

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2018-46337

(P2018-46337A)

(43) 公開日 平成30年3月22日(2018.3.22)

(51) Int.Cl.	F I	テーマコード (参考)
HO4N 5/232 (2006.01)	HO4N 5/232	H 2H011
GO2B 7/28 (2006.01)	GO2B 7/28	N 2H151
GO3B 13/36 (2006.01)	GO3B 13/36	5B029
GO6K 9/22 (2006.01)	HO4N 5/232	Z 5C122
	GO6K 9/22	

審査請求 未請求 請求項の数 14 O L (全 16 頁)

(21) 出願番号 特願2016-177878 (P2016-177878)
 (22) 出願日 平成28年9月12日 (2016.9.12)

(特許庁注：以下のものは登録商標)

1. WINDOWS

(71) 出願人 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
 (74) 代理人 110001243
 特許業務法人 谷・阿部特許事務所
 (72) 発明者 大久保 譲
 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
 Fターム(参考) 2H011 AA01 AA02 AA03 BA31 BB04 CA24
 2H151 BA47 DA07 DA18 DA19 DA33
 5B029 AA01 BB02 BB17 CC02 CC25
 5C122 DA09 DA28 EA68 FA05 FD05
 FD13 FH14 HB01 HB05

(54) 【発明の名称】 情報処理装置、プログラム及び制御方法

(57) 【要約】

【課題】カメラを搭載したモバイル端末等を用いて文書を接近して撮影する際のフォーカス調整に要する時間を短くすることを目的とする。

【解決手段】本発明に係る情報処理装置は、撮影手段を有する携帯可能な情報処理装置であって、前記撮影手段で撮影された画像の特徴点を抽出する抽出手段と、抽出された特徴点に基づいて、前記撮影手段におけるフォーカス調整を行う制御手段と、を備え、被写体の全体に対応する全体画像を撮影した後、当該被写体の一部に対応する複数の部分画像を取得するための撮影を、前記撮影手段で行う場合において、前記制御手段は、前記全体画像から抽出された特徴点と、前記複数の部分画像のうち第1の部分画像から抽出された特徴点とのマッチング判定を行って、当該マッチング判定の結果に基づき、前記撮影手段におけるフォーカス調整のための設定値を決定し、当該決定された設定値に従ったフォーカス調整を、前記第1の部分画像よりも後に取得される第2の部分画像の撮影に適用する、ことを特徴とする。

【選択図】 図7

ピクセル距離	フォーカス値
1000	0.00
900	0.04
800	0.09
700	0.16
600	0.25
500	0.36
400	0.49
300	0.64
200	0.81

【特許請求の範囲】**【請求項 1】**

撮影手段を有する携帯可能な情報処理装置であって、
前記撮影手段で撮影された画像の特徴点を抽出する抽出手段と、
抽出された特徴点に基づいて、前記撮影手段におけるフォーカス調整を行う制御手段と

、
を備え、

被写体の全体に対応する全体画像を撮影した後、当該被写体の一部に対応する複数の部分画像を取得するための撮影を、前記撮影手段で行う場合において、

前記制御手段は、前記全体画像から抽出された特徴点と、前記複数の部分画像のうち第 1 の部分画像から抽出された特徴点とのマッチング判定を行って、当該マッチング判定の結果に基づき、前記撮影手段におけるフォーカス調整のための設定値を決定し、当該決定された設定値に従ったフォーカス調整を、前記第 1 の部分画像よりも後に取得される第 2 の部分画像の撮影に適用する、

ことを特徴とする情報処理装置。

【請求項 2】

前記複数の部分画像を取得するための前記撮影は、一定時間内に複数の静止画像を連続して取得する連写又は動画による撮影であることを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

OCR処理を行うOCR手段をさらに備え、

前記制御手段は、取得された前記複数の部分画像のうち、最も鮮鋭性の高い部分画像を、前記OCR処理の対象とする部分画像として決定する、

ことを特徴とする請求項 1 又は 2 に記載の情報処理装置。

【請求項 4】

前記制御手段は、前記複数の部分画像を取得するための前記撮影の際に、前記被写体のうち前記OCR処理を掛けたい部分の位置を示すガイド表示を行なうことを特徴とする請求項 3 に記載の情報処理装置。

【請求項 5】

前記OCR処理の結果が、一定の精度に達していない場合、前記制御手段は、前記部分画像の撮影をやり直すことを特徴とする請求項 3 に記載の情報処理装置。

【請求項 6】

前記制御手段は、前記撮影のやり直しを行う際、撮影条件を変更することを特徴とする請求項 5 に記載の情報処理装置。

【請求項 7】

前記全体画像を撮影する際の解像度は、前記複数の部分画像を撮影する際の解像度よりも高いことを特徴とする請求項 1 乃至 6 のいずれか 1 項に記載の情報処理装置。

【請求項 8】

前記抽出手段は、特徴ベクトルが互いに似通っている特徴点を、抽出結果から除外することを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の情報処理装置。

【請求項 9】

前記被写体における基準となる辺の長さとの複数の前記設定値との対応関係を規定するテーブルをさらに備え、

前記制御手段は、

前記マッチング判定においてマッチしたと判定された特徴点の組合せが 2 組以上ある場合、前記全体画像における特徴点同士の距離と、前記部分画像における特徴点同士の距離との比を表すスケール変化値を求め、

前記全体画像の前記基準となる辺の長さに前記スケール変化値を乗算して前記部分画像における前記基準となる辺の長さを求め、当該求めた長さに対応する設定値を、前記テーブルを参照して決定する

10

20

30

40

50

ことを特徴とする請求項 1 乃至 8 のいずれか 1 項に記載の情報処理装置。

【請求項 10】

前記制御手段は、前記マッチング判定においてマッチしたと判定された特徴点の組合せが3組以上ある場合は、複数得られる前記部分画像における特徴点同士の距離の平均値を用いて、前記スケール変化値を求めることを特徴とする請求項 9 に記載の情報処理装置。

【請求項 11】

前記撮影手段から前記被写体までの距離と複数の前記設定値との対応関係を規定するテーブルをさらに備え、

前記制御手段は、

前記マッチング判定においてマッチしたと判定された特徴点の組合せが4組以上ある場合、カメラ外部パラメータを推定し、当該推定したカメラ外部パラメータを用いて、前記全体画像上の任意の座標をカメラ空間座標に変換することにより、当該任意の座標に対応する、前記部分画像における座標の奥行きを推定し、

推定された前記部分画像における奥行きに対応する設定値を、前記テーブルを参照して決定する

ことを特徴とする請求項 1 乃至 8 のいずれか 1 項に記載の情報処理装置。

【請求項 12】

前記撮影手段から前記被写体までの距離と複数の前記設定値との対応関係を規定するテーブルをさらに備え、

前記制御手段は、

前記マッチング判定においてマッチしたと判定された特徴点の組合せが4組以上ある場合、ホモグラフィ行列を推定し、当該推定したホモグラフィ行列を用いて、前記全体画像上の任意の座標をカメラ空間座標に変換することにより、当該任意の座標に対応する、前記部分画像における座標の奥行きを推定し、

推定された前記部分画像における奥行きに対応する設定値を、前記テーブルを参照して決定する

ことを特徴とする請求項 1 乃至 8 のいずれか 1 項に記載の情報処理装置。

【請求項 13】

撮影手段を有する携帯可能な情報処理装置の制御方法であって、

前記情報処理装置は、前記撮影手段で撮影された画像の特徴点を抽出する抽出手段と、抽出された特徴点に基づいて、前記撮影手段におけるフォーカス調整を行う制御手段と、を備え、

前記制御方法は、

前記撮影手段を用いて、被写体の全体に対応する全体画像を撮影する第 1 の撮影ステップと、

前記第 1 の撮影ステップに続いて、前記撮影手段を用いて、前記被写体の一部に対応する複数の部分画像を撮影する第 2 の撮影ステップと、

前記全体画像から抽出された特徴点と、前記複数の部分画像のうち第 1 の部分画像から抽出された特徴点とのマッチング判定を行って、当該マッチング判定の結果に基づき、前記撮影手段におけるフォーカス調整のための設定値を決定し、当該決定された設定値に従ったフォーカス調整を、前記第 2 の撮影ステップにおける撮影に適用するステップと、を含むことを特徴とする情報処理装置の制御方法。

【請求項 14】

コンピュータを、請求項 1 乃至 12 のいずれか 1 項に記載の情報処理装置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、情報処理装置、プログラム及び制御方法に関する。

【背景技術】

【 0 0 0 2 】

現在ではカメラ機能を持つモバイル端末が広く普及している。従来、ユーザは文書を電子的に取り込むためにスキャナなどを利用していたが、モバイル端末のカメラ機能を利用することで簡単に文書を画像データとして取り込むことが出来るようになった。

【 先行技術文献 】

【 特許文献 】

【 0 0 0 3 】

【 特許文献 1 】 特開 2 0 0 4 3 1 7 6 9 9 号 公 報

【 発明の概要 】

【 発明が解決しようとする課題 】

10

【 0 0 0 4 】

モバイル端末のカメラで文書を画像データとして取り込み、OCRを使用し文字列データを取得してそのテキスト情報を利用する際には、OCR精度が非常に重要となる。名刺や免許証といった対象物に接近して取り込めるケースと異なり、A4やA3のような比較的大きな用紙等に印刷された文書を対象とする場合、OCR精度が悪化することになる。これは対象物との距離が離れることにより、識別しようとするテキストの1文字当たりの解像度が低下するためである。そこでA4やA3サイズの文書を対象とするケースにおいては、文書内の目的のテキスト領域に接近して撮影することがOCR精度向上には望ましい。しかしながら、文書の一部に接近して撮影する際には、カメラと被写体との距離が変動するため、フォーカスを調整し直す必要がある。そして、OCRに適したフォーカスの合った画像の撮影には時間が掛かりやすく、特に十分な光量の無い屋内での撮影するケースでは、フォーカス調整に多くの時間を通常要することになる。この点、例えば特許文献1には、被写体の顔の大きさや目幅などの特徴部位の情報に基づいて、被写体までの距離を算出し、フォーカスレンズの移動範囲を調整するという手法が開示されている。

20

【 0 0 0 5 】

しかし、上記特許文献1の手法は、主として顔の特徴解析から被写体までの距離を算出するもので、被写体が紙等の記録媒体に印刷された文書であるユースケースには馴染まない。顔の特徴に比べ、文書は罫線が交差する箇所など特徴的ではあるが類似する点が多数存在するためである。また、顔の特徴解析では、顔の各パーツの特徴に関して区別が付きやすく、その個数も予め分かっているが、文書の場合は特徴の区別やその個数を予め把握することが困難である。仮に上記特許文献1の手法を文書の撮影に適用すると、特徴点抽出の際に、本来は異なる点を同一の特徴点として認識してしまうなどの弊害が予想される。

30

【 0 0 0 6 】

そこで、本発明は、カメラを搭載したモバイル端末等を用いて文書を接近して撮影する際のフォーカス調整に要する時間を短くすることを目的とする。

【 課題を解決するための手段 】

【 0 0 0 7 】

本発明に係る情報処理装置は、撮影手段を有する携帯可能な情報処理装置であって、前記撮影手段で撮影された画像の特徴点を抽出する抽出手段と、抽出された特徴点に基づいて、前記撮影手段におけるフォーカス調整を行う制御手段と、を備え、被写体の全体に対応する全体画像を撮影した後、当該被写体の一部に対応する複数の部分画像を取得するための撮影を、前記撮影手段で行う場合において、前記制御手段は、前記全体画像から抽出された特徴点と、前記複数の部分画像のうち第1の部分画像から抽出された特徴点とのマッチング判定を行って、当該マッチング判定の結果に基づき、前記撮影手段におけるフォーカス調整のための設定値を決定し、当該決定された設定値に従ったフォーカス調整を、前記第1の部分画像よりも後に取得される第2の部分画像の撮影に適用する、ことを特徴とする。

40

【 発明の効果 】

【 0 0 0 8 】

50

本発明によれば、カメラを搭載したモバイル端末等を用いて文書を接近して撮影する際のフォーカス調整に要する時間を短くできる。

【図面の簡単な説明】

【0009】

【図1】モバイル端末の外観の一例を示す図である。

【図2】モバイル端末のハードウェア構成の一例を示す図である。

【図3】モバイル端末のソフトウェア構成の一例を示す図である。

【図4】モバイルアプリのUI画面の一例を示す図である。

【図5】全体画像と部分画像の一例を示す図である。

【図6】実施例1に係る、文字情報取り込み処理の流れを示すフローチャートである。

10

【図7】ピクセル距離フォーカス値テーブルの一例を示す図である。

【図8】文書面と撮影面とが平行になっていない状態を示す図である。

【図9】実施例2に係る、文字情報取り込み処理の流れを示すフローチャートである。

【発明を実施するための形態】

【0010】

以下、本発明を実施するための形態について図面などを参照して説明する。なお、実施形態は、本発明を限定するものではなく、また、実施形態で説明されている全ての構成が本発明の課題を解決するための手段に必須であるとは限らない。

【実施例1】

【0011】

20

本実施形態に係る携帯可能な情報処理装置の一例として、以下ではカメラ付きのモバイル端末を例に説明する。いわゆるタブレットPCやスマートフォンは、無線通信機能などの装備によって自由な場所で撮影やデータ通信などができる、モバイル端末の代表例である。

【0012】

図1は、モバイル端末の外観の一例を示す図である。図1(a)は、タッチパネル101があるモバイル端末100の表側(正面)を示している。タッチパネル101は、液晶ディスプレイ等の表示部の一例であり、情報の出力(表示)と入力との2つの機能を備えている。図1(b)は、モバイル端末100の裏側(背面)を示している。モバイル端末の背面には、画像を取り込むための撮像用レンズ(以下、単に「レンズ」とする)102を備える。本実施例では、モバイル端末100のユーザは、被写体を後述のモバイルアプリケーション(以下、「モバイルアプリ」と呼ぶ。)を使って撮影する。本実施例における被写体は、所定サイズの用紙に所定の様式で作成された注文書110である。なお、被写体は、注文書110のような文書の他、名刺や写真、或いはホワイトボードに書き込まれた文字等であっても良い。後述のモバイルアプリは、被写体の画像を取り込み、タッチパネル101に当該取り込んだ画像を表示することができる。

30

【0013】

[ハードウェア構成]

続いて、モバイル端末100のハードウェア構成について説明する。図2は、モバイル端末100のハードウェア構成の一例を示す図である。モバイル端末100は、CPU201、RAM202、ROM203、入出力I/F204、NIC205、カメラ部206、加速度/ジャイロセンサ207で構成され、これら各部はバス208で相互に接続されている。

40

【0014】

CPU201は、各種のプログラムを実行して、様々な機能を実現する演算処理装置である。RAM202は、各種の情報を記憶する読み書き可能なメモリである。また、RAM202は、CPU201のワークエリアとしても利用される。ROM203は、OSや上述の撮影アプリ等の各種プログラムを記憶するメモリである。例えば、CPU201は、ROM203に記憶されているプログラムをRAM202にロードしてプログラムを実行する。また、CPU201は、フラッシュメモリ、HDD、SSDといった外部記憶装置(不図示)に記憶されているプログラムをRAM202に読み込んで実行することもできる。なお、モバイル端末100の機

50

能及び後述するシーケンスに係る処理の全部又は一部については専用のハードウェアを用いて実現してもよい。

【0015】

入出力I/F204は、タッチパネル101に対して表示データを出力したり、タッチパネル101からの入力情報を受け付けるインタフェースである。NIC(Network Interface Card)205は、モバイル端末100をネットワーク(不図示)に接続するためのインタフェースである。カメラ部206は、レンズ102を介して被写体の画像をモバイル端末100に取り込む。バス208は、上述した各部を繋ぐデータ通信路である。加速度/ジャイロセンサ207は、モバイル端末100の姿勢情報を検出するセンサである。

【0016】

[ソフトウェア構成]

次に、モバイル端末100のソフトウェア構成について説明する。図3は、モバイル端末100のソフトウェア構成の一例を示す図である。モバイル端末100のソフトウェアは、データ管理モジュール300とモバイルアプリ310で構成される。そして、モバイルアプリ310は、メイン制御、情報表示、操作情報取得、画像解析、OCR、記憶処理、座標処理、画像特徴点抽出の各機能に対応する複数のモジュール311~318で構成される。前述の通り、これら各モジュールに相当するプログラムは、ROM203等に記憶されている。

【0017】

データ管理モジュール300は、画像データやモバイルアプリ310における処理データ(アプリデータ)などを管理する。モバイルアプリ310は、不図示のOSが提供する制御API(Application Programming Interface)を利用することで、データ管理モジュール300が管理する画像データやアプリデータの取得や保存を行う。

【0018】

ユーザは、モバイル端末100のOSのインストール機能を利用することによって、モバイルアプリ310のダウンロードとインストールが可能である。モバイルアプリ310は、カメラ部206を介して取り込んだ画像データに対して各種の処理を行う。

【0019】

メイン制御モジュール311は、モバイルアプリ310を統括的に制御するモジュールであり、以下の各モジュール312~318に対する指示及び管理を行う。情報表示モジュール312は、メイン制御モジュール311からの指示に従い、モバイルアプリ310のユーザインタフェース用画面(UI画面)を提供する。図4は、モバイルアプリ310のUI画面の一例を示す図である。UI画面400は、モバイル端末100のタッチパネル101に表示される。UI画面400の表示・操作領域401にはカメラ部206を介して取り込んだ画像が表示され、ユーザは当該画像に対して各種操作を行うことができる。なお、モバイルアプリ310のUIの形態(位置、大きさ、範囲、配置、表示内容など)は、図示するものに限定されないことはいうまでもない。

【0020】

操作情報取得モジュール313は、上記UI画面400を介したユーザ操作に係る入力情報を取得し、取得した入力情報をメイン制御モジュール311に渡す。例えば、表示・操作領域401をユーザが手で触れると、操作情報取得モジュール313は、触れられたUI画面400上の位置を検知し、当該検知した位置の情報をメイン制御モジュール311に送信する。

【0021】

画像解析モジュール314は、カメラ部206を介して取得した画像データを解析する。この解析には、被写体が例えば上述の注文書110である場合には、撮影した画像内の注文書110の領域を検出し、文書の種類やサイズを特定する文書検出処理や、画像内で被写体が傾いている場合に修正する歪み補正処理(台形補正処理)などが含まれる。なお、カメラ部206は、単にスチール撮影を行えるだけでなく、静止画像を高速で連続して取得する連写撮影機能(或いは動画撮影機能)も備える。

10

20

30

40

50

【 0 0 2 2 】

OCRモジュール315は、画像内のテキスト領域に対して、OCR（光学文字認識）処理を行う。記憶処理モジュール316は、ユーザにより入力された各種設定値の他、ユーザ選択に係る画像領域の座標、画像の特徴点、画像のエッジ度といった各種データをRAM202やHDD等の記憶装置に保存し、必要に応じてデータベースとして登録する処理を行う。座標処理モジュール317は、矩形領域の座標情報をデータベースとして登録する際の座標変換処理や画像の位置特定処理を行う。特徴点抽出モジュール318は、入力された画像データに対して特徴点の抽出処理を行う。本実施例における特徴点は、文書画像上の際立った点（自然特徴点）である。

【 0 0 2 3 】

[全体画像と部分画像]

次に、本実施例においてユーザが撮影を行って取得する、全体画像と部分画像について説明する。図5は、全体画像と部分画像の一例を示す図である。図5において、画像500は、被写体としての注文書110全体が包含されるように撮影された画像に対して被写体以外の領域を除外する処理や歪みを補正する処理を施して整形して得られた、被写体全体に対応する全体画像である。そして、画像510は、OCR処理の対象としたい注文書110の一部に対して、レンズ102を接近して撮影（もしくは光学ズームで拡大撮影）することにより取得した、被写体の一部に対応する部分画像である。すなわち、部分画像510は、被写体の一部を高解像度の画像として取得したものに等しい。なお、全体画像500内の破線で示す領域501は、部分画像510として撮影された部分の相対位置を示している。

【 0 0 2 4 】

[文字情報取り込み処理]

続いて、モバイル端末100を用いて文書等のテキスト領域を撮影した画像にOCR処理を行い、文字情報を保存するまでの処理（文字情報取り込み処理）の流れについて説明する。図6は、本実施例に係る、文字情報取り込み処理の流れを示すフローチャートである。本フローは、ユーザがモバイルアプリ310を起動させることをトリガーに開始する。

【 0 0 2 5 】

ステップ601では、メイン制御モジュール311において、全体画像データが取得される。具体的には、まず、ユーザがモバイル端末100の撮影機能を用いて、被写体となる文書等（ここでは、注文書110）の全体を撮影する。この際、ユーザが手動（撮影ボタンの押下）で撮影を行ってもよいし、オートフォーカス機能を使用するなどしてフォーカスが合う等の所定条件の充足を検出したタイミングで自動撮影してもよい。全体画像のデータがメイン制御モジュール311に入力されると、画像解析モジュール314において、前述の文書検出処理や歪み補正処理がなされる。例えば、文書検出処理では、撮影された画像と予め記憶した文書毎（注文書、請求書、納品書など）の特徴的なパターンとを比較するなどして、撮影画像内の文書領域（＝全体画像）に加え、当該文書の種類やサイズが特定される。例えば、図1のように注文書110が収まるよう撮影がなされた場合において、カメラ部206の解像度が640×480ピクセルであったとする。この場合の全体画像のサイズは、表示・操作画面401に占める検出された文書領域の割合から、例えば512×430（単位：ピクセル数）といった値が求まる。また、文書サイズに関しては、凡そピントが合っているとの前提で、全体画像撮影時のカメラ部206のフォーカス値に基づき推測してもよい。この推測に際しては、例えば、A4やA3といった所定サイズに対応する最良のフォーカス値を規定したテーブルを予め用意し、これを参照すればよい。なお、全体画像を取得する際の解像度を、後述の部分画像の取得する際の解像度よりも高解像度に（例えば、部分画像を640×480ピクセル、全体画像を1024×768ピクセルなど）してもよい。これは、全体画像を取得するための撮影は本ステップの1回のみで部分画像のように繰り返し行わないこと、高解像度にすることで全体画像特徴点の抽出時に細かな特徴を捉えることが可能になりマッチング精度が上がること、等がその理由である。上述のような解析によって取得された全体画像のデータは、特徴点抽出モジュール318に送られる。

10

20

30

40

50

【0026】

ステップ602では、特徴点抽出モジュール318において、全体画像データから特徴点（全体画像特徴点）が抽出される。特徴点の抽出にはHarrisのコーナー検出やSIFT(Scale-Invariant feature Transform)、SURF(Speeded Up Robust Features)などを適用すればよい。なお、被写体が注文書等の帳票である場合には、一般的に罫線の交わる部分やコーナーが特徴点として考えられるが、これらは似たような箇所が多く文書内に存在し、特徴点として適さないことがある。そこで、上述の手法によって特徴点が抽出出来たとしても、例えばその特徴ベクトルが互いに似通っている特徴点を、抽出結果から除外してもよい。その結果、相互に類似しない特徴点、具体的には、特殊なイラストや文字といったような、罫線やコーナー以外の箇所が特徴点として残る。特徴ベクトルが互いに似通っているかどうかの判定には、ユークリッド距離やマンハッタン距離、相関係数などを用いればよい。前述の図5の例では、全体画像500内の2つの丸印502及び503が、全体画像500から抽出された特徴点を示している。

10

【0027】

ステップ603では、特徴点抽出モジュール318において、抽出された全体画像特徴点間の距離が導出される。この場合において、距離の単位はピクセル数である。図5に示した全体画像特徴点502及び503の場合、両特徴点の位置座標(x,y)を結ぶ両方向矢印504の長さに相当するピクセル数が、全体画像特徴点間の距離として導出されることになる。導出されたピクセル数は、記憶処理モジュール316によって、全体画像特徴点間の距離としてRAM202に保存される。

20

【0028】

ステップ604では、メイン制御モジュール311において、部分画像データが取得される。具体的には、ユーザがモバイル端末100の連写（若しくは動画）撮影機能を用いて、被写体となる文書等のうち、OCR処理の対象となり得る部分を所定枚数（所定フレーム数）撮影することで、部分画像のデータがメイン制御モジュール311に入力される。このように連写又は動画で部分画像を撮影する理由は2つある。1つ目は、ピンボケや手ブレがあると特徴点の抽出精度が落ちてしまうところ、一定時間内（例えば1.0sec）に高速で連続撮影を行って複数の部分画像を取得して、特徴点抽出精度の良い1枚をその中から選択するためである。2つ目は、複数の部分画像を撮影する場合において、次に撮影して欲しい部分に誘導するためのトラッキング処理を行うためである。この部分画像の撮影に際しては、例えば被写体である文書等のうちOCR処理を掛けたい所望の部分に、レンズ102を接近又は光学ズームで拡大し、例えばオートフォーカス機能によってフォーカスが合ったタイミングで撮影が自動で開始するようにする。また、被写体となる文書のフォーマットが既知で、OCR処理を掛けたい部分が事前に把握できている場合には、当該部分の位置を表示・操作領域401上に示すガイド表示を行ってもよい。このガイド表示としては、例えば、当該部分に対応する枠を表示させたり、当該部分の輝度を他の部分よりも上げるといった方法が考えられる。もちろん、全体画像の撮影時と同様、ユーザが撮影ボタンを押下することによって部分画像の撮影を行ってもよい。このようにして、図5に示すような、全体画像500の一部を拡大した部分画像510が取得される。

30

【0029】

ステップ605では、特徴点抽出モジュール318において、部分画像データから特徴点（部分画像特徴点）が、前述のステップ602と同様の手法にて抽出される。前述の図5の例では、部分画像510内の2つの丸印511及び512が、部分画像510から抽出された特徴点を示している。

40

【0030】

ステップ606では、メイン制御モジュール311において、ステップ602で抽出した全体画像特徴点とステップ605で抽出した部分画像特徴点との間のマッチングが判定される。すなわち、特徴点同士の類似度を求め、類似度が一定の閾値以上かどうか判定される。類似度を求める手法としては、例えば、2つの特徴点の特徴ベクトルに対してユークリッド距離やマンハッタン距離を算出する手法や、相関係数を算出する手法がある。

50

また、誤判定を減らすために、マッチング結果を1対1に絞る、クロスチェックを取り入れてもよい。図5の例では、全体画像特徴点502と部分画像特徴点511との組合せと、全体画像特徴点503と部分画像特徴点512との組合せの2組について、それぞれマッチしたと判定されることになる。

【0031】

ステップ607では、メイン制御モジュール311において、マッチング判定の結果、マッチしたと判定された特徴点の組合せが少なくとも2組以上あるかどうか判定される。マッチしたと判定された特徴点の組合せの数が2組以上の場合は、ステップ608に進む。一方、マッチしたと判定された特徴点の組合せの数が2組未満の場合は、ステップ614に進む。

10

【0032】

ステップ608では、メイン制御モジュール311において、部分画像特徴点間の距離が導出される。図5に示した部分画像特徴点511及び512の場合、両特徴点の位置座標(x,y)を結ぶ両方向矢印513の長さに相当するピクセル数が、部分画像特徴点間の距離として導出されることになる。導出されたピクセル数は、記憶処理モジュール316によって、部分画像特徴点間の距離としてRAM202に保存される。

【0033】

ステップ609では、メイン制御モジュール311において、特徴点間距離のスケール変化値、すなわち、部分画像特徴点間の距離に対応する全体画像特徴点間の距離の比が導出される。

20

【0034】

ステップ610では、メイン制御モジュール311において、導出したスケール変化値に基づいて、部分画像の撮影に適するフォーカス値が決定される。この際、対象文書における基準となる縦又は横の一辺(基準辺)の長さを表すピクセル数(ピクセル距離)とフォーカス値との対応関係を示すテーブル(ピクセル距離フォーカス値テーブル)が参照される。ここで、フォーカス値とは、カメラ部206内のフォーカスレンズ(不図示)の位置を制御してフォーカス調整を行うための設定値であり、カメラ部206の光学構成、OS、ドライバによって変わり得る。このフォーカス値は、例えば、マイクロソフト社の提供するOS(Windows)であれば「CameraControl_Focus」、アップル社の提供するOS(iOS)であれば「LensPosition」にそれぞれ相当する。ピクセル距離フォーカス値テーブルは、

30

【0035】

図5のケースを例に具体的に説明する。いま、全体画像特徴点間距離504が250ピクセル、部分画像特徴点間距離513が415ピクセル、全体画像500における基準辺(ここでは縦方向の長辺)の長さが512ピクセルであったとする。まず、ステップ609において、スケール変化値1.66(=部分画像特徴点間距離:415ピクセル/全体画像特徴点間距離:250ピクセル)が導出される。そして、全体画像500の基準辺の長さが512ピクセルであるので、この512ピクセルにスケール変化値1.66を乗算した値=850ピクセルが、部分画像501における基準辺のピクセル距離となる。最後に、図7のテーブルを参照して、部分画像501の基準辺のピクセル距離(850ピクセル)に対応するフォーカス値を取得する。この際、本実施例のように基準片のピクセル距離に直接対応するフォーカス値がテーブル内にはない場合には、補間処理によって求めればよい。例えば、ピクセル距離が840の場合、ピクセル距離:900に対応するフォーカス値:0.04と、ピクセル距離:800に対応するフォーカス値:0.09との中間値である0.065が、最適なフォーカス値として決定さ

40

50

れる。なお、補間処理には、線形補間、多項式補間、キュービック補間など公知の手法をOSやドライバ等も考慮して適宜適用すればよい。以上のようにして決定されたフォーカス値によって、カメラ部206のフォーカス値は更新（新たなフォーカス値として設定）される。この際、単に新たなフォーカス値を固定値で設定するだけでなく、決定されたフォーカス値に余裕（例えば±10%）を持たせ、幅のあるフォーカス値を設定してもよい。これにより、カメラ部206に備わるオートフォーカス機能も併用することができる。

【0036】

ステップ611では、画像解析モジュール314において、ステップ604で取得したフレームの部分画像についてエッジ度が検出される。このエッジ度は、部分画像の鮮鋭性を評価するためのもので、エッジ度が高いほど鮮鋭性が高く、OCR処理に適した画像と判断することができる。エッジ度の検出には公知のフィルタ処理を適用すればよい。

10

【0037】

ステップ612では、メイン制御モジュール311において、今回検出されたエッジ度が最良のエッジ度であるかが判定される。具体的には、過去に検出されたエッジ度のうち最も高いエッジ度を最良のエッジ度として保存するようにする。そして、エッジ度が検出される度に当該保存されたエッジ度と比較して、新たに検出されたエッジ度の方が高い場合には、それを新たな最良のエッジ度として保存するようにする。なお、エッジ度が最初に検出された場合には、比較対象となる保存されたエッジ度が存在しないので、判定を行うことなく最良のエッジ度として扱われる。判定の結果、今回検出されたエッジ度が最良のエッジ度である場合は、ステップ613に進む。一方、今回検出されたエッジ度が最良のエッジ度でない場合は、ステップ614に進む。

20

【0038】

ステップ613では、記憶処理モジュール316によって、ステップ612で最良と判定されたエッジ度を持つ部分画像が、OCR処理用の画像としてRAM202に保存される。

【0039】

ステップ614では、メイン制御モジュール311において、取得した部分画像が所定枚数に達したかが判定される。ここで、所定枚数は、カメラ部206の連写或いは動画モードでの撮影性能にも拠るが、例えば所定の撮影時間（例えば1sec）に取得可能なフレーム数（例えば10~30フレーム）となる。取得した部分画像が所定枚数に達していれば、ステップ615に進む。一方、取得した部分画像が所定枚数に達していなければ、ステップ604に戻って次のフレームの撮影を行って、ステップ605以降の処理を繰り返す。なお、部分画像の撮影の終了を、取得枚数ではなく撮影開始からの経過時間（例えば1sec）で判断してもよい。

30

【0040】

ステップ615では、OCRモジュール315において、撮影された部分画像のうち最良のエッジ度を持つ部分画像として保存された画像をRAM202から呼び出し、OCR処理を行う。

【0041】

ステップ616では、メイン制御モジュール311において、OCR処理の結果（文字認識率）が十分な精度、例えば一定のOCR確信度に対応する閾値以上であるかが判定される。文字認識率が一定の閾値未満であれば、ステップ604に戻って部分画像の撮影からやり直す。この際、ユーザに所定の認識率が得られなかった旨を通知してもよい。また、露出を変更するなど異なる撮影条件の下で部分画像の撮影を開始するようにしてもよい。さらには、点線で示すようにステップ601に戻って全体画像の撮影からやり直してもよい。

40

【0042】

ステップ617では、メイン制御モジュール311において、ユーザによる目視チェックの結果を受け付け、処理の切り分けがなされる。例えば、表示・操作領域401に表示されたOCR処理の結果をユーザが実際に確認して、問題がなければOKボタン、問題があればNGボタン（いずれのボタンも不図示）を押下するようにする。こうして、ユーザによる

50

目視チェックの結果に従い、問題がなければステップ618に進む。一方、問題があればステップ604（若しくは601）に戻り、上述のとおり撮影をやり直す。なお、ユーザによる目視チェックが困難なユースケースなどでは本ステップを省略しても構わない。

【0043】

ステップ618では、データ管理モジュール300によって、ステップ615でのOCR処理の結果が、その基となった撮影画像（少なくとも部分画像を含み、さらには全体画像を含んでもよい。）と共に保存される。なお、保存先は、モバイル端末100が有する不揮発性の外部記憶装置（例えばフラッシュメモリ）でもよいし、LANやインターネットといったネットワーク上のPCやサーバにアップロードしてもよい。また、複数の文書についてのOCR処理の結果をまとめて、所定時間経過後に、他のPCやサーバにアップロードしてもよい。

10

【0044】

以上が、文字情報取り込み処理の内容である。

【0045】

なお、マッチしたと判定された特徴点の組合せが3組以上ある場合には、部分画像特徴点間距離を複数得ることができ、その結果、ステップ608では基準辺のピクセル距離が複数得られることになる。この場合、得られた複数のピクセル距離の平均値を取ることにより、特徴点の誤抽出に対して耐性のある処理とすることができる。また、部分画像を所定フレーム数取得後（或いは撮影開始から所定時間経過後）もマッチする特徴点の組合せが2組以上でない場合には、その旨を知らせる通知等のフォールバック処理を行ってもよい。

20

【0046】

また、特徴点の抽出方法によっては、LoG(Laplacian of Gaussian)やDoG(Difference of Gaussian)画像から極値を求めるものがある。極値の存在するスケール係数を、決定したフォーカス値の補正や、不正なフォーカス値を除外する判断に使用してもよい。

【0047】

さらに、ピクセル距離フォーカス値テーブルは、カメラ部206のズーム機能に応じて複数用意してもよい。これは光学ズームの性能・構造によって、最適なフォーカス値が変化し得るためである。また、決定されたフォーカス値を設定する際には、単に新たなフォーカス値を固定値として設定するのではなく、決定されたフォーカス値に余裕（例えば±10%）を持たせ、幅のあるフォーカス値を設定してもよい。これにより、カメラ部206に備わるオートフォーカス機能を併用することができる。

30

【0048】

以上のように本実施例によれば、カメラ付きのモバイル端末等を用いて文書等に接近して画像を取り込む際に、特徴点間距離のスケール変化に基づき、素早く最適なフォーカス値を決定することができる。

【実施例2】

【0049】

実施例1では、文書等の全体画像における特徴点間距離と部分画像における特徴点間距離との比（スケール変化値）に基づいて、最適なフォーカス値を決定する態様を説明した。次に、撮影面と文書面とが平行ではない場合に、全体画像における任意座標に対応する、部分画像における座標の奥行きを推定して、最適なフォーカス値を決定する態様について、実施例2として説明する。なお、実施例1と共通する内容については説明を省略ないしは簡略化し、以下では差異点を中心に説明を行うものとする。

40

【0050】

図8は、被写体である注文書110の面（文書面）と撮影された全体画像800の面（撮影面）とが平行ではなく大きく傾いている状態を示している。そして、図8において、破線で示す枠810は、図5における部分画像510に相当する位置を示している。枠810内の4つの印は特徴点を示し、1つの×印はフォーカスを合わせたい位置を示している。

50

【 0 0 5 1 】

最初に、本実施例で用いる奥行推定手法の基礎となる知識を確認しておく。以下の式(1)で示す行列式において、(X,Y,Z)は単位をミリメートルとするワールド空間における1つの点の座標を表している。また、(x,y)は当該1つの点を、カメラを通して射影された画像空間上の点の座標を表している。本実施例の場合、(x,y)の単位はピクセルである。つまり式(1)は、カメラからワールド空間上の或る点を見た場合、画像空間上でどの位置に表示されるかを表している。

【 0 0 5 2 】

【 数 1 】

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \dots \text{式 (1)}$$

10

【 0 0 5 3 】

上記式(1)におけるsは、同次座標のスケール不変を表すスケールファクタである。また、 f_x および f_y は、ピクセル単位の焦点距離である。また、 c_x 及び c_y は、画像の中心が原点では無い場合に、画像の中心をピクセル単位で表している。また、 $r_{11} \sim r_{33}$ は、ワールド空間から見た時のカメラの姿勢を表す回転成分である。そして、 t_x と t_y と t_z は、ワールド空間から見た時のカメラの位置を表す、平行移動成分である。上記式(1)は2つの行列の積で表現されており、以下の式(2)のように表現することもできる。

20

【 0 0 5 4 】

【 数 2 】

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = M_i M_e \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \dots \text{式 (2)}$$

30

【 0 0 5 5 】

上記式(2)において、 M_i はカメラ空間座標から画像空間座標へと変換するパラメータ(カメラ内部パラメータ)を表し、 M_e はワールド空間座標をカメラ空間座標に変換するパラメータ(カメラ外部パラメータ)を表す。なお、 M_i にはカメラの歪み成分を含める場合もある。

【 0 0 5 6 】

カメラ内部パラメータ M_i を求めることをカメラキャリブレーションと呼び、様々な方法がすでに提案されている。なお、本実施例ではカメラキャリブレーションは事前に行われており M_i は既知であるものとする。また、カメラ外部パラメータ M_e はワールド空間から見たカメラの位置及び姿勢を表すものである。したがって、 M_e を推定することが出来れば、カメラを動かした際のカメラの三次元位置、またはカメラを固定した状態で物体が移動した場合の物体の三次元位置が求められる。

40

【 0 0 5 7 】

ワールド空間上の点と画像空間上の点との複数の組、及びカメラ内部パラメータ M_i が既知である時、カメラ外部パラメータ M_e を求める問題は、一般にPerspective-n-Point問題として知られ、様々な解法が考えられている。特に、推定に用いるワールド空間上の点が全て同一平面上にあるとき、少なくとも4つ以上の点の組の位置が既知であれば、カメラ外部パラメータ M_e が推定可能であることが知られている。

50

【 0 0 5 8 】

そこで本実施例では、マッチング判定の結果、マッチすると判定された特徴点の組合せが4組以上ある場合に、まずカメラ外部パラメータ M_e の推定を行なう。カメラ外部パラメータ M_e を推定した後、被写体である文書上の任意の点座標を、推定したカメラ外部パラメータ M_e で変換することで、カメラ空間上における座標を求める。こうして、カメラ空間上に変換した際のz座標の値が、単位をミリメートルとしたカメラから文書上の任意の点までの距離、すなわち、奥行となる。

【 0 0 5 9 】

図9は、本実施例に係る、文字情報取り込み処理の流れを示すフローチャートである。ステップ901及び902は、実施例1の図6のフローにおけるステップ601及び602に相当する。すなわち、被写体となる文書の全体画像の取得(ステップ901)と特徴点の抽出がなされる(ステップ902)。続くステップ903~905は、実施例1の図6のフローにおけるステップ604~606に相当する。すなわち、連写機能等を用いた部分画像の取得(ステップ903)、部分画像特徴点の抽出(ステップ904)、全体画像特徴点と部分画像特徴点とのマッチング判定(ステップ905)がなされる。

10

【 0 0 6 0 】

そして、ステップ906では、メイン制御モジュール311において、上述の通り、マッチング判定の結果、マッチしたと判定された特徴点の組合せが少なくとも4組以上あるかどうか判定される。マッチしたと判定された特徴点の組合せの数(マッチング数)が4組以上の場合は、ステップ907に進む。一方、マッチしたと判定された特徴点の組合せの数が4組未満の場合は、カメラ外部パラメータステップ M_e の推定ができないので、ステップ913に進む。

20

【 0 0 6 1 】

ステップ907では、メイン制御モジュール311において、マッチしたと判定された4組以上の特徴点の組合せと、事前に推定したカメラ内部パラメータ M_i から、カメラ外部パラメータ M_e の推定がなされる。この際、マッチしたと判定された特徴点の組合せが5組以上ある場合には、特徴点の外れ値を推定に用いないようにしてもよい。このような手法は一般にRANSACと呼ばれており、推定の外れ値に対するロバストネスを高めることができる。

30

【 0 0 6 2 】

ステップ908では、メイン制御モジュール311において、全体画像における任意座標に対応する、部分画像における座標の奥行を推定する。ここで、任意座標は、OCR処理を掛けたい領域内における、上述の×印820で示したフォーカスを合わせたい位置である。ステップ907で推定したカメラ外部パラメータ M_e は、ワールド空間座標をカメラ空間座標へ変換する座標変換行列である。そのため、フォーカスを合わせたい全体画像上の任意の座標(ワールド空間上の座標)を、カメラ外部パラメータでカメラ空間座標へ変換し、変換後のz値を見ることで、この奥行を推定することができる。

【 0 0 6 3 】

ステップ909では、メイン制御モジュール311において、推定された部分画像における座標の奥行きに基づき、カメラ部206のフォーカス値を決定する。この決定に際しては、カメラ(=レンズ102)から被写体までの距離(単位:mm)と最適なフォーカス値とを対応付けたテーブルを事前に用意し不図示のフラッシュメモリ等に保持しておくものとする。例えば、カメラからの距離が200mmの時はフォーカス値0.2、400mmの時はフォーカス値0.4といったように、カメラからの距離と最適フォーカス値とが1対1で対応づけられている。なお、カメラからの距離とフォーカス値との関係は、使用するカメラ、OS、カメラのドライバに依存するものである。以上のようにして決定されたフォーカス値によって、カメラ部206のフォーカス値は更新(新たなフォーカス値として設定)される。

40

【 0 0 6 4 】

続くステップ910~917は、実施例1の図6のフローにおけるステップ611~618に相当し、特に異なるところはないので省略する。

50

【0065】

なお、本実施例では、奥行き推定にカメラ外部パラメータを用いたが、他の手法、例えばホモグラフィ行列を用いてもよい。

【0066】

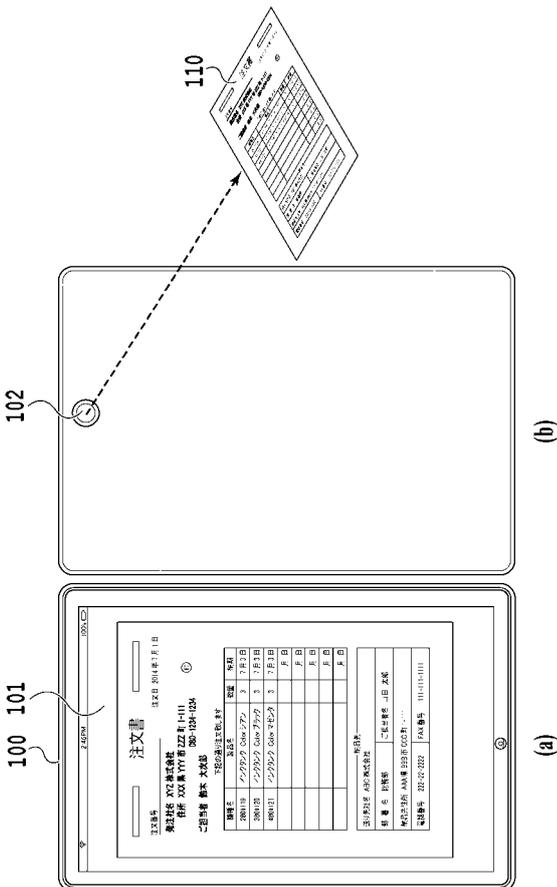
以上のように本実施例によれば、モバイル端末等を用いて文書等に接近してその一部の画像を取り込む際に、文書面と撮影面とが平行になっていなくても、素早く任意の点にフォーカス値を合わせることができる。

【0067】

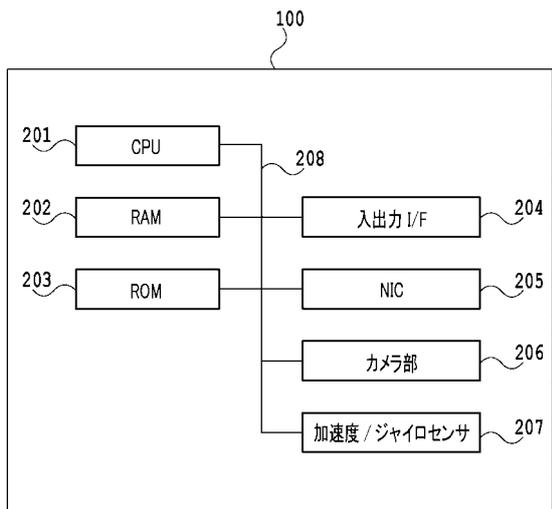
[その他の実施例]

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。

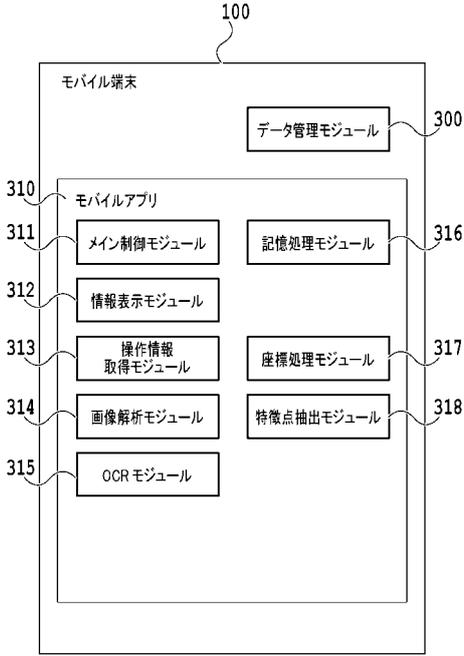
【図1】



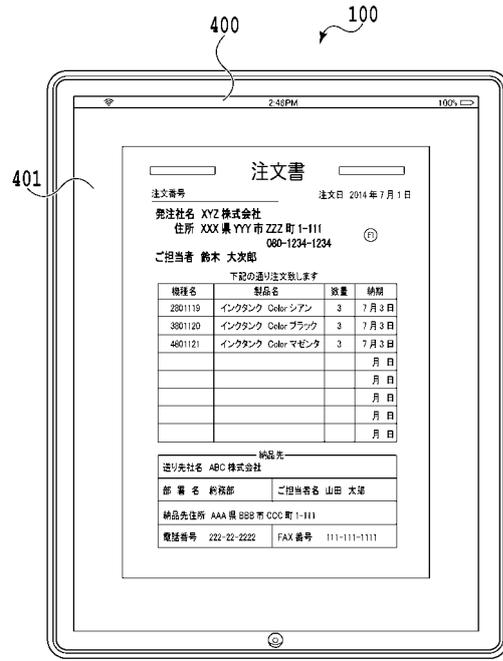
【図2】



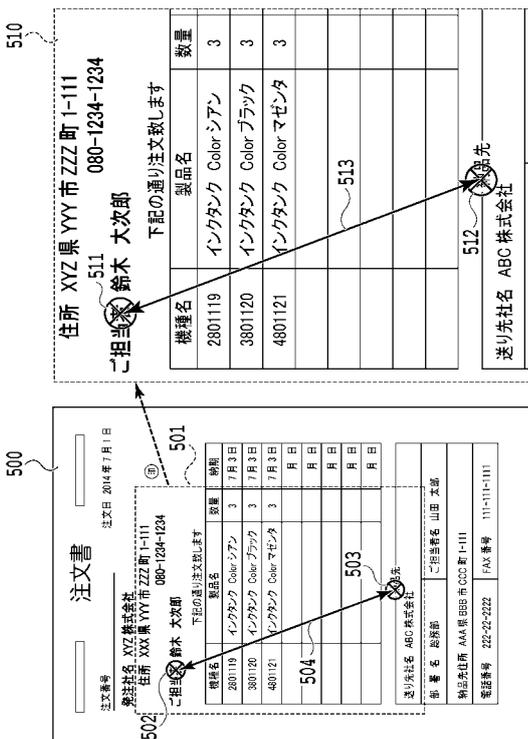
【図3】



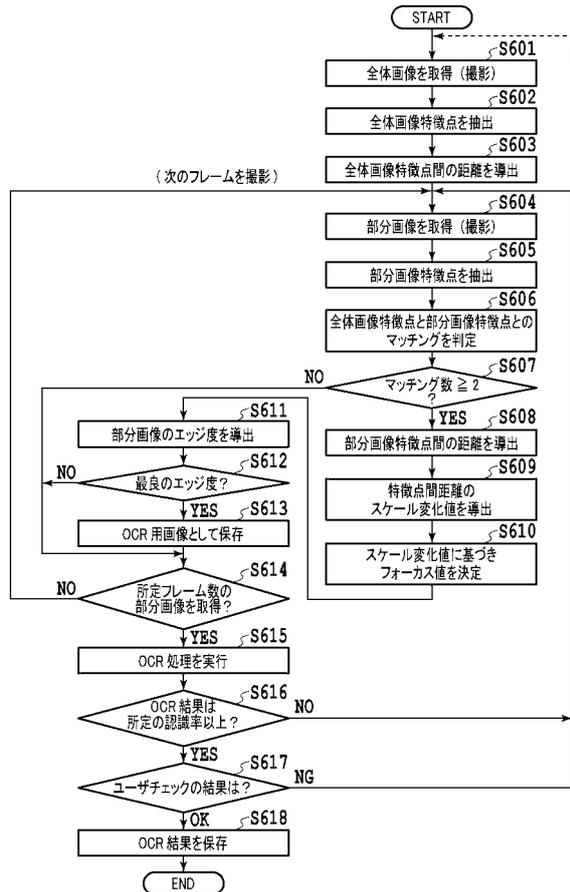
【図4】



【図5】



【図6】



【 図 7 】

ピクセル距離	フォーカス値
1000	0.00
900	0.04
800	0.09
700	0.16
600	0.25
500	0.36
400	0.49
300	0.64
200	0.81

【 図 8 】

800

注文書

注文番号 _____ 注文日 2014年7月1日

発注社名 XYZ株式会社
住所 XXX県YYY市ZZZ町1-111
080-1234-1234

ご担当者 鈴木 大次郎

下記の通り注文致します

機種名	製品名	数量	納期
2801119	インクタンク Color シアン	3	7月3日
3801120	インクタンク Color ブラック	3	7月3日
4801121	インクタンク Color マゼンタ	3	7月3日
			月 日
			月 日
			月 日

820

先社名 ABC株式会社
部署名 総務部
納品先住所 AAA県BBB市CCC町1-111
電話番号 222-22-2222 FAX番号 111-111-1111

【 図 9 】

