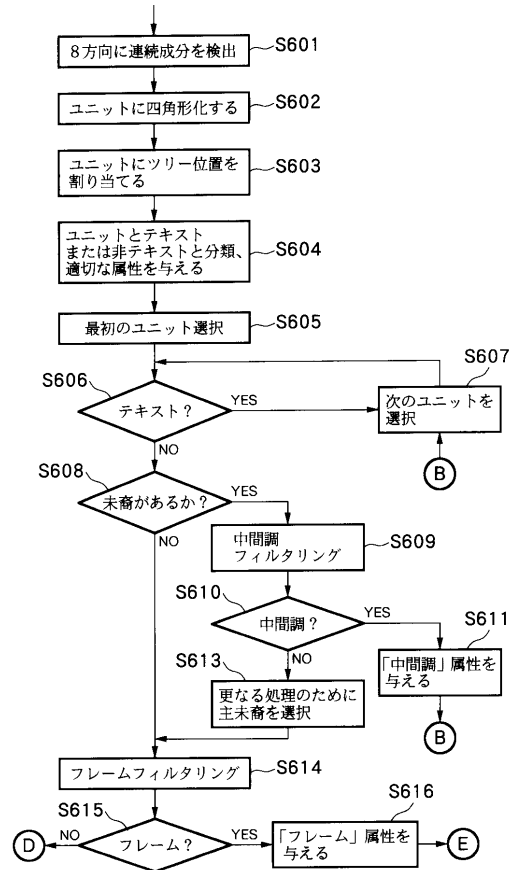
【図 4】



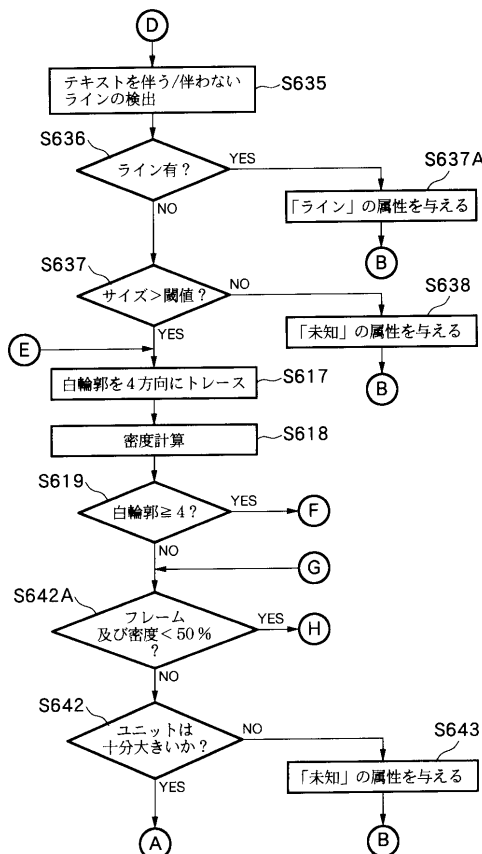
【図 5】



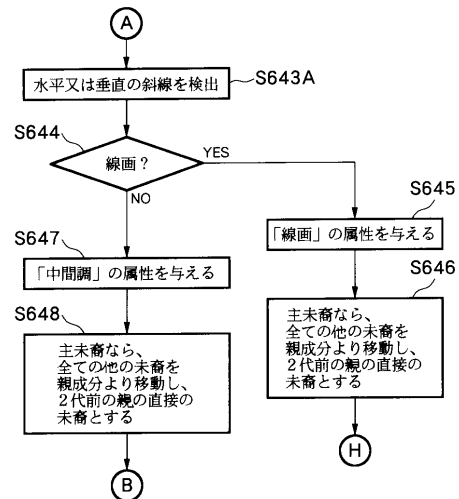
【図 6 A】



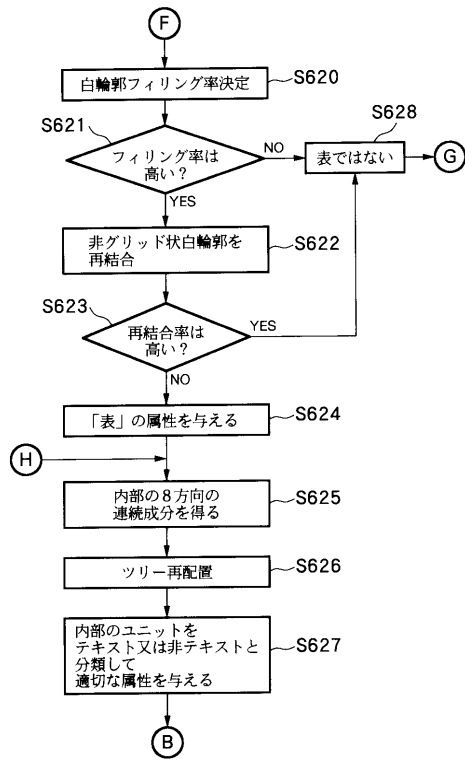
【図 6 B】



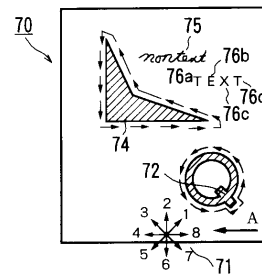
【図 6 C】



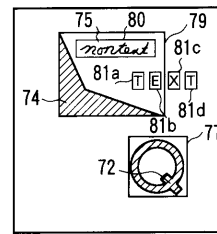
【図 6 D】



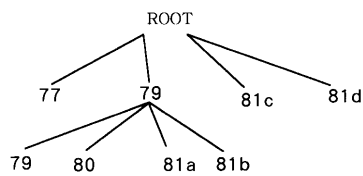
【図 7 A】



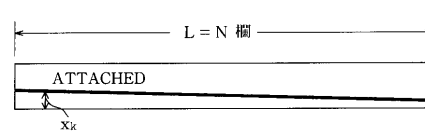
【図 7 B】



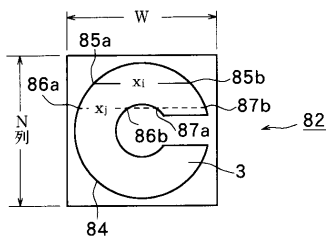
【図 7 C】



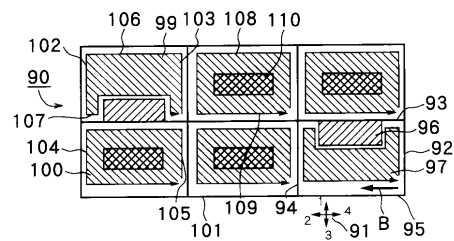
【図 8 C】



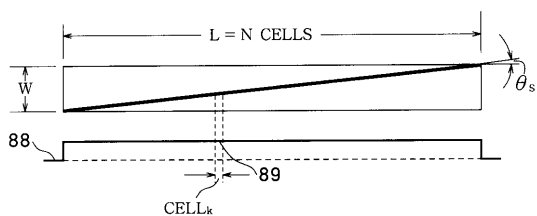
【図 8 A】



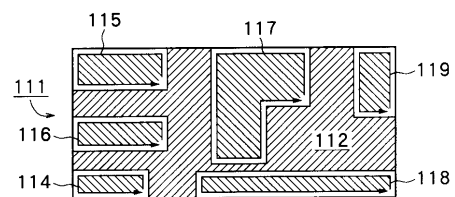
【図 9 A】



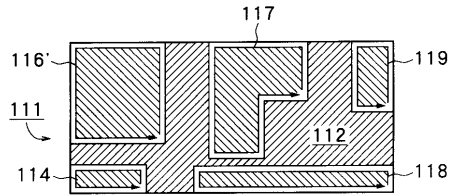
【図 8 B】



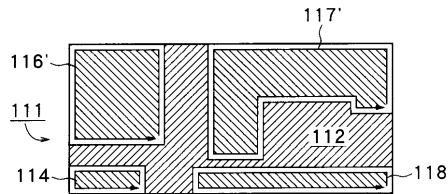
【図 9 B】



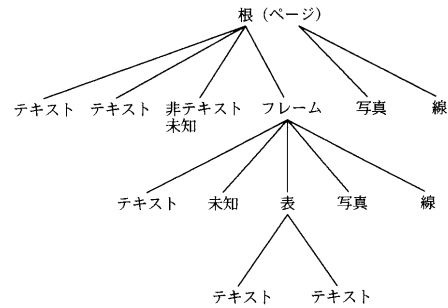
【図 9 C】



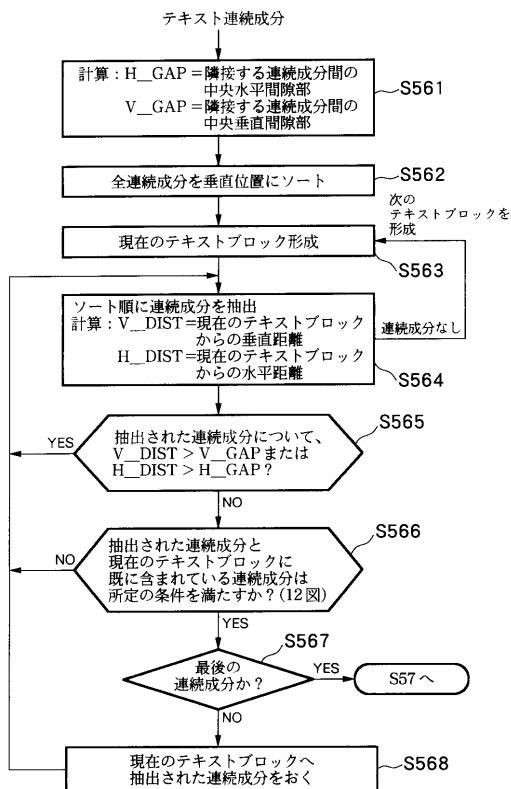
【図 9 D】



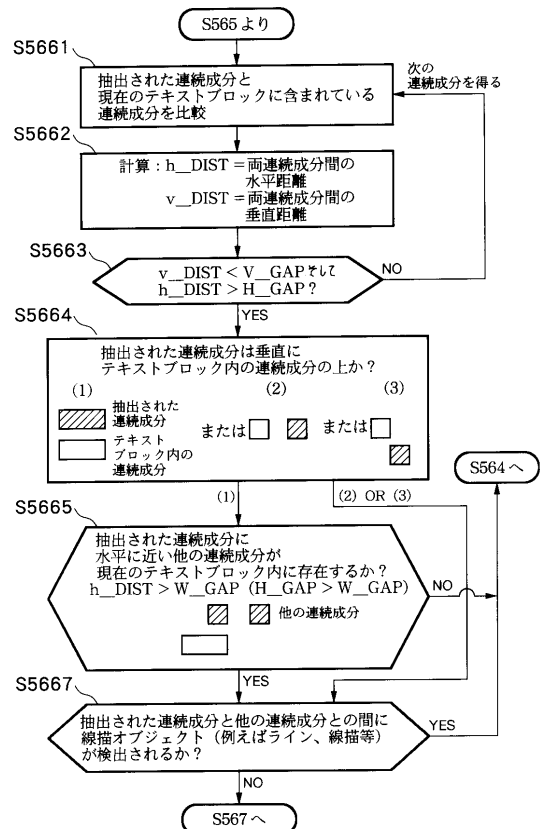
【図 10】



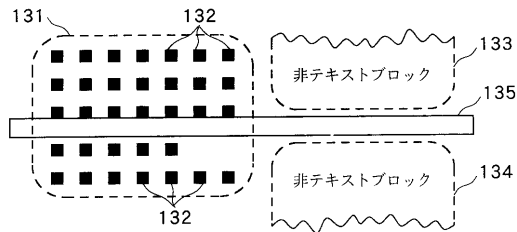
【図 11】



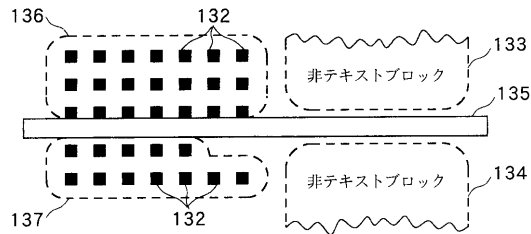
【図 12】



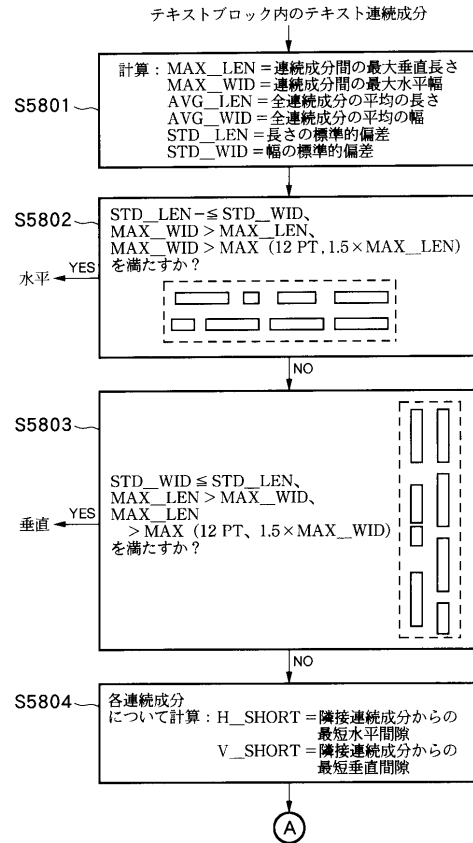
【図 13 A】



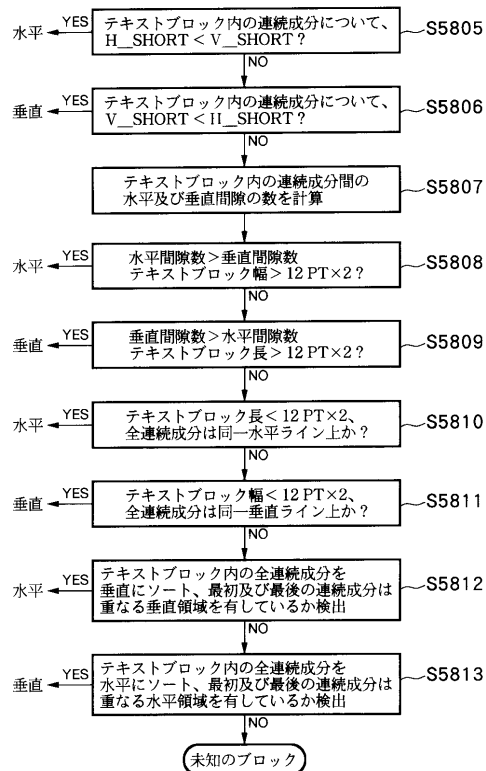
【図 13 B】



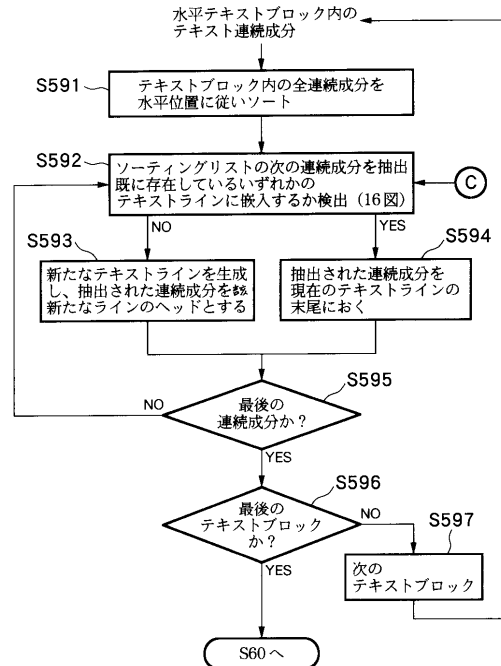
【図 14 A】



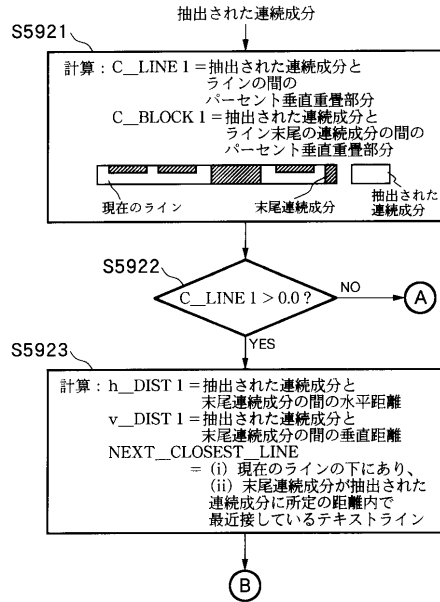
【図 14 B】



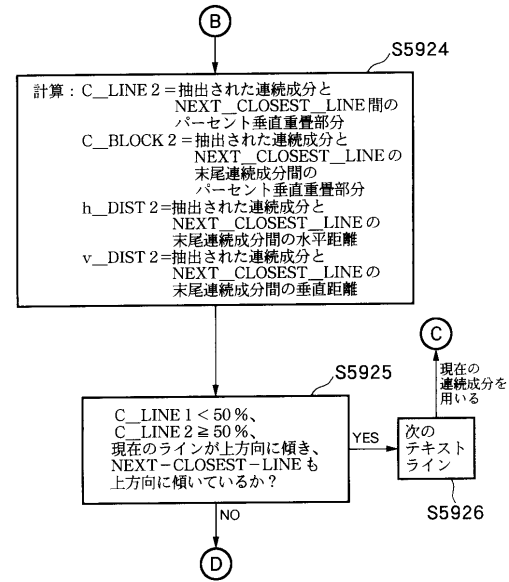
【図 15】



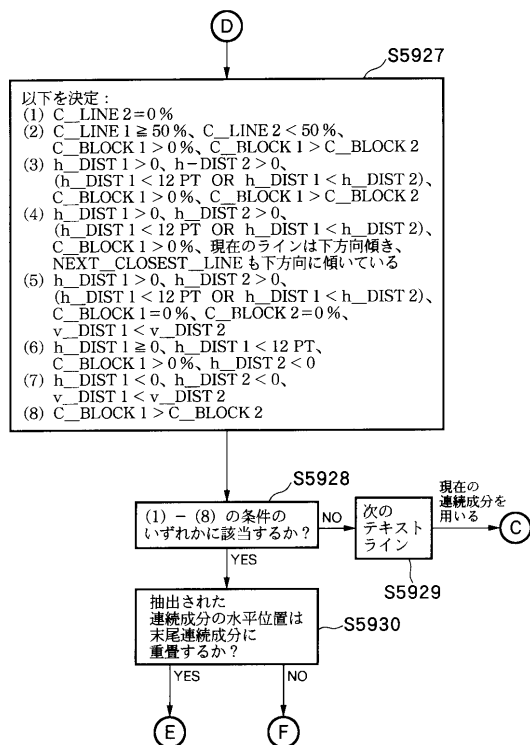
【図 16 A】



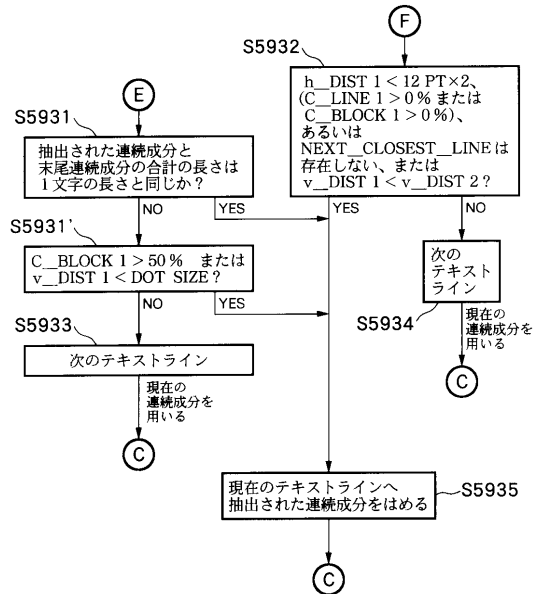
【図 16 B】



【図 16 C】

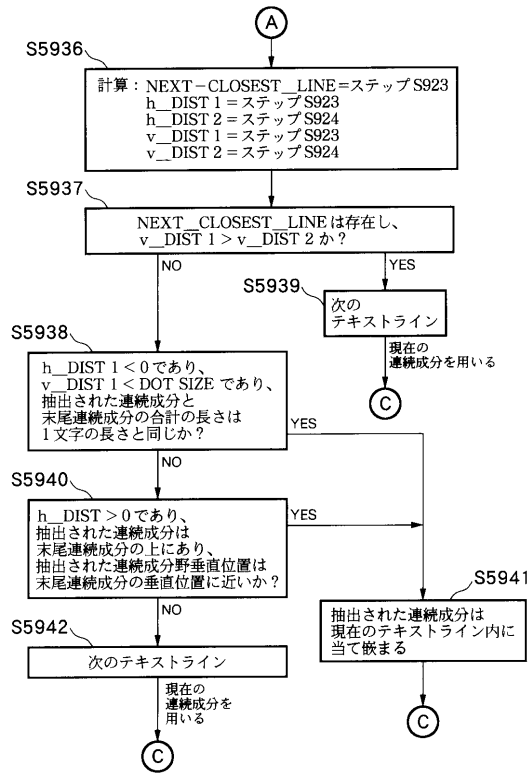


【図 16 D】

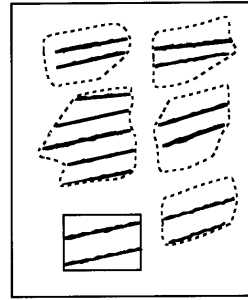




【図 16 E】



【図 17】



---

フロントページの続き

審査官 新井 則和

(56)参考文献 欧州特許出願公開第00567344(E P , A 1 )  
特開平06 - 068301(J P , A )  
特開昭57 - 52971(J P , A )

(58)調査した分野(Int.Cl. , D B名)  
G06K 9/00~9/76