

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7379063号
(P7379063)

(45)発行日 令和5年11月14日(2023.11.14)

(24)登録日 令和5年11月6日(2023.11.6)

(51)国際特許分類

F I

G 0 6 V 30/148 (2022.01)

G 0 6 V 30/14 (2022.01)

G 0 6 T 7/00 (2017.01)

G 0 6 V 30/148

G 0 6 V 30/14 3 4 0 Z

G 0 6 T 7/00 3 5 0 C

請求項の数 13 (全42頁)

(21)出願番号	特願2019-183923(P2019-183923)	(73)特許権者	000001007
(22)出願日	令和1年10月4日(2019.10.4)		キヤノン株式会社
(65)公開番号	特開2021-60729(P2021-60729A)		東京都大田区下丸子3丁目30番2号
(43)公開日	令和3年4月15日(2021.4.15)	(74)代理人	100126240
審査請求日	令和4年9月30日(2022.9.30)		弁理士 阿部 琢磨
		(74)代理人	100124442
			弁理士 黒岩 創吾
		(72)発明者	池田 元気
			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
		審査官	新井 則和

最終頁に続く

(54)【発明の名称】 画像処理システム、画像処理方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

手書き文字の画素を抽出するための手書き抽出処理を原稿の画像に対して実行することにより、前記原稿の画像から手書き文字の画素を抽出し、当該抽出した手書き文字の画素に基づいて手書き抽出画像を生成する抽出手段と、
手書き領域を推定するための領域推定処理を前記原稿の画像に対して実行することにより、前記原稿の画像から、手書き文字を含む手書き領域を推定する推定手段と、
前記抽出手段で生成された前記手書き抽出画像に対して、前記推定手段で推定された手書き領域ごとに文字認識処理を実行する処理手段と
を有することを特徴とする画像処理システム。

10

【請求項2】

前記手書き抽出処理は、原稿の画像から手書き文字の画素を抽出するための第1のニューラルネットワークの学習モデルを用いて行われる処理である、ことを特徴とする請求項1に記載の画像処理システム。

【請求項3】

前記領域推定処理は、原稿の画像から手書き領域を推定するための第2のニューラルネットワークの学習モデルを用いて行われる処理である、ことを特徴とする請求項1または2に記載の画像処理システム。

【請求項4】

前記第2のニューラルネットワークの学習モデルは、原稿の画像における記入欄内を手

20

書き領域とする正解データに基づいて学習処理を行うことにより得られたものであることを特徴とする請求項 3 に記載の画像処理システム。

【請求項 5】

前記第 2 のニューラルネットワークの学習モデルは、原稿の画像における記入欄内に記入された手書き文字を内包する領域を手書き領域とする正解データに基づいて学習処理を行うことにより得られたものであることを特徴とする請求項 3 に記載の画像処理システム。

【請求項 6】

前記第 2 のニューラルネットワークの学習モデルは、桁線で区切られた複数の手書き文字を内包する領域を手書き領域とする正解データに基づいて学習処理を行うことにより得られたものであることを特徴とする、請求項 3 に記載の画像処理システム。

10

【請求項 7】

手書き文字の前景サンプル画像と、手書き文字を含まず背景のみを含む背景サンプル画像とを合成して学習データを生成する生成手段を有し、

前記第 1 のニューラルネットワークの学習モデルは、前記生成手段によって生成された学習データに基づいて学習処理を行うことにより得られたものであることを特徴とする請求項 2 に記載の画像処理システム。

【請求項 8】

手書き文字の前景サンプル画像と、手書き文字を含まず背景のみを含む背景サンプル画像とを合成して学習データを生成する生成手段を有し、

前記第 2 のニューラルネットワークの学習モデルは、前記生成手段によって生成された学習データに基づいて学習処理を行うことにより得られたものであることを特徴とする請求項 3 に記載の画像処理システム。

20

【請求項 9】

前記手書き抽出処理は、前記原稿の画像内に含まれる輝度のばらつきが大きい細線を構成する画素を、前記手書き文字の画素として抽出する処理である、ことを特徴とする請求項 1 に記載の画像処理システム。

【請求項 10】

前記領域推定処理は、前記原稿の画像内に含まれる罫線を検出し、当該検出した罫線に囲まれる領域に基づいて前記手書き領域を推定する処理である、ことを特徴とする請求項 1 または 9 に記載の画像処理システム。

30

【請求項 11】

前記推定手段によって推定された手書き領域を、前記抽出手段によって抽出された手書き文字の画素に基づいて補正する補正手段をさらに有することを特徴とする請求項 1 に記載の画像処理システム。

【請求項 12】

手書き文字の画素を抽出するための手書き抽出処理を原稿の画像に対して実行することにより、前記原稿の画像から手書き文字の画素を抽出し、当該抽出した手書き文字の画素に基づいて手書き抽出画像を生成する抽出工程と、

手書き領域を推定するための領域推定処理を前記原稿の画像に対して実行することにより、前記原稿の画像から、手書き文字を含む手書き領域を推定する推定工程と、

40

前記抽出工程で生成された前記手書き抽出画像に対して、前記推定工程で推定された手書き領域ごとに文字認識処理を実行する処理工程と、
を有することを特徴とする画像処理方法。

【請求項 13】

コンピュータを、請求項 1 乃至 11 のいずれか 1 項に記載の画像処理システムの各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像処理システム、画像処理方法、及びプログラムに関する。

50

【背景技術】

【0002】

近年、原稿に手書きされた文字（手書き文字）を認識する技術が知られている。

【0003】

特許文献1には、原稿を読み取って得られた画像中の画素の輝度の特徴に基づいて、画素単位で手書き文字を認識する技術が開示されている。

【0004】

また、特許文献2には、帳票を読み取って得られた画像に含まれる罫線を認識し、罫線で囲われた内部の領域を手書き文字が記入された領域として抽出する技術が開示されている。

10

【先行技術文献】

【特許文献】

【0005】

【文献】特開2010-218106号公報

【文献】特開2017-138703号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

特許文献1の技術によれば、手書き文字を抽出した後の画像は、手書き抽出された文字しか含まない場合、帳票内に記入された文字間や項目間の区切りが不明瞭になる場合がある。例えば、図22(a)は帳票への氏名の記入例を示したものであり、図22(b)は図22(a)から抽出した手書き抽出画素を示したものである。図22(a)を参照すれば、罫線の存在により、記入された内容は「小木 才一郎」であることが分かる。しかし、その手書き抽出された画像である図22(b)では、「木」と「才」の文字間の区切りが不明瞭であるため、「木」と「才」が「材」の一文字として認識され、文字認識処理の結果、「小材一郎」と認識されてしまう場合がある。

20

【0007】

また、別の例を説明する。図22(c)は帳票への氏名の記入例を示したものであり、図22(d)は図22(c)から抽出した手書き抽出画素を示したものである。図22(c)を参照すれば、記入内容は「ALIVAIL」と分かる。しかし、図22(d)においては、「I」と「V」の文字間の区切りが不明瞭であり、「I」と「V」がくっつき、ローマ数字の4を示す文字として認識されてしまうため、文字認識処理の結果が誤って認識されてしまう場合がある。また、図22(e)のように、「001」と「8051」という複数の項目が記入されているような場合には、その手書き抽出画像は図22(f)のようになり、項目間の区切りが不明瞭となる。そして、手書きOCRにより「0018051」のように認識され、「001」「8051」なのか「0018」「051」なのか分からなくなってしまう。

30

【0008】

また、図22(g)のように、「マツ」「イクミ」と2つの項目が記入されている場合には、手書き抽出画像は手書きOCRにより「マツイクミ」と認識され、「マツ」「イクミ」なのか「マツイ」「クミ」なのか分からなくなってしまう。

40

【0009】

また、図22(h)のように、「LON」「GRAY」と2つの項目が記入されている場合には、手書き抽出画像は手書きOCRにより「LOGRAY」と認識され、「LON」「GRAY」なのか「LONG」「RAY」なのか分からなくなってしまう。

【0010】

また、特許文献2の技術によれば、罫線内を手書き領域とするが、例えば、図22(i)や図22(j)のように、罫線内には原稿に元々印刷された注記や記号などが含まれている場合がある。こうした注記や記号が手書き文字と共に文字認識処理の結果に含まれてしまい、正しく認識結果が得られなくなることがある。

50

【 0 0 1 1 】

特に、読み取られる原稿の罫線が予め認識されていない場合、どの範囲が一つの項目に該当する記入であるのかが分からないため、課題が顕著になる。

【 0 0 1 2 】

本発明は上記の課題に鑑みてなされたものである。本発明の目的は、原稿に手書きされた文字が不適切に結合してしまうことによる文字認識精度の低下を抑制することを目的とする。

【課題を解決するための手段】

【 0 0 1 3 】

本発明の画像処理システムは、手書き文字の画素を抽出するための手書き抽出処理を原稿の画像に対して実行することにより、前記原稿の画像から手書き文字の画素を抽出し、当該抽出した手書き文字の画素に基づいて手書き抽出画像を生成する抽出手段と、手書き領域を推定するための領域推定処理を前記原稿の画像に対して実行することにより、前記原稿の画像から、手書き文字を含む手書き領域を推定する推定手段と、前記抽出手段で生成された前記手書き抽出画像に対して、前記推定手段で推定された手書き領域ごとに文字認識処理を実行する処理手段とを有することを特徴とする。

10

【発明の効果】

【 0 0 1 4 】

本発明によれば、原稿に手書きされた文字が不適切に結合してしまうことによる文字認識精度の低下を抑制することができる。

20

【図面の簡単な説明】

【 0 0 1 5 】

【図 1】画像処理システムの構成を示した図である。

【図 2】画像処理装置の構成、学習装置の構成、画像処理サーバの構成、OCRサーバの構成を示す図である。

【図 3】画像処理システムの学習シーケンス、画像処理システムの利用シーケンスを示す図である。

【図 4】帳票の例を示す図である。

【図 5】本発明の実施形態における画面の例を示す図である。

【図 6】本発明の実施形態におけるフローチャートを示す図である。

30

【図 7】本発明の実施形態におけるフローチャートを示す図である。

【図 8】本発明の実施形態における学習データの構成例を示す図である。

【図 9】本発明の実施形態におけるフローチャートを示す図である。

【図 10】本発明の実施形態における帳票テキスト化処理におけるデータ生成処理の概要を示す図である。

【図 11】本発明の実施形態における画像処理システムの利用シーケンスを示す図である。

【図 12】本発明の実施形態におけるフローチャートを示す図である。

【図 13】本発明の実施形態におけるフローチャートを示す図である。

【図 14】本発明の実施形態における画面の例を示す図である。

【図 15】本発明の実施形態における画像処理システムの学習シーケンスを示す図である。

40

【図 16】本発明の実施形態における画面の例を示す図である。

【図 17】本発明の実施形態における合成領域定義情報のデータ構成を示す図である。

【図 18】本発明の実施形態におけるフローチャートを示す図である。

【図 19】本発明の実施形態におけるフローチャートを示す図である。

【図 20】本発明の実施形態における学習データの構成例を示す図である。

【図 21】処理対象画像と、手書き抽出および手書き領域推定のニューラルネットワークの推論結果の一例を示す図である。

【図 22】帳票の記入例を示す図である。

【発明を実施するための形態】

【 0 0 1 6 】

50

以下、本発明を実施するための形態について実施例にて具体的な構成を挙げ、図面を用いて説明する。なお、本発明を実現するための構成は実施例に記載された構成のみに限定されるものではない。同様の効果を得られる範囲で実施例に記載の構成の一部を省略または均等物に置き換えてもよい。

【0017】

なお、以降では、画素毎に手書き文字を抽出することを「手書き抽出」と呼ぶ。また、手書き抽出によって抽出された画素を「手書き抽出画素」と呼ぶ。また、手書き抽出により得た画像を「手書き抽出画像」と呼ぶ。手書き抽出画像に対して文字認識処理（OCR）を実行することを「手書きOCR」と呼ぶ。手書きOCRによって手書きされた文字は認識され、テキスト化（データ化）することができる。

10

【0018】

また、手書き文字を含む領域を推定することを「手書き領域推定」と呼ぶ。また、手書き領域推定によって得た領域を「手書き領域」と呼ぶ。スキャン画像中の手書き領域を、手書きOCRによって認識してテキスト化することができる。

【0019】

（実施例1）

本実施例では、手書き抽出および手書き領域推定を、ニューラルネットワークを用いて構成する例を示す。

【0020】

< 画像処理システム >

20

図1は画像処理システムの構成を示した図である。画像処理システム100は、画像処理装置101、学習装置102、画像処理サーバ103、OCRサーバ104より構成される。画像処理装置101と学習装置102、画像処理サーバ103、OCRサーバ104は、ネットワーク105を介して接続されている。

【0021】

画像処理装置101は、MFP（Multi Function Peripheral）等と呼ばれるデジタル複合機などであり、印刷機能やスキャン機能（画像取得部111としての機能）を有する。画像処理装置101は、帳票などの原稿をスキャンして画像データを生成する（以降この画像データを「原稿サンプル画像」と呼称する）。複数枚の原稿をスキャンして原稿サンプル画像を複数得る。これら原稿には、手書き記入が成されたものを含む。ネットワーク105を介して、原稿サンプル画像を学習装置102に送信する。また、画像処理装置101は、帳票のテキスト化を行う際に、手書き文字（手書き記号、手書き図形）が含まれる原稿をスキャンして処理対象とする画像データを得る（以降この画像データを「処理対象画像」と呼称する）。そして、画像処理装置101は、得られた処理対象画像を、ネットワーク105を介して画像処理サーバ103に送信する。

30

【0022】

学習装置102は、画像処理装置101が生成した原稿サンプル画像を蓄積する画像蓄積部115として機能する。またこのようにして蓄積した画像から学習データを生成する学習データ生成部112として機能する。学習データは、手書き抽出および手書き領域推定を行うニューラルネットワークを学習するために用いられるデータである。学習装置102は、生成した学習データを用いて、ニューラルネットワークの学習をおこなう学習部113として機能する。学習部113の学習処理により、学習結果（ニューラルネットワークのパラメータなど）が生成される。学習装置102は、学習結果（学習モデル）をネットワーク105を介して、画像処理サーバ103に送信する。ニューラルネットワークを用いる機械学習手法の一つとして、多層ニューラルネットワークを用いるディープラーニングが挙げられる。

40

【0023】

画像処理サーバ103は、処理対象画像を変換する画像変換部114として機能する。画像変換部114は、処理対象画像から手書きOCRの対象とする画像を生成する。すなわち、画像変換部114は、画像処理装置101が生成した処理対象画像に対して手書き

50

抽出を行う。この際、画像処理装置 101 は、学習装置 102 が生成した学習結果を用いることで、ニューラルネットワークにより推論して処理対象画像中の手書きの画素（画素位置）を抽出（特定）する。そして、手書き抽出画像を得る。さらに、画像処理サーバ 103 は、画像処理装置 101 が生成した処理対象画像に対して手書き領域推定を行う。画像処理装置 101 は、学習装置 102 が生成した学習結果を用いることで、ニューラルネットワークにより推論して処理対象画像中の手書き文字が記入された領域を推定（特定）する。これにより手書き領域を得る。ここで、この手書き領域の実態は、処理対象画像中の部分領域を示す情報であり、例えば、処理対象画像上の特定の画素位置（座標）と、当該画素位置からの幅や高さから成る情報として表現される。また、手書き領域は、帳票に記入された項目の数に応じて、複数得られる場合がある。ここで、手書き抽出画像は、処理対象画像中の手書きのみを抜き出した画像である。一方、手書き領域は、処理対象画像における個々の記入を各々区分した領域である。よって、手書き領域に基づいて手書き抽出画像上の部分領域を決定することで、手書きを個々の記入に分割して扱うことができる。そして、画像変換部 114 は、手書き抽出画像と手書き領域とを OCR サーバ 104 に送信する。これにより、OCR サーバ 104 に、手書き抽出画像上の領域であって、推定した手書き領域に該当する領域を、それぞれ手書き OCR の対象領域とすることを指示するものである。また、画像変換部 114 は、処理対象画像から手書き抽出画像に含まれる手書き画素を除去した画像（以降「活字画像」と呼称する）を生成する。そして、画像変換部 114 は、活字画像上の領域であって、活字 OCR の対象とする活字を含む領域（以降この領域を「活字領域」と呼称する）の情報を生成する。活字領域の生成については後述する。そして、画像変換部 114 は、活字画像と活字領域とを OCR サーバ 104 に送信する。これにより、OCR サーバ 104 に、活字画像上の活字領域を、それぞれ活字 OCR することを指示するものである。画像変換部 114 は、OCR サーバ 104 から手書き OCR 結果および活字 OCR 結果を受信する。そしてこれを統合し、テキストデータとして画像処理装置 101 に送信する。以降このテキストデータを「帳票テキストデータ」と呼称する。

10

20

【0024】

OCR サーバ 104 は、手書き OCR 部 116、活字 OCR 部 117 としての機能を備える。OCR サーバ 104 は、手書き抽出画像と手書き領域とを受信すると、手書き抽出画像中の手書き領域に対して OCR 処理を行ってテキストデータ（OCR 結果）を取得する。手書き OCR 部 116 は、当該テキストデータを画像処理サーバ 103 に送信する。また、OCR サーバ 104 は、活字画像と活字領域を受信すると、活字画像中の活字領域に対して OCR 処理を行ってテキストデータを取得する。活字 OCR 部 117 は、当該テキストデータを画像処理サーバ 103 に送信する。

30

【0025】

<学習シーケンス>

本システムにおける学習シーケンスについて説明する。図 3（a）は画像処理システムの学習シーケンスを示す図である。

【0026】

ステップ 301（以降の説明において S301 等と表記する）において、ユーザが原稿の読取指示を行うと、画像取得部 111 は、原稿を読みとって原稿サンプル画像を生成する（S302）。

40

【0027】

上述のように生成された原稿サンプル画像は、学習データ生成部 112 に送信される（S303）。なお、このとき、原稿サンプル画像に ID 情報を付与するとよい。この ID 情報は例えば、画像取得部 111 として機能する画像処理装置 101 を識別するための情報である。なお、ID 情報として、画像処理装置 101 を操作するユーザを識別するためのユーザ識別情報や、ユーザが所属するグループを識別するためのグループ識別情報であってもよい。

【0028】

50

画像が送信されてくると、学習データ生成部 112 は、画像蓄積部 115 に原稿サンプル画像を蓄積する (S304)。

【0029】

ユーザが学習装置 102 に原稿サンプル画像に対して正解データの付与指示を行うと (S305)、学習データ生成部 112 は当該正解データを取得する。そして原稿サンプル画像に紐づけて画像蓄積部 115 に蓄積する (S306)。正解データは、ニューラルネットワークの学習に用いるデータである。正解データの付与方法については後述する。そして、学習データ生成部 112 は、このようにして蓄積したデータに基づいて学習データを生成する (S307)。このとき、特定の ID 情報に基づく原稿サンプル画像のみを用いて学習データを生成してもよい。その後、学習データ生成部 112 は、学習部 113 に学習データを送信する (S308)、特定の ID 情報に基づく画像のみで学習データを生成した場合は、ID 情報も併せて送信する。学習部 113 は、受信した学習データに基づき学習処理を行い、学習モデルを更新する (S309)。学習部 113 は、ID 情報ごとに学習モデルを保持し、対応する学習データのみで学習をおこなってもよい。このように ID 情報と学習モデルを紐づけることで、特定の利用環境に特化した学習モデルを構築することができる。

10

【0030】

< 利用シーケンス >

本システムにおける利用シーケンスについて説明する。図 3 (b) は画像処理システムの利用シーケンスを示した図である。

20

【0031】

S351において、ユーザが原稿 (帳票) の読取指示を行うと、画像取得部 111 は、原稿を読み取って処理対象画像を生成する (S352)。ここで読み取られる画像は、例えば図 4 に示すような帳票 400 や帳票 490 である。これらの帳票は氏名の記入欄 (氏記入欄 401 と名記入欄 402 および氏名記入欄 451) や、住所記入欄 403 および 452、電話番号記入欄 404 および 453 を備え、それぞれ、氏名や住所、電話番号が手書きで記入されている。しかし、これら記入欄の配置 (帳票のレイアウト) は、帳票作成元により決定されるため、帳票毎に異なる (非定型帳票)。

【0032】

上述のように読み取られた処理対象画像は、画像変換部 114 に送信される (S353)。なお、このとき、送信データに ID 情報を付与するとよい。

30

【0033】

データを受信すると、画像変換部 114 は、処理対象画像のテキスト化指示を受け付ける (S354)。このとき、画像変換部 114 は、画像取得部 111 をデータの返信先として記憶する。手書き文字の加工指示を受け付けた画像変換部 114 は、最新の学習モデルを学習部 113 に要求する (S355)。これに応じて、学習部 113 は最新の学習モデルを画像変換部 114 に送信する (S356)。画像変換部 114 からの要求時に ID 情報が指定されていた場合は、ID 情報に対応する学習モデルを送信する。画像変換部 114 は、取得した学習モデルに基づいて、処理対象画像に対して、手書き抽出および手書き領域推定を行う (S357)。そして、これにより生成された手書き抽出画像と手書き領域とを手書き OCR 部 116 に送信する (S358)。手書き OCR 部 116 は、手書き抽出画像上の手書き領域に該当する部分領域それぞれについて手書き OCR 処理を施し、テキストデータ (手書き) を取得する (S359)。手書き OCR 部 116 は、取得したテキストデータ (手書き) を画像変換部 114 に送信する (S360)。続けて、画像変換部 114 は処理対象画像から活字画像と活字領域とを生成する (S361)。そして、活字 OCR 部 117 に活字画像と活字領域を送信する (S362)。活字 OCR 部 117 は、活字画像に活字 OCR 処理を施し、テキストデータ (活字) を取得する (S363)。そして、取得したテキストデータ (活字) を画像変換部 114 に送信する (S364)。その後、画像変換部 114 は、テキストデータ (手書き)、テキストデータ (活字) に少なくとも基づいて帳票テキストデータを生成する (S365)。画像変換部 114 は

40

50

、帳票テキストデータを画像取得部 111 に送信する (S366)。帳票テキストデータを取得した画像取得部 111 は、帳票テキストデータの利用画面をユーザに提示する (S367)。このあと、画像取得部 111 は、帳票テキストデータの利用用途に応じて、帳票テキストデータを出力する。例えば、別体外部の業務システム (不図示) に送信したり、印刷したりして出力する。

【0034】

<装置構成>

上述したシステムを実現するために、各装置は次のような構成を備える。図 2 (a) は画像処理装置の構成を示す図である。図 2 (b) は学習装置の構成を示す図である。図 2 (c) は画像処理サーバの構成を示す図である。図 2 (d) は OCR サーバの構成を示す図である。

10

【0035】

図 2 (a) に示すように、画像処理装置 101 は、次を備える。CPU 201、ROM 202、RAM 204、プリンタデバイス 205、スキャナデバイス 206、原稿搬送デバイス 207、ストレージ 208、入力デバイス 209、表示デバイス 210、及び外部インタフェース 211 を備える。各デバイスは、データバス 203 によって相互通信可能に接続されている。

【0036】

CPU 201 は、画像処理装置 101 を統括的に制御するためのコントローラである。CPU 201 は、ROM 202 に格納されているブートプログラムにより OS (オペレーティングシステム) を起動する。この OS 上で、ストレージ 208 に記憶されているコントローラプログラムが実行される。コントローラプログラムは、画像処理装置 101 を制御するためのプログラムである。CPU 201 は、データバス 203 によって接続されている各デバイスを統括的に制御する。RAM 204 は、CPU 201 の主メモリやワークエリア等の一時記憶領域として動作する。

20

【0037】

プリンタデバイス 205 は、画像データを用紙 (記録材、シート) 上に印刷するものである。これには感光体ドラムや感光体ベルトなどを用いた電子写真印刷方式や、微小ノズルアレイからインクを吐出して用紙上に直接画像を印字するインクジェット方式などがあるが、どの方式でもかまわない。スキャナデバイス 206 は、CCD などの光学読取装置を用いて紙などの原稿上の走査を行い、電気信号データを得てこれを変換し、画像データを生成する。また、ADF (オート・ドキュメント・フィーダ) などの原稿搬送デバイス 207 は、原稿搬送デバイス 207 上の原稿台に載置された原稿を 1 枚ずつスキャナデバイス 206 に搬送する。

30

【0038】

ストレージ 208 は、HDD や SSD などの、読み出しと書き込みが可能な不揮発メモリであり、ここには、前述のコントローラプログラムなど、様々なデータが記録される。入力デバイス 209 は、タッチパネルやハードキーなどから構成される入力装置である。入力デバイス 209 は、ユーザの操作指示を受け付ける。そして、指示位置を含む指示情報を CPU 201 に伝達する。表示デバイス 210 は、LCD や CRT などの表示装置である。表示デバイス 210 は、CPU 201 が生成した表示データを表示する。CPU 201 は、入力デバイス 209 より受信した指示情報と、表示デバイス 210 に表示させている表示データとから、いずれの操作が成されたかを判定する。そしてこの判定結果に応じて、画像処理装置 101 を制御するとともに、新たな表示データを生成し表示デバイス 210 に表示させる。

40

【0039】

外部インタフェース 211 は、LAN や電話回線、赤外線といった近接無線などのネットワークを介して、外部機器と、画像データをはじめとする各種データの送受信を行う。外部インタフェース 211 は、学習装置 102 や PC (不図示) などの外部機器より、PDL データを受信する。CPU 201 は、外部インタフェース 211 が受信した PDL デ

50

ータを解釈し、画像を生成する。生成した画像は、プリンタデバイス 205 により印刷したり、ストレージ 108 に記憶したりする。また、外部インタフェース 211 は、画像処理サーバ 103 などの外部機器より画像データを受信する。受信した画像データをプリンタデバイス 205 により印刷したり、ストレージ 108 に記憶したり、外部インタフェース 211 により、他の外部機器に送信したりする。

【0040】

図 2 (b) の学習装置 102 は、CPU 231、ROM 232、RAM 234、ストレージ 235、入力デバイス 236、表示デバイス 237、外部インタフェース 238、GPU 239 を備える。各部は、データバス 233 を介して相互にデータを送受信することができる。

10

【0041】

CPU 231 は、学習装置 102 の全体を制御するためのコントローラである。CPU 231 は、不揮発メモリである ROM 232 に格納されているブートプログラムにより OS を起動する。この OS の上で、ストレージ 235 に記憶されている学習データ生成プログラムおよび学習プログラムを実行する。CPU 231 が学習データ生成プログラムを実行することにより、学習データを生成する。また、CPU 231 が学習プログラムを実行することにより、手書き抽出を行うニューラルネットワークを学習する。CPU 231 は、データバス 233 などのバスを介して各部を制御する。

【0042】

RAM 234 は、CPU 231 のメインメモリやワークエリア等の一時記憶領域として動作するものである。ストレージ 235 は、読み出しと書き込みが可能な不揮発メモリであり、前述の学習データ生成プログラムや学習プログラムを記録する。

20

【0043】

入力デバイス 236 は、マウスやキーボードなどから構成される入力装置である。表示デバイス 237 は、図 2 (a) を用いて説明した表示デバイス 210 と同様である。

【0044】

外部インタフェース 238 は、図 2 (a) を用いて説明した外部インタフェース 211 と同様である。

【0045】

GPU 239 は、画像処理プロセッサであり、CPU 231 と協調して画像データの生成やニューラルネットワークの学習を行う。

30

【0046】

図 2 (c) の画像処理サーバ 103 は、CPU 261、ROM 262、RAM 264、ストレージ 265、入力デバイス 266、表示デバイス 267、外部インタフェース 268 を備える。各部は、データバス 263 を介して相互にデータを送受信することができる。

【0047】

CPU 261 は、画像処理サーバ 103 の全体を制御するためのコントローラである。CPU 261 は、不揮発メモリである ROM 262 に格納されているブートプログラムにより OS を起動する。この OS の上で、ストレージ 265 に記憶されている画像処理サーバプログラムを実行する。CPU 261 がこの画像処理サーバプログラムを実行することにより、処理対象画像に対して手書き抽出および手書き領域推定を行う。CPU 261 は、データバス 263 などのバスを介して各部を制御する。

40

【0048】

RAM 264 は、CPU 261 のメインメモリやワークエリア等の一時記憶領域として動作するものである。ストレージ 265 は、読み出しと書き込みが可能な不揮発メモリであり、前述の画像処理プログラムを記録する。

【0049】

入力デバイス 266 は、図 2 (b) を用いて説明した入力デバイス 236 と同様である。表示デバイス 267 は、図 2 (a) を用いて説明した表示デバイス 210 と同様である。

【0050】

50

外部インタフェース 268 は、図 2 (a) を用いて説明した外部インタフェース 211 と同様である。

【 0051 】

図 2 (d) の OCR サーバ 104 は、CPU 291、ROM 292、RAM 294、ストレージ 295、入力デバイス 296、表示デバイス 297、外部インタフェース 298 を備える。各部は、データバス 293 を介して相互にデータを送受信することができる。

【 0052 】

CPU 291 は、OCR サーバ 104 の全体を制御するためのコントローラである。CPU 291 は、不揮発メモリである ROM 292 に格納されているブートプログラムにより OS を起動する。この OS の上で、ストレージ 295 に記憶されている OCR サーバプログラムを実行する。CPU 291 がこの OCR サーバプログラムを実行することより、手書き抽出画像や活字画像の手書き文字や活字を認識してテキスト化する。CPU 291 は、データバス 293 などのバスを介して各部を制御する。

【 0053 】

RAM 294 は、CPU 291 のメインメモリやワークエリア等の一時記憶領域として動作するものである。ストレージ 295 は、読み出しと書き込みが可能な不揮発メモリであり、前述の画像処理プログラムを記録する。

【 0054 】

入力デバイス 296 は、図 2 (b) を用いて説明した入力デバイス 236 と同様である。表示デバイス 297 は、図 2 (a) を用いて説明した表示デバイス 210 と同様である。

【 0055 】

外部インタフェース 298 は、図 2 (a) を用いて説明した外部インタフェース 211 と同様である。

【 0056 】

< 操作画面 >

S301 に示したユーザの指示は、次のような操作画面で行われる。図 5 (a) は学習原稿スキャン画面を示す図である。

【 0057 】

学習原稿スキャン画面 500 は、表示デバイス 210 に表示される画面の一例である。図 5 (a) に示すように、学習原稿スキャン画面 500 は、プレビュー領域 501、スキャンボタン 502、送信開始ボタン 503 を備える。

【 0058 】

スキャンボタン 502 はスキャナデバイス 206 にセットされた原稿の読取を開始するためのボタンである。スキャンが完了すると、原稿サンプル画像が生成され、プレビュー領域 501 に表示される。スキャナデバイス 206 に別の原稿をセットし、スキャンボタン 502 を再び押すことで、複数の原稿サンプル画像をまとめて保持しておくこともできる。

【 0059 】

原稿が読み取られると、送信開始ボタン 503 が指示可能となる。送信開始ボタン 503 を指示すると、原稿サンプル画像が学習装置 102 に送信される。

【 0060 】

S305 に示したユーザの指示は、次のような操作画面で行われる。図 5 (b) は手書き抽出正解データ作成画面を示す図である。図 5 (c) は手書き領域推定正解データ作成画面を示す図である。ユーザは、手書き抽出正解データ作成画面および手書き領域推定正解データ作成画面の表示内容に基づいて操作し、正解データを作成する。

【 0061 】

手書き抽出正解データ作成画面 520 は、表示デバイス 237 に表示される画面の一例である。図 5 (b) に示すように、手書き抽出正解データ作成画面 520 は、画像表示領域 521、画像選択ボタン 522、拡大ボタン 523、縮小ボタン 524、抽出ボタン 525、推定ボタン 526、保存ボタン 527 を備える。

10

20

30

40

50

【 0 0 6 2 】

画像選択ボタン 5 2 2 は、画像処理装置 1 0 1 から受信し、画像蓄積部 1 1 5 に蓄積されている原稿サンプル画像を選択するためのボタンである。画像選択ボタン 5 2 2 を指示すると、選択画面（不図示）が表示され、原稿サンプル画像を選択することができる。原稿サンプル画像を選択すると、画像表示領域 5 2 1 に選択した原稿サンプル画像が表示される。ユーザは画像表示領域 5 2 1 に表示された原稿サンプル画像に対して操作し、正解データを作成する。

【 0 0 6 3 】

拡大ボタン 5 2 3 と縮小ボタン 5 2 4 は、画像表示領域 5 2 1 の表示を拡大あるいは縮小するためのボタンである。拡大ボタン 5 2 3 や縮小ボタン 5 2 4 を指示することにより、正解データの作成が行いやすいよう、画像表示領域 5 2 1 に表示されている原稿サンプル画像を拡大および縮小して表示することができる。

【 0 0 6 4 】

抽出ボタン 5 2 5 および推定ボタン 5 2 6 は、手書き抽出と手書き領域推定のいずれの正解データを作成するかを選択するためのボタンである。いずれかを選択すると、選択したボタンは強調して表示される。抽出ボタン 5 2 5 を選択すると、手書き抽出の正解データを作成する状態となる。このボタンを選択した場合、ユーザは、手書き抽出の正解データを次のように操作して作成する。ユーザは、図 5（c）に示すように、入力デバイス 2 3 6 を介してマウスカースルを操作し、画像表示領域 5 2 1 に表示されている原稿サンプル画像中の手書き文字をなぞって選択する。学習データ生成部 1 1 2 は、この操作を受信すると、上記操作により選択された原稿サンプル画像上の画素位置を記録する。すなわち、手書き抽出の正解データは、原稿サンプル画像上の、手書きに該当する画素の位置である。一方、推定ボタン 5 2 6 を選択すると、手書き領域推定の正解データを作成する状態となる。このボタンを選択した場合、ユーザは、手書き領域推定の正解データを次のように操作して作成する。ユーザは、図 5（c）に点線枠で示すように、入力デバイス 2 3 6 を介してマウスカースルを操作し、画像表示領域 5 2 1 に表示されている原稿サンプル画像中の手書き文字が記入されている罫線に囲まれた領域内（記入欄内であり罫線は含まない）を選択する。これはすなわち、帳票の記入欄毎に領域を選択する操作である。学習データ生成部 1 1 2 は、この操作を受信すると、上記操作により選択された領域を記録する。すなわち、手書き領域推定の正解データは、原稿サンプル画像上の、記入欄内の領域（手書きが記入され得る領域）である（以降、手書きが記入された領域を「手書き領域」と呼称する）。

【 0 0 6 5 】

保存ボタン 5 2 7 は、作成された正解データを保存するためのボタンである。手書き抽出の正解データは、次のような画像として、画像蓄積部 1 1 5 に蓄積される。原稿サンプル画像と同じサイズ（幅および高さ）を有する。ユーザにより選択された手書き文字位置の画素の値は、手書きを示す値（例えば 2 5 5、以降も同様）である。それ以外の画素の値は、手書きではないことを示す値（例えば 0、以降も同様）である。以降、このような手書き抽出の正解データである画像を「手書き抽出正解画像」と呼称する。手書き抽出正解画像の例を図 4（c）に示す。また、手書き領域推定の正解データは、次のような画像として、画像蓄積部 1 1 5 に蓄積される。原稿サンプル画像と同じサイズ（幅および高さ）を有する。ユーザにより選択された手書き領域に該当する画素の値は、手書き領域であることを示す値（例えば 2 5 5、以降も同様）である。また、それ以外の画素の値は手書き領域ではないことを示す値（例えば 0、以降も同様）である。以降、このような手書き領域推定の正解データである画像を「手書き領域推定正解画像」と呼称する。手書き領域推定正解画像の例を図 4（d）に示す。

【 0 0 6 6 】

S 3 5 1 に示したユーザの指示は、次のような操作画面で行われる。図 5（d）は帳票処理画面を示す図である。図 5（d）に示すように、帳票処理画面 5 2 0 は、プレビュー領域 5 4 1、スキャンボタン 5 4 2、送信開始ボタン 5 4 3 を備える。

10

20

30

40

50

【 0 0 6 7 】

スキャンボタン 5 4 2 はスキャナデバイス 2 0 6 にセットされた原稿の読取を開始するためのボタンである。スキャンが完了すると、処理対象画像が生成され、プレビュー領域 5 4 1 に表示される。

【 0 0 6 8 】

原稿が読み取られると、送信開始ボタン 5 4 3 が指示可能となる。送信開始ボタン 5 4 3 を指示すると、処理対象画像が学習サーバ 1 0 3 に送信される。

【 0 0 6 9 】

< 原稿サンプル画像生成処理 >

次に、画像処理装置 1 0 1 による原稿サンプル画像生成処理について説明する。図 6 (a) は原稿サンプル画像生成処理のフローを示す図である。この処理は、C P U 2 0 1 が、ストレージ 2 0 8 に記録されているコントローラプログラムを読み出し、R A M 2 0 4 に展開して実行することで実現される。これは、ユーザが、画像処理装置 1 0 1 の入力デバイス 2 0 9 を操作することにより開始される。

10

【 0 0 7 0 】

C P U 2 0 1 は、S 6 0 1 において、原稿のスキャン指示が成されたか否かを判定する。ユーザが、入力デバイス 2 0 9 を介して、原稿をスキャンするための所定の操作（スキャンボタン 5 0 2 の指示）を行った場合には、Y E S と判定し、S 6 0 2 に遷移する。そうでなければ、N O と判定し、S 6 0 4 に遷移する。

【 0 0 7 1 】

C P U 2 0 1 は、S 6 0 2 において、スキャナデバイス 2 0 6 や原稿搬送デバイス 2 0 7 を制御して、原稿をスキャンして原稿サンプル画像を生成する。原稿サンプル画像は、グレースケールの画像データとして生成される。

20

【 0 0 7 2 】

C P U 2 0 1 は、S 6 0 3 において、S 6 0 2 で生成した原稿サンプル画像を、外部インタフェース 2 1 1 を介して、学習装置 1 0 2 に送信する。

【 0 0 7 3 】

C P U 2 0 1 は、S 6 0 4 において、処理を終了するか否かを判定する。ユーザが、原稿サンプル画像生成処理を終了する所定の操作を行った場合には、Y E S と判定して、処理を終了する。そうでなければ、N O と判定し、S 6 0 1 に遷移する。

30

【 0 0 7 4 】

以上の処理によって、画像処理装置 1 0 1 は、原稿サンプル画像を生成して学習装置 1 0 2 に送信する。ユーザの操作や、原稿搬送デバイス 2 0 7 に載置した原稿枚数に応じて、原稿サンプル画像が複数取得される。

【 0 0 7 5 】

< 原稿サンプル画像受信処理 >

次に、学習装置 1 0 2 による原稿サンプル画像受信処理について説明する。図 6 (b) は原稿サンプル画像受信処理のフローを示す図である。この処理は、C P U 2 3 1 が、ストレージ 2 3 5 に記録されている学習データ生成プログラムを読み出し、R A M 2 3 4 に展開して実行することで実現される。これは、ユーザが、学習装置 1 0 2 の電源を O N (オン) にすると開始される。

40

【 0 0 7 6 】

C P U 2 3 1 は、S 6 2 1 において、原稿サンプル画像を受信したか否かを判定する。C P U 2 3 1 は、外部インタフェース 2 3 8 を介して画像データを受信していたならば、Y E S と判定し、S 6 2 2 に遷移する。そうでなければ、N O と判定し、S 6 2 3 に遷移する。

【 0 0 7 7 】

C P U 2 3 1 は、S 6 2 2 において、受信した原稿サンプル画像を、ストレージ 2 3 5 の所定の領域に記録する。

【 0 0 7 8 】

50

C P U 2 3 1 は、S 6 2 3 において、処理を終了するか否かを判定する。ユーザが、学習装置 1 0 2 の電源の O F F などの、原稿サンプル画像受信処理を終了する所定の操作を行った場合には、Y E S と判定して、処理を終了する。そうでなければ、N O と判定し、S 6 2 1 に遷移する。

【 0 0 7 9 】

< 正解データ生成処理 >

次に、学習装置 1 0 2 による正解データ生成処理について説明する。図 6 (c) は正解データ生成処理のフローを示す図である。

【 0 0 8 0 】

この処理は、学習装置 1 0 2 の学習データ生成部 1 1 2 により実現される。これは、ユーザが、学習装置 1 0 2 の入力デバイス 2 3 6 を介して、所定の操作を行うことで開始される。

10

【 0 0 8 1 】

C P U 2 3 1 は、S 6 4 1 において、原稿サンプル画像の選択指示が成されたか否かを判定する。ユーザが、入力デバイス 2 3 6 を介して、原稿サンプル画像を選択するための所定の操作（画像選択ボタン 5 2 2 の指示）を行った場合には、Y E S と判定し、S 6 4 2 に遷移する。そうでなければ、N O と判定し、S 6 4 3 に遷移する。

【 0 0 8 2 】

C P U 2 3 1 は、S 6 4 2 において、S 6 4 1 でユーザが選択した原稿サンプル画像を、ストレージ 2 3 5 から読み出してユーザに対して出力する（画像表示領域 5 2 1 に表示）。

20

【 0 0 8 3 】

C P U 2 3 1 は、S 6 4 3 において、ユーザが正解データの入力指示を行ったか否かを判定する。ユーザが、入力デバイス 2 3 6 を介して、前述したように、原稿サンプル画像上の手書き文字をなぞる、あるいは、手書き文字が記入されている罫線枠をなぞる操作を行っていたならば、Y E S と判定し、S 6 4 4 に遷移する。そうでなければ、N O と判定し、S 6 4 7 に遷移する。

【 0 0 8 4 】

C P U 2 3 1 は、S 6 4 4 において、ユーザが入力した正解データは、手書き抽出の正解データであるか否かを判定する。C P U 2 3 1 は、ユーザが手書き抽出の正解データ作成を指示する操作を行っていたならば（抽出ボタン 5 2 5 の選択）、Y E S と判定し、S 6 4 5 に遷移する。そうでない場合、すなわち、ユーザが入力した正解データは手書き領域推定の正解データである場合（推定ボタン 5 2 6 を選択している）、S 6 4 6 に遷移する。

30

【 0 0 8 5 】

C P U 2 3 1 は、S 6 4 5 において、ユーザが入力した手書き抽出の正解データを、R A M 2 3 4 に一時的に記憶する。前述のとおり、手書き抽出の正解データは、原稿サンプル画像中の手書きに該当する画素の位置情報である。

【 0 0 8 6 】

C P U 2 3 1 は、S 6 4 6 において、ユーザが入力した手書き領域推定の正解データを R A M 2 3 4 に一時的に記憶する。前述のとおり、手書き領域推定の正解データは、原稿サンプル画像上の、手書き領域に該当する領域情報である。

40

【 0 0 8 7 】

C P U 2 3 1 は、S 6 4 7 において、正解データの保存指示が成されたか否かを判定する。ユーザが、入力デバイス 2 3 6 を介して、正解データを保存するための所定の操作（保存ボタン 5 2 7 の指示）を行った場合には、Y E S と判定し、S 6 4 8 に遷移する。そうでなければ、N O と判定し、S 6 5 0 に遷移する。

【 0 0 8 8 】

C P U 2 3 1 は、S 6 4 8 において、手書き抽出正解画像を生成し、手書き抽出の正解データとして保存する。C P U 2 3 1 は、次のようにして手書き抽出正解画像を生成する

50

。CPU231は、手書き抽出正解画像として、S642で読み出した原稿サンプル画像と同じサイズの画像を生成する。当該画像の全ての画素を、手書きではないことを示す値にする。次いで、S645においてRAM234に一時的に記憶した位置情報を参照し、手書き抽出正解画像上の該当する位置の画素の値を、手書きであることを示す値に変更する。このようにして生成した手書き抽出正解画像を、S642で読み出した原稿サンプル画像と関連付けて、ストレージ235の所定の領域に保存する。

【0089】

CPU231は、S649において、手書き領域推定正解画像を生成し、手書き領域推定の正解データとして保存する。CPU231は、次のようにして手書き領域推定正解画像を生成する。CPU231は、手書き領域推定正解画像として、S642で読み出した原稿サンプル画像と同じサイズの画像を生成する。当該画像の全ての画素を、手書き領域ではないことを示す値にする。次いで、S646においてRAM234に一時的に記憶した領域情報を参照し、手書き領域推定正解画像上の該当する領域内の画素の値を、手書き領域であることを示す値に変更する。このようにして生成した手書き領域推定正解画像を、S642で読み出した原稿サンプル画像と関連付けて、ストレージ235の所定の領域に保存する。

【0090】

CPU231は、S650において、処理を終了するか否かを判定する。ユーザが、正解データ生成処理を終了する所定の操作を行った場合には、YESと判定して、処理を終了する。そうでなければ、NOと判定し、S641に遷移する。

【0091】

<学習データ生成処理>

次に、学習装置102による学習データ生成処理について説明する。図7(a)は学習データ生成処理のフローを示す図である。この処理は、学習装置102の学習データ生成部112により実現される。これは、ユーザが、画像処理装置101の入力デバイス209を介して、所定の操作を行うことで開始される。

【0092】

まずCPU231は、S701において、ストレージ235に記憶している原稿サンプル画像を選択して読み出す。図6(b)のフローチャートのS622の処理ステップにより、ストレージ235には複数の原稿サンプル画像が記録されているので、その中からランダムにひとつを選択する。

【0093】

CPU231は、S702において、ストレージ235に記憶している手書き抽出正解画像を読み出す。S648の処理によって、S701で読み出した原稿サンプル画像に関連付けられた手書き抽出正解画像がストレージ235に記憶されているので、これを読み出す。

【0094】

CPU231は、S703において、ストレージ235に記憶している手書き領域推定正解画像を読み出す。S649の処理によって、S701で読み出した原稿サンプル画像に関連付けられた手書き領域推定正解画像がストレージ235に記憶されているので、これを読み出す。

【0095】

CPU231は、S704において、S701で読み出した原稿サンプル画像中の一部（例えば縦×横＝256×256の大きさ）を切り出して、学習データに用いる入力画像を生成する。切り出す位置はランダムに決定する。

【0096】

CPU231は、S705において、S702で読み出した手書き抽出正解画像中の一部を切り出して、手書き抽出の学習データに用いる正解ラベル画像（教師データ、正解画像データ）を生成する。以降この正解ラベル画像を「手書き抽出正解ラベル画像」と呼称する。切り出す位置およびサイズは、S704で原稿サンプル画像から入力画像を切り出

10

20

30

40

50

した位置およびサイズと同様とする。

【0097】

CPU231は、S706において、S703で読み出した手書き領域推定正解画像中の一部を切り出して、手書き領域推定の学習データに用いる正解ラベル画像を生成する（以降この正解ラベル画像を「手書き領域推定正解ラベル画像」と呼称する）。切り出す位置およびサイズは、S704で原稿サンプル画像から入力画像を切り出した位置およびサイズと同様とする。

【0098】

CPU231は、S707において、S704で生成した入力画像と、S706で生成した手書き抽出正解ラベル画像とを対応付け、手書き抽出の学習データとしてストレージ235の所定の領域に保存する。本実施例では、図8（a）のような学習データが保存される。

10

【0099】

CPU231は、S708において、S704で生成した入力画像と、S707で生成した手書き領域推定正解ラベル画像とを対応付け、手書き領域推定の学習データとしてストレージ235の所定の領域に保存する。本実施例では、図8（b）のような学習データが保存される。

【0100】

CPU231は、S709において、学習データ生成処理を終了するか否かを判定する。CPU231は、予め決定した学習データの数（本フローチャートの開始時に、学習置102の入力デバイス236を介して、ユーザが指定するなどして決定）だけ学習データを生成していたならば、YESと判定し、処理を終了する。そうでなければ、NOと判定し、S701に遷移する。

20

【0101】

以上により、手書き抽出を行うニューラルネットワークの学習データと、手書き領域推定を行うニューラルネットワークの学習データが生成される。ニューラルネットワークの汎用性を高めるために、学習データの加工を行っても良い。例えば、入力画像を所定の範囲（例えば、50%～150%の間）からランダムに選択して決定する変倍率で変倍する。手書き抽出および手書き領域推定の正解ラベル画像も同様に変倍する。あるいは、入力画像を所定の範囲（例えば、-10度～10度の間）からランダムに選択して決定する回転角度で回転する。手書き抽出および手書き領域推定の正解ラベル画像も同様に回転する。変倍や回転を考慮すれば、S704やS705、S706で入力画像や手書き抽出および手書き領域推定の正解ラベル画像を切り出す際に、少し大きめのサイズ（例えば、縦×横＝512×512の大きさ）で切り出す。そして、変倍および回転後に、最終的な入力画像や手書き抽出および手書き領域推定の正解ラベル画像のサイズ（例えば、縦×横＝256×256）となるよう、中心部分から切り出す。あるいは、入力画像の各画素の輝度を変更して加工してもよい。すなわち、ガンマ補正を用いて入力画像の輝度を変更する。ガンマ値は所定の範囲（例えば、0.1～10.0の間）からランダムに選択して決定する。

30

【0102】

<学習処理>

次に、学習装置102による学習処理について説明する。図7（b）は学習処理のフローを示す図である。この処理は、学習装置102の学習部113により実現される。これは、ユーザが、学習装置102の入力デバイス236を介して、所定の操作を行うことで開始される。なお、本実施例において、ニューラルネットワークの学習には、ミニバッチ法を用いるものとする。

40

【0103】

まずCPU231は、S731において、手書き抽出と手書き領域推定のニューラルネットワークをそれぞれ初期化する。すなわち、CPU231は、2つのニューラルネットワークを構築し、これらニューラルネットワークに含まれる各パラメタの値を、ランダム

50

に決定して初期化する。これらニューラルネットワークの構造は、様々なものを用いることができるが、例えば、公知技術であるFCN(Fully Convolutional Networks)の形態を取り得る。手書き領域推定のニューラルネットワークについては、他にも、例えば、公知技術であるYOLO(You Only Look Once)の形態を取り得る。

【0104】

CPU231は、S732において、学習データを取得する。CPU231は、図7(a)のフローチャートに示した、学習データ生成処理を実行して、所定の数(ミニバッチサイズ、例えば10)の学習データを取得する。

【0105】

CPU231は、S733において、手書き抽出のニューラルネットワークの誤差を算出する。すなわち、手書き抽出の各学習データに含まれる入力画像を手書き抽出のニューラルネットワークに入力して出力を得る。当該出力は、入力画像と同じ画像サイズであり、予測結果として、手書きであると判定された画素は、画素の値が手書きを示す値、そうではないと判定された画素は、画素の値が手書きではないことを示す値である画像である。そして、当該出力と学習データに含まれる手書き抽出正解ラベル画像との差を評価して誤差を求める。当該評価には指標として交差エントロピーを用いることができる。

【0106】

CPU231は、S734において、手書き抽出のニューラルネットワークのパラメタを調整する。すなわち、S733において算出した誤差をもとに、バックプロパゲーション法によって手書き抽出のニューラルネットワークのパラメタ値を変更するものである。

【0107】

CPU231は、S735において、手書き領域推定のニューラルネットワークの誤差を算出する。すなわち、手書き領域推定の各学習データに含まれる入力画像を手書き領域推定のニューラルネットワークに入力して出力を得る。当該出力は、入力画像と同じ画像サイズであり、予測結果として、手書き領域であると判定された画素は、画素の値が手書き領域を示す値、そうではないと判定された画素は、画素の値が手書き領域ではないことを示す値である画像である。そして、当該出力と学習データに含まれる手書き領域推定正解ラベル画像との差を評価して誤差を求める。当該評価の指標には、手書き抽出と同様、交差エントロピーを用いることができる。

【0108】

CPU231は、S736において、手書き領域推定のニューラルネットワークのパラメタを調整する。すなわち、S735において算出した誤差をもとに、バックプロパゲーション法によって手書き領域推定のニューラルネットワークのパラメタ値を変更するものである。

【0109】

CPU231は、S737において、学習を終了するか否かを判定する。これは次のようにして行う。CPU231は、S732～S736の処理を、所定回数(例えば、60000回)行ったか否かを判定する。当該所定回数は、本フローチャートの開始時にユーザが操作入力するなどして決定することができる。所定回数行った場合には、YESと判定し、S738に遷移する。そうでない場合は、S732に遷移し、ニューラルネットワークの学習を続ける。

【0110】

CPU231は、S738において、学習結果として、S734とS736において調整した手書き抽出および手書き領域推定のニューラルネットワークのパラメタを、それぞれ、画像処理サーバ103に送信する。

【0111】

< 帳票テキスト化依頼処理 >

次に、画像処理装置101による、帳票テキスト化依頼処理について説明する。画像処理装置101は、手書き記入が成された帳票をスキャンして処理対象画像を生成する。そ

10

20

30

40

50

して、処理対象画像データを画像処理サーバ１０３に送信して、帳票テキスト化を依頼する。図９（ａ）は帳票テキスト化依頼処理のフローを示す図である。この処理は、画像処理装置１０１のＣＰＵ２０１が、ストレージ２０８に記録されているコントローラプログラムを読み出し、ＲＡＭ２０４に展開して実行することにより実現される。これは、ユーザが、画像処理装置１０１の入力デバイス２０９を介して、所定の操作を行うことで開始される。

【０１１２】

まずＣＰＵ２０１は、Ｓ９０１において、スキャナデバイス２０６や原稿搬送デバイス２０７を制御して、原稿をスキャンして処理対象画像を生成する。処理対象画像は、グレースケールの画像データとして生成される。

【０１１３】

ＣＰＵ２０１は、Ｓ９０２において、Ｓ９０１で生成した処理対象画像を、外部インタフェース２１１を介して、画像処理サーバ１０３に送信する。

【０１１４】

ＣＰＵ２０１は、Ｓ９０３において、画像処理サーバ１０３から、処理結果を受信したか否かを判定する。画像処理サーバ１０３から、外部インタフェース２１１を介して、処理結果を受信していた場合には、ＹＥＳと判定し、Ｓ９０４に遷移する。そうでない場合には、ＮＯと判定し、Ｓ９０３の処理ステップを繰り返す。

【０１１５】

ＣＰＵ２０１は、Ｓ９０４において、画像処理サーバ１０３から受信した処理結果、すなわち、Ｓ９０１で生成した処理対象画像に含まれる手書き文字や活字を認識して生成した帳票テキストデータを出力する。例えば、ユーザが、入力デバイス２０９を操作して設定した送信宛先に、外部インタフェース２１１を介して、帳票テキストデータを送信することができる。

【０１１６】

< 帳票テキスト化処理 >

次に、画像処理サーバ１０３による帳票テキスト化処理について説明する。図９（ｂ）は帳票テキスト化処理のフローを示す図である。図１０は、帳票テキスト化処理における、データ生成処理の概要を示す図である。画像変換部１１４として機能する画像処理サーバ１０３は、画像処理装置１０１から処理対象画像を受信し、当該スキャン画像データに含まれる活字や手書き文字をＯＣＲしてテキストデータを得る。活字に対するＯＣＲは、活字ＯＣＲ部１１７に実行させる。手書き文字に対するＯＣＲは、手書きＯＣＲ部１１６に実行させる。帳票テキスト化処理は、ＣＰＵ２６１が、ストレージ２６５に記憶されている画像処理サーバプログラムを読み出し、ＲＡＭ２６４に展開して実行することで実現される。これは、ユーザが、画像処理サーバ１０３の電源をＯＮ（オン）にすると開始される。

【０１１７】

まずＣＰＵ２６１は、Ｓ９５１において、手書き抽出を行うニューラルネットワークと手書き領域推定を行うニューラルネットワークをロードする。ＣＰＵ２６１は、図７（ｂ）のフローチャートのＳ７３１の場合と同一のニューラルネットワークを構築する。そして、Ｓ７３８において、学習装置１０２から送信された学習結果（手書き抽出を行うニューラルネットワークのパラメタと手書き領域推定を行うニューラルネットワークのパラメタ）を、構築したニューラルネットワークにそれぞれ反映する。

【０１１８】

ＣＰＵ２６１は、Ｓ９５２において、処理対象画像を、画像処理装置１０１より受信したかを判定する。外部インタフェース２６８を介して、処理対象画像を受信していたならば、ＹＥＳと判定し、Ｓ９５３に遷移する。そうでなければ、ＮＯと判定し、Ｓ９６３に遷移する。例として、ここでは、処理対象画像として、図１０の帳票４００（図４に示した帳票４００）を受信したものとする。

【０１１９】

10

20

30

40

50

CPU261は、S953において、画像処理装置101から受信した処理対象画像から手書き画素を抽出する。CPU261は、処理対象画像を、S951で構築した手書き抽出を行うニューラルネットワークに入力して、手書き画素を推定させる。ニューラルネットワークの出力として、次のような画像データが得られる。処理対象画像と同じ画像サイズであり、予測結果として手書きであると判定された画素には、手書きであることを示す値、手書きではないと判定された画素には、手書きではないことを示す値が、それぞれ記録された画像データ。そして、当該画像データ中の手書きであることを示す値の画素と同じ位置の画素を、処理対象画像が抽出して手書き抽出画像を生成する。これにより図10の手書き抽出画像1001が得られる。図示のように、手書きのみが含まれた画像である。
【0120】

10

CPU261は、S954において、画像処理装置101から受信した処理対象画像から手書き領域を推定する。CPU261は、処理対象画像を、S951で構築した手書き領域推定を行うニューラルネットワークに入力して、手書き領域を推定させる。ニューラルネットワークの出力として、次のような画像データが得られる。処理対象画像と同じ画像サイズであり、予測結果として手書き領域であると判定された画素には、手書き領域であることを示す値、手書き領域ではないと判定された画素には、手書き領域ではないことを示す値が、それぞれ記録された画像データ。S305において、ユーザは、手書き領域推定の正解データを、罫線枠（記入欄）を考慮して、帳票の記入項目ごとに作成した。手書き領域推定を行うニューラルネットワークはこれを学習しているため、記入欄（記入項目）ごとに手書き領域であることを示す画素を出力する。なお、手書き領域推定のニューラルネットワークの出力は、画素毎の予測結果であるため、予測された領域は必ずしも正確な矩形ではなく扱いにくいいため、当該領域を内包する外接矩形を設定する。外接矩形の設定には公知の技術を適用し実現することができる。それぞれの外接矩形は、処理対象画像上における左上端点と幅および高さから成る情報として表現することができる。このようにして得た矩形情報群を手書き領域とする。図10の1002に、処理対象画像（帳票400）に対して推定した手書き領域を、点線枠で示して例示する。

20

【0121】

CPU261は、S955において、S953で生成した手書き抽出画像と、S954で生成した手書き領域とを、外部インタフェース268を介して、手書きOCR部116に送信する。そして、手書き文字のみが含まれる手書き抽出画像に対して、推定した手書き領域毎に手書きOCRを実行させる。手書きOCRには公知の技術を適用し実現することができる。手書き抽出画像を手書き領域毎に手書きOCRすることにより、例えば、図4の住所記入欄403や電話番号記入欄404に図示したように、手書き記入欄内の活字やマークなどの印字内容が、手書きOCRの対象となってしまうことを低減する。また、氏記入欄401や名記入欄402のように隣接した項目でも、別々の手書き領域と推定させて手書きOCRすることで、文字や項目が連結されてしまうことを低減する。

30

【0122】

CPU261は、S956において、手書きOCR部116から、手書きOCR結果を受信したか否かを判定する。手書きOCR結果とは、手書きOCR部116が、手書き領域に含まれていた手書き文字を認識して得たテキストデータである。外部インタフェース268を介して、手書きOCR部116から、手書きOCR結果を受信していたならば、YESと判定し、S957に遷移する。そうでなければ、S956の処理を繰り返す。CPU261は、ここまでの処理によって、手書き領域（座標情報）とそこに含まれていた手書き文字を認識して得たテキストデータを得た。CPU261は、これらを手書き情報テーブル1003としてRAM246に記憶しておく。

40

【0123】

CPU261は、S957において、S953で得た手書き抽出のニューラルネットワークの出力に基づいて、処理対象画像から手書きを除去して活字画像を生成する。CPU261は、処理対象画像の画素であって、ニューラルネットワークが出力した画像データにおいて画素値が手書きを示す値である画素と同位置の画素を、白（RGB = (255,

50

255, 255))に変更する。これにより、図10の活字画像1004が得られる。

【0124】

CPU261は、S958において、S957で生成した活字画像から活字領域を抽出する。CPU261は、活字領域として、活字を内包する活字画像上の部分領域を抽出する。ここで部分領域とは、印刷内容のまとまり(オブジェクト)であり、例えば、複数の文字からなる文字行や、複数の文字行からなる文章、あるいは、図や写真、表、グラフ、といったオブジェクトである。この部分領域の抽出方法として、例えば次のような手法を取り得る。活字画像を白黒に二値化して二値画像を生成する。この二値画像において黒画素が連結する部分(連結黒画素)を抽出し、これに外接する矩形を作成していく。当該矩形の形状や大きさを評価することで、文字ないし文字の一部である矩形群を得ることができる。これら矩形群について、矩形間の距離を評価し、予め定めた閾値以下の距離である矩形の統合を行うことで、文字である矩形群を得ることができる。同様の大きさの文字の矩形が近くに並んでいる場合には、それらを統合して文字行の矩形群を得ることができる。短辺長が同様の文字行の矩形が等間隔に並んでいる場合は、それらを統合して文章の矩形群を得ることができる。また、図や写真、表、グラフなど、文字や行、文章以外のオブジェクトを内包する矩形も得ることができる。以上で抽出した矩形から、単独の文字あるいは文字の一部である矩形を除外する。残った矩形を部分領域とする。図10の1005に、活字画像に対して抽出した活字領域を、点線枠で例示する。本処理ステップでは、背景サンプル画像から複数の背景部分領域を抽出し得る。

10

【0125】

CPU261は、S959において、S956で生成した処理対象画像と、S957で取得した活字領域とを、外部インターフェース268を介して、活字OCR部117に送信し、活字OCRを実行させる。活字OCRには公知の技術を適用し実現することができる。

20

【0126】

CPU261は、S960において、活字OCR部117から、活字OCR結果を受信したか否かを判定する。活字OCR結果とは、活字OCR部117が、活字領域に含まれていた活字を認識して得たテキストデータである。外部インターフェース268を介して、活字OCR部117から、活字OCR結果を受信していたならば、YESと判定し、S961に遷移する。そうでなければ、S960の処理を繰り返す。ここまでの処理によって、活字領域(座標情報)とそこに含まれていた活字を認識して得たテキストデータを得た。CPU261は、これらを活字情報テーブル1006としてRAM246に記憶しておく。

30

【0127】

CPU261は、S961において、手書きOCR部116ないし活字OCR部117から受信した手書きOCR結果および活字OCR結果を統合する。CPU261は、元となった手書き領域および活字領域の位置関係や、手書きOCR結果ないし活字OCR結果であるテキストデータの意味的な関係性を評価することで、手書きOCR結果および活字OCR結果の関連性を推定する。この推定は、手書き情報テーブル1003や活字情報テーブル1006に基づき行われる。そして、氏見出し411内の活字領域に対して、最も距離が近い手書き領域として、氏記入欄401内の手書き領域が特定される。距離のみならず、上下・左右等の、手書き領域と活字領域の位置関係も考慮してよい。かつ、当該手書き領域の手書きOCR結果であるテキストデータが、氏名の氏を含む文字列であれば、これは氏名の氏であると特定できる。よって、これら活字OCR結果と手書きOCR結果とを、氏名の氏に関する項目と値のペアの関係性と評価する。同様に氏名の名についても、活字領域と手書き領域の距離、および手書きOCR結果に氏名の名を含む文字列であることから、両者の関係性評価され、氏名の名に関する項目と値のペアの関係性と評価する。氏と名は統合し、氏名の項目と値のペアとすることができる。さらに同様に、住所についても、活字領域と手書き領域の距離、および、手書きOCR結果に地名を含むことから、両者の関係性が評価される。電話番号についても、活字領域と手書き領域の距離、および、手書きOCR結果が数字から構成されることから、両者の関係性が評価される。こう

40

50

して、手書きOCR結果と活字OCR結果とが、ひとつ以上の項目と値のペアとして統合され、帳票テキストデータ1007が生成される。

【0128】

CPU261は、S962において、生成した帳票データを、画像取得部111に送信する。

【0129】

CPU261は、S963において、処理を終了するか否かを判定する。ユーザが、画像処理サーバ103の電源のOFFなどの所定の操作を行った場合には、YESと判定し、処理を終了する。そうでなければ、NOと判定し、S952に遷移する。

【0130】

以上、本実施例に示したとおり、手書き記入を含む帳票のスキャン画像から、手書きのみを含む手書き抽出画像を生成する。一方で、当該スキャン画像から手書き領域を推定する。そして、手書き抽出画像中の手書きを、手書き領域毎に手書きOCRによって認識し、テキスト化することができる。これにより、手書きOCRの対象に、活字やマークなどの印字内容が含まれることによる、手書きOCRの認識精度低下を低減することができる。また、記入枠により区別される記入項目ごとに手書きOCRするため、不要な文字や項目の連結を低減することができる。

【0131】

本実施例では、処理対象画像から生成した手書き抽出画像と手書き領域とを手書きOCR部に送信したがこの限りではない。手書き抽出画像を手書き領域毎に切り出して複数の部分画像を生成して送信してもよい。

【0132】

また、なお、本実施例では、処理対象画像に対して手書き抽出を行い、手書き抽出画像を生成したが、この限りではない。手書き領域推定により推定した領域内に限定して手書き抽出を行ってもよい。このようにしても、手書きOCRの対象となる画像は、記入項目を示す手書き領域であり、かつ手書き画素のみを含むものである。

【0133】

また、なお、本実施例では、手書き領域推定のニューラルネットワークは、画素の値を手書き領域であることを示す値と、手書き領域ではないことを示す値との、2クラス分類を行う例を示した。これに限らず、例えば、手書き領域を複数のクラスとして詳細に分類して推定してもよい。例えば、氏名が記入された手書き領域や、住所が記入された手書き領域、電話番号が記入された手書き領域、手書き領域ではない領域、のように、記入された内容に応じて分類（多クラス分類）してもよい。この場合、手書き領域推定正解画像を作成する際に、正解データとして含まれる手書き領域それぞれの画素を、分類対象の内容を示す値とする。例えば、氏名の手書き領域は255、住所の手書き領域は127、電話番号の手書き領域は64、その他の手書き領域は0。このようにして作成した手書き領域推定正解画像を、学習時にニューラルネットワークに参照させる。図9のS955では、手書き領域推定のニューラルネットワークが推定した手書き領域が、いずれのクラス（氏名／住所／電話番号）であっても、手書き抽出画像を推定した手書き領域毎に手書きOCRに入力する。S961では、手書きOCRおよび活字OCRの意味的な関係性に加えて、手書き領域推定が推定した内容に加味して帳票データ1007を生成することができる。

【0134】

（実施例2）

本実施例では、手書き抽出および手書き領域推定の別の方法として、実施例1とは別の方法を採用するケースについて説明する。本実施例では、手書き抽出および手書き領域推定をニューラルネットワークではなく、ルールベースのアルゴリズム設計により実現する。実施例2の画像処理システムの構成は、特徴部分を除いて実施例1の構成と同様である。そのため、同様の構成については、同様の符号を付し、その詳細な説明を省略する。

【0135】

< 画像処理システム >

10

20

30

40

50

本実施例における、学習処理システムについて説明する。画像処理システムは、図 1 に示した画像処理装置 101 と、画像処理サーバ 103、OCRサーバ 104 より構成される。

【0136】

<利用シーケンス>

本システムにおける、利用シーケンスについて説明する。図 11 は実施例 2 における画像処理システムの利用シーケンスを示した図である。

【0137】

S351 と S352 の処理は、図 3 (b) のシーケンスにおける同符合のステップと同様のステップである。

【0138】

画像取得部 111 が帳票原稿を読み取って生成した処理対象画像は、画像変換部 113 に送信される (S1101)。

【0139】

S354 の処理は、図 3 (b) のシーケンスにおける同符合のステップと同様のステップである。

【0140】

画像変換部 114 は、アルゴリズム設計に基づいて、処理対象画像に対して、手書き抽出および手書き領域推定を行う (S1102)。

【0141】

S358 から S367 の処理は、図 3 (b) のシーケンスにおける同符合のステップと同様のステップである。

【0142】

<帳票テキスト化処理>

本実施例における、画像処理サーバ 103 による帳票テキスト化処理について説明する。図 12 (a) は実施例 2 における帳票テキスト化処理のフローを示す図である。この処理は、CPU261 が、ストレージ 265 に記憶されている画像処理サーバプログラムを読み出し、RAM264 に展開して実行することで実現される。これは、ユーザが、画像処理サーバ 103 の電源を ON (オン) にすると開始される。

【0143】

S952 の処理は、図 9 (b) のフローチャートにおける同符合の処理ステップと同様の処理ステップである。本実施例では、S952 で、YES と判定する場合には、S1201 に遷移する。

【0144】

CPU261 は、S1201 において、手書き抽出処理を実行し、画像処理装置 101 から受信した処理対象画像から手書き画素を抽出する。この手書き抽出処理について、図 12 (b) を用いて説明する。図 12 (b) は実施例 2 における手書き抽出処理のフローを示す図である (この処理は特許文献 1 などに示される方法と同様の方法である)。

【0145】

CPU261 は、S1231 において、処理対象画像を 2 値化し、2 値画像を得る。

【0146】

CPU261 は、S1232 において、2 値画像をラスタスキャンして、白画素と黒画素の境界 (黒エッジ) 部を検出する。

【0147】

CPU261 は、S1233 において、S1232 で検出した各黒エッジに対して、水平方向および垂直方向にスキャンして、黒画素と白画素の境界 (白エッジ) を検出する。スキャンは、水平方向と垂直方向のいずれかの方向で白エッジを検出するまで行う。

【0148】

CPU261 は、S1234 において、それぞれの黒エッジと白エッジについて、両者の間の画素数を線幅として確定する。

10

20

30

40

50

【0149】

CPU261は、S1235において、S1234で取得した線幅が閾値以下であるスキャン部を細線として抽出する。これにより、印字ないし手書きの細線を複数抽出する。

【0150】

CPU261は、S1236において、各細線について、当該細線に該当する処理対象画像上の画素群からヒストグラムを算出する。

【0151】

CPU261は、S1237において、S1236で算出したヒストグラムの分散が、閾値以上（輝度のばらつきが大きい）の細線を手書きとして推定する。これは、一般に、印字の細線に比して、手書きの細線には輝度のバラつきが生じやすいことを利用した手法である。処理対象画像の当該手書きに該当する画素のみを抽出して、手書き抽出画像を生成する。

10

【0152】

図12(a)のフローチャートに戻って、CPU261は、S1202において、手書き領域推定処理を実行し、画像処理装置101から受信した処理対象画像から手書き領域を推定する。この手書き領域推定処理について、図12(c)を用いて説明する。図12(c)は実施例2における手書き領域処理のフローを示す図である（この処理は特許文献2などに示される方法と同様の方法である）。

【0153】

CPU261は、S1261において、処理対象画像を2値化し、2値画像を得る。

20

【0154】

CPU261は、S1262において、2値画像から水平方向および垂直方向の罫線（直線）を検出する。これは、2値画像内に含まれる個々の黒画素の連結部分について幅や高さを評価して検出するなど、公知の技術を適用し実現することができる。

【0155】

CPU261は、S1263において、検出した罫線に囲まれた領域内を、手書き領域として決定する。

【0156】

図12(a)のフローチャートに戻って、CPU261は、S1203において、S1201で得た手書き抽出画像と、S1202で得た手書き領域とを評価する。すなわち、CPU261は、S1202で得た手書き領域それぞれについて、手書き抽出画像上の対応する領域内に含まれる黒画素（有色画素）の数を評価し、これが閾値以下であるものを除外する。S1202で得た手書き領域は、罫線に囲まれた領域であり、これらの中には活字を含むものが含まれる可能性がある。よって、この処理ステップにおいて、手書き画素のみを含む手書き抽出画像を用いて評価することにより、手書き領域から活字を含んだものを除外するものである。

30

【0157】

S955～S963の処理は、図9(b)のフローチャートにおける同符合の処理ステップと同様の処理ステップである。

【0158】

以上、本実施例に示したとおり、手書き抽出および手書き領域推定は、ニューラルネットワークではなく、ルールベースのアルゴリズム設計によっても実現することができる。一般に、ニューラルネットワークを用いた手法は、処理計算量多くなりがちであるため、比較的高価な処理プロセッサ（CPUやGPU）が用いられる。コスト等の理由により、こうした計算資源を用意できないような場合に、本実施例で示した方法が有効である。

40

【0159】

（実施例3）

本実施例では、実施例1で説明した帳票テキスト化処理の方法に加えて、手書き抽出画像を参照しながら手書き領域を補正する方法について説明する。図14に示すように、帳票への手書き記入が罫線をはみ出してしまう場合がある。これに対して、実施例1では、

50

ニューラルネットワークに、手書き文字が記入されている罫線に囲まれた領域内（記入欄）を手書き領域とする正解データを含んだ学習データを学習させ、そして、当該ニューラルネットワークを用いて手書き領域を推定した。よって、手書き記入のうち、罫線をはみ出した部分は、手書き領域に含まれなかったり、当該部分が他の手書き領域に混入してしまう可能性がある。また、必ずしも記入欄の全域に手書き文字を記入するわけではないため、手書き記入（手書き画素）に対して推定した手書き領域が無駄に大きい（広い余白を有する）場合が生ずる可能性がある。本実施例ではこうしたケースに対して、手書き領域を補正する方法を説明する。なお、実施例３の画像処理システムの構成は、特徴部分を除いて実施例１の構成と同様である。そのため、同様の構成については、同様の符号を付し、その詳細な説明を省略する。

10

【０１６０】

< 帳票テキスト化処理 >

本実施例における、画像処理サーバ１０３による帳票テキスト化処理について説明する。図１３（ａ）は実施例２における帳票テキスト化処理のフローを示す図である。この処理は、ＣＰＵ２６１が、ストレージ２６５に記憶されている画像処理サーバプログラムを読み出し、ＲＡＭ２６４に展開して実行することで実現される。これは、ユーザが、画像処理サーバ１０３の電源をＯＮ（オン）にすると開始される。

【０１６１】

Ｓ９５１～Ｓ９５４の処理は、図９（ｂ）のフローチャートにおける同符合の処理ステップと同様の処理ステップである。

20

【０１６２】

ＣＰＵ２６１は、Ｓ１３０１において、手書き領域補正処理を実行し、Ｓ９５３で生成した手書き抽出画像を参照しながらＳ９５４で推定した手書き領域を補正する。この手書き領域補正処理について、図１３（ｂ）を用いて説明する。図１３（ｂ）は実施例３における手書き領域補正処理のフローを示す図である。

【０１６３】

ＣＰＵ２６１は、Ｓ１３５１において、Ｓ９５４で推定した手書き領域のなかから一つを選択する。

【０１６４】

ＣＰＵ２６１は、Ｓ１３５２において、Ｓ１３５１で選択した手書き領域を、手書き画素に合わせて縮小する。すなわち、ＣＰＵ２６１は、手書き抽出画像上の当該手書き領域に該当する画素群を参照する。当該画素群に含まれる手書き画素の全てを含む外接矩形を算出する。そして、手書き領域を当該外接矩形の領域で更新する（当該外接矩形よりも少し大きい矩形で更新してもよい）。これにより、当該手書き領域について、これに含まれる手書き記入に対する余計な余白を除去することができる。

30

【０１６５】

ＣＰＵ２６１は、Ｓ１３５３において、全ての手書き領域を処理したか否かを判定する。ＣＰＵ２６１は、Ｓ９５４で推定した手書き領域について、全てＳ１３５３の処理を行った場合にはＹＥＳと判定し、Ｓ１３５４に遷移する。そうでない場合はＮＯと判定し、Ｓ１３５１に遷移し、次の手書き領域を選択して処理する。

40

【０１６６】

ＣＰＵ２６１は、Ｓ１３５４において、Ｓ９５４で推定した手書き領域の中から一つ、改めて選択する。

【０１６７】

ＣＰＵ２６１は、Ｓ１３５５において、手書き領域の端部と手書き記入が交差するか否かを判定する。すなわち、ＣＰＵ２６１は、手書き領域の端部に手書き画素が存在し、当該手書き領域外に、当該手書き画素と連結する別の手書き画素が存在するか否かを判定する。存在する場合にはＹＥＳと判定し、Ｓ１３５６に遷移する。そうでない場合は、ＮＯと判定し、Ｓ１３５８に遷移する。

【０１６８】

50

CPU261は、S1356において、手書き領域が拡張可能であるか否かを判定する。すなわち、CPU261は、S1355で検出した手書き領域の端部と交差する手書き記入（手書き抽出画素群）に対して、これを内包するように手書き領域を拡張した場合に、当該手書き領域が他の手書き領域と交差しないか否かを判定する。交差しない場合には、YESと判定し、S1357に遷移する。そうでない場合には、NOと判定し、S1358に遷移する。

【0169】

CPU261は、S1357において、手書き領域が、S1355で検出した手書き領域の端部と交差する手書き記入を包含するよう、当該手書き領域を拡張する。

【0170】

CPU261は、S1358において、全ての手書き領域を処理したか否かを判定する。CPU261は、S954で推定した手書き領域について、全てS1355、S1356ないしS1357の処理を行った場合にはYESと判定し、処理を終了する。そうでない場合はNOと判定し、S1354に遷移し、次の手書き領域を選択して処理する。

【0171】

以上、本実施例に示したとおり、推定した手書き領域を、抽出した手書き画素に応じて補正することができる。これにより、不要な余白が除去され、手書きOCRのためのデータ送信や処理の負荷を低減することができる。また、手書き記入が記入枠をはみ出してしまった場合でも、はみ出した部分が手書きOCRの対象から欠損してしまうことや、他の記入項目に混入してしまうことを低減して、手書きOCR認識率を向上させることができる。

【0172】

なお、本実施例および実施例1におけるS953により得られる手書き抽出画像と、実施例2におけるS1201により得られる手書き抽出画像とは同様のものである。また、本実施例および実施例1におけるS954により得られる手書き領域と、実施例2におけるS1202により得られる手書き領域とは同様のものである。よって、本実施例は、実施例1の差分として記載したが、本実施例で説明した手書き抽出画像を参照しながら手書き領域を補正する方法は、実施例2の方法に対しても適用することができる。

【0173】

（実施例4）

本実施例では、手書き領域推定の別の方法として、実施例1とは別の方法を採用するケースについて説明する。本実施例では、手書き領域推定を行うニューラルネットワークの学習に用いる正解データを、実施例1とは異なる条件で作成し、より高度な推定を行うことができるようにする。実施例4の画像処理システムの構成は、特徴部分を除いて実施例1の構成と同様である。そのため、同様の構成については、同様の符号を付し、その詳細な説明を省略する。

【0174】

< 操作画面 >

実施例4における手書き領域推定の学習データの作成について説明する。図14は、実施例1において図5(c)に示した手書き領域推定正解データ作成画面の、画像表示領域521の表示例を示す図である。実施例4における手書き領域推定の正解データの作成指示は、実施例1同様、ユーザが、図5(c)に示したが画面を参照しながら、入力デバイス236を操作して行う。実施例1においては、ユーザは、画像表示領域521に表示されている原稿サンプル画像中の手書き文字が記入されている罫線に囲まれた領域内（記入欄内）を選択するよう操作指示を行った。実施例4では、次のように操作指示を行い、手書き領域推定の正解データを作成する。

【0175】

ユーザは、正解データが、記入欄内であり（正解データの領域に罫線が含まれない）かつ記入された手書きを内包し、かつ余白が最小限になるよう操作指示する（例：図14「ヤマダ」および「山田」「太郎」部分）。同一の記入欄内にあっても、意味的に別の記入

10

20

30

40

50

内容については、別々の正解データを作成する（例：図 1 4 「0 3 - 9 9 9 - 9 9 9」と「0 9 0 - 8 8 8 8 - 8 8 8 8」部分）。

【0 1 7 6】

また、手書き記入が罫線をはみ出している場合に限り、正解データが記入欄外にはみ出る（正解データの領域に罫線が含まれる）ことを許容して作成する（例：図 1 4 「タロウ」）。

【0 1 7 7】

また、罫線のなかでも桁線にいては、正解データの領域に桁線が含まれることを許容し、桁線で区切られた各数値の手書き記入を同一の正解データに含めるように作成する（例：「0 0 1」「8 0 5 1」）。ここで桁線とは、数値を区切る縦の点線等の非直実線である罫線である。

【0 1 7 8】

このような条件のもと、手書き領域推定の正解データの作成を操作指示したあと、保存ボタン 5 2 7 を指示することにより、手書き領域推定正解画像が生成される。学習装置 1 0 2 は、当該手書き領域推定正解画像を学習データの一部として用い、手書き領域推定を行うニューラルネットワークを学習することができる。

【0 1 7 9】

以上、本実施例に示したとおり、本実施例の画像変換システムによれば種々の条件のもと、手書き領域推定の正解データを作成することができる。手書き領域推定の正解データが、記入欄内であり（正解データの領域に罫線が含まれない）かつ記入された手書きを内包し、かつ余白が最小限になるよう作成することができる。これにより、ニューラルネットワークに、記入欄内であってかつ、余計な余白を含まないよう手書き領域を推定させる。また、手書き記入が罫線をはみ出している場合に限り、正解データが記入欄外にはみ出ることを許容して作成することができる。これにより、ニューラルネットワークに、手書き記入の画素のうち、記入欄外にはみ出た部分をも手書き領域に含めるようこれを推定させる。また、桁線で区切られた各数値の手書き記入を同一の正解データに含めるように作成することができる。これにより、ニューラルネットワークに、桁線で区切られた複数の数値記入を、ひとつの項目として単一手書き領域に含めるよう推定させる。

【0 1 8 0】

（実施例 5）

本実施例では、手書き抽出および手書き領域推定の学習データの作成方法として、実施例 1、3、4 とは別の方法を採用するケースについて説明する。本実施例では画像合成により学習データを作成する。このような処理により、学習データを効率的に作成することができる。実施例 5 の画像処理システムの構成は、特徴部分を除いて実施例 1、3、4 の画像処理システムの構成と同様である。そのため、同様の構成については、同様の符号を付し、その詳細な説明を省略する。

【0 1 8 1】

< 画像処理システム >

本実施例における、学習処理システムについて説明する。画像処理システムの構成は、実施例 1 において図 1 に示したものと同様である。

【0 1 8 2】

画像処理装置 1 0 1 は、白紙に手書きのみが記入された原稿をスキャン機能によりスキャンして画像データを生成する（以降この画像を「前景サンプル画像」と呼称する）。複数枚の原稿をスキャンして前景サンプル画像を複数得る。一方、画像処理装置 1 0 1 は、電子文書（帳票）を印刷し印刷原稿を出力する。さらにこの印刷原稿（手書き記入なし）をスキャンして画像データを生成する（以降この画像を「背景サンプル画像」と呼称する）。複数枚の印刷原稿をスキャンして、背景サンプル画像を複数得る。画像処理装置 1 0 1 は、ネットワーク 1 0 5 を介して、前景サンプル画像と背景サンプル画像を学習装置 1 0 2 に送信する。

【0 1 8 3】

学習装置 102 は、画像処理装置 101 が生成した前景サンプル画像と背景サンプル画像を蓄積する画像蓄積部 115 として機能する。またこのようにして蓄積した画像から学習データを生成する学習データ生成部 112 として機能する。学習データ生成部 112 は、手書き抽出を行うニューラルネットワークと手書き領域推定を行うニューラルネットワークを学習するために用いられる学習データを生成する。学習装置 102 は、生成した学習データを用いてニューラルネットワークの学習をおこなう学習部 113 として機能する。学習部 113 の学習により、学習結果（ニューラルネットワークのパラメータなど）が生成される。学習装置 102 は、学習結果（学習モデル）を、ネットワーク 105 を介して、画像処理サーバ 103 に送信する。

【0184】

<学習シーケンス>

本システムにおける学習シーケンスについて説明する。図 15 は実施例 5 における画像処理システムの学習シーケンスを示す図である。

【0185】

S1501 において、ユーザが手書きのみが記入された原稿の読取指示を行うと、画像取得部 111 は、これを読みとり、前景サンプル画像を生成する（S1502）。そして、この画像に前景フラグを付与する。また、S1503 において、ユーザが帳票の印刷原稿の読取指示をおこなうと、画像取得部 111 は、原稿を読みとり、背景サンプル画像を生成する（S1504）。そして、この画像に背景フラグを付与する。

【0186】

上述のように読み取られた前景サンプル画像、背景サンプル画像は、学習データ生成部 112 に送信される（S1505）。なお、このとき、前景サンプル画像、背景サンプル画像に ID 情報を付与するとよい。この ID 情報は例えば、画像取得部 111 として機能する画像処理装置 101 を識別するための情報である。なお、ID 情報として、画像処理装置 101 を操作するユーザを識別するためのユーザ識別情報や、ユーザが所属するグループを識別するためのグループ識別情報であってもよい。

【0187】

画像が送信されてくると、学習データ生成部 112 は、画像蓄積部 115 に画像を蓄積する（S1506）。

【0188】

ユーザが学習装置 102 に合成領域データの付与指示を行うと（S1507）、学習データ生成部 112 は当該合成領域データを取得して画像蓄積部 115 に蓄積する（S1508）。後述するように、学習データ生成部 112 は、前景サンプル画像と背景サンプル画像を合成して学習データを生成する。上述の合成領域データとは、当該合成処理において、背景サンプル画像上であって、手書きサンプルを合成する領域のことである。学習データの生成に際し、予め合成領域データを作成しておくものである。合成領域データの付与方法については後述する。

【0189】

そして、学習データ生成部 112 は、このようにして蓄積したデータに基づいて学習データを生成する（S1509）。このとき、特定の ID 情報に基づく画像のみを用いて学習データを生成してもよい。その後、学習データ生成部 112 は、学習部 113 に学習データを送信する（S1510）、特定の ID 情報に基づく画像のみで学習データを生成した場合は、ID 情報も併せて送信する。学習部 113 は、受信した学習データに基づき、学習モデルを更新する（S1511）。学習部 113 は、ID 情報ごとに学習モデルを保持し、対応する学習データのみで学習をおこなってもよい。このように ID 情報と学習モデルを紐づけることで、特定の利用環境に特化した学習モデルを構築することができる。

【0190】

<操作画面>

S1501、S1503 に示したユーザの指示は次のような操作画面で行われる。図 16（a）は、実施例 5 における学習原稿スキャン画面を示す図である。

10

20

30

40

50

【 0 1 9 1 】

学習原稿スキャン画面 1 6 0 0 は、表示デバイス 2 1 0 に表示される画面の一例である。図 1 6 (a) に示すように、学習原稿スキャン画面 1 6 0 0 は、プレビュー領域 1 6 0 1、画像属性指定ボタン 1 6 0 2、1 6 0 3、スキャンボタン 1 6 0 4、送信開始ボタン 1 6 0 5 を備える。

【 0 1 9 2 】

スキャンボタン 1 6 0 4 は、スキャナデバイス 2 0 6 にセットされた原稿の読取を開始するためのボタンである。スキャンが完了すると、プレビュー領域 1 6 0 1 に読み取られた画像のプレビューが表示される。スキャナデバイス 2 0 6 に別の原稿をセットし、スキャンボタン 1 6 0 4 を再び押すことで、複数の画像データをまとめて保持しておくこともできる。

10

【 0 1 9 3 】

画像属性指定ボタン 1 6 0 2、1 6 0 3 は、読み取られた画像の属性を指定するためのボタンである。画像属性指定ボタン 1 6 0 2 を選択すると読取画像に前景フラグが付与される。画像属性指定ボタン 1 6 0 3 を選択すると読取画像に背景フラグが付与される。選択中のボタンは識別可能に強調表示される。

【 0 1 9 4 】

画像が読み取られ、画像種別が選択されると、送信開始ボタン 1 6 0 5 が選択可能となる。

【 0 1 9 5 】

送信開始ボタン 1 6 0 5 を選択すると、選択された画像属性が付与された読取画像が学習装置 1 0 2 に送信される。

20

【 0 1 9 6 】

S 1 5 0 7 に示したユーザの指示は、次のような操作画面で行われる。図 1 6 (b) および (c)、(d)、(e)、(f) は、合成領域データ作成画面を示す図である。

【 0 1 9 7 】

合成領域データ作成画面 1 6 5 0 は、表示デバイス 2 6 7 に表示される画面の一例である。図 1 6 (b) に示すように、合成領域データ作成画面 1 6 5 0 は、画像表示領域 1 6 5 1、画像選択ボタン 1 6 5 2、拡大ボタン 1 6 5 3、縮小ボタン 1 6 5 4、保存ボタン 1 6 5 6 を備える。

30

【 0 1 9 8 】

画像選択ボタン 1 6 5 2 は、画像処理装置 1 0 1 から受信し、画像蓄積部 1 1 5 に蓄積されている背景サンプル画像を選択するためのボタンである。画像選択ボタン 1 6 5 2 を指示すると、選択画面 (不図示) が表示され、背景サンプル画像を選択することができる。背景サンプル画像を選択すると、画像表示領域 1 6 5 1 に当該画像が表示される。ユーザは画像表示領域 1 6 5 1 に表示された背景サンプル画像に対して操作し、合成領域データを作成する。

【 0 1 9 9 】

拡大ボタン 1 6 5 3 と縮小ボタン 1 6 5 4 は、画像表示領域 1 6 5 1 の表示を拡大あるいは縮小するためのボタンである。拡大ボタン 1 6 5 3 や縮小ボタン 1 6 5 4 を指示することにより、合成領域データの作成が行いやすいよう、画像表示領域 1 6 5 1 に表示されている背景サンプル画像を拡大および縮小して表示することができる。

40

【 0 2 0 0 】

ユーザは、合成領域データを次のように操作して作成する。ユーザは、図 1 6 (b) に点線枠で示すように、入力デバイス 2 3 6 を介してマウスカーソルを操作し、画像表示領域 1 6 5 1 に表示されている背景サンプル画像中の罫線に囲まれた領域内 (記入欄内) を選択する。この選択した領域が合成領域 (合成領域データ) である。すなわち、後述する、学習データ生成部 1 1 2 による学習データの生成時に、背景サンプル画像上の、当該合成領域内に前景サンプル画像の一部を合成するよう指示するものである。尚、図 1 6 (c) に示すように、単一の記入欄内であっても、複数の項目を含む場合には、当該記入欄内

50

に複数の合成領域データを作成してよい。

【0201】

統合ボタン1655は、作成済みの複数の合成領域をグルーピングするためのボタンである。図16(d)に点線で示すように、例えば桁線で区切られた複数の記入欄(数値記入欄)にそれぞれ合成領域データを作成する。そしてこれらを、入力デバイス236を介してマウ斯卡ーソルを操作して選択し、その状態で統合ボタン1655を指示すると、選択した複数の合成領域データをグループとして記録する。後述するように、手書き領域推定の正解データを生成する際には、グルーピングされた合成領域を統合してひとつの正解データとして扱う。

【0202】

以上のようにして、ユーザは、個々の背景サンプル画像に対して、記入欄の存在に従って、複数の合成領域を作成することができる。なお、図16(e)や図16(f)に示すように、帳票の中には、記入欄が罫線に囲まれないものも存在し得る。このような場合には、図16(e)や図16(f)に点線で示すように、記入者が記入するであろう領域を選択する。

【0203】

保存ボタン1606は、作成された合成領域データを保存するためのボタンである。合成領域データは、作成対象であった背景サンプル画像に対してひとつのXML形式のデータとして記録される(以降、このXML形式で保存された合成領域データを「合成領域定義情報」と呼称する)。合成領域定義情報の例を図17に示す。合成領域定義情報の詳細は後述する。合成領域定義情報は、対応する背景サンプル画像と紐づけて画像蓄積部115に蓄積される。

【0204】

<サンプル画像生成処理>

次に、画像処理装置101によるサンプル画像生成処理について説明する。図18(a)は、サンプル画像生成処理のフローを示す図である。この処理は、CPU201が、ストレージ208に記録されているコントローラプログラムを読み出し、RAM204に展開して実行することで実現される。これは、ユーザが、画像処理装置101の入力デバイス209を操作することにより開始される。

【0205】

まずCPU201は、S1801において、手書きのみの原稿のスキャン指示が成されたか否かを判定する。ユーザが、入力デバイス209を介して、手書きのみの原稿をスキャンするための所定の操作(スキャンボタン1604への指示および手書き原稿ボタン1602の選択)を行った場合には、YESと判定し、S1802に遷移する。そうでなければ、NOと判定し、S1804に遷移する。

【0206】

CPU201は、S1802において、スキャナデバイス206や原稿搬送デバイス207を制御して、原稿をスキャンして前景サンプル画像を生成する。この原稿は、白紙に、人が手書きの文や図形(線分や矢印など)を記入したものであり、手書きのみが含まれる。特に、帳票に記入されるような文や図形が記入されたものである。前景サンプル画像は、グレースケールの画像データとして生成される。

【0207】

CPU201は、S1803において、S1802で生成した前景サンプル画像を、外部インタフェース211を介して、学習装置102に送信する。このとき、送信する画像データが、前景サンプル画像であることを示す情報(前景フラグ)を付与して送信する。

【0208】

CPU201は、S1804において、印刷内容のみの原稿をスキャンする指示が成されたか否かを判定する。ユーザが、入力デバイス209を介して、印刷内容のみの原稿をスキャンするための所定の操作(スキャンボタン1604への指示および帳票原稿ボタン1603の選択)を行った場合には、YESと判定し、S1805に遷移する。そうでな

10

20

30

40

50

ければ、N Oと判定し、S 1 8 0 7に遷移する。

【 0 2 0 9 】

C P U 2 0 1は、S 1 8 0 5において、スキャナデバイス2 0 6や原稿搬送デバイス2 0 7を制御して、原稿をスキャンして背景サンプル画像を生成する。この原稿は、用紙に帳票などを印刷したものであり、印刷内容のみが含まれる。背景サンプル画像は、グレースケールの画像データとして生成される。

【 0 2 1 0 】

C P U 2 0 1は、S 1 8 0 6において、S 1 8 0 5で生成した背景サンプル画像を、外部インタフェース2 1 1を介して、学習装置1 0 2に送信する。このとき、送信する画像データが、背景サンプル画像であることを示す情報（背景フラグ）を付与して送信する。

10

【 0 2 1 1 】

C P U 2 0 1は、S 1 8 0 7において、処理を終了するか否かを判定する。ユーザが、サンプル画像生成処理を終了する所定の操作を行った場合には、Y E Sと判定して、処理を終了する。そうでなければ、N Oと判定し、S 1 8 0 1に遷移する。

【 0 2 1 2 】

以上の処理によって、画像処理装置1 0 1は、前景サンプル画像および背景サンプル画像を生成して学習装置1 0 2に送信する。ユーザの操作や、原稿搬送デバイス2 0 7に載置した原稿枚数に応じて、前景サンプル画像および背景サンプル画像が、それぞれ複数取得される。

【 0 2 1 3 】

20

< サンプル画像受信処理 >

次に、学習装置1 0 2によるサンプル画像取受信処理について説明する図1 8 (b)はサンプル画像受信処理のフローを示す図である。この処理は、C P U 2 3 1が、ストレージ2 3 5に記録されている学習データ生成プログラムを読み出し、R A M 2 3 4に展開して実行することで実現される。これは、ユーザが、学習装置1 0 2の電源をO N (オン) にすると開始される。

【 0 2 1 4 】

C P U 2 3 1は、S 1 8 3 1において、前景サンプル画像を受信したか否かを判定する。C P U 2 3 1は、外部インタフェース2 3 8を介して画像データを受信し、かつ、当該画像データに前景フラグが付与されていたならば、Y E Sと判定し、S 1 8 3 2に遷移する。そうでなければ、N Oと判定し、S 1 8 3 3に遷移する。

30

【 0 2 1 5 】

C P U 2 3 1は、S 1 8 3 2において、受信した前景サンプル画像を、ストレージ2 3 5の所定の領域に記録する。

【 0 2 1 6 】

C P U 2 3 1は、S 1 8 3 3において、背景サンプル画像を受信したか否かを判定する。C P U 2 3 1は、外部インタフェース2 3 8を介して画像データを受信し、かつ、当該画像データに背景フラグが付与されていたならば、Y E Sと判定し、S 1 8 3 4に遷移する。そうでなければ、N Oと判定し、S 1 8 3 5に遷移する。

【 0 2 1 7 】

40

C P U 2 3 1は、S 1 8 3 4において、受信した背景サンプル画像を、ストレージ2 3 5の所定の領域に記録する。

【 0 2 1 8 】

C P U 2 3 1は、S 1 8 3 5において、処理を終了するか否かを判定する。ユーザが、学習装置1 0 2の電源のO F Fなどの、サンプル画像受信処理を終了する所定の操作を行った場合には、Y E Sと判定して、処理を終了する。そうでなければ、N Oと判定し、S 1 8 3 1に遷移する。

【 0 2 1 9 】

< 合成領域定義情報生成処理 >

次に、学習装置1 0 2による合成領域定義情報生成処理について説明する。図1 8 (c

50

）は合成領域定義情報生成処理のフローを示す図である。この処理は、学習装置 102 の学習データ生成部 112 により実現される。これは、ユーザが、学習装置 102 の入力デバイス 236 を介して、所定の操作を行うことで開始される。

【0220】

CPU 231 は、S1861において、背景サンプル画像の選択指示が成されたか否かを判定する。ユーザが、入力デバイス 236 を介して、背景サンプル画像を選択するための所定の操作（画像選択ボタン 1602 の指示）を行った場合には、YES と判定し、S1862 に遷移する。そうでなければ、NO と判定し、S1863 に遷移する。

【0221】

CPU 231 は、S1862において、S1861でユーザが選択した背景サンプル画像を、ストレージ 235 から読み出してユーザに対して出力する（画像表示領域 1601 に表示）。

10

【0222】

CPU 231 は、S1863において、ユーザが合成領域データの入力指示を行ったか否かを判定する。ユーザが、入力デバイス 236 を介して、前述したように、記入欄内の領域を選択する操作を行っていたならば、YES と判定し、S1864 に遷移する。そうでなければ、NO と判定し、S1865 に遷移する。

【0223】

CPU 231 は、S1864において、ユーザが入力した合成領域データを、RAM 234 に一時的に記憶する。

20

【0224】

CPU 231 は、S1865において、ユーザが合成領域のグルーピング入力指示を行ったか否かを判定する。ユーザが、入力デバイス 236 を介して、所定の操作（入力済みの合成領域の選択、および統合ボタン 1655 の選択）を行っていたならば、YES と判定し、S1866 に遷移する。そうでなければ、NO と判定し、S1867 に遷移する。

【0225】

CPU 231 は、S1866において、合成領域のグルーピングの情報（選択された複数の合成領域の紐付け情報）を、RAM 234 に一時的に記憶する。

【0226】

CPU 231 は、S1867において、合成領域データの保存指示が成されたか否かを判定する。ユーザが、入力デバイス 236 を介して、合成領域データを保存するための所定の操作（保存ボタン 1656 の指示）を行った場合には、YES と判定し、S1868 に遷移する。そうでなければ、NO と判定し、S1869 に遷移する。

30

【0227】

CPU 231 は、S1868において、合成領域定義情報を生成して保存する。CPU 231 は、次のようにして合成領域定義情報を生成する。CPU 231 は、合成領域定義情報を示すXMLデータを生成する。そして、S1864でRAM 234に一時的に記憶した合成領域データを参照し、その数分だけ<area>タグを挿入する。各々の<area>タグにはx属性およびy属性を追加し、該当する合成領域の背景サンプル画像上の画素位置を示すx座標およびy座標の値を記録する。また、width属性およびheight属性を追加し、該当する合成領域の幅の画素数および高さの画素数を記録する。さらに、S1866でRAM 234に一時的に記憶した合成領域のグルーピングの情報を参照し、該当する合成領域の<area>タグにgroup属性を付与する。そして、同一グループに属する合成領域について、同一の値（グループのID）をgroup属性の値に記録する。このようにして、図17に例示した合成領域定義情報が生成される。CPU 231 は、S1861で選択し現在処理対象としている背景サンプル画像と、生成した合成領域定義情報とを紐付けて、ストレージ 235 に記憶する。

40

【0228】

CPU 231 は、S1869において、処理を終了するか否かを判定する。ユーザが、合成領域定義情報生成処理を終了する所定の操作を行った場合には、YES と判定して、

50

処理を終了する。そうでなければ、N Oと判定し、S 1 8 6 1に遷移する。

【 0 2 2 9 】

< 学習データ生成処理 >

次に、実施例 5 における、学習装置 1 0 2 による学習データ生成処理について説明する。図 1 9 は実施例 5 における学習データ生成処理のフローを示す図である。図 2 0 (a) は画像合成により生成する入力画像の構成例を示す図である。図 2 0 (b) は手書き抽出の学習データの構成例を示す図である。図 2 0 (c) は手書き領域推定の学習データの構成例を示す図である。学習データ生成処理は、学習装置 1 0 2 の学習データ生成部 1 1 2 により実現される。これは、ユーザが、学習装置 1 0 2 の入力デバイス 2 3 6 を介して、所定の操作を行うことで開始される。

10

【 0 2 3 0 】

まず、C P U 2 3 1 は、S 1 9 0 1 において、ストレージ 2 3 5 に記憶している前景サンプル画像を選択して読み出す。図 1 8 の S 1 8 3 2 の処理ステップにより、ストレージ 2 3 5 には複数の前景サンプル画像が記憶されているので、その中からランダムにひとつを選択する。

【 0 2 3 1 】

C P U 2 3 1 は、S 1 9 0 2 において、ストレージ 2 3 5 に記憶している背景サンプル画像を選択して読み出す。図 1 8 の S 1 8 3 4 の処理ステップにより、ストレージ 2 3 5 には複数の背景サンプル画像が記憶されているので、その中からランダムにひとつを選択する。

20

【 0 2 3 2 】

C P U 2 3 1 は、S 1 9 0 3 において、背景サンプル画像の一部（例えば縦×横 = 2 5 6 × 2 5 6 の大きさ）を切り出した画像データを生成する（以降この画像データを「背景パッチ」と呼称する）。切り出し位置はランダムに決定する。背景パッチの例を図 2 0 (a) に示す。

【 0 2 3 3 】

C P U 2 3 1 は、S 1 9 0 4 において、手書き抽出正解ラベル画像を生成する。すなわち、C P U 2 3 1 は、背景パッチと同サイズの画像を生成し、当該画像の全ての画素の値を、手書きではないことを示す値とし、これを手書き抽出正解ラベル画像とする。

【 0 2 3 4 】

C P U 2 3 1 は、S 1 9 0 5 において、手書き領域推定正解ラベル画像を生成する。すなわち、C P U 2 3 1 は、背景パッチと同サイズの画像を生成し、当該画像の全ての画素の値を、手書き領域ではないことを示す値とし、これを手書き領域推定正解ラベル画像とする。

30

【 0 2 3 5 】

C P U 2 3 1 は、S 1 9 0 6 において、背景パッチにおける合成領域を特定する。これは次のように処理する。C P U 2 3 1 は、ストレージ 2 3 5 から、S 1 9 0 2 で選択した背景サンプル画像と紐付けて記憶されている合成領域定義情報を読み出す。当該合成領域定義情報内に < a r e a > タグとして記憶されている各合成領域（矩形情報）と、背景サンプル画像上の背景パッチ切り出し領域（矩形情報）との交差部分（交差矩形）を求める。その中で、交差矩形のサイズが縦と横それぞれについて 0 以上のものを背景パッチにおける合成領域として採用する。採用した合成領域は背景サンプル画像上の左上端点を原点としているので、背景パッチ切り出し領域の左上端点を基準としたものに補正する。以上のようにして背景パッチ上の合成領域を特定した。後の処理のために、算出した背景パッチ上の合成領域は R A M 2 3 4 に一時的に記憶しておく。また、交差矩形算出時に、もとなつた合成領域定義情報の < a r e a > タグに g r o u p 属性が付与されていたならば、この値（グルーピング情報）も合成領域と関連付けて R A M 2 3 4 に一時的に記憶しておく。図 2 0 (a) に、背景パッチ上の合成領域を、斜線で例示する。C P U 2 3 1 は、背景パッチ上の合成領域のひとつを選択する。

40

【 0 2 3 6 】

50

C P U 2 3 1 は、S 1 9 0 7 において、前景サンプル画像の一部を切り出した画像データを生成する（以降この画像データを「前景パッチ」と呼称する）。切り出し位置はランダムに決定する。切り出すサイズは、現在処理対象として選択している背景パッチ上の合成領域のサイズと同一とする。なお、前景パッチに対して二値化処理を行い、黒画素が予め定めた所定数に満たない場合は、前景パッチを切り出し直す。

【 0 2 3 7 】

C P U 2 3 1 は、S 1 9 0 8 において、切り出した前景パッチの余白を除去する。これは次のように処理する。C P U 2 3 1 は、前景パッチに対して二値化処理を行い、二値画像を生成する。二値画像中の全ての黒画素を含む外接矩形を求める。当該外接矩形に該当する領域を前景パッチから切り出して、切り出した画像で前景パッチを更新する。

10

【 0 2 3 8 】

C P U 2 3 1 は、S 1 9 0 9 において、前景パッチの合成位置を決定する。これは次のように処理する。C P U 2 3 1 は、合成領域の中心点を基準に、合成領域を拡大した領域（以降この領域を「拡大合成領域」と呼称する）を決定する。拡大率は、1 0 0 % から予め定めた閾値（例えば 1 0 5 % ）までの間からランダムに決定する。そして、拡大合成領域内に収まることを条件に、前景パッチの合成位置を拡大合成領域内においてランダムに決定する。なお、合成領域が背景パッチの端部にある場合は、拡大合成領域は端部方向へは拡大しない。図 2 0 (a) に、拡大合成領域を、破線で例示する。前述のとおり、合成領域は背景サンプル画像中の記入欄内の領域を選択するようにして作成した。故に、本処理ステップのように処理することで、後述する背景パッチに前景パッチを合成する処理の際に、記入欄周辺の罫線と手書き記入が交差する場合を生じさせる。これはすなわち、記入欄をはみ出して記入した場合を再現する学習データを生成することである。

20

【 0 2 3 9 】

C P U 2 3 1 は、S 1 9 1 0 において、背景パッチ上の、S 1 9 0 9 で決定した合成位置に、前景パッチを合成する。

【 0 2 4 0 】

C P U 2 3 1 は、S 1 9 1 1 において、手書き抽出正解ラベル画像を更新する。これは次のように処理する。C P U 2 3 1 は、前景パッチに対して二値化処理を行う。そして、予め定めた閾値よりも低い値である画素の値を、手書きを示す値とする。また、他の画素の値を、手書きではないことを示す値とする。そして、手書き抽出正解ラベル画像上であって、S 1 9 0 9 で決定した合成位置に該当する部分に、上述のようにして画素の値を定めた画像データを合成する。

30

【 0 2 4 1 】

C P U 2 3 1 は、S 1 9 1 2 において、手書き領域推定正解ラベル画像を更新する。これは次のように処理する。C P U 2 3 1 は、S 1 9 0 9 で生成した、手書きを示す値と手書きではないことを示す値からなる画像データを再度参照する。そして、当該画像内の全ての手書きを示す画素を内包する外接矩形を算出する。当該画像データ中の外接矩形に内包される画素の値を、手書き領域であることを示す値に更新する。また、他の画素を、手書き領域ではないことを示す値に更新する。そして、手書き領域推定正解ラベル画像上であって、S 1 9 0 9 で決定した合成位置に該当する部分に、上述のようにして画素の値を定めた画像データを合成する。

40

【 0 2 4 2 】

C P U 2 3 1 は、S 1 9 1 3 において、現在処理対象としている背景パッチ上の合成領域について、全て処理を行ったか否かを判定する。当該背景パッチ上の全ての合成領域について、S 1 9 0 7 ~ S 1 9 1 2 の処理を実行しているならば、Y E S と判定し、S 1 9 1 4 に遷移する。そうでなければ、N O と判定し、S 1 9 0 7 に遷移して、背景パッチ上の次の合成領域について処理していく。

【 0 2 4 3 】

C P U 2 3 1 は、S 1 9 1 4 において、S 1 9 1 6 で特定した背景パッチ上の合成領域について、グルーピング情報を参照して手書き領域推定正解ラベル画像を更新する。これ

50

は次のように処理する。CPU 231は、S 1916においてRAM 234に一時的に記憶した合成領域とグルーピングに関するデータを読み出す。同一のgroup属性の合成領域を選択する。これら合成領域の矩形の結合矩形（これら合成矩形をすべて内包する最小矩形）を求める。手書き領域推定正解ラベル画像上であって、当該結合矩形に該当する領域内の画素について、手書き領域であることを示す値である画素をすべて内包する外接矩形を求める。そして、手書き領域推定正解ラベル上であって当該外接矩形内の全ての画素の値を、手書き領域であることを示す値に変更する。このように処理することで、グルーピングされた、桁線等で区切られた複数の数値記入は、単一の記入領域内に記入されたことを示す正解データを生成する。

【0244】

CPU 231は、S 1915において、手書き抽出の学習データを保存する。すなわち、S 1910で前景パッチを合成した背景パッチを入力画像とする。この入力画像と、S 1911で合成して作成した手書き抽出正解ラベル画像とを対応付け、手書き抽出のニューラルネットワークを学習するための学習データとして、ストレージ235の所定の領域に保存する。本実施例では、図20(b)のような手書き抽出の学習データが保存される。

【0245】

CPU 231は、S 1916において、手書き領域推定の学習データを保存する。すなわち、S 1910で前景パッチを合成した背景パッチを入力画像とする。この入力画像と、S 1912で合成して作成した手書き領域推定正解ラベル画像とを対応付け、手書き領域推定のニューラルネットワークを学習するための学習データとして、ストレージ235の所定の領域に保存する。本実施例では、図20(c)のような手書き領域推定の学習データが保存される。

【0246】

CPU 231は、S 1917において、学習データ生成処理を終了するか否かを判定する。CPU 231は、予め決定した学習データの数（本フローチャートの開始時に、画像処理装置101の入力デバイス209を介して、ユーザが指定するなどして決定）だけ学習データを生成していたならば、YESと判定し、処理を終了する。そうでなければ、NOと判定し、S 1901に遷移する。

【0247】

以上により、手書き抽出を行うニューラルネットワークの学習データと、手書き領域推定を行うニューラルネットワークの学習データを、前景サンプル画像と背景サンプル画像、合成領域定義情報から、画像合成により生成される。ニューラルネットワークの汎用性を高めるために、学習データの加工を行っても良い。例えば、前景パッチや背景パッチを所定の範囲（例えば、50%～150%の間）からランダムに選択して決定する変倍率で変倍する。あるいは、前景パッチや背景パッチの各画素の輝度を変更して加工してもよい。すなわち、ガンマ補正を用いてこれら画像の輝度を変更する。ガンマ値は所定の範囲（例えば、0.1～10.0の間）からランダムに選択して決定する。あるいは、前景パッチや、前景パッチを合成した背景パッチを所定の範囲（例えば、-10度～10度の間）からランダムに選択して決定する回転角度で回転する。変倍や回転を考慮すれば、前景パッチや背景パッチは少し大きめのサイズ（例えば、合成領域の縦横2倍の大きさ／縦×横＝512×512の大きさ）で切り出す。そして、最終的な入力画像や手書き抽出および手書き領域推定の正解ラベル画像のサイズ（例えば、縦×横＝256×256）となるよう、中心部分から切り出す。

【0248】

このようにして生成した学習データを用いて、学習装置102は、手書き抽出のニューラルネットワークおよび手書き領域推定のニューラルネットワークを学習することができる。

【0249】

以上、本実施例に示したとおり、本実施例の画像変換システムによれば、手書き抽出を行うニューラルネットワークおよび手書き領域推定を行うニューラルネットワークの学習

10

20

30

40

50

データを合成して生成することができる。ユーザは、手書きのみを記入した紙をスキャンした前景サンプル画像と印刷された未記入の帳票をスキャンした背景サンプル画像と、背景サンプル上の合成領域の情報のみを作成すればよい。実施例 1、3、4 では、ユーザによる操作指示に基づいてこれら学習データを生成する例を説明したが、大量の学習データを作成する場合には、時間/コストを要してしまう。本実施例に示した方法によれば、手書きのみを記入した紙と未記入の帳票の原稿をスキャンすればよい。そして、最低限の操作指示として、合成領域を定義すればよい。ため、効率的に学習データを生成することができる。

【0250】

なお、S1912では、手書き領域推定の正解ラベル画像中の手書き領域を示す値の画素を、前景パッチをもとに決定したが、合成領域をもとに決定してもよい。こうした場合は、実施例 1 で示した方法と同様の正解ラベル画像が生成される。

【0251】

(その他の実施例)

本発明は、上述の実施例の 1 以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける 1 つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、1 以上の機能を実現する回路(例えば、ASIC)によっても実現可能である。

【0252】

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よい。例えば、実施例では、学習データ生成部 112 と学習部 113 は、学習装置 102 において実現されるものとしたが、それぞれ別体の装置において実現されるようにしてもよい。この場合、学習データ生成部 112 を実現する装置は、学習データ生成部 112 が生成した学習データを、学習部 113 を実現する装置に送信する。そして、学習部 113 は受信した学習データをもとにニューラルネットワークを学習する。画像処理装置 101 と画像処理サーバ 103 とを別体の装置として説明したが、画像処理装置 101 が画像処理サーバ 103 の機能を具備してもよい。画像処理サーバ 103 と OCR サーバ 104 とを別体の装置として説明したが、画像処理サーバ 103 の OCR サーバ 104 としての機能を具備してもよい。

【0253】

本発明は上記実施例に限定されるものではなく、本発明の趣旨に基づき種々の変形(各実施例の有機的な組合せを含む)が可能であり、それらを本発明の範囲から除外するものではない。即ち、上述した各実施例及びその変形例を組み合わせた構成も全て本発明に含まれるものである。

【0254】

実施例では、活字領域の抽出を S958 に示したように、画素の連結性に基づいて判定する方法を示したが、手書き領域推定と同様にニューラルネットワークを用いて推定してもよい。手書き領域推定正解画像を作成したのと同じ要領で、活字領域をユーザが選択し、それに基づいて正解データを作成、活字 OCR 領域推定を行うニューラルネットワークを新たに構成し、当該正解データを参照させて学習する。

【0255】

実施例では、学習処理時に学習データ生成処理より学習データを生成した。しかしながら、事前に学習データ生成処理によって学習データを大量に生成しておいて、学習処理時にそこから随時、ミニバッチサイズ分をサンプルするようにしてもよい。

【0256】

実施例では、入力画像をグレースケール画像として生成したが、フルカラー画像など他の形式として生成してもよい。

【0257】

なお、各実施例中に登場する略称の定義は次の通りである。MFP とは、Multi Function Peripheral のことである。ASIC とは、Applica

10

20

30

40

50

tion Specific Integrated Circuitのことである。CPUとは、Central Processing Unitのことである。RAMとは、Random Access Memoryのことである。ROMとは、Read Only Memoryのことである。HDDとはHard Disk Driveのことである。SSDとはSolid State Driveのことである。LANとは、Local Area Networkのことである。PDLとは、Page Description Languageのことである。OSとは、Operating Systemのことである。PCとは、Personal Computerのことである。OCRとは、Optical Character Recognition/Readerのことである。CCDとは、Charge-Coupled Deviceのことである。LCDとはLiquid Crystal Displayのことである。ADFとはAuto Document Feederのことである。CRTとはCathode Ray Tubeのことである。GPUとは、Graphics Processing Unitのことである。IDとは、Identificationのことである。

10

【0258】

なお、最後の、本発明に示した方法により学習した手書き抽出のニューラルネットワークと手書き領域推定のニューラルネットワークの推論結果を例示する。図21(a)は処理対象画像の一例である。図21(b)は手書き抽出および手書き領域推定のニューラルネットワークの推論結果の一例である。図21(b)は図21(a)の処理対象画像を入力とした場合の推論結果である。図21(b)では、手書き抽出のニューラルネットワークにより抽出された手書き文字(文字部分)と手書き領域推定のニューラルネットワークにより推定された手書き領域(白ハッチ部分)とを示している。

20

【0259】

<その他の実施例>

本発明は、上述の実施形態の1以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける1つ以上のプロセッサがプログラムを読出し実行する処理でも実現可能である。また、1以上の機能を実現する回路(例えば、ASIC)によっても実現可能である。

【符号の説明】

【0260】

201 CPU
231 CPU
261 CPU

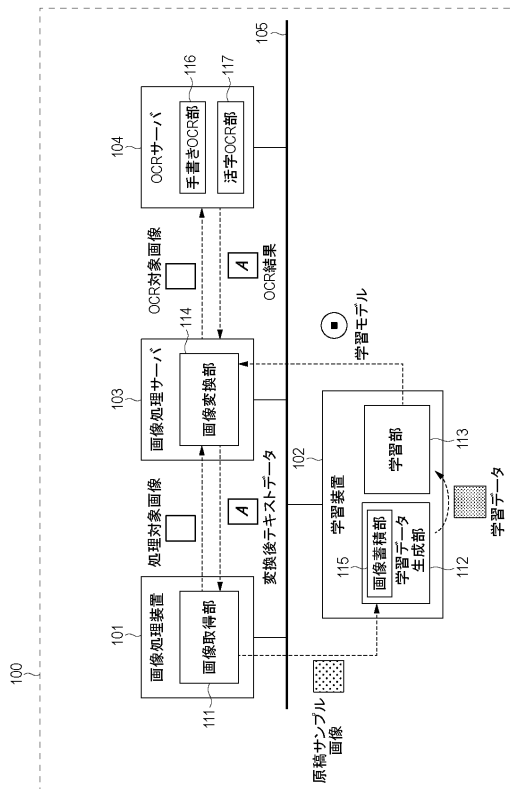
30

40

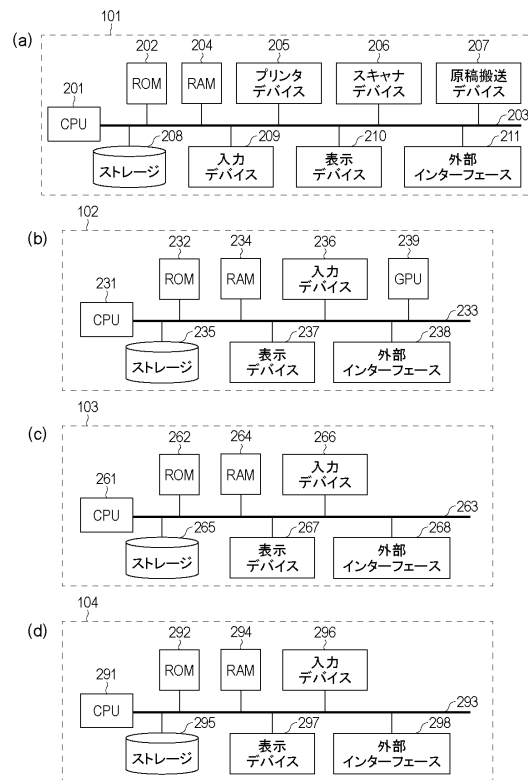
50

【図面】

【図 1】



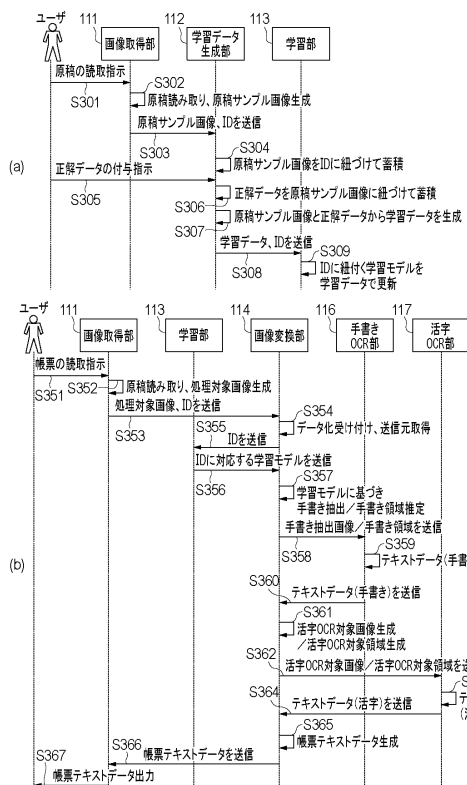
【図 2】



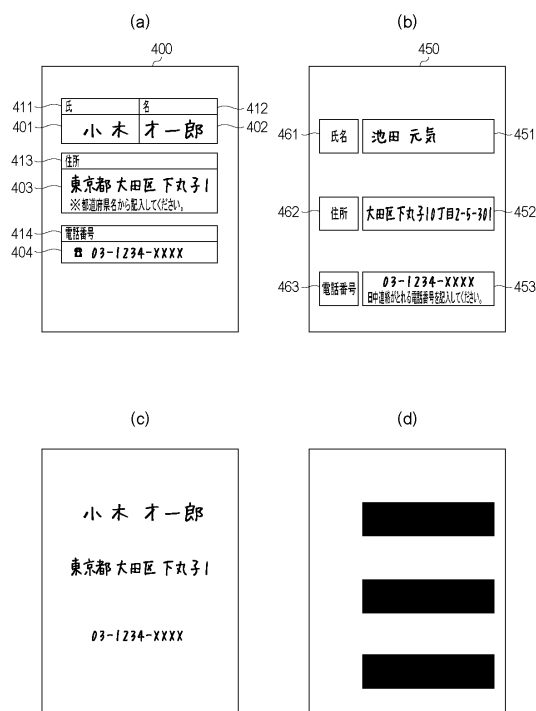
10

20

【図 3】



【図 4】

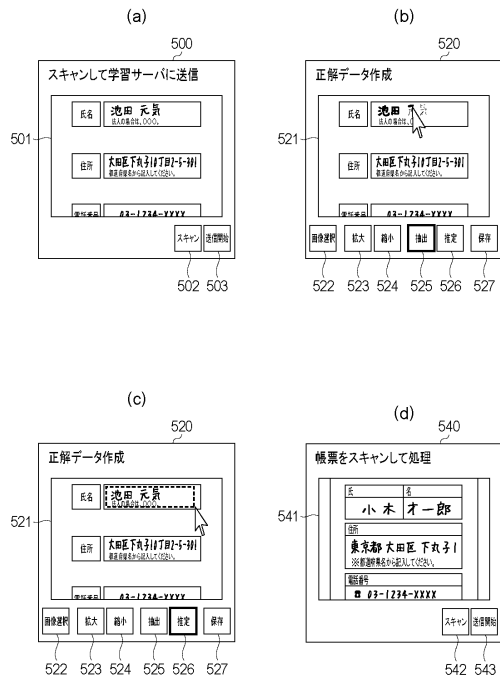


30

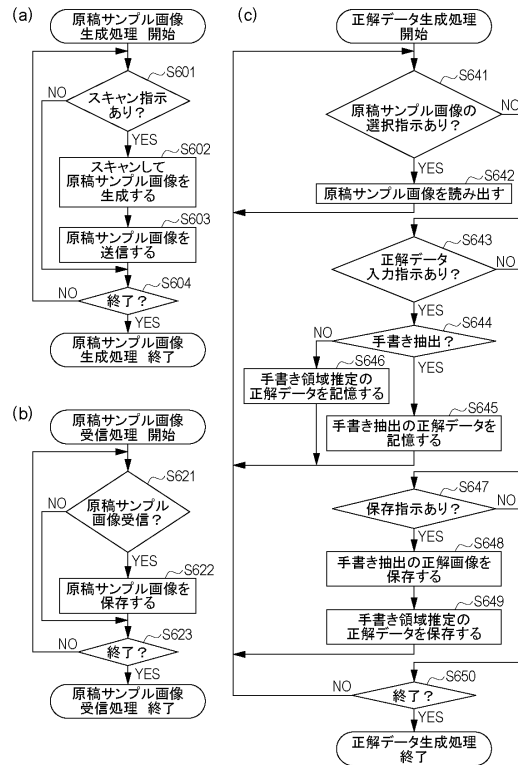
40

50

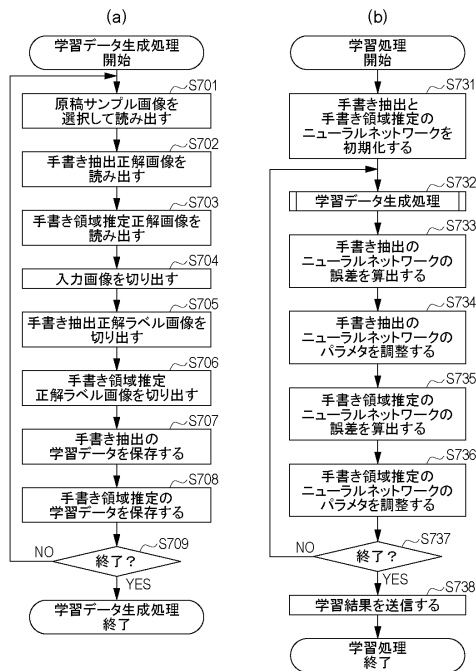
【図 5】



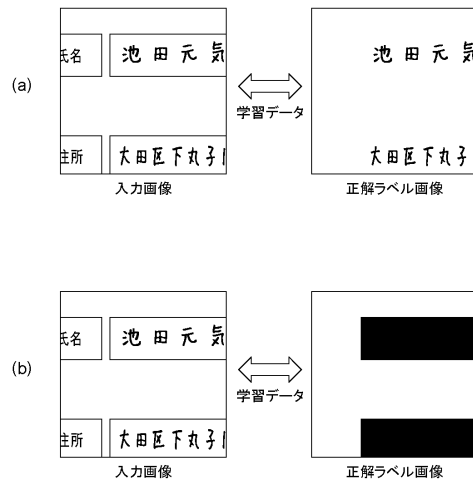
【図 6】



【図 7】



【図 8】



10

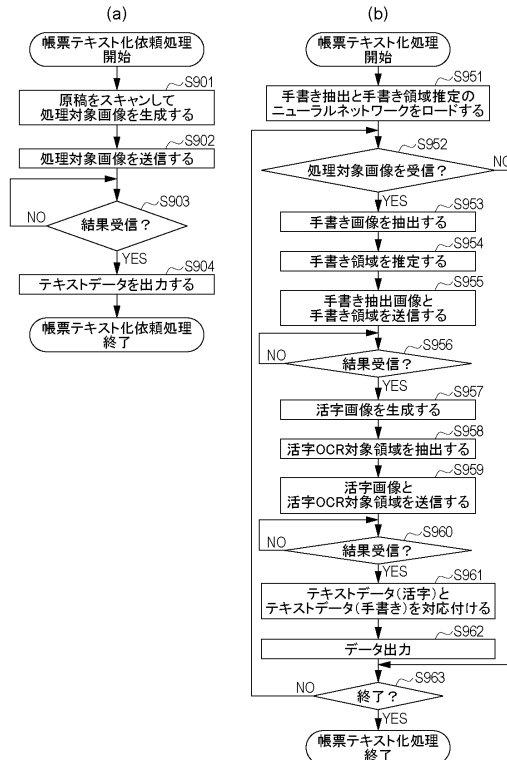
20

30

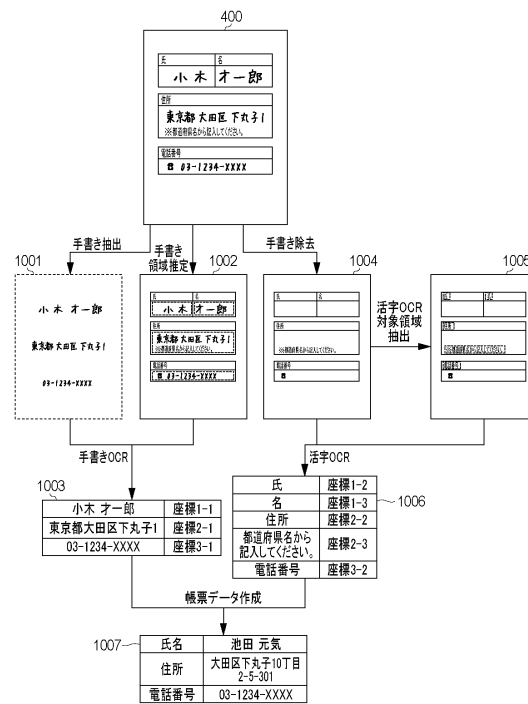
40

50

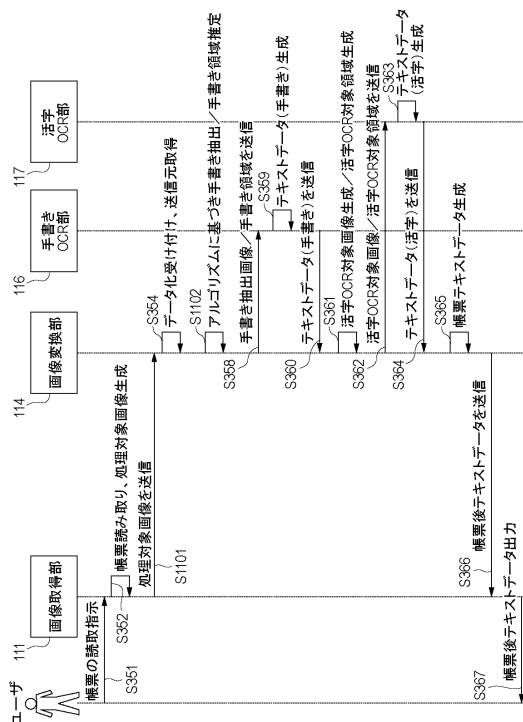
【図 9】



【図 10】



【図 11】



【図 12】



10

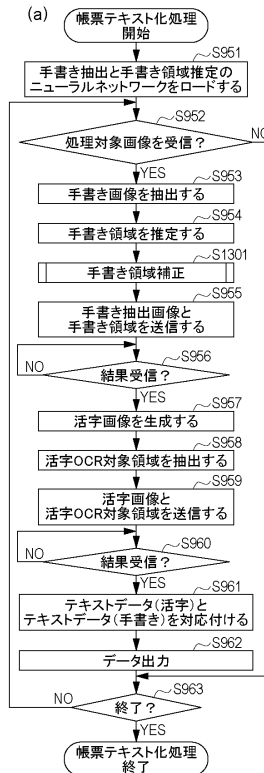
20

30

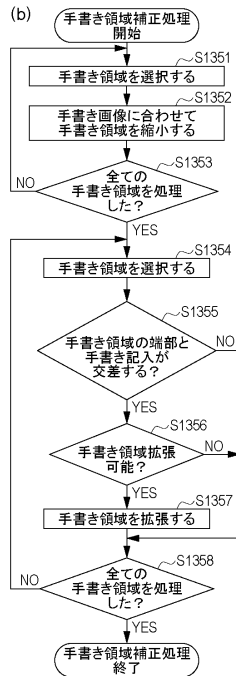
40

50

【図 13】



【図 14】



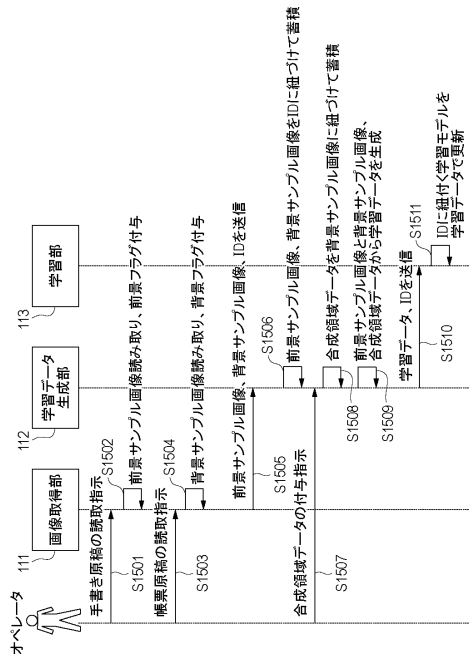
521

フリガナ	ヤマダ	タロウ
姓	山田	名 太郎
電話番号	03-999-9999 携帯 090-888-8888	
生年月日	0018051	

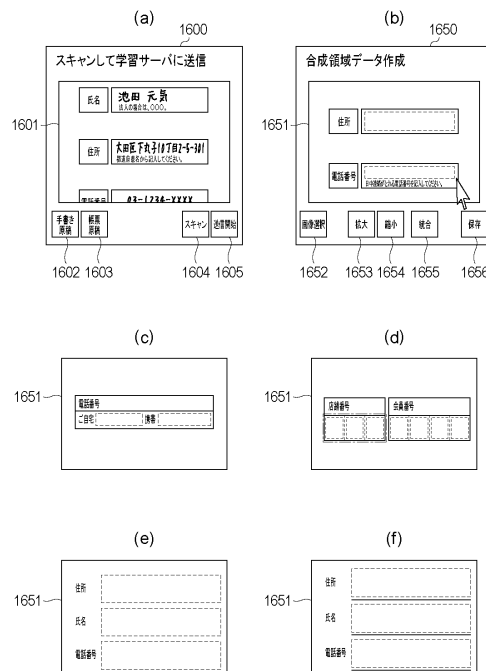
10

20

【図 15】



【図 16】



30

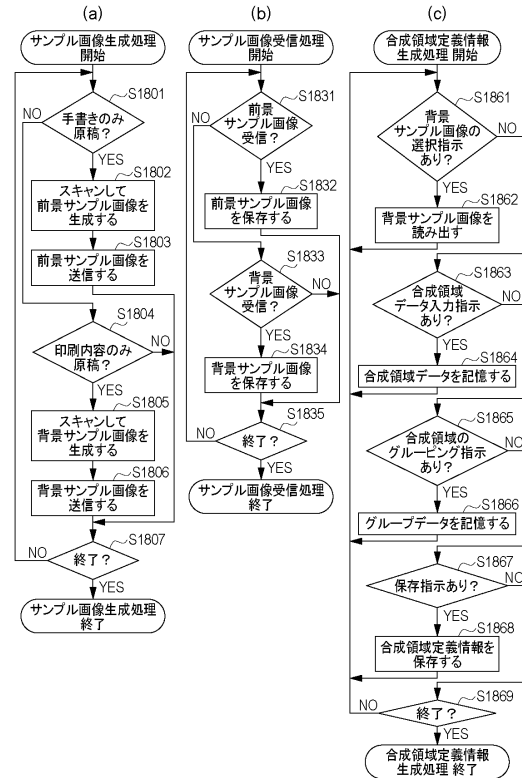
40

50

【 図 1 7 】

```
<?xml version="1.0" encoding="utf-8" ?>
<areas>
  <area x="125" y="202" width="233" height="30"/>
  <area x="109" y="550" width="244" height="42"/>
  ...
  <area x="125" y="700" width="40" height="40" group="0"/>
  <area x="165" y="700" width="40" height="40" group="0"/>
  <area x="205" y="700" width="40" height="40" group="0"/>
  ...
</areas>
```

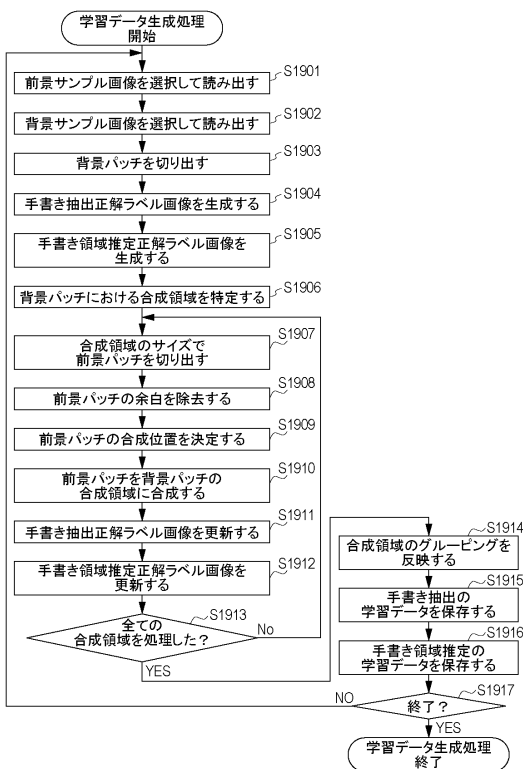
【 図 1 8 】



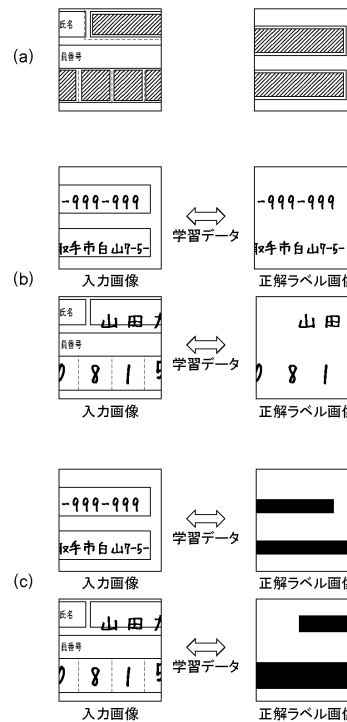
10

20

【 図 1 9 】



【 図 2 0 】



30

40

【 図 2 1 】

(a)

氏 名	オ ー 様	〒 100-0001
小 木 太 郎		
住所		
東京都秋田区下丸子1-1-1 1F 101		
※郵便番号を記入してください。		
電話番号		
特 090-XXXX-XXXX 宅 03-XXXX-XXXX		

(b)

氏 名	オ ー 様	〒 100-0001
小 木 太 郎		
住所		
東京都秋田区下丸子1-1-1 1F 101		
※郵便番号を記入してください。		
電話番号		
特 090-XXXX-XXXX 宅 03-XXXX-XXXX		

【 図 2 2 】

(a)

姓	名
小 木	太 郎

(b)

小 木	太 郎
-----	-----

(c)

姓	名
ALIV	AIL

(d)

ALIV	AIL
------	-----

(e)

店舗番号	会員番号
0 0 1 8 0 5 1	

(f)

0 0 1 8 0 5 1

(g)

フリガナ	マ ヅ イ ク ミ
------	-----------

(h)

姓	名
L O N G	G R A Y

(i)

電話番号
0 3 - 1 2 3 4 - XXXX

(j)

0 3 - 1 2 3 4 - XXXX
日中連絡がとれる電話番号を記入してください。

10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2 0 1 0 - 2 1 8 1 0 6 (J P , A)
特開 2 0 1 7 - 1 3 8 7 0 3 (J P , A)
特開平 0 9 - 1 1 4 9 2 4 (J P , A)
- (58)調査した分野 (Int.Cl. , D B 名)
- G 0 6 V 3 0 / 1 4
G 0 6 T 7 / 0 0