

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4405831号  
(P4405831)

(45) 発行日 平成22年1月27日(2010.1.27)

(24) 登録日 平成21年11月13日(2009.11.13)

(51) Int.Cl.		F I
<b>G06T 7/00</b>	<b>(2006.01)</b>	G06T 7/00 300F
<b>G06F 17/30</b>	<b>(2006.01)</b>	G06F 17/30 170B
<b>G06T 1/00</b>	<b>(2006.01)</b>	G06F 17/30 310C
		G06T 1/00 200A

請求項の数 18 (全 33 頁)

(21) 出願番号	特願2004-69487(P2004-69487)	(73) 特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成16年3月11日(2004.3.11)	(74) 代理人	100076428 弁理士 大塚 康德
(65) 公開番号	特開2005-4724(P2005-4724A)	(74) 代理人	100112508 弁理士 高柳 司郎
(43) 公開日	平成17年1月6日(2005.1.6)	(74) 代理人	100115071 弁理士 大塚 康弘
審査請求日	平成18年12月7日(2006.12.7)	(74) 代理人	100116894 弁理士 木村 秀二
(31) 優先権主張番号	特願2003-142408(P2003-142408)	(72) 発明者	東條 洋 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
(32) 優先日	平成15年5月20日(2003.5.20)		
(33) 優先権主張国	日本国(JP)		

最終頁に続く

(54) 【発明の名称】 画像処理装置及びその制御方法、プログラム

(57) 【特許請求の範囲】

【請求項1】

複数の電子データを比較先画像として記憶する記憶手段と、  
印刷物を電子的に読み取り、その印刷物の電子データを比較元画像として入力する入力手段と、

前記比較元画像に含まれるページ画像の数に基づいて、前記比較元画像に含まれるページ画像を含む処理対象領域の特徴量を抽出する抽出手段と、

前記抽出手段で抽出した特徴量を用いて、前記比較元画像に対応する比較先画像を前記記憶手段から検索する検索手段と

を備えることを特徴とする画像処理装置。

10

【請求項2】

前記検索手段の検索結果である画像に対して処理を行う処理手段と

を更に備えることを特徴とする請求項1に記載の画像処理装置。

【請求項3】

前記比較元画像に含まれるページ画像のページ数を指定する指定手段を更に備え、

前記抽出手段は、前記指定手段で指定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出する

ことを特徴とする請求項1または2に記載の画像処理装置。

【請求項4】

前記抽出手段は、前記処理対象領域の画像特徴量及び文字特徴量のいずれかまたは両方

20

を抽出する

ことを特徴とする請求項 1 に記載の画像処理装置。

【請求項 5】

前記画像特徴量は、色特徴量、輝度特徴量、テクスチャ特徴量、形状特徴量のいずれか 1 つ、或いは、それらの組み合わせである

ことを特徴とする請求項 4 に記載の画像処理装置。

【請求項 6】

前記文字特徴量は、文字コードである

ことを特徴とする請求項 4 に記載の画像処理装置。

【請求項 7】

前記検索手段による検索結果を表示する表示手段と、

前記検索手段による検索結果として、前記表示手段に複数の比較先画像が表示された場合、該複数の比較先画像から所望の画像を選択する選択手段と

を更に備えることを特徴とする請求項 1 または 2 に記載の画像処理装置。

【請求項 8】

前記比較元画像が複数のページ画像から構成されている場合、該複数のページ画像から所望のページ画像を指定する指定手段を更に備え、

前記抽出手段は、前記指定手段で指定されたページ画像を含む処理対象領域の特徴量を抽出する

ことを特徴とする請求項 1 または 2 に記載の画像処理装置。

【請求項 9】

前記処理手段は、前記検索手段で検索された画像を、指定された印刷条件で印刷する

ことを特徴とする請求項 2 に記載の画像処理装置。

【請求項 10】

前記検索結果の画像に対する処理の種類を指示する指示手段を更に備え、

前記処理の種類は、印刷、配信、蓄積、編集のいずれかである

ことを特徴とする請求項 2 に記載の画像処理装置。

【請求項 11】

前記比較元画像に複数のページ画像が存在するか否かを判定する第 1 判定手段と、

前記第 1 判定手段による判定の結果、前記比較元画像に複数のページ画像が存在する場合、前記比較元画像に含まれるページ画像とのページ数を判定する第 2 判定手段とを更に備え、

前記抽出手段は、前記第 2 判定手段で判定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出する

ことを特徴とする請求項 1 または 2 に記載の画像処理装置。

【請求項 12】

前記第 1 判定手段は、前記比較元画像中に含まれるページ番号を検出し、前記ページ番号が特定領域内で等間隔に存在しているか否かを判定することで、前記比較元画像に複数のページ画像が存在するか否かを判定する

ことを特徴とする請求項 11 に記載の画像処理装置。

【請求項 13】

前記第 1 判定手段は、前記比較元画像の最右端余白または最左端余白と、それ以外の中間余白を検出し、検出した中間余白が、前記最右端余白または前記最左端余白よりも大きい否かを判定することで、前記比較元画像に複数のページ画像が存在するか否かを判定する

ことを特徴とする請求項 11 に記載の画像処理装置。

【請求項 14】

前記第 1 判定手段の判定結果を表示する表示手段と、

前記判定結果を修正する修正手段と

を更に備えることを特徴とする請求項 11 に記載の画像処理装置。

10

20

30

40

50

**【請求項 15】**

前記比較先画像を入力する入力手段と、  
前記入力手段で入力される比較先画像に含まれるページ画像のページ数を指定する指定手段とを更に備え、  
前記抽出手段は、前記指定手段で指定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出し、  
前記記憶手段は、前記比較先画像と、前記抽出手段で抽出した特徴量とを対応づけて記憶することを特徴とする請求項 1 に記載の画像処理装置。

10

**【請求項 16】**

前記比較先画像を入力する入力手段と、  
前記比較先画像に複数のページ画像が存在するか否かを判定する第 1 判定手段と、  
前記第 1 判定手段による判定の結果、前記比較元画像に複数のページ画像が存在する場合、前記比較元画像に含まれるページ画像とのページ数を判定する第 2 判定手段とを更に備え、  
前記抽出手段は、前記第 2 判定手段で判定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出し、  
前記記憶手段は、前記比較先画像と、前記抽出手段で抽出した特徴量とを対応づけて記憶することを特徴とする請求項 1 に記載の画像処理装置。

20

**【請求項 17】**

印刷物を電子的に読み取ることによって得た電子データを比較元画像として入力する入力工程と、  
前記比較元画像に含まれるページ画像の数に基づいて、前記比較元画像に含まれるページ画像を含む処理対象領域の特徴量を抽出する抽出工程と、  
前記抽出工程で抽出した特徴量を用いて、記憶媒体に比較先画像として記憶されている複数の電子データから、前記比較元画像に対応する比較先画像を検索する検索工程と、  
を備えることを特徴とする画像処理装置の制御方法。

**【請求項 18】**

画像処理の制御をコンピュータで実行させるためのコンピュータプログラムであって、印刷物を電子的に読み取ることによって得た電子データを比較元画像として入力する入力工程と、  
前記比較元画像に含まれるページ画像の数に基づいて、前記比較元画像に含まれるページ画像を含む処理対象領域の特徴量を抽出する抽出工程と、  
前記抽出工程で抽出した特徴量を用いて、記憶媒体に比較先画像として記憶されている複数の電子データから、前記比較元画像に対応する比較先画像を検索する検索工程と、  
を備えることを特徴とするコンピュータプログラム。

30

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、複写機などの画像入力装置で、読み取った紙文書から対応するオリジナルの電子データを検索し、オリジナルの電子データを印刷、配信、蓄積、編集などに活用することを可能とする画像処理技術に関するものである。

40

**【背景技術】****【0002】**

近年、複写機のデジタル化及びネットワーク化が進んでいる。印刷機能も多様化しており、同一の、或いは、複数の文書に含まれる複数のページを、1枚の紙に配置して印刷することが可能となっている。

**【0003】**

例えば、会議などで多用されている配布資料などは参照用であるため、1枚の紙に1文

50

書の2ページ分、4ページ分といったように、Nページ分の文書を1枚の紙にレイアウトして印刷されたものであることが多い(以降、Nページ印刷と呼ぶ)。また、複数の文書のページを一度に参照する必要があるときは、結合させて印刷する場合もあり得る(以降、結合印刷と呼ぶ)。更に、印刷後に2つ折にして製本できるように、ページ番号が製本時の順に印刷される場合もあり得る(以降、製本印刷と呼ぶ)。

【0004】

尚、以降、Nページ印刷、結合印刷、製本印刷等の1枚の紙にNページ分の文書をレイアウトして印刷する印刷を総称してNup印刷と呼ぶことにする。

【0005】

一方、紙文書をスキャナで読み取り、その読み取った紙文書画像のオリジナルの電子データを検索し、オリジナルの電子データそのものから印刷したり、ネットワーク内の所望のPC(パーソナルコンピュータ)等の端末へ配信したり、ネットワーク内の所望の蓄積媒体へ保存したり、または編集したりすることを可能とするシステムが提案されはじめて

10

【0006】

これを実現するための方法として、例えば、特許文献1では、紙文書をスキャンしたスキャン電子データと、記憶媒体中の電子データをビット単位で比較することにより、スキャン電子データのオリジナルの電子データを検索することを可能としている。

【0007】

また、例えば、特許文献2では、書類を電子データにする際に、その電子データに識別コードを付与して、その識別コードを含む印刷物を生成する。これにより、別途、印刷物に対応する電子データを検索したり印刷したりする場合には、その印刷物を読み込み、印刷されている識別コードを認識することで、所望の電子データの検索や印刷が可能となっている。

20

【特許文献1】特許第3017851号

【特許文献2】特開2001-257862号

【発明の開示】

【発明が解決しようとする課題】

【0008】

しかしながら、特許文献1、2では、Nup印刷の文書については何ら考慮していないので、Nup印刷が使用されるような環境においては、次のような問題が生じる。

30

【0009】

例えば、特許文献1では、Nup印刷された文書をスキャンして得た電子データを1ページ分と解釈する。これに対し、オリジナルの電子データは、1ページを単位とした電子データで記憶媒体に管理されているので、どの電子データと比較しても高い類似度が得られない。よって、オリジナルの電子データを検索できなくなる。

【0010】

また、例えば、特許文献2では、識別コードは各ページに対して印字位置が決まっているので、Nup印刷した際には印字位置がずれてしまい、各識別コードが読み取れなくなってしまうので、やはりオリジナルの電子データを検索できなくなってしまう。

40

【0011】

本発明は上記の課題に鑑みてなされたものであり、Nup印刷された紙文書から、オリジナルの電子データを検索し、活用することを可能とする画像処理装置及びその制御方法、プログラムを提供することを目的とする。

【課題を解決するための手段】

【0012】

上記の目的を達成するための本発明による画像処理装置は以下の構成を備える。即ち、複数の電子データを比較先画像として記憶する記憶手段と、印刷物を電子的に読み取り、その印刷物の電子データを比較元画像として入力する入力手段と、

50

前記比較元画像に含まれるページ画像の数に基づいて、前記比較元画像に含まれるページ画像を含む処理対象領域の特徴量を抽出する抽出手段と、

前記抽出手段で抽出した特徴量を用いて、前記比較元画像に対応する比較先画像を前記記憶手段から検索する検索手段と、  
を備える。

【0013】

また、好ましくは、前記検索手段の検索結果である画像に対して処理を行う処理手段と更に備える。

【0014】

また、好ましくは、前記比較元画像に含まれるページ画像のページ数を指定する指定手段を更に備え、

前記抽出手段は、前記指定手段で指定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出する。

【0015】

また、好ましくは、前記抽出手段は、前記処理対象領域の画像特徴量及び文字特徴量のいずれかまたは両方を抽出する。

【0016】

また、好ましくは、前記画像特徴量は、色特徴量、輝度特徴量、テクスチャ特徴量、形状特徴量のいずれか1つ、或いは、任意の組み合わせである。

【0017】

また、好ましくは、前記文字特徴量は、文字コードである。

【0018】

また、好ましくは、前記検索手段による検索結果を表示する表示手段と、  
前記検索手段による検索結果として、前記表示手段に複数の比較先画像が表示された場合、該複数の比較先画像から所望の画像を選択する選択手段と  
を更に備える。

【0019】

また、好ましくは、前記比較元画像が複数のページ画像から構成されている場合、該複数のページ画像から所望のページ画像を指定する指定手段を更に備え、

前記抽出手段は、前記指定手段で指定されたページ画像を含む処理対象領域の特徴量を抽出する。

【0020】

また、好ましくは、前記処理手段は、前記検索手段で検索された画像を、指定された印刷条件で印刷する。

【0021】

また、好ましくは、前記検索結果の画像に対する処理の種類を指示する指示手段を更に備え、

前記処理の種類は、印刷、配信、蓄積、編集のいずれかである。

【0022】

また、好ましくは、前記比較元画像に複数のページ画像が存在するか否かを判定する第1判定手段と、

前記第1判定手段による判定の結果、前記比較元画像に複数のページ画像が存在する場合、前記比較元画像に含まれるページ画像とのページ数を判定する第2判定手段と、

前記抽出手段は、前記第2判定手段で判定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出する。

【0023】

また、好ましくは、前記第1判定手段は、前記比較元画像中に含まれるページ番号を検出し、前記ページ番号が所定の領域内で等間隔に存在しているか否かを判定することで、前記比較元画像に複数のページ画像が存在するか否かを判定する。

【0024】

10

20

30

40

50

また、好ましくは、前記第1判定手段は、前記比較元画像の最右端余白または最左端余白と、それ以外の中間余白を検出し、検出した中間余白が、前記最右端余白または前記最左端余白よりも大きいかなかを判定することで、前記比較元画像に複数のページ画像が存在するか否かを判定する。

【0025】

また、好ましくは、前記第1判定手段の判定結果を表示する表示手段と、前記判定結果を修正する修正手段とを更に備える。

【0026】

また、好ましくは、前記比較先画像を入力する入力手段と、前記入力手段で入力される比較先画像に含まれるページ画像のページ数を指定する指定手段とを更に備え、

前記抽出手段は、前記指定手段で指定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出し、

前記記憶手段は、前記比較先画像と、前記抽出手段で抽出した特徴量とを対応づけて記憶する。

【0027】

また、好ましくは、前記比較先画像を入力する入力手段と、

前記比較先画像に複数のページ画像が存在するか否かを判定する第1判定手段と、

前記第1判定手段による判定の結果、前記比較元画像に複数のページ画像が存在する場合、前記比較元画像に含まれるページ画像とのページ数を判定する第2判定手段とを更に備え、

前記抽出手段は、前記第2判定手段で判定されたページ数に基づいて決定される処理対象領域毎に、その処理対象領域の特徴量を抽出し、

前記記憶手段は、前記比較先画像と、前記抽出手段で抽出した特徴量とを対応づけて記憶する。

【0028】

上記の目的を達成するための本発明による画像処理装置の制御方法は以下の構成を備える。即ち、

印刷物を電子的に読み取ることによって得た電子データを比較元画像として入力する入力工程と、

前記比較元画像に含まれるページ画像の数に基づいて、前記比較元画像に含まれるページ画像を含む処理対象領域の特徴量を抽出する抽出工程と、

前記抽出工程で抽出した特徴量を用いて、記憶媒体に比較先画像として記憶されている複数の電子データから、前記比較元画像に対応する比較先画像を検索する検索工程とを備える。

【0029】

上記の目的を達成するための本発明によるプログラムは以下の構成を備える。即ち、

画像処理の制御をコンピュータで実現するプログラムであって、

印刷物を電子的に読み取ることによって得た電子データを比較元画像として入力する入力工程のプログラムコードと、

前記比較元画像に含まれるページ画像の数に基づいて、前記比較元画像に含まれるページ画像を含む処理対象領域の特徴量を抽出する抽出工程のプログラムコードと、

前記抽出工程で抽出した特徴量を用いて、記憶媒体に比較先画像として記憶されている複数の電子データから、前記比較元画像に対応する比較先画像を検索する検索工程のプログラムコードと、

を備える。

【発明の効果】

【0030】

以上説明したように、本発明によれば、Nup印刷された紙文書から、オリジナルの電

10

20

30

40

50

子データを検索し、活用することを可能とする画像処理装置及びその制御方法、プログラムを提供できる。

【発明を実施するための最良の形態】

【0031】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

【0032】

<<実施形態1>>

図1は本発明の実施形態1の画像処理システムの構成を示すブロック図である。

【0033】

この画像処理システムは、オフィス10とオフィス20とをインターネット等のネットワーク104で接続された環境で実現する。

【0034】

オフィス10内に構築されたLAN107には、複数種類の機能を実現する複合機であるMFP(Multi Function Peripheral)100、MFP100を制御するマネジメントPC101、クライアントPC102、文書管理サーバ106及びそのデータベース105、及びプロキシサーバ103が接続されている。

【0035】

オフィス10内のLAN107及びオフィス20内のLAN108は、双方のオフィスのプロキシサーバ103を介してネットワーク104に接続されている。

【0036】

MFP100は、特に、紙文書を電子的に読み取る画像読取部と、画像読取部から得られる画像信号に対する画像処理を実行する画像処理部を有し、この画像信号はLAN109を介してマネジメントPC101に送信することができる。

【0037】

マネジメントPC101は、通常のPCであり、内部に画像記憶部、画像処理部、表示部、入力部等の各種構成要素を有するが、その構成要素の一部はMFP100に一体化して構成されている。

【0038】

尚、ネットワーク104は、典型的にはインターネットやLANやWANや電話回線、専用デジタル回線、ATMやフレームリレー回線、通信衛星回線、ケーブルテレビ回線、データ放送用無線回線等のいずれか、またはこれらの組み合わせにより実現されるいわゆる通信ネットワークであり、データの送受信が可能であれば良い。

【0039】

また、マネジメントPC101、クライアントPC102、文書管理サーバ等の各種端末はそれぞれ、汎用コンピュータに搭載される標準的な構成要素(例えば、CPU、RAM、ROM、ハードディスク、外部記憶装置、ネットワークインタフェース、ディスプレイ、キーボード、マウス等)を有している。

【0040】

次に、MFP100の詳細構成について、図2を用いて説明する。

【0041】

図2は本発明の実施形態1のMFPの詳細構成を示すブロック図である。

【0042】

図2において、原稿台とオートドキュメントフィーダ(ADF)を含む画像読取部110は、束状のあるいは1枚の原稿画像を光源(不図示)で照射し、原稿反射像をレンズで固体撮像素子上に結像し、固体撮像素子からラスタ状の画像読取信号を所定密度(例えば、600DPI)のラスタ画像として得る。

【0043】

また、MFP100は、画像読取信号に対応する画像を印刷部112で記録媒体に印刷する複写機能を有し、原稿画像を1つ複写する場合には、この画像読取信号をデータ処理部115で画像処理して記録信号を生成し、これを印刷部112によって記録媒体上に印

10

20

30

40

50

刷させる。一方、原稿画像を複数複写する場合には、記憶部 1 1 1 に一旦一つ分の記録信号を記憶保持させた後、これを印刷部 1 1 2 に順次出力して記録媒体上に印刷させる。

【 0 0 4 4 】

また、N u p 印刷を行う場合は、データ処理部 1 1 5 で指定されたページ分の画像信号を縮小し、1 枚の紙で印刷可能となるように各ページの画像信号をレイアウトする。その後、印刷部 1 1 2 に出力して記録媒体上に印刷させる。

【 0 0 4 5 】

一方、クライアント P C 1 0 2 から出力される記録信号は、L A N 1 0 7 及びネットワーク I F 1 1 4 を介してデータ処理部 1 1 5 が受信し、データ処理部 1 1 5 は、その記録信号を印刷部 1 1 2 で記録可能なラスターデータに変換した後、印刷部 1 1 2 によって記録媒体上に印刷させる。

10

【 0 0 4 6 】

M F P 1 0 0 への操作者の指示は、M F P 1 0 0 に装備されたキー操作部とマネジメント P C 1 0 1 に接続されたキーボード及びマウスからなる入力部 1 1 3 から行われ、これら一連の動作はデータ処理部 1 1 5 内の制御部（不図示）で制御される。また、操作入力の状態表示及び処理中の画像データの表示は、表示部 1 1 6 で行われる。

【 0 0 4 7 】

記憶部 1 1 1 は、マネジメント P C 1 0 1 から制御され、M F P 1 0 0 とマネジメント P C 1 0 1 とのデータの送受信及び制御は、ネットワーク I F 1 1 7 及び L A N 1 0 9 を介して行われる。

20

【 0 0 4 8 】

尚、M F P 1 0 0 では、後述する各種処理を実行するための各種操作・表示をユーザに提供するユーザインタフェースを、表示部 1 1 6 及び入力部 1 1 3 によって実現している。

【 0 0 4 9 】

本発明による画像処理システムで実行する処理としては、大きく分けて画像データを登録する登録処理と、所望の画像データを検索する検索処理の 2 つがある。

【 0 0 5 0 】

尚、実施形態 1 では、画像処理システム全体の処理効率を向上するために、以下に説明する各種処理を、画像処理システムを構成する各種端末に分散させて実行するようにしているが、1 つの端末（例えば、M F P 1 0 0 ）上で実行するようにしても構わない。

30

【 0 0 5 1 】

まず、登録処理について説明する。

【 0 0 5 2 】

[ 登録処理の概要 ]

登録対象の画像データの登録方法としては、紙文書をスキャン入力して生成した画像データを登録する場合と、文書作成アプリケーション等で作成された電子文書をラスター画像に展開した画像データを登録する場合がある。

【 0 0 5 3 】

そこで、オリジナル文書を登録する登録処理の概要について、図 3 A を用いて説明する。

40

【 0 0 5 4 】

図 3 A は本発明の実施形態 1 の登録処理を示すフローチャートである。

【 0 0 5 5 】

尚、この処理は、画像読取部 1 1 0 の A D F に、登録対象の紙文書がセットされ、入力部 1 1 3 の登録ボタンが操作された時点で開始される。また、登録対象の紙文書は、1 枚でも複数枚でも可能であるが、複数枚の場合は、その紙文書から得られる画像データ群（ページ画像群）を 1 つのファイルとして管理することになる。

【 0 0 5 6 】

まず、ステップ S 3 0 1 0 で、登録対象のオリジナル文書を入力する。また、この入力

50



に伴って、オリジナル文書を管理するための各種情報を生成して記憶部 1 1 1 に記憶する。

【 0 0 5 7 】

尚、オリジナル文書を登録する際のオリジナル文書の入力方法には、2種類存在する。

【 0 0 5 8 】

オリジナル文書が電子データである場合は、クライアント P C 1 0 2 内のハードディスク内、あるいはオフィス 1 0 や 2 0 内の文書管理サーバ 1 0 6 内のデータベース 1 0 5 内、あるいは M F P 1 0 0 の記憶部 1 1 1 のいずれかに格納されており、これらの記憶元から登録対象のオリジナル文書の電子データを読み出してネットワーク I F 1 1 4 を介してデータ処理部 1 1 5 に入力し、データ処理部 1 1 5 でその電子データをラスタ画像に変換する。

10

【 0 0 5 9 】

一方、オリジナル文書が紙文書である場合は、M F P 1 0 0 の画像読取部 1 1 0 で、その紙文書をラスタ状に走査しラスタ画像を得る。

【 0 0 6 0 】

このように、実施形態 1 では、登録対象のオリジナル文書に、電子データあるいは紙文書のどちらも扱うことが可能である。その後、ラスタ画像をデータ処理部 1 1 5 で前処理を施し記憶部 1 1 1 に保存する（尚、これ以降、紙文書の場合はラスタ画像がオリジナル文書の電子データとなる）。このとき、登録対象のオリジナル文書毎に固有の文書 I D を発行し、オリジナル文書の電子データのアドレスと対応付けて記憶部 1 1 1 にアドレス情報として保存する。

20

【 0 0 6 1 】

ここで、アドレスとは、U R L や、サーバ名とディレクトリ、ファイル名からなる電子データの格納先を示すフルパス情報である。また、アドレス情報の一例を示すと、図 4 のようになる。また、アドレス情報の格納先は、データベース 1 0 5 や記憶部 1 1 1 等が挙げられる。

【 0 0 6 2 】

尚、オリジナル文書が紙文書である場合の電子データのファイル形式は、例えば、B M P 形式とするが、これに限定されるものではなく、色情報を保存しておくことが可能なファイル形式（例えば、G I F、J P E G）であればどのようなものでも良い。

30

【 0 0 6 3 】

一方、オリジナル文書が電子データである場合のその電子データのファイル形式は、その電子データを作成したアプリケーション（例えば、マイクロソフト（登録商標）社の M S - W o r d や、アドビシステム（登録商標）のアクロバット等）で作成されたファイル形式（\* . d o c や \* . p d f）となる。

【 0 0 6 4 】

次に、ステップ S 3 0 1 1 で、ステップ S 3 0 1 0 で入力された文書の数を入力文書数 P に設定する。次に、ステップ S 3 0 1 2 で、処理文書数を示す変数 a と、処理対象の 1 文書中の処理ページ数を示す変数 b にそれぞれ初期値として 1 を設定する。次に、ステップ S 3 0 1 3 で、1 文書のページ数 Q に a 番目の文書中のページ数を設定する。

40

【 0 0 6 5 】

次に、終了判定として、P が a 以上であるか否かを判定する。a 未満である場合（ステップ S 3 0 1 4 で N O）、処理を終了する。一方、a 以上である場合（ステップ S 3 0 1 4 で Y E S）、未処理の文書が存在するので、処理を続行する。

【 0 0 6 6 】

次に、ステップ S 3 0 1 5 で、文書単位の終了判定として、Q が b 以上であるか否かを判定する。b 未満である場合（ステップ S 3 0 1 5 で N O）、ステップ S 3 0 3 2 へ進み、a を 1 インクリメントして、ステップ S 3 0 1 4 へ戻り、処理対象を次の文書に移す。一方、b 以上である場合（ステップ S 3 0 1 5 で Y E S）、未処理のページが存在するので、処理を続行する。

50

## 【 0 0 6 7 】

次に、P番目の文書のQページに関して、ブロックセレクション（BS）処理を行う。この処理は、マネージメントPC101の制御によって実行する。

## 【 0 0 6 8 】

具体的には、マネージメントPC101のCPUは、記憶部111に格納された処理対象のオリジナル文書のラスト画像を、まず、文字／線画部分とハーフトーン画像部分とに領域分割し、文字／線画部分は更に段落で塊として纏まっているブロック毎に、あるいは線で構成された表、図形毎に分割する。

## 【 0 0 6 9 】

一方、ハーフトーン画像部分は、矩形に分離されたブロックの画像部分、背景部分等のブロックに分割する。

10

## 【 0 0 7 0 】

そして、処理対象のページのページ番号、そのページ中の各ブロックを特定するブロックIDを発行し、各ブロックの属性（画像、文字等）、サイズやオリジナル文書内の位置（ページ内の座標）と各ブロックを関連付けて記憶部111にブロック情報として記憶する。このブロック情報の一例を示すと、図5のようになる。

## 【 0 0 7 1 】

次に、ステップS3030で、データ処理部115において、各ブロックの種別に応じて、各ブロックの特徴量情報を抽出する特徴量情報抽出処理を行う。

## 【 0 0 7 2 】

特に、文字ブロックについては、OCR処理を施して文字コードを抽出し、これを文字特徴量とする。また、画像ブロックについては、色に関する画像特徴量を抽出する。このとき、それぞれのブロックに対応する特徴量をオリジナル文書単位にまとめ、文書ID、ページ番号、ブロックIDに関連付けて記憶部111に特徴量情報として記憶する。この特徴量情報の一例を示すと、図6及び図7のようになる。

20

## 【 0 0 7 3 】

次に、ステップS3031で、bを1インクリメントし、ステップS3014へ戻り、処理対象の次のページに移す。

## 【 0 0 7 4 】

## 〔 検索処理の概要 〕

オリジナル文書の電子データを検索する検索処理の概要について、図3Bを用いて説明する。

30

## 【 0 0 7 5 】

図3Bは本発明の実施形態1の検索処理を示すフローチャートである。

## 【 0 0 7 6 】

まず、ステップS3110で、検索条件となる紙文書の入力を行う。この処理は、ステップS3010の処理と同様であるので説明は省略する。但し、この処理によって生成するラスト画像は一時保存するだけであり、その紙文書に対するアドレス情報を記憶部111に保存しておく必要はない。

## 【 0 0 7 7 】

次に、ステップS3111で、ユーザによるNup印刷の指定の有無を判定する。Nup印刷の指定がある場合（ステップS3111でYES）、ステップS3112に進み、ページ数Lにユーザが指定した1枚中のページ数を設定する。一方、Nup印刷の指定がない場合（ステップS3111でNO）、ステップS3115に進み、1枚中1ページとみなし、ページ数Lに1を設定する。

40

## 【 0 0 7 8 】

次に、ステップS3113で、処理対象の1文書中の処理ページ数を示す変数bに初期値として1を設定する。次に、比較処理の終了判定として、Lがb以上であるか否かを判定する。b未満である場合（ステップS3116でNO）、ステップS3150に進む。一方、b以上である場合（ステップS3116でYES）、未処理のページが存在するの

50

で、処理を続行するために、ステップS 3 1 2 0に進む。

【0079】

次に、ステップS 3 1 2 0で、ラスト画像中のbページ目の画像領域に対して、ブロックセレクション(B S)処理を行う。

【0080】

この時のブロックセレクション処理の具体例については、図8を用いて説明する。

【0081】

図8では、2ページ分をN u p印刷した例を示している。図8において、8 1 0は1枚の紙全体の領域であり、8 1 1と8 1 2は各ページのページ画像を示している。8 1 3と8 1 4は各ページのページ番号である。ここで、b = 1である場合、1ページ目のページ画像8 1 1を含む処理対象領域8 1 5に対してのみ、ブロックセレクション処理を行うことになる。そして、図8の例では、b = 2の場合は、2ページ目のページ画像8 1 2を含む処理対象領域に対してブロックセレクション処理を行うことになる。

10

【0082】

尚、このブロックセレクション処理自体は、ステップS 3 0 2 0のブロックセレクション処理と同様であるので説明は省略する。但し、このブロックセレクション処理によって生成する各ブロックの属性、サイズ、位置は一時保存するだけであり、その紙文書に対するブロック情報は記憶部1 1 1に保存しておく必要はない。

【0083】

また、処理対象領域は、Lの値と、紙文書の向きに基づいて、1枚の紙文書をL個の領域に分割し、更に所定の処理順に基づいて決定される。

20

【0084】

次に、ステップS 3 1 3 0で、各ブロックの特徴量情報を抽出する特徴量情報抽出処理を行う。この処理は、ステップS 3 0 3 0の処理と同様であるので説明は省略する。但し、この処理によって生成する各ブロックの特徴量は一時保存するだけであり、その紙文書に対する特徴量情報を記憶部1 1 1に保存しておく必要はない。

【0085】

次に、ステップS 3 1 4 0で、入力した紙文書中のbページ目の画像(比較元画像)に対する特徴量情報と、クライアントP C 1 0 2内のハードディスク内、あるいはオフィス1 0や2 0内の文書管理サーバ1 0 6内のデータベース1 0 5内、あるいはM F P 1 0 0の記憶部1 1 1のいずれかに格納されている電子データ(比較先画像)の特徴量情報を比較して、その類似度を算出し、その類似度に基づいて、検索結果とするオリジナル文書候補を決定する。

30

【0086】

次に、ステップS 3 1 4 1で、bを1インクリメントし、ステップS 3 1 1 6へ戻り、処理対象を次のページに移す。

【0087】

そして、ステップS 3 1 1 6において、Lがb未満である場合(ステップS 3 1 1 6でN O)、ステップS 3 1 5 0に進み、ユーザ確認モードであるか否かを判定する。

【0088】

40

尚、ユーザ確認モードとは、比較処理によって得られたオリジナル文書候補の中から、ユーザが所望するオリジナル文書が検索されたか否かを確認するモードである。具体的には、オリジナル文書候補群を含むユーザインタフェースを表示部1 1 6・入力部1 1 3で実現し、このユーザインタフェースによって、オリジナル文書候補群の内容をユーザが確認することができるモードである。また、このユーザインタフェースの構成の詳細については後述する。

【0089】

ステップS 3 1 5 0において、ユーザ確認モードでない場合(ステップS 3 1 5 0でN O)、比較処理で最も類似していると判断された文書を自動的に選択して、ステップS 3 1 7 0に進む。一方、ユーザ確認モードである場合(ステップS 3 1 5 0でY E S)、ス

50

ステップ S 3 1 6 0 に進み、オリジナル文書候補の表示 / 選択を行う。特に、この選択は、オリジナル文書候補のサムネイル画像を表示部 1 1 6 に表示し、複数のオリジナル文書候補の中からユーザが所望のオリジナル文書候補のサムネイル画像を選択させることで実現する。

【 0 0 9 0 】

次に、ステップ S 3 1 7 0 では、選択されたオリジナル文書に対し、表示部 1 0 6 ・入力部 1 1 4 で実現されるユーザインタフェースを介するユーザからの操作に基いて、そのオリジナル文書の印刷、配信、蓄積、編集のいずれかの処理を実行する。

【 0 0 9 1 】

以上説明したように、N u p 印刷時は、1 文書に含まれる各ページ毎に、ブロックセレクション処理、特徴量情報抽出処理を行い、これらを用いて、各ページ毎に順に類似画像検索を実行することで、各ページに対応する電子データを全て検索し、活用することができる。

10

【 0 0 9 2 】

[ 各処理の詳細 ]

以下、各処理の詳細について説明する。

【 0 0 9 3 】

まず、ステップ S 3 0 2 0 及びステップ S 3 1 2 0 のブロックセレクション処理の詳細について説明する。

【 0 0 9 4 】

20

ブロックセレクション処理とは、例えば、図 9 ( a ) のラスト画像を、図 9 ( b ) のように、意味のあるブロック毎の塊として認識し、該ブロック各々の属性 ( 文字 ( T E X T ) / 図画 ( P I C T U R E ) / 写真 ( P H O T O ) / 線 ( L I N E ) / 表 ( T A B L E ) 等 ) を判定し、異なる属性を持つブロックに分割する処理である。

【 0 0 9 5 】

ブロックセレクション処理の実施形態を以下に説明する。

【 0 0 9 6 】

まず、入力画像を白黒に二値化し、輪郭線追跡を行って黒画素輪郭で囲まれる画素の塊を抽出する。面積の大きい黒画素の塊については、内部にある白画素に対しても輪郭線追跡を行って白画素の塊を抽出、さらに一定面積以上の白画素の塊の内部からは再帰的に黒画素の塊を抽出する。

30

【 0 0 9 7 】

このようにして得られた黒画素の塊を、大きさ及び形状で分類し、異なる属性を持つブロックへ分類していく。例えば、縦横比が 1 に近く、大きさが一定の範囲のブロックは文字相当の画素塊とし、さらに近接する文字が整列良くグループ化可能な部分を文字ブロック、扁平な画素塊を線ブロック、一定大きさ以上でかつ矩形の白画素塊を整列よく内包する黒画素塊の占める範囲を表ブロック、不定形の画素塊が散在している領域を写真ブロック、それ以外の任意形状の画素塊を図画ブロックとする。

【 0 0 9 8 】

次に、ステップ S 3 0 3 0 及びステップ S 3 1 3 0 の特徴量情報抽出処理の詳細について説明する。

40

【 0 0 9 9 】

尚、特徴量情報抽出処理は、画像ブロック及び文字ブロックで処理方法が異なるので、それぞれ別に説明する。

【 0 1 0 0 】

ここで、画像ブロックは、図 9 ( b ) の例の場合、写真ブロックと図画ブロックとするが、用途や目的に応じて、画像ブロックを写真ブロック及び図画ブロックの少なくとも一方にすることも可能である。

【 0 1 0 1 】

まず、画像ブロックに対する特徴量情報抽出処理について説明する。

50

## 【0102】

尚、1文書に複数の画像ブロックが存在する場合は、その総数分、以下の処理を繰り返す。

## 【0103】

実施形態1では、一例として、画像の色に関する色特徴量を抽出する色特徴量情報抽出処理を行う。

## 【0104】

この色特徴量情報抽出処理の詳細について、図10を用いて説明する。

## 【0105】

図10は本発明の実施形態1の色特徴量情報抽出処理の詳細を示すフローチャートである。

10

## 【0106】

尚、この処理では、処理対象画像を複数のメッシュブロックに分割した各メッシュブロックの色ヒストグラム中の最頻色を有する色と各メッシュブロックの位置情報を対応づけた情報を色特徴情報として抽出する。

## 【0107】

まず、ステップS1020で、画像を複数のメッシュブロックに分割する。実施形態1では、図11に示すように、画像を縦横をそれぞれ9メッシュブロックに分割する。特に、実施形態1では、表記の都合上 $9 \times 9 = 81$ メッシュブロックに分割している例を示しているが、実際には、 $15 \times 15 = 225$ メッシュブロック程度であることが好ましい。

20

## 【0108】

次に、ステップS1030で、処理対象となる着目メッシュブロックを左上端のブロックに設定する。尚、この着目メッシュブロックの設定は、例えば、図12に示すように、予め処理順序が決定された順序決定テーブルを参照して行う。

## 【0109】

ステップS1040で、未処理の着目メッシュブロックの有無を判定する。未処理の着目メッシュブロックがない場合(ステップS1040でNO)、処理を終了する。一方、未処理の着目メッシュブロックがある場合(ステップS1040でYES)、ステップS1050に進む。

## 【0110】

ステップS1050で、着目メッシュブロックの全画素の各濃度値を、図13の色空間を分割して作った部分空間である色ビンへ射影し、色ビンに対する色ヒストグラムを生成する。

30

## 【0111】

尚、実施形態1では、図13に示すように、RGB色空間を $3 \times 3 \times 3 = 27$ に分割した色ビンへ着目メッシュブロックの全画素の濃度値を射影する場合を示しているが、実際には、RGB色空間を $6 \times 6 \times 6 = 216$ に分割した色ビンへ着目メッシュブロックの全画素の濃度値を射影するほうが好ましい。

## 【0112】

ステップS1060で、色ヒストグラムの最頻色ビンの色ビンIDをその着目メッシュブロックの代表色と決定し、その着目メッシュブロックとその位置に対応づけて記憶部111に記憶する。

40

## 【0113】

ステップS1070で、図12の順序決定テーブルを参照して、次の処理対象となる着目メッシュブロックを設定する。その後、ステップS1040に戻り、未処理の着目メッシュブロックがなくなるまで、ステップS1040~ステップS1070の処理を再帰的に繰り返す。

## 【0114】

以上の処理によって、処理対象画像(画像ブロック)のメッシュブロック毎の代表色と各メッシュブロックの位置情報が対応付けられた情報を色特徴量情報として抽出すること

50

ができる。

【 0 1 1 5 】

次に、文字ブロックに対する特徴量情報抽出処理について説明する。

【 0 1 1 6 】

尚、1文書に複数の文字ブロックが存在する場合は、その総数分、以下の処理を繰り返す。

【 0 1 1 7 】

文字ブロックに対する文字特徴量情報は、その文字ブロックにOCR（文字認識）処理を施して得られる文字コードとする。

【 0 1 1 8 】

OCR（文字認識）処理は、文字ブロックから文字単位で切り出された文字画像に対し、パターンマッチングの一手法を用いて文字認識を行い、対応する文字コードを取得する。

【 0 1 1 9 】

この文字認識処理は、文字画像から得られる特徴を数十次元の数値列に変換した観測特徴ベクトルと、あらかじめ字種毎に求められている辞書特徴ベクトルとを比較し、最も距離の近い字種を認識結果とするものである。

【 0 1 2 0 】

特徴ベクトルの抽出には種々の公知手法があり、例えば、文字をメッシュ状に分割し、各メッシュブロック内の文字線を方向別に線素としてカウントしたメッシュ次元ベクトルを特徴とする方法がある。

【 0 1 2 1 】

ブロックセレクション処理（ステップS3020あるいはステップS3120）で抽出された文字ブロックに対して文字認識を行う場合は、まず、該当文字ブロックに対し横書き／縦書きの判定を行い、各々対応する方向に文字列を切り出し、その後、文字列から文字を切り出して文字画像を取得する。

【 0 1 2 2 】

横書き／縦書きの判定は、該当文字ブロック内で画素値に対する水平／垂直の射影を取り、水平射影の分散が大きい場合は横書き、垂直射影の分散が大きい場合は縦書きと判定する。文字列及び文字への分解は、横書きの文字ブロックである場合には、その水平方向の射影を利用して行を切り出し、さらに切り出された行に対する垂直方向の射影から、文字を切り出すことで行う。一方、縦書きの文字ブロックに対しては、水平と垂直を逆にすれば良い。

【 0 1 2 3 】

次に、図3BのステップS3111のNup印刷の指定方法の詳細について説明する。

【 0 1 2 4 】

Nup印刷指定は、ユーザが、MFP100のユーザインタフェースを介して、Nup印刷であるか否か、更に、Nup印刷の場合は、1枚の文書に含まれるページ数を指定するものである。

【 0 1 2 5 】

このユーザインタフェースの一例について、図14を用いて説明する。

【 0 1 2 6 】

図14は本発明の実施形態1のユーザインタフェースの一例を示す図である。

【 0 1 2 7 】

1411は表示・操作パネルである。1412～1415は各種機能ボタンであり、それぞれの機能ボタン1412～1415は、処理対象の画像の印刷指示、配信指示、蓄積指示及び編集指示を行うためのものである。

【 0 1 2 8 】

1416はスタートボタンであり、押下することで、機能ボタンで選択した機能を実行させることができる。1425はテンキーであり、印刷時の枚数の指定や、Nup印刷す

10

20

30

40

50

る場合の 1 枚の紙に含まれるページ数の指定を行うことが可能である。

【 0 1 2 9 】

1 4 1 7 は表示領域であり、タッチパネルで構成され、ユーザが直接画面に触れることで選択指示が可能である。1 4 1 8 は紙文書確認用領域であり、画像読取部 1 1 0 で読み取った紙文書画像を、領域内に収まるサイズに縮小して表示する。ユーザは紙文書画像の状態を、この領域 1 4 1 8 で確認することが可能となる。

【 0 1 3 0 】

1 4 1 9 は入力原稿の種類を確認する領域であり、読み取った紙文書の種類の内容を確認することができる。この領域 1 4 1 9 内において、1 4 2 0 は N u p 印刷の指定を行う指定領域である。また、N u p 印刷が指定された場合には、その指定された状態を示すために、その表示形態が、例えば、色付き表示、ブリンク表示、ハイライト表示等の表示形態に変更される。

10

【 0 1 3 1 】

1 4 2 4 は N u p 印刷を指定する場合に、1 枚の紙に含まれるページ数を表示する領域であり、特に、N u p 印刷が指定されていない場合には、1 が設定されている。ページ数の指定は、テンキー 1 4 2 5 を使用する。

【 0 1 3 2 】

このようなユーザインタフェースを構成することで、読み取った紙文書に対する N u p 印刷の指定、N u p 印刷である場合には、1 枚の紙に何ページ分の文書に含まれるかを、その状態を表示しながら指定することができる。

20

【 0 1 3 3 】

次に、ステップ S 3 1 4 0 の比較処理の詳細について、図 1 5 を用いて説明する。

【 0 1 3 4 】

図 1 5 は本発明の実施形態 1 の比較処理の詳細を示すフローチャートである。

【 0 1 3 5 】

まず、ステップ S 1 5 1 0 で、アドレス情報を参照し、未比較のオリジナル文書の有無を判定する。未比較のオリジナル文書がない場合（ステップ S 1 5 1 0 で N O ）、ステップ S 1 5 5 0 に進む。一方、未比較のオリジナル文書がある場合（ステップ S 1 5 1 0 で Y E S ）、ステップ S 1 5 2 0 に進む。

【 0 1 3 6 】

30

次に、レイアウトの比較を行う。ここで、レイアウトとは、ブロック情報にあるブロックの属性、サイズ、位置のことである。比較元画像（紙文書）と比較先画像（電子データ）のレイアウトが同じである場合（ステップ S 1 5 2 0 で Y E S ）、ステップ S 1 5 3 0 に進む。一方、比較元画像と比較先画像のレイアウトが同じでない場合（ステップ S 1 5 2 0 で N O ）、ステップ S 1 5 1 0 に戻る。

【 0 1 3 7 】

次に、ステップ S 1 5 3 0 で、比較元画像（紙文書）と比較先画像（電子データ）のページ同士の比較を行うページ比較処理を実行する。この比較は、ブロックの属性に合わせ、文字、画像それぞれに応じた特徴量を用いて、複合的に比較を行い、類似度を算出する。この処理の詳細については後述する。

40

【 0 1 3 8 】

次に、ステップ S 1 5 4 0 で、比較先の文書 I D、ページ番号に対応させて文書の類似度を記憶部 1 1 1 に一時記憶する。

【 0 1 3 9 】

ステップ S 1 5 1 0 において、全ての文書との比較が終了した場合（ステップ S 1 5 1 0 で N O ）、ステップ S 1 5 5 0 に進み、類似度の降順に文書 I D、ページ番号をソートし出力する。

【 0 1 4 0 】

次に、ステップ S 1 5 3 0 のページ比較処理の詳細について、図 1 6 を用いて説明する。

50

## 【0141】

図16は本発明の実施形態1のページ比較処理の詳細を示すフローチャートである。

## 【0142】

まず、ステップS1610で、ブロック情報を参照し、処理対象となる文書ID、ページ番号に対応する電子データ中で、未比較のブロックの有無を判定する。未比較のブロックがない場合(ステップS1610でNO)、ステップS1670に進む。一方、未比較のブロックがある場合(ステップS1610でYES)、ステップS1620に進む。

## 【0143】

次に、ステップS1620で、比較対象のブロックの属性を判定する。属性が画像ブロックである場合、ステップS1640へ進む。一方、属性が文字ブロックである場合、ス

10

## 【0144】

属性が画像ブロックである場合、ステップS1640で、色に関する特徴量情報で比較先ブロックとの類似比較である色特徴量情報比較処理を行う。この処理の詳細については後述する。これによって得られる類似度は、比較先の文書ID、ページ番号、ブロックIDに対応させて記憶部111に一時記憶する。

## 【0145】

一方、属性が文字ブロックである場合、ステップS1660で、文字の特徴量情報での比較元ブロックと比較先ブロックとの類似比較である文字特徴量情報比較処理を行う。この処理の詳細については後述する。また、これによって得られる類似度は、比較先の文書

20

## 【0146】

次に、ステップS1610において、全てのブロックとの比較が終了した場合(ステップS1610でNO)、ステップS1670に進み、ステップS1640及びステップS1660の処理によって記憶部111に記憶されている、比較先文書(電子データ)のページに含まれる全てのブロックの類似度を統合し、検索条件である紙文書とオリジナル文書中のページとの類似度を算出する統合処理を行う。この処理の詳細については後述する。

## 【0147】

次に、ステップS1640の色特徴量情報比較処理の詳細について、図17を用いて説明する。

30

## 【0148】

図17は本発明の実施形態1の色特徴量情報比較処理の詳細を示すフローチャートである。

## 【0149】

まず、ステップS1710で、比較元画像ブロックと比較先画像ブロックの色特徴量を色特徴量情報から読み出す。

## 【0150】

次に、ステップS1720で、処理対象とする画像ブロック中の着目メッシュブロックを先頭に設定する。ステップS1730で、比較元画像ブロックの色特徴量と、比較対象の色特徴量の類似度を示す類似距離を0にリセットする。

40

## 【0151】

ステップS1740で、未比較の着目メッシュブロックの有無を判定する。未比較の着目メッシュブロックがない場合(ステップS1740でNO)、ステップS1780に進む。一方、未比較の着目メッシュブロックがある場合(ステップS1740でYES)、ステップS1750に進む。

## 【0152】

ステップS1750で、比較元画像と比較先画像のそれぞれの色特徴量から、それぞれの着目メッシュブロックの色ピンIDを取得する。

## 【0153】

50



ステップS 1760で、図18の色ピンペナルティマトリックスを参照して、取得した色ピンID間に対応する着目メッシュブロックの局所的類似距離を取得し、これを直前の処理で取得している類似距離に累積加算する。そして、この類似距離は記憶部111に記憶する。

【0154】

ここで、色ピンペナルティマトリックスについて、図18を用いて説明する。

【0155】

図18は本発明の実施形態1の色ピンペナルティマトリックスの構成を示す図である。

【0156】

色ピンペナルティマトリックスは、色ピンID同士の局所的類似距離を管理するマトリックスである。図18によれば、色ピンペナルティマトリックスは、同一色ピンIDではその類似距離は0となり、色ピンID同士の差が大きくなるほど、つまり、類似度が低くなるほど、その類似距離は大きくなるように構成されている。また、同一色ピンIDの対角位置は全て、その類似距離は0で、それを境に対象性を持っている。

【0157】

このように、実施形態1では、色ピンペナルティマトリックスを参照するだけで、色ピンID同士の類似距離を取得することができるので、処理の高速化を図ることができる。

【0158】

そして、ステップS 1770で、図12の順序決定テーブルを参照して、次の処理対象となる着目メッシュブロックを設定する。その後、ステップS 1740に戻る。

【0159】

そして、ステップS 1740で、未比較の着目メッシュブロックがない場合(ステップS 1740でNO)、ステップS 1780に進み、記憶部111に記憶されている類似距離を類似度に変換し、ブロックIDと対にして出力する。

【0160】

尚、類似度への変換は、例えば、類似距離が最小値のときを類似度100%、類似距離が最大値のときを類似度0%として、その範囲内の類似距離に対する類似度は、最小値あるいは最大値に対する差に基づいて算出するようにすれば良い。

【0161】

次に、ステップS 1660の文字特徴量情報比較処理の詳細について説明する。

【0162】

この処理では、比較元画像と比較先画像中のそれぞれの文字ブロック内の各文字コード同士の比較を行い、その一致度から類似度を算出する。

【0163】

尚、検索条件とする紙文書とオリジナル文書との比較である場合、類似度は100%となるのが理想的であるが、実際には、検索条件となる紙文書中の文字ブロックに対するOCR処理では誤認識が発生する場合があるので、オリジナル文書との比較であっても、類似度は100%にならないことはあるが、かなり100%に近い値となる。

【0164】

次に、ステップS 1670の統合処理の詳細について説明する。

【0165】

この統合処理では、比較先画像であるオリジナル文書内で占めている割合の大きいブロックの類似度が、オリジナル文書全体の類似度としてより大きく反映されるような、算出されたブロック毎の類似度の統合を行う。

【0166】

例えば、オリジナル文書中のブロックB1~B6に対し、ブロック毎の類似率がn1~n6と算出されたとする。このときオリジナル文書全体の総合類似率Nは、以下の式で表現される。

【0167】

$$N = w_1 * n_1 + w_2 * n_2 + w_3 * n_3 + \dots + w_6 * n_6 \quad (1)$$

10

20

30

40

50

ここで、 $w_1 \sim w_6$  は、各ブロックの類似率を評価する重み係数である。重み係数  $w_1 \sim w_6$  は、ブロックのオリジナル文書内の占有率により算出する。例えば、ブロック 1 ~ 6 のサイズを  $S_1 \sim S_6$  とすると、ブロック 1 の占有率  $w_1$  は、

$$w_1 = S_1 / (S_1 + S_2 + \dots + S_6) \quad (2)$$

として算出することができる。

#### 【0168】

このような占有率を用いた重み付け処理により、オリジナル文書内で大きな領域を占めるブロックの類似度がより、オリジナル文書全体の類似度に反映することができる。

#### 【0169】

次に、ステップ S 3 1 5 0 及びステップ S 3 1 6 0 に示す確認モード時の処理の詳細について説明する。

10

#### 【0170】

確認モードは、ユーザが予めユーザインタフェースから指定してもよいし、確認モードにするべきか否かを自動で判定しても良い。自動判定の方法としては、次の方法がある。例えば、検索されたオリジナル文書候補が 1 つの場合、または、1 位のオリジナル文書候補と 2 位以降のオリジナル文書候補のそれぞれ類似度の差が所定値以上で、1 位のオリジナル文書候補が、所望とするオリジナル文書である可能性が高い場合は、「非確認モード」としてステップ S 3 1 7 0 に進み、そうでない場合は、「確認モード」とする。

#### 【0171】

但し、Nup 印刷の紙文書をスキャンした場合は、スキャン画像中の各ページに対応するそれぞれの候補について、1 つでも上記の条件を満足しない場合は、「確認モード」となり、上記の条件を満足しなかったページのみの確認を行う。

20

#### 【0172】

「確認モード」の際は、MFP 1 0 0 の表示部 1 1 0 と入力部 1 1 3 で実現されるユーザインタフェースに、オリジナル文書候補群を類似度の高い順に表示して、その中から所望のオリジナル文書の選択をユーザに行ってもらおう。

#### 【0173】

このように、確認モードの実行の有無を自動判定する場合は、ユーザによるオリジナル文書の選択操作が不要となるので、操作工数を低減することができる。

#### 【0174】

30

ここで、確認モード時のユーザインタフェースの一例について、図 1 9 を用いて説明する。

#### 【0175】

図 1 9 は本発明の実施形態 1 のユーザインタフェースの一例を示す図である。

#### 【0176】

1 9 1 7 は表示領域であり、タッチパネルで構成され、ユーザが直接画面に触れることで選択指示が可能である。尚、この表示領域 1 9 1 7 は、図 1 4 の表示領域 1 4 1 7 と同様のものである。

#### 【0177】

1 9 1 8 はモード表示領域であり、図 1 9 では、確認モードであることを示している。通常は自動判定されたモードを表示する。また、これに触れることで、「確認モード」と「非確認モード」をユーザがサイクリックに指定することができる。

40

#### 【0178】

1 9 1 9 ~ 1 9 2 8 は、検索結果として出力するオリジナル文書候補のサムネイル画像群である。このサムネイル画像の表示は、1 9 1 9 から番号順に類似度の高い順で表示されている。

#### 【0179】

この例では、最大 1 0 のサムネイル画像が表示され、オリジナル文書候補が 1 0 以上である場合には、上位 1 0 までのサムネイル画像が表示される。そして、このサムネイル画像群 1 9 1 9 ~ 1 9 2 8 から、所望するサムネイル画像を選択することで、所望のオリジ

50

ナル文書を選択することが可能であり、その選択したオリジナル文書に対する各種処理を実行することが可能となる。

【0180】

1929はNup印刷の紙文書をスキャンした時に表示され、文書中の何ページ目の候補が表示されているかを示すものである。また、これを選択することで、他のページの候補を切り替えて表示させることもできる。

【0181】

次に、ステップS3170の印刷/配信/蓄積/編集処理の詳細について説明する。

【0182】

例えば、Nup印刷した紙文書をスキャンして、これを用いて、類似画像検索を行った場合は、その文書中の各ページに対する複数ページ分の検索結果（オリジナル文書）が得られる。

10

【0183】

そして、これらの複数ページ分の検索結果を用いて印刷を行う場合には、その印刷条件としては、例えば、1枚につき1ページ分を印刷してもよいし、もとのNup印刷とはレイアウトを変更したNup印刷を行ってもよいし、1枚の両面にそれぞれのページを印刷するようにしてもよいし、各ページを1枚の紙に重ねて（オーバーレイして）印刷するようにしてもよいし、順番を変えて製本印刷するようにしてもよい。また、もちろん、ユーザが指定したページのみを1枚の紙に印刷してもよい。

【0184】

20

また、印刷処理以外に、検索結果の全てについて、配信、蓄積、編集等の処理を行うようにしても、ユーザが指定したページのみを配信、蓄積、編集等の処理を行うようにしてもよい。

【0185】

以上説明したように、実施形態1によれば、Nup印刷した紙文書に含まれる全てのページに対して、オリジナルの電子データを検索することができ、かつその検索した電子データを用いた様々な活用を実現することができる。

【0186】

<<実施形態2>>

実施形態1では、入力する紙文書がNup印刷であるか否か、更に、Nup印刷の場合は、1枚の文書に含まれるページ数の指定をユーザが行う構成としたが、実施形態2では、入力した紙文書の内容を解析して、その解析結果に基づいて、紙文書がNup印刷であるか否か、更に、Nup印刷の場合は、1枚の文書に含まれるページ数を自動判定する構成について説明する。また、この自動判定のタイミングは、実施形態1の図3BのステップS3110の紙文書の入力後に実行する。

30

【0187】

以下、実施形態2の検索処理について、図20を用いて説明する。

【0188】

図20は本発明の実施形態2の検索処理を示すフローチャートである。

【0189】

40

尚、図20において、実施形態1の図3Bの検索処理と同一の処理ステップには、同一のステップ番号を付加し、その詳細については省略する。

【0190】

ステップS3110で、紙文書の入力後、ステップS2011で、入力されたスキャン画像がNup印刷であるか否かを判定するNup印刷判定処理を行う。そして、この判定結果に基づいて、ステップS3111の処理を実行することになる。

【0191】

ここで、紙文書に1ページ分の文書が印刷されている通常印刷の場合には、紙文書の上端か下端にページ番号が印刷される。一方、Nup印刷の場合は、紙文書内に複数のページ番号が等間隔に印刷されることになる。そこで、実施形態2では、このことを利用して

50

、処理対象のスキャン画像がN u p印刷であるか否かを判定する。

【0192】

例として、図21に示すような1枚の紙に4ページ分の文書をN u p印刷した場合について説明する。

【0193】

図21は本発明の実施形態2の1枚の紙に4ページ分の文書をN u p印刷した例を示す図である。

【0194】

図21において、2110はN u p印刷時の紙全体の領域を示している。2111~2114は各ページのページ画像を示している。2115~2118は各ページに付与されたページ番号である。2119と2120の太枠内は、後述するページ番号探索領域である。

10

【0195】

次に、ステップS2011のN u p印刷判定処理の詳細について、図21及び図22を用いて説明する。

【0196】

図22は本発明の実施形態2のN u p印刷判定処理の詳細を示すフローチャートである。

【0197】

まず、ステップS2210で、紙全体の領域2110の上端と下端の領域に対して、O C R処理を施す。次に、ステップS2220で、O C R処理の処理結果として、ページ番号(例えば、アラビア数字、もしくは英数字)が2個所以上、上端領域または下端領域の同じ領域内に等間隔で存在するか否かを判定する。ページ番号が2個所以上存在しない場合(ステップS2220でN O)、ステップS2260に進み、通常印刷であると判定する。一方、ページ番号が2個所以上存在する場合(ステップS2220でY E S)、ステップS2230に進む。

20

【0198】

図21の例では、ページ番号2117(「3」と)と2118(「4」と)が下端領域に2箇所検出されることになる。

【0199】

ステップS2230で、検出されたページ番号に基づいて、他のページ番号を探索するためのページ番号探索領域を設定し、その設定したページ番号探索領域に対してO C R処理を施す。

30

【0200】

図21の例では、ページ番号画像2117と2118で、それぞれのページ番号を含む紙全体の垂直方向にページ番号探索領域2119及び2120を設定する。そして、各ページ番号探索領域2119、2120に対してO C R処理を施す。

【0201】

次に、ステップS2240で、各ページ番号探索領域でページ番号が検出され、各ページ番号探索領域中のページ番号同士の間隔が同一であるか否かを判定する。同一でない場合(ステップS2240でN O)、ステップS2260に進み、通常印刷であると判定する。一方、同一である場合(ステップS2240でY E S)、ステップS2250に進み、N u p印刷であると判定する。

40

【0202】

尚、ここでの間隔が同一であるか否かの判定は、ある程度の誤差を考慮した誤差範囲を含むページ番号同士の間隔に対して実行する。つまり、この誤差範囲内にページ番号同士の間隔が含まれる場合には、ページ番号同士の間隔が同一であると判定する。

【0203】

図21の例では、ページ番号探索領域2119にページ番号2115(「1」と)と2117(「3」と)が検出され、ページ番号探索領域2120にページ番号2116(「2」と)

50

)と2118(「4」)が検出される。そして、ページ番号探索領域2119とページ番号探索領域2120中のページ番号の間隔は、同じ間隔となっている。そのため、この場合は、Nup印刷と判定される。

【0204】

また、この際、ステップS2210で検出されたページ番号の数と、ステップS2230の1つのページ番号探索領域で検出されたページ番号の数を乗算することで、1枚に含まれるページ数を算出することができ、このページ数を記憶部111に一時保存する。

【0205】

以上説明したように、実施形態2によれば、実施形態1で説明した効果に加えて、スキャン画像がNup印刷であるか否かを自動的に判定することで、ユーザの手を煩わすことなく、Nup印刷であるか否か、また、Nup印刷である場合に何ページ分含んでいるかを判定することが可能となる。

10

【0206】

尚、実施形態2において、ステップS2011のNup印刷判定処理の処理結果を、ユーザに一旦提示して確認させる構成としても良い。このときは、図14の紙文書確認用領域1418のように判定結果を縮小画像として表示すればよい。また、確認の結果、誤判定である場合には、領域1419を用いて判定内容を修正する機会を与えることが可能である。

【0207】

このような構成にすることにより、Nup印刷判定処理に誤判定があっても、その判定内容を修正することができ、誤ったオリジナルの電子データが出力される、といったことを防止することができる。

20

【0208】

<<実施形態3>>

実施形態2では、ステップS2011のNup印刷判定処理において、OCR処理を用いて、ページ番号を検出する方法について説明したが、これに限定されるものではない。例えば、次のようにして、ページ番号を検出するようにしても良い。

【0209】

ここでは、図23に示すように、1枚の紙に4ページ分の文書をNup印刷した場合について説明する。

30

【0210】

尚、2310は図21の2110と、以下同様に、2311は2111と、2312は2112と、2313は2113と、2314は2114と、2315は2115と、2316は2116と、2317は2117と、2318は2118に対応するので、その説明は省略する。

【0211】

図23において、2320は左側の2つのページの左マージン(最左端余白)である。2321は右側の2つのページの右マージン(最右端余白)である。2319は左側の2つのページと右側の2つのページの間マージン(中間余白)である。ここで、中間マージン2319は、左側の2つのページの右マージンと、右側の2つのページの左マージンとが連続しているため、左マージン2320や右マージン2321より長くなる。

40

【0212】

そして、この中間マージン2319が検出される場合には、Nupページ印刷であると判定することができる。

【0213】

尚、図23では、横方向のマージンに着目して説明したが、縦方向に関しても同様のことが言える。

【0214】

次に、より詳細な具体例について、図24を用いて説明する。

【0215】

50

2 4 1 0 は紙全体の領域である。まず、これに対してブロックセレクション処理を行うと、ブロック 2 4 1 1 ~ 2 4 2 1 が、ブロックセレクション処理結果として得られる。まず、検出されたブロック群の最左端を検出し、紙全体領域の左端 2 3 1 0 の差分である左マージン 2 4 2 2 を算出する。そして、この左マージン 2 4 2 2 が、図 2 3 の左マージン 2 3 2 0 に相当する。

【 0 2 1 6 】

同様に、ブロック群の最右端を検出し、紙全体領域の右端 2 3 1 0 の差分である右マージン 2 4 2 3 を算出する。そして、この右マージン 2 4 2 3 が、図 2 3 の右マージン 2 3 2 1 に相当する。

【 0 2 1 7 】

その後、各ブロック群の横方向の間隔を算出し、中間マージン 2 3 1 9 の中間マージン候補 2 4 2 4、2 4 2 5 を算出する。中間マージン候補 2 4 2 4、2 4 2 5 の中で、左マージン 2 4 2 2、右マージン 2 4 2 3 よりも長い中間マージン候補 2 4 2 4 を中間マージンとする。

【 0 2 1 8 】

このようにして検出した中間マージンの数 + 1 が、横方向のページ数となる。同様の処理を縦方向にも行い、それぞれの結果を乗算することで、ページの総数を算出することができる。

【 0 2 1 9 】

以上説明したように、実施形態 3 によれば、実施形態 2 で説明した効果と同様に、スキャン画像が N u p 印刷であるか否かを自動的に判定することで、ユーザの手を煩わすことなく、N u p 印刷であるか否か、また、N u p 印刷である場合に何ページ分含んでいるかを判定することが可能となる。

【 0 2 2 0 】

尚、実施形態 1 において、図 1 4 の紙文書確認用領域 1 4 1 8 で、ユーザが所望のページが表示されている範囲に触れて、ページを指定することで、処理対象のページを選択するようにしても構わない。このときは、ステップ S 3 1 2 0 からステップ S 3 1 4 0 の処理は、指定されたページ領域についてのみ処理が行われる。

【 0 2 2 1 】

また、実施形態 2、3 においても、N u p 印刷判定結果をユーザに提示し、同様な手続きを行うことによって、指定されたページ領域についてのみ処理を行うようにしても構わない。こうすることによって、ユーザは N u p 印刷された紙文書中の特定のページについてのみ、印刷等の処理を行うことができる。

【 0 2 2 2 】

また、図 1 0 の色特徴量情報抽出処理では、処理対象画像の最頻色を色特徴情報として抽出する例を説明したが、これに限定されるものではなく、例えば、平均色を色特徴情報として抽出するようにしても良い。

【 0 2 2 3 】

また、画像特徴量として色特徴量を用いたが、これに限定されるものではなく、例えば、最頻輝度、平均輝度等の輝度特徴量、共起行列、コントラスト、エントロピ、G a b o r 変換等で表現されるテクスチャ特徴量、エッジ、フーリエ記述子等の形状特徴量等の複数種類の画像特徴量を 1 つ、或いは、任意に組み合わせた画像特徴量を用いても良い。

【 0 2 2 4 】

また、ブロックセレクション処理を行い、処理対象の文書を文字ブロックと画像ブロックに分割し、これらの各ブロックの特徴量を複合的に用いて検索を行ったが、文書全体を 1 つの画像とみなし、オリジナル文書の検索を行うようにしても構わない。また、精度が許容される範囲ならば、文書中の画像ブロックのみを利用して、オリジナル文書の検索を行うようにしても構わない。

【 0 2 2 5 】

また、電子データを、一旦、ラスタ画像に変換したが、電子データから直接、文字コー

10

20

30

40

50

ドや画像を抜き出して比較してももちろん構わない。

【0226】

また、文字特徴量としては文字コードを採用したが、例えば、単語辞書とのマッチングを予め行って単語の品詞を抽出しておき、名詞である単語を文字特徴量としても良い。

【0227】

<<実施形態4>>

上記実施形態1では、登録対象のオリジナル文書は、Nup印刷された紙文書（あるいはその電子データ）でない、通常印刷された紙文書（あるいはその電子データ）としているが、Nup印刷された紙文書（あるいはその電子データ）を登録対象のオリジナル文書とすることもできる。この場合、実施形態1の検索処理と同様に、登録処理時に、ユーザがNup印刷の指定及び1枚の文書（あるいは電子データ）に含まれるページ数の指定を行えば良い。

10

【0228】

この時の登録処理は、例えば、図25のようになる。但し、図25では、説明を簡単にするために、1枚のNup印刷された紙文書を登録する場合の登録処理を示している。

【0229】

尚、図25の各ステップは、実施形態1の図3Aと図3Bの一部のステップとほぼ同様のステップの組み合わせで実現される。

【0230】

簡単に説明すると、ステップS3010で、登録対象のオリジナル文書を入力する。ステップS3111で、ユーザによってオリジナル文書がNup印刷文書であるかどうかの指定がなされたかを判定する。

20

【0231】

そして、Nup印刷文書の指定がなされたと判定した場合（ステップS3111でYES）、ステップS3112において、ページ数Lにユーザが指定した1枚中のページ数を設定する。一方、Nup印刷文書の指定がなされないと判定した場合（ステップS3111でNO）、ステップS3115において、ページ数Lに1を設定する。

【0232】

次に、ステップS3113で、処理対象の1文書中の処理ページ数を示す変数bに初期値として1を設定する。次に、比較処理の終了判定として、Lがb以上であるか否かを判定する。Lがb未満である場合（ステップS3116でNO）、登録処理を終了する。

30

【0233】

一方、Lがb以上である場合（ステップS3116でYES）、未処理のページが存在するので、ステップS3120に進み、Nupのオリジナル文書中のbページ目の領域に対してブロックセレクション処理を行う。次に、ステップS3130で、各ブロックの特徴量情報を抽出して記憶部111に記憶する。そして、ステップS3141で、bを1インクリメントし、ステップS3116に戻る。

【0234】

もちろん、複数枚のNup印刷された紙文書を登録することも可能であり、この場合も、実施形態1の図3Aと図3Bのステップの内、必要なステップを組みあわせることで実現できる。

40

【0235】

尚、Nup印刷された紙文書を登録する場合には、登録対象の文書画像に含まれているそれぞれのページ画像について、紙1枚に1ページを印刷したときと同じになるように解像度変換を行い、これを紙1枚に1ページを印刷された文書画像として登録しても良いし、Nup印刷された紙文書に対応する文書画像をそのまま登録するようにしても良い。

【0236】

<<実施形態5>>

実施形態3のNup印刷の指定に代えて、Nup印刷であるか否か、更に、Nup印刷の場合は、1枚の文書に含まれるページ数を自動判定して、登録処理を行うようにしても

50

よい。

【0237】

この時の登録処理は、例えば、図26によくなる。但し、図26では、説明を簡単にするために、1枚のNup印刷された紙文書を登録する場合の登録処理を示している。

【0238】

尚、図26の各ステップは、実施形態1の図3Aと実施形態2の図20の一部のステップとほぼ同様のステップの組み合わせで実現される。

【0239】

簡単に説明すると、ステップS3010で登録対象のオリジナル文書を入力する。ステップS2011で、オリジナル文書がNup印刷文書であるかどうかを自動判定する。そして、Nup印刷文書であると判定した場合、ステップS3112において、判定されたページ数をページ数Lに設定する。一方、Nup印刷文書でないと判定した場合、ステップS3115において、ページ数Lに1を設定する。

10

【0240】

次に、ステップS3113で、処理対象の1文書中の処理ページ数を示す変数bに初期値として1を設定する。次に、比較処理の終了判定として、Lがb以上であるか否かを判定する。Lがb未満である場合(ステップS3116でNO)、登録処理を終了する。

【0241】

一方、Lがb以上である場合(ステップS3116でYES)、未処理のページが存在するので、ステップS3120に進み、Nupのオリジナル文書中のbページ目の領域に対してブロックセレクション処理を行う。次に、ステップS3130で、各ブロックの特徴量情報を抽出して記憶部111に記憶する。そして、ステップS3141で、bを1インクリメントし、ステップS3116に戻る。

20

【0242】

もちろん、複数枚のNup印刷された紙文書を登録することも可能であり、この場合も、実施形態1の図3Aと実施形態2の図20のステップの内、必要なステップを組みあせることで実現できる。

【0243】

以上、実施形態例を詳述したが、本発明は、例えば、システム、装置、方法、プログラムもしくは記憶媒体等としての実施態様をとることが可能であり、具体的には、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

30

【0244】

尚、本発明は、前述した実施形態の機能を実現するソフトウェアのプログラム(実施形態では図に示すフローチャートに対応したプログラム)を、システムあるいは装置に直接あるいは遠隔から供給し、そのシステムあるいは装置のコンピュータが該供給されたプログラムコードを読み出して実行することによっても達成される場合を含む。

【0245】

従って、本発明の機能処理をコンピュータで実現するために、該コンピュータにインストールされるプログラムコード自体も本発明を実現するものである。つまり、本発明は、本発明の機能処理を実現するためのコンピュータプログラム自体も含まれる。

40

【0246】

その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、OSに供給するスクリプトデータ等の形態であっても良い。

【0247】

プログラムを供給するための記録媒体としては、例えば、フロッピー(登録商標)ディスク、ハードディスク、光ディスク、光磁気ディスク、MO、CD-ROM、CD-R、CD-RW、磁気テープ、不揮発性のメモ리카ード、ROM、DVD(DVD-ROM, DVD-R)などがある。

【0248】

50



その他、プログラムの供給方法としては、クライアントコンピュータのブラウザを用いてインターネットのホームページに接続し、該ホームページから本発明のコンピュータプログラムそのもの、もしくは圧縮され自動インストール機能を含むファイルをハードディスク等の記録媒体にダウンロードすることによっても供給できる。また、本発明のプログラムを構成するプログラムコードを複数のファイルに分割し、それぞれのファイルを異なるホームページからダウンロードすることによっても実現可能である。つまり、本発明の機能処理をコンピュータで実現するためのプログラムファイルを複数のユーザに対してダウンロードさせるWWWサーバも、本発明に含まれるものである。

【0249】

また、本発明のプログラムを暗号化してCD-ROM等の記憶媒体に格納してユーザに配布し、所定の条件をクリアしたユーザに対し、インターネットを介してホームページから暗号化を解く鍵情報をダウンロードさせ、その鍵情報を使用することにより暗号化されたプログラムを実行してコンピュータにインストールさせて実現することも可能である。

【0250】

また、コンピュータが、読み出したプログラムを実行することによって、前述した実施形態の機能が実現される他、そのプログラムの指示に基づき、コンピュータ上で稼動しているOSなどが、実際の処理の一部または全部を行ない、その処理によっても前述した実施形態の機能が実現され得る。

【0251】

さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によっても前述した実施形態の機能が実現される。

【図面の簡単な説明】

【0252】

【図1】本発明の実施形態1の画像処理システムの構成を示すブロック図である。

【図2】本発明の実施形態1のMFPの詳細構成を示すブロック図である。

【図3A】本発明の実施形態1の登録処理を示すフローチャートである。

【図3B】本発明の実施形態1の検索処理を示すフローチャートである。

【図4】本発明の実施形態1のアドレス情報の一例を示す図である。

【図5】本発明の実施形態1のブロック情報の一例を示す図である。

【図6】本発明の実施形態1の特徴量情報の一例を示す図である。

【図7】本発明の実施形態1の特徴量情報の一例を示す図である。

【図8】本発明の実施形態1の2ページ分をNup印刷した例を示す図である。

【図9】本発明の実施形態1の画像ブロック抽出の一例を示す図である。

【図10】本発明の実施形態1の色特徴量情報抽出処理の詳細を示すフローチャートである。

【図11】本発明の実施形態1の画像メッシュブロック分割の一例を示す図である。

【図12】本発明の実施形態1の順序決定テーブルの一例を示す図である。

【図13】本発明の実施形態1の色空間上の色ピンの構成の一例を示す図である。

【図14】本発明の実施形態1のユーザインタフェースの一例を示す図である。

【図15】本発明の実施形態1の比較処理の詳細を示すフローチャートである。

【図16】本発明の実施形態1のページ比較処理の詳細を示すフローチャートである。

【図17】本発明の実施形態1の色特徴量情報比較処理の詳細を示すフローチャートである。

【図18】本発明の実施形態1の色ピンパネルティマトリックスの構成の一例を示す図である。

【図19】本発明の実施形態1のユーザインタフェースの一例を示す図である。

【図20】本発明の実施形態2の検索処理を示すフローチャートである。

10

20

30

40

50

【図 2 1】本発明の実施形態 2 の 1 枚の紙に 4 ページ分の文書を N u p 印刷した例を示す図である。

【図 2 2】本発明の実施形態 2 の N u p 印刷判定処理の詳細を示すフローチャートである。

【図 2 3】本発明の実施形態 3 の 1 枚に 4 ページ分を N u p 印刷した例を示す図である。

【図 2 4】本発明の実施形態 3 の N u p 印刷判定処理を説明するための図である。

【図 2 5】本発明の実施形態 4 の登録処理を示すフローチャートである。

【図 2 6】本発明の実施形態 5 の登録処理を示すフローチャートである。

【符号の説明】

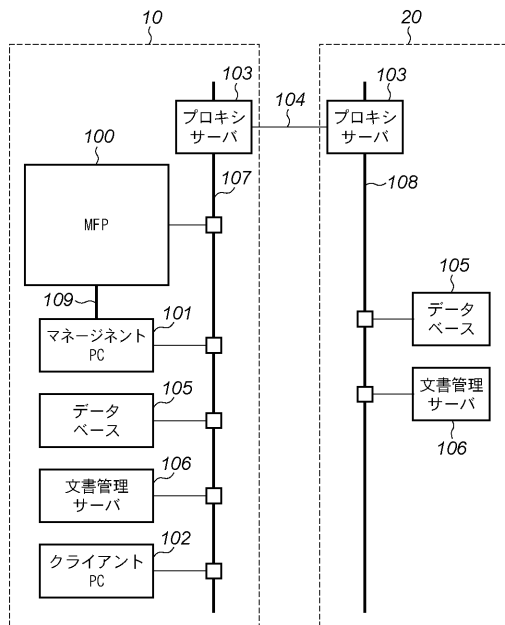
【 0 2 5 3 】

- 1 0 0 M F P
- 1 0 1 マネージメント P C
- 1 0 2 クライアント P C
- 1 0 3 プロキシサーバ
- 1 0 4 ネットワーク
- 1 0 5 データベース
- 1 0 6 文書管理サーバ
- 1 0 7 L A N
- 1 1 0 画像読取部
- 1 1 1 記憶部
- 1 1 2 印刷部
- 1 1 3 入力部
- 1 1 4、1 1 7 ネットワーク I / F
- 1 1 5 データ処理部
- 1 1 6 表示部

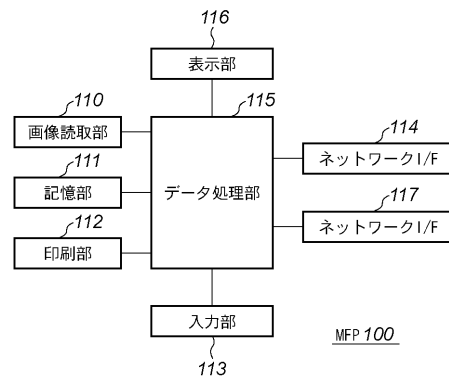
10

20

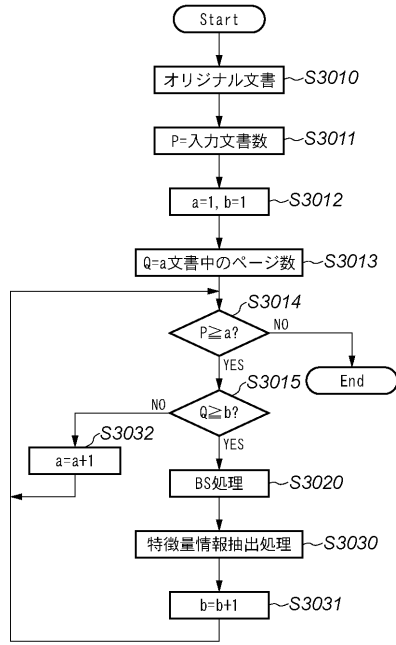
【図 1】



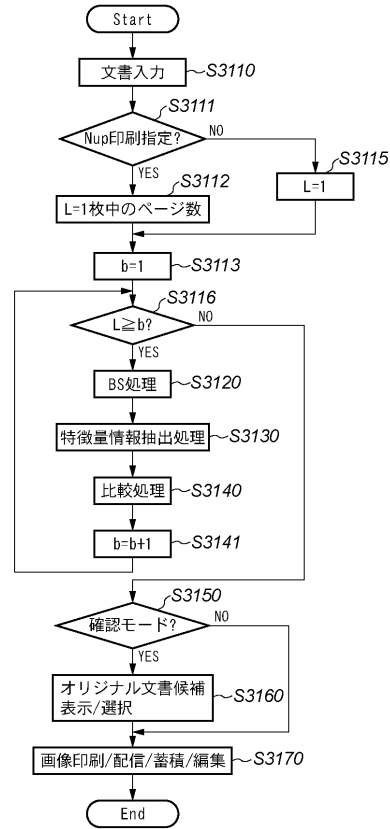
【図 2】



【図3A】



【図3B】



【図4】

文書ID	アドレス
0000001	¥¥abc¥doc¥ship.doc
0000002	C:¥img¥car.bmp

【図5】

文書ID	ページ番号	ブロックID	属性	サイズ	位置
0000001	1	0001	画像	30	5, 5
0000001	1	0002	画像	40	5, 50
0000001	1	0003	文字	30	80, 5
0000001	2	0001	文字	40	5, 5
0000001	2	0002	画像	60	5, 45
0000002	1	0001	画像	30	15, 5
0000002	1	0002	画像	35	15, 50
0000002	1	0003	文字	35	75, 5

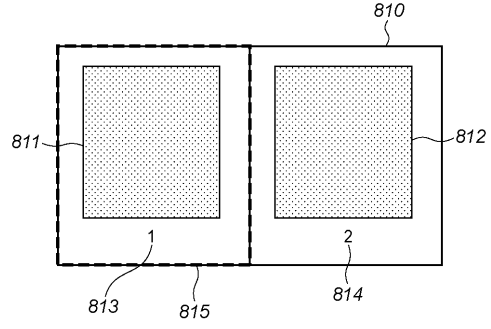
【図6】

文書ID	頁番号	ブロックID	色特徴量
0000001	1	0001	.....
0000001	1	0002	.....
0000001	2	0002	.....
0000002	1	0001	.....
0000002	1	0002	.....

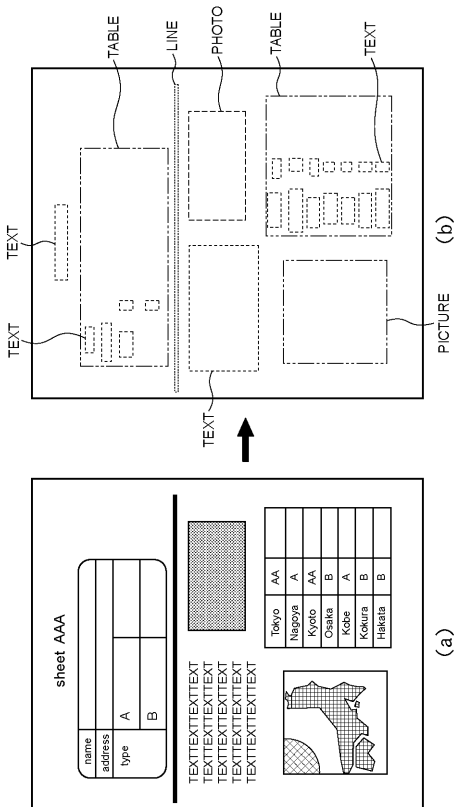
【図7】

文章ID	頁番号	ブロックID	文字特徴量
0000001	1	0003	.....
0000001	2	0001	.....
0000002	1	0003	.....

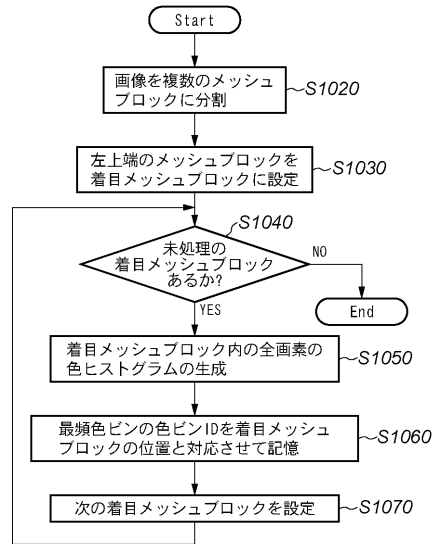
【図8】



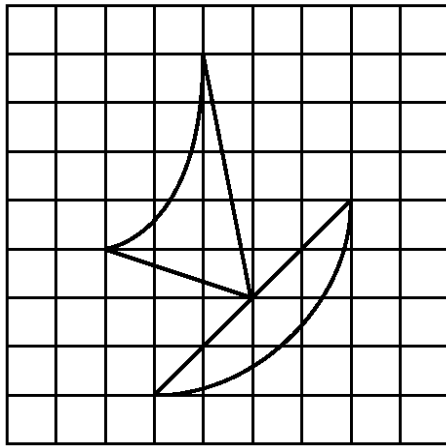
【図9】



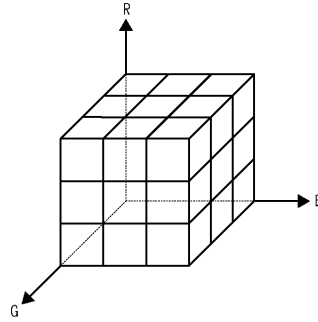
【図10】



【図 1 1】



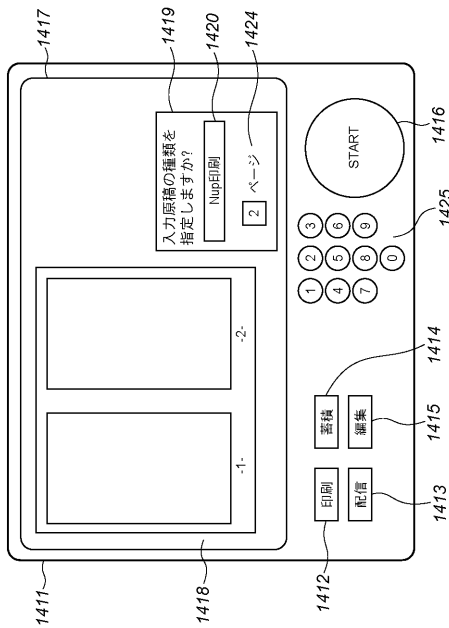
【図 1 3】



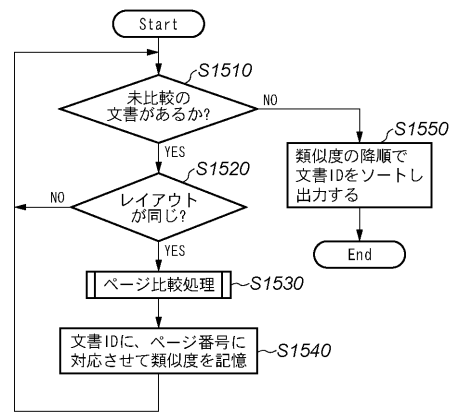
【図 1 2】

1	2	4
3	5	7
6	8	9

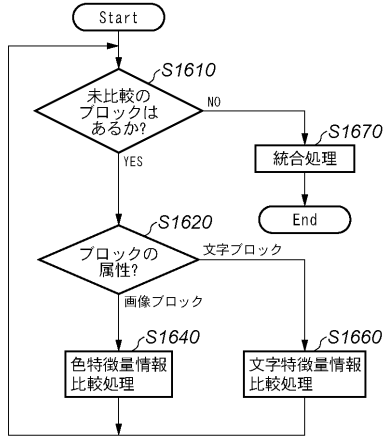
【図 1 4】



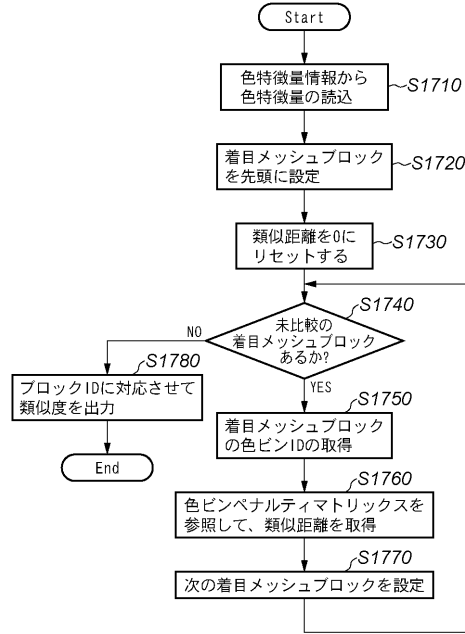
【図 1 5】



【図16】



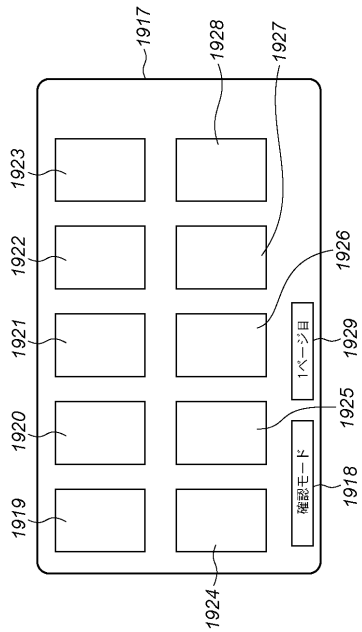
【図17】



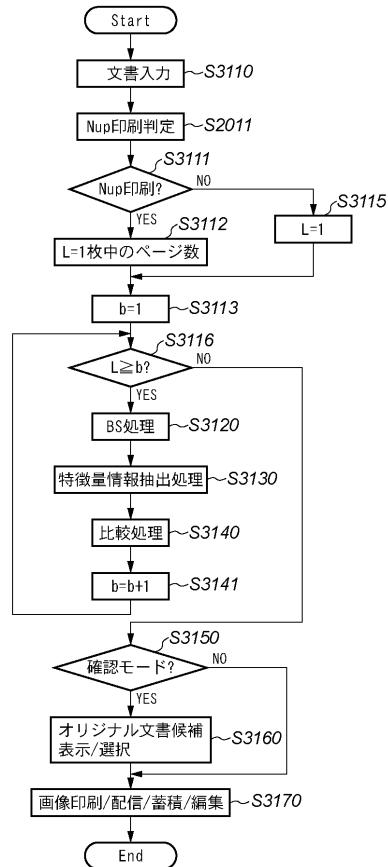
【図18】

	1	2	3	4	5	6	7	8	...
1	0	1	1	5	1	5	7	7	...
2		0	2	1	1	7	2	5	...
3			0	7	1	1	5	2	...
4				0	1	9	1	7	...
:									

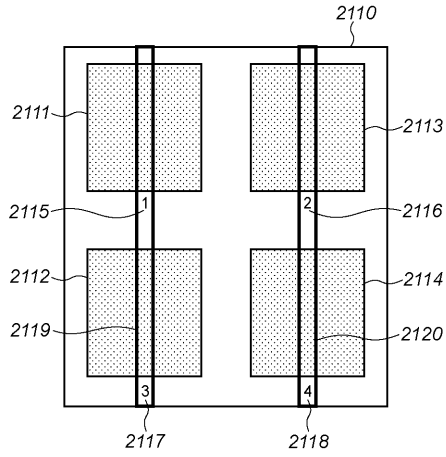
【図19】



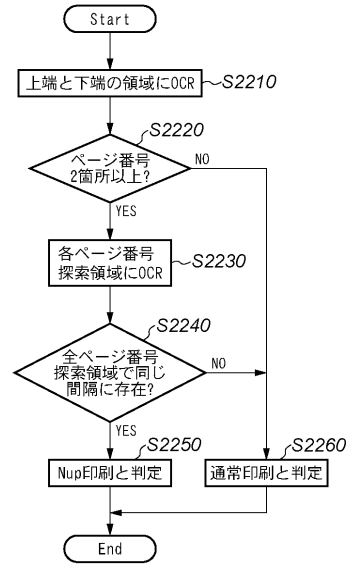
【図20】



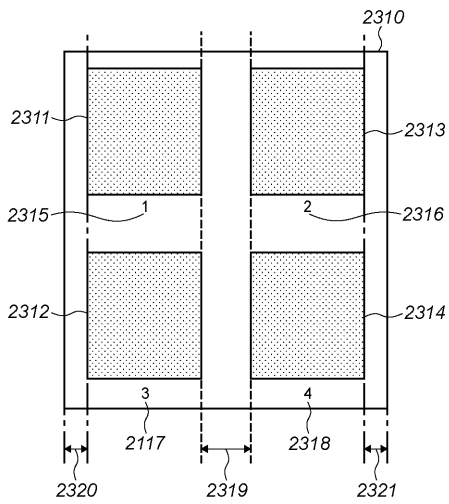
【図 2 1】



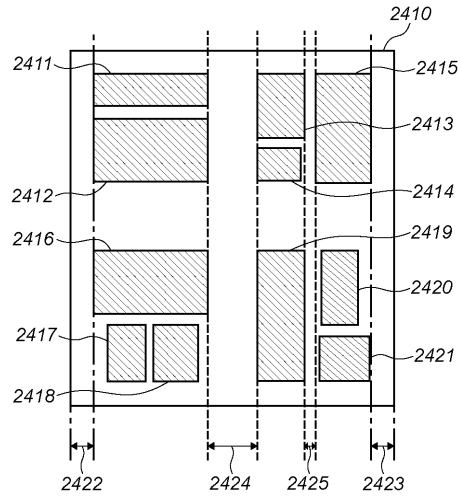
【図 2 2】



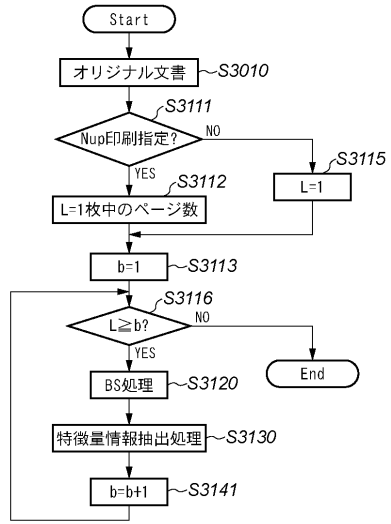
【図 2 3】



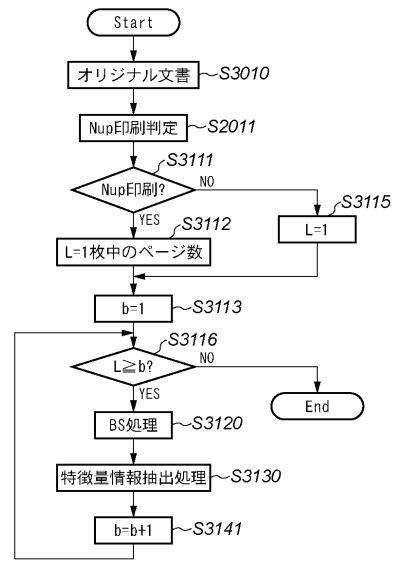
【図 2 4】



【図 25】



【図 26】





---

フロントページの続き

審査官 松永 稔

- (56)参考文献 特開2002-007413(JP,A)  
特開平08-022535(JP,A)  
特開2004-348706(JP,A)  
特開2004-334336(JP,A)

- (58)調査した分野(Int.Cl., DB名)  
G06T 7/00