

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6970553号
(P6970553)

(45) 発行日 令和3年11月24日 (2021. 11. 24)

(24) 登録日 令和3年11月2日 (2021. 11. 2)

(51) Int. Cl.		F I	
G06T 7/70	(2017.01)	G06T 7/70	Z
G06T 7/00	(2017.01)	G06T 7/00	350C
G06N 3/04	(2006.01)	G06N 3/04	
G06N 3/08	(2006.01)	G06N 3/08	

請求項の数 25 (全 21 頁)

(21) 出願番号	特願2017-157618 (P2017-157618)	(73) 特許権者	000001007
(22) 出願日	平成29年8月17日 (2017. 8. 17)		キヤノン株式会社
(65) 公開番号	特開2019-36167 (P2019-36167A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成31年3月7日 (2019. 3. 7)	(74) 代理人	110003281
審査請求日	令和2年7月22日 (2020. 7. 22)		特許業務法人大塚国際特許事務所
		(72) 発明者	稲葉 正樹
			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	山田 辰美

最終頁に続く

(54) 【発明の名称】 画像処理装置、画像処理方法

(57) 【特許請求の範囲】

【請求項 1】

入力画像における対象物体の位置として認識器が認識した複数の認識位置に対してクラスタリングを行うことで1以上の代表位置を取得する取得手段と、

前記代表位置に対するユーザからの編集指示に応じて該代表位置を編集する編集手段と、

前記入力画像と前記代表位置とを前記認識器の学習に用いるための学習データとして保存する保存手段と

を備えることを特徴とする画像処理装置。

【請求項 2】

前記取得手段は、前記入力画像の各画素位置のうち前記対象物体の位置として前記認識器が認識した複数の画素位置を前記複数の認識位置として取得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 3】

前記取得手段は、前記入力画像の画素位置を含む画像領域内の画像を入力した前記認識器からの出力が、該画像領域内に前記対象物体が写っていることを示す場合に、該画素位置を前記認識位置として取得することを特徴とする請求項 1 又は 2 に記載の画像処理装置。

【請求項 4】

前記取得手段は、前記複数の認識位置に対してクラスタリングを行うことで該複数の認

識位置を予め定められたクラス数のクラスに分類し、該クラスに属する認識位置に基づいて該クラスに対応する代表位置を求めることを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の画像処理装置。

【請求項 5】

更に、ユーザによる指示に応じて前記クラス数を設定する設定手段を備え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】

前記取得手段は、前記クラスに属する認識位置の平均位置を該クラスに対応する代表位置として求めることを特徴とする請求項 4 又は 5 に記載の画像処理装置。

【請求項 7】

前記取得手段は、前記クラスに属する認識位置の中央値を該クラスに対応する代表位置として求めることを特徴とする請求項 4 又は 5 に記載の画像処理装置。

【請求項 8】

前記取得手段は、前記クラスに属する認識位置をガウス分布に当てはめた際のピーク位置を該クラスに対応する代表位置として求めることを特徴とする請求項 4 又は 5 に記載の画像処理装置。

【請求項 9】

更に、

前記クラスに対する信頼度を求める算出手段を備え、

前記編集手段は、前記信頼度が閾値未満となるクラスの数規定数以上、若しくは全クラス数に対する前記信頼度が閾値未満となるクラスの数割合規定数以上となった場合に動作することを特徴とする請求項 4 乃至 8 の何れか 1 項に記載の画像処理装置。

【請求項 10】

前記算出手段は、前記クラスに属する認識位置の数に応じて該クラスに対する信頼度を求めることを特徴とする請求項 9 に記載の画像処理装置。

【請求項 11】

前記算出手段は、前記クラスに属する認識位置に対応する前記認識器の出力に応じて該クラスに対する信頼度を求めることを特徴とする請求項 9 に記載の画像処理装置。

【請求項 12】

前記算出手段は、前記クラスに属する認識位置と該クラスに対応する代表位置との間の距離に応じて該クラスに対する信頼度を求めることを特徴とする請求項 9 に記載の画像処理装置。

【請求項 13】

前記編集手段は、前記入力画像に前記代表位置を示す情報を重ねて表示し、該代表位置に対するユーザからの編集操作を受け付けることを特徴とする請求項 1 乃至 12 の何れか 1 項に記載の画像処理装置。

【請求項 14】

前記編集手段は、前記入力画像に前記代表位置を示す情報と前記信頼度を示す情報とを重ねて表示し、該代表位置に対するユーザからの編集操作を受け付けることを特徴とする請求項 9 乃至 12 の何れか 1 項に記載の画像処理装置。

【請求項 15】

前記編集手段は、前記入力画像に前記信頼度が閾値以上となるクラスの代表位置を示す情報を重ねて表示し、該代表位置に対するユーザからの編集操作を受け付けることを特徴とする請求項 9 乃至 12 の何れか 1 項に記載の画像処理装置。

【請求項 16】

前記編集手段は、前記代表位置に対するユーザからの移動の指示に応じて該代表位置を移動させることを特徴とする請求項 1 乃至 15 の何れか 1 項に記載の画像処理装置。

【請求項 17】

前記編集手段は、前記代表位置に対するユーザからの削除の指示に応じて該代表位置を削除することを特徴とする請求項 1 乃至 16 の何れか 1 項に記載の画像処理装置。

10

20

30

40

50

【請求項 18】

前記編集手段は、ユーザからの新たな代表位置の追加の指示に応じて代表位置を追加することを特徴とする請求項 1 乃至 17 の何れか 1 項に記載の画像処理装置。

【請求項 19】

更に、

前記保存手段が保存した学習データを用いて前記認識器の学習を行う学習手段を備えることを特徴とする請求項 1 乃至 18 の何れか 1 項に記載の画像処理装置。

【請求項 20】

更に、

前記認識器を、前記学習を行った認識器に更新する更新手段を備えることを特徴とする請求項 19 に記載の画像処理装置。

10

【請求項 21】

前記更新手段は、学習の前の認識器に対する評価値が学習後の認識器に対する評価値よりも高い場合に、該学習の前の認識器を該学習後の認識器に更新することを特徴とする請求項 20 に記載の画像処理装置。

【請求項 22】

入力画像における対象物体の位置姿勢として認識器が認識した複数の認識位置姿勢に対してクラスタリングを行うことで 1 以上の代表位置姿勢を取得する取得手段と、

前記代表位置姿勢に対するユーザからの編集指示に応じて該代表位置姿勢を編集する編集手段と、

20

前記入力画像と前記代表位置姿勢とを前記認識器の学習に用いるための学習データとして保存する保存手段と

を備えることを特徴とする画像処理装置。

【請求項 23】

入力画像における対象物体の位置として認識器が認識した複数の認識位置に対してクラスタリングを行うことで 1 以上の代表位置を取得する取得工程と、

前記代表位置に対するユーザからの編集指示に応じて該代表位置を編集する編集工程と、

前記入力画像と前記代表位置とを前記認識器の学習に用いるための学習データとして保存する保存工程と

30

を備えることを特徴とする画像処理方法。

【請求項 24】

入力画像における対象物体の位置姿勢として認識器が認識した複数の認識位置姿勢に対してクラスタリングを行うことで 1 以上の代表位置姿勢を取得する取得工程と、

前記代表位置姿勢に対するユーザからの編集指示に応じて該代表位置姿勢を編集する編集工程と、

前記入力画像と前記代表位置姿勢とを前記認識器の学習に用いるための学習データとして保存する保存工程と

を備えることを特徴とする画像処理方法。

【請求項 25】

40

コンピュータを、請求項 1 乃至 22 の何れか 1 項に記載の画像処理装置の各手段として機能させるためのコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像認識に係る技術に関するものである。

【背景技術】

【0002】

従来、対象物体を撮影したカラー画像やデプス画像から特徴やパターンを学習し、対象物体を認識する手法がある。このような手法では、特徴やパターンを学習するために、カ

50

ラー画像やデプス画像に対して、人間がどの対象物体が写っているかを指定したり、対象物体が写っている位置や位置姿勢を指定する（以下では「ラベル付け」と称する）必要がある。学習にはこのようなデータを大量に用意しなければならないが、人間が全てのラベル付けを正確に行うのは難しいという問題がある。

【 0 0 0 3 】

このような問題を解決するために、特許文献 1 のように、「現状の認識器による認識結果の修正」、「修正したデータを用いた認識器の学習・更新」を繰り返すことによって、認識器の精度を高めていく方法がある。

【先行技術文献】

【特許文献】

10

【 0 0 0 4 】

【特許文献 1】特許第 5 9 5 3 1 5 1 号

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 5 】

しかしながら、対象物体が写っている位置や位置姿勢を認識する認識器の場合、画像内の大量の画素が認識結果として出力され得るため、人間が各認識結果を修正するコストが高いという問題がある。

【 0 0 0 6 】

本発明はこのような問題に鑑みてなされたものであり、画像中の対象物体の位置の指定に係る負荷を軽減させるための技術を提供する。

20

【課題を解決するための手段】

【 0 0 0 7 】

本発明の一様態は、入力画像における対象物体の位置として認識器が認識した複数の認識位置に対してクラスタリングを行うことで 1 以上の代表位置を取得する取得手段と、

前記代表位置に対するユーザからの編集指示に応じて該代表位置を編集する編集手段と

、

前記入力画像と前記代表位置とを前記認識器の学習に用いるための学習データとして保存する保存手段と

を備えることを特徴とする。

30

【発明の効果】

【 0 0 0 8 】

本発明の構成によれば、画像中の対象物体の位置の指定に係る負荷を軽減させることができる。

【図面の簡単な説明】

【 0 0 0 9 】

【図 1】物体認識装置の機能構成例を示すブロック図。

【図 2】物体認識処理のフローチャート。

【図 3】認識部 1 0 2 及び分類部 1 0 3 の処理を説明する図。

【図 4】画面の表示例を示す図。

40

【図 5】ユーザによるラベルの編集について説明する図。

【図 6】物体認識装置の機能構成例を示すブロック図。

【図 7】ステップ S 2 及び S 3 における処理の詳細を示すフローチャート。

【図 8】物体認識装置の機能構成例を示すブロック図。

【図 9】ステップ S 2 及び S 3 における処理の詳細を示すフローチャート。

【図 1 0】位置だけでなく姿勢も認識する場合の処理を説明する図。

【図 1 1】コンピュータ装置のハードウェア構成例を示すブロック図。

【発明を実施するための形態】

【 0 0 1 0 】

以下、添付図面を参照し、本発明の実施形態について説明する。なお、以下説明する実

50

施形態は、本発明を具体的に実施した場合の一例を示すもので、特許請求の範囲に記載した構成の具体的な実施例の１つである。

【 0 0 1 1 】

〔 第 1 の実施形態 〕

本実施形態に係る画像処理装置は、入力画像に対して認識器を用いて認識した対象物体の認識位置をクラスタリングし、該クラスタリングにより得られる該認識位置の代表位置をラベルとしてユーザの修正対象にするラベル付け機能を有するものである。このような機能により、ユーザは全ての認識位置を確認、修正する必要がなくなり、代表位置に基づくラベルを確認、修正すればよいため、ラベル付けの手間が低減される。ここで、ラベルとは、画像中の対象物体の位置または位置姿勢に関する情報を意味しており、ラベルは認識器を構築するための教師データに利用される。

10

【 0 0 1 2 】

先ず、本実施形態に係る物体認識装置の機能構成例について、図 1 のブロック図を用いて説明する。図 1 に示す如く、本実施形態に係る物体認識装置 1 0 0 0 は、撮像装置 1、表示装置 2、操作装置 3、画像処理装置 1 1 0 0、を有する。

【 0 0 1 3 】

先ず、撮像装置 1 について説明する。撮像装置 1 は、静止画像若しくは動画像を撮像する装置である。撮像画像は特定の画像に限らず、カラー画像（例えば R G B 画像）であっても良いし、モノクロ画像であっても良いし、赤外線画像であっても良いし、深度情報であるデプス画像であっても良い。撮像装置 1 によって撮像された撮像画像は画像処理装置 1 1 0 0 に対して送出される。

20

【 0 0 1 4 】

次に、表示装置 2 について説明する。表示装置 2 は、C R T や液晶画面等により構成されており、画像処理装置 1 1 0 0 による処理結果を画像や文字などでもって表示することができる。なお、表示装置 2 はタッチパネル画面であってもよく、この場合、タッチパネル画面に対するユーザからの各種の操作の内容は操作情報として画像処理装置 1 1 0 0 に対して送出される。

【 0 0 1 5 】

次に、操作装置 3 について説明する。操作装置 3 は、マウスやキーボードなどのユーザインターフェースにより構成されており、ユーザが操作することで各種の指示を画像処理装置 1 1 0 0 に対して入力することができる。以下では、操作装置 3 はマウスであるものとして説明するが、以下に説明する同様の指示を画像処理装置 1 1 0 0 に入力可能なユーザインターフェースであれば、如何なるユーザインターフェースを操作装置 3 に適用しても良い。

30

【 0 0 1 6 】

次に、画像処理装置 1 1 0 0 について説明する。画像取得部 1 0 1 は、撮像装置 1 から送出された撮像画像を取得する。なお、撮像画像の取得元は撮像装置 1 に限らず、L A N やインターネットなどのネットワークを介してデータベース、サーバ装置、タブレット端末機器、スマートフォンなどの機器から撮像画像を取得するようにしても良い。然るに撮像画像は撮像装置 1 によって撮像された撮像画像であっても良いし、公開されている画像データベースに登録されている画像であっても良いし、W e b にアップロードされている画像であっても良い。

40

【 0 0 1 7 】

認識部 1 0 2 は先ず、認識器記憶部 1 0 8 に保存されている C N N (C o n v o l u t i o n a l N e u r a l N e t w o r k) (認識器) のモデル構造と C N N のパラメータとを用いて C N N を構成する。そして認識部 1 0 2 は、画像取得部 1 0 1 が取得した撮像画像に対して該構成した C N N を適用することで、該撮像画像における対象物体の位置を認識位置として認識する。

【 0 0 1 8 】

分類部 1 0 3 は、認識部 1 0 2 が認識した認識位置をクラスタリングすることで該認識

50

位置を１以上のクラスに分類し、該クラスに属する認識位置に基づいて該クラスに対応する代表位置（ラベル）を求める（計算する）。

【００１９】

操作受付部１０４は、ユーザが操作装置３を操作したことで入力した操作内容に対応する命令を入力部１０５に送出する。入力部１０５は、操作受付部１０４から受けた命令に基づいて、上記の代表位置の編集（ラベル付け）を行う。そして入力部１０５は、撮像画像と代表位置とを上記のＣＮＮの学習に用いる学習データとして記憶部１０６に格納する。

【００２０】

学習部１０７は、記憶部１０６に格納されている学習データを用いて教師データを作成する。そして学習部１０７は、認識器記憶部１０８に保存されているＣＮＮのモデル構造とＣＮＮのパラメータとを用いてＣＮＮを構成し、上記作成した教師データを用いて、該構成したＣＮＮのパラメータの学習処理を行う。そして学習部１０７は、認識器記憶部１０８に格納されているＣＮＮのパラメータを、該学習処理によって得られたパラメータに更新する。

10

【００２１】

認識器記憶部１０８は、ＣＮＮのモデル構造やパラメータを保持している。本実施形態では、ＣＮＮのモデル構造は、撮像画像を入力とし、コンボリューション層や全結合層、Softmax層を経て、該撮像画像における各画像領域内に対象物体が写っているか否かの２クラス判別を出力するモデルとして設計されているとする。ここで、ＣＮＮのパラメータとは、コンボリューション層や全結合層内の各々の重みのことを指している。

20

【００２２】

制御部１９９は、画像取得部１０１、認識部１０２、分類部１０３、学習部１０７、操作受付部１０４、入力部１０５、の各機能部の動作制御を行うと共に、画像処理装置１１００全体の動作制御を行う。

【００２３】

なお、図１では、撮像装置１、表示装置２、操作装置３、画像処理装置１１００はそれぞれ別個の装置として示しているが、このうち２つ以上の機器を１つの機器に一体化させても良い。

【００２４】

30

次に、画像処理装置１１００による物体認識処理について、図２（ａ）のフローチャートに従って説明する。ステップＳ１では、制御部１９９は、ＣＮＮ（認識器）の学習が必要であるか否かを判断する。ＣＮＮの学習が必要であるか否かを判断する条件は特定の条件に限らない。例えば制御部１９９は、ユーザが操作装置３を操作してＣＮＮの学習指示を入力したことを検知した場合には、ＣＮＮの学習が必要であると判断する。また、制御部１９９は、例えば、規定のタイミングごと若しくは規定の条件が満たされるたびに、ＣＮＮの学習が必要であると判断する。

【００２５】

このような判断の結果、ＣＮＮの学習が必要であると判断した場合には、処理はステップＳ２に進む。一方、ＣＮＮの学習が必要であると判断していない場合には、処理はステップＳ４に進む。

40

【００２６】

ステップＳ２では、上記のようにして学習データを生成して記憶部１０６に格納する。そしてステップＳ３では、学習データを用いて教師データを作成し、該教師データを用いてＣＮＮのパラメータの学習を行い、認識器記憶部１０８に格納されているＣＮＮのパラメータを該学習後のパラメータに更新する。

【００２７】

ステップＳ２及びＳ３における処理の詳細について、図２（ｂ）のフローチャートに従って説明する。図２（ｂ）においてステップＳ１１～Ｓ１５が上記のステップＳ２において実行され、ステップＳ１６、Ｓ１７が上記のステップＳ３において実行される。

50

【 0 0 2 8 】

ステップ S 1 1 では、画像取得部 1 0 1 は、撮像装置 1 から送出された撮像画像を取得する。ステップ S 1 2 では、認識部 1 0 2 は先ず、認識器記憶部 1 0 8 に保存されている C N N (認識器) のモデル構造と C N N のパラメータを用いて C N N を構成する。そして認識部 1 0 2 は、ステップ S 1 1 で画像取得部 1 0 1 が取得した撮像画像に対して該構成した C N N を適用することで、該撮像画像における対象物体の位置を認識位置として認識する。

【 0 0 2 9 】

例えば、画像取得部 1 0 1 が図 3 (a) に示す画像 3 1 を撮像画像として取得した場合、認識部 1 0 2 は、図 3 (b) に示す如く、画像 3 1 における着目画素位置 3 2 を中心とする画像領域 3 3 を設定する。画像領域 3 3 は着目画素位置 3 2 を含んでいれば良く、必ずしも着目画素位置 3 2 を中心とする画像領域であることに限らない。そして認識部 1 0 2 は、画像領域 3 3 内の画像を上記の構築した C N N に入力して該 C N N の演算処理を行うことで、該 C N N の出力を求める。そして認識部 1 0 2 は、C N N の出力が「画像領域 3 3 内に対象物体が写っている」ことを示す場合には、着目画素位置 3 2 を画像 3 1 における対象物体の位置 (認識位置) と認識する。一方、認識部 1 0 2 は、C N N の出力が「画像領域 3 3 内に対象物体が写っている」ことを示していない場合には、着目画素位置 3 2 を画像 3 1 における対象物体の位置 (認識位置) とは認識しない。認識部 1 0 2 は、着目画素位置 3 2 を点線 3 4 で示す如くラスタスキャン順で移動させながら各画素位置について同様の処理を行うことで、画像 3 1 上の各画素位置に対し、該画素位置を中心とする画像領域内に対象物体が写っているか否かを判断することができる。これにより、画像 3 1 上における対象物体の位置を認識することができる。なお、画像 3 1 上の各画素位置に対して、該画素位置を中心とする画像領域内に対象物体が写っているか否かを判断することができるのであれば、着目画素位置 3 2 の移動順はラスタスキャン順に限らない。

【 0 0 3 0 】

画像 3 1 中に写っているオブジェクトのうち円柱のオブジェクトを対象物体とした場合に、画像 3 1 上における対象物体の認識位置に「 x 」のマーカ (以下、認識位置マーカと称する) を付した結果を図 3 (c) に示す。近傍画素では似たような画像が切り抜かれることになるため、例えば認識位置マーカ群 3 5 で示す如く、認識位置が密集することが多い。また、対象物体と似たような特徴を持つ位置は誤認識が起こりやすい。たとえば、立方体はエッジ (画像中の輝度勾配) に囲まれている点で円柱と共通しているため、例えば認識位置マーカ群 3 6 で示す如く誤認識が起こることがある。

【 0 0 3 1 】

ステップ S 1 3 で分類部 1 0 3 は、ステップ S 1 2 で認識部 1 0 2 が認識した対象物体の位置 (認識位置) の集合を初期集合とし、規定のクラスタリング条件に基づいて該初期集合をクラスタリングすることで、該集合に含まれる認識位置を 1 以上のクラスに分類する。そして分類部 1 0 3 は、分類した 1 以上のクラスのそれぞれについて、該クラスに属する認識位置に基づいて該クラスに対応する代表位置を求める。例えば、初期集合を特徴として K m e a n s アルゴリズムを用いて該初期集合をクラスタリングすると、近い認識位置が一つのクラスとしてまとめられる。そしてクラスごとに、該クラスに属する認識位置の平均位置を該クラスの代表位置とする。たとえば、図 3 (c) に示した認識位置マーカ群の位置をクラスタリングすると 5 つのクラスに分類される。この分類結果を図 3 (d) に示す。図 3 (d) では、分類された 5 つのクラスのそれぞれの位置 (該クラスの代表位置) に「 」のマーカ (以下、代表位置マーカと称する) を付している。代表位置マーカ 3 7 は、認識位置マーカ群 3 5 に対応するクラスの代表位置に位置しており、代表位置マーカ 3 8 は、認識位置マーカ群 3 6 に対応するクラスの代表位置に位置している。なお、クラスタリングに必要なパラメータ (クラス数、クラスとしてまとめてよい範囲 (クラス範囲) など) は、ユーザによって予め指定されており、分類部 1 0 3 が保持しているものとする。

【 0 0 3 2 】

そして分類部 103 は、ステップ S1 で画像取得部 101 が取得した撮像画像上（入力画像上）の各クラスの代表位置に代表位置マーカを合成してから、該撮像画像を含む画面を表示装置 2 の表示画面に表示させる。この画面の表示例を図 4 に示す。画面 41 において表示領域 42 には、ステップ S1 で画像取得部 101 が取得した撮像画像が表示されており、この撮像画像上の各クラスの代表位置には代表位置マーカ 401 ~ 405 が配されている。

【0033】

スライダ 43 は、クラスタリングに必要なパラメータを設定するためのものである。例えば、ユーザが操作装置 3 を操作してカーソル 44 をスライダ 43 の位置に移動させてからスライダ 43 の位置を変更する操作を行うことでスライダ 43 をより右端に近づけると、より大きいクラス数やより広いクラス範囲を指定することができる。また、ユーザが操作装置 3 を操作してカーソル 44 をスライダ 43 の位置に移動させてからスライダ 43 の位置を変更する操作を行うことでスライダ 43 をより左端に近づけると、より小さいクラス数やより狭いクラス範囲を指定することができる。現在指定されているクラス数やクラス範囲は領域 43a 内に数値として表示される。なお、領域 43a 内の数値をユーザが操作装置 3 を操作して直接指定することでクラス数やクラス範囲を設定しても良い。このように、クラス数やクラス範囲の指定方法は特定の指定方法に限らない。クラスタリングに必要なパラメータが新たに指定される度に分類部 103 は、該新たに指定されたパラメータを用いて上記初期集合をクラスタリングし、該クラスタリングにより得られる各クラスの代表位置に代表位置マーカを配置した撮像画像を表示領域 42 内に表示する。分類部 103 は、最新の代表位置のリスト（代表位置リスト）を管理している。なお、図 4 の画面は、本実施形態に適用可能な画面の一例であり、このような画面に限らない。

【0034】

次に、ステップ S14a では、操作受付部 104 は、代表位置に係る編集操作入力（編集指示）を受け付ける。ユーザが操作装置 3 を操作して代表位置に係る編集の操作入力を行うと、操作受付部 104 は該編集の為の命令を作成して入力部 105 に入力する。

【0035】

ステップ S14b では、入力部 105 は、操作受付部 104 から命令を受け取ると、該命令に応じて、分類部 103 により分類された代表位置（ラベル）、つまり、上記の代表位置リストに登録されている代表位置に係る編集処理を行う。例えばユーザは操作装置 3 を操作して以下のような操作を行うことで、ラベルの削除、追加、修正、といったラベル付けを実現する。

【0036】

ここで、ユーザによるラベルの編集について、図 5 を用いて説明する。なお、以下に説明するユーザ操作に応じたラベルの編集処理は何れも、入力部 105 が行うものとする。

【0037】

ユーザは操作装置 3 を操作してカーソル 44 を代表位置マーカ 401 ~ 405 のうち所望の代表位置マーカの位置に移動させて、そこで右クリック操作を行うことで、該所望の代表位置マーカを削除することができる。代表位置マーカを削除すると、上記の代表位置リストから該代表位置マーカに対応する代表位置が削除される。例えば、図 5(a) に示す如く、表示領域 42 内に表示されている撮像画像上に代表位置マーカ 401 ~ 405 が配置されているとする。代表位置マーカ 401 ~ 405 のうち代表位置マーカ 401, 403 は円柱のオブジェクト、すなわち対象物体の位置を表している。しかし、代表位置マーカ 402, 404 は円柱のオブジェクトではなく立方体のオブジェクトの位置を表しており、対象物体の位置を表していない。また、代表位置マーカ 405 はそもそも、オブジェクトの位置を表していない。このような場合、ユーザは操作装置 3 を操作してカーソル 44 を代表位置マーカ 402, 404, 405 のそれぞれの位置に移動させて、そこで右クリック操作を行うことで、該代表位置マーカを削除する。その結果、代表位置マーカのレイアウトは図 5(a) から図 5(b) のように変化する。図 5(b) では、図 5(a) の代表位置マーカ 401 ~ 405 のうち代表位置マーカ 402, 404, 405 が削除さ

れている。然るにこの場合、上記の代表位置リストから代表位置マーカ 4 0 2 , 4 0 4 , 4 0 5 のそれぞれに対応する代表位置が削除され、その結果、代表位置リストには、代表位置マーカ 4 0 1 , 4 0 3 に対応する代表位置のみが登録されている状態となっている。

【 0 0 3 8 】

図 5 (b) では、3 つの対象物体のうち 2 つの対象物体のそれぞれには代表位置マーカ 4 0 1 , 4 0 3 が配されているが、残りの 1 つの対象物体 4 0 6 には代表位置マーカは配されていない。そこでユーザは操作装置 3 を操作してカーソル 4 4 を対象物体 4 0 6 の位置に移動させてから左クリック操作を行うことで、図 5 (c) に示す如く、該位置に新たな代表位置マーカ 4 1 0 を配置する。新たな代表位置マーカ 4 1 0 を配置した場合、上記の代表位置リストには、この代表位置マーカ 4 1 0 の配置位置が代表位置として登録される。

10

【 0 0 3 9 】

また、ユーザは操作装置 3 を操作してカーソル 4 4 を所望の代表位置マーカの位置に移動させて、そこで左クリック操作を行いながらドラッグ&ドロップ操作を行うことで、該所望の代表位置マーカの位置を移動させることができる。代表位置マーカの位置を移動させると、上記の代表位置リストにおいて該代表位置マーカの位置が移動後の位置に変更される。図 5 (c) においてユーザが操作装置 3 を操作して代表位置マーカ 4 0 3 の位置を移動させた結果を図 5 (d) に示す。

【 0 0 4 0 】

ステップ S 1 4 c では、制御部 1 9 9 は、ユーザが操作装置 3 を操作して終了ボタン 4 5 を指定したか否かを判断する。この判断の結果、ユーザが操作装置 3 を操作して終了ボタン 4 5 を指定した場合には、処理はステップ S 1 5 に進む。一方、ユーザが操作装置 3 を操作して終了ボタン 4 5 を指定していない場合には、処理はステップ S 1 4 a に進む。

20

【 0 0 4 1 】

ステップ S 1 5 で入力部 1 0 5 は、代表位置リストと、ステップ S 1 1 で画像取得部 1 0 1 が取得した撮像画像と、を C N N のパラメータの学習に用いる学習データとして記憶部 1 0 6 に格納する。また、分類部 1 0 3 は、図 4 の画面においてクラスタリングに必要なパラメータが変更された場合には、この変更されたパラメータを保持する。

【 0 0 4 2 】

ステップ S 1 6 では、学習部 1 0 7 は、記憶部 1 0 6 に格納されている学習データを用いて教師データを作成する。例えば学習部 1 0 7 は、認識部 1 0 2 と同様にして、学習データに含まれている撮像画像における各画素位置に対して画像領域を設定する。そして学習部 1 0 7 は、撮像画像における各画素位置に対し、該画素位置が該学習データに含まれている代表位置リストに登録されている代表位置の何れかに該当しているのか否か (正負 (T r u e 、 F a l s e)) 、を示す符号データを生成する。ここで、「 A が B に該当している」とは、 A と B とが全く同じであることに限らず、 A と B との差が規定値以内である、ことも含むものとする。学習部 1 0 7 は、学習データに含まれている撮像画像における着目画素位置が、該学習データに含まれている代表位置リストに登録されている代表位置の何れかに該当している場合には、該着目画素位置に対して符号データ「 T r u e 」を割り当てる。一方、学習部 1 0 7 は、学習データに含まれている撮像画像における着目画素位置が、該学習データに含まれている代表位置リストに登録されている代表位置の何れにも該当していない場合には、該着目画素位置に対して符号データ「 F a l s e 」を割り当てる。そして学習部 1 0 7 は、撮像画像における画素位置ごとに、該画素位置に対して設定した画像領域内の画像 (学習画像) と、該画素位置に割り当てた符号データと、を関連づけて保持する教師データを生成する。

30

40

【 0 0 4 3 】

そして学習部 1 0 7 は、認識器記憶部 1 0 8 に保存されている C N N のモデル構造と C N N のパラメータとを用いて C N N を構成する。学習部 1 0 7 は、該構成した C N N にそれぞれの学習画像を入力することで得られる該 C N N の出力が、該学習画像に対応する符号データとなるように、ディープラーニングで一般的に用いられている誤差逆伝播法で C

50

ＮＮのパラメータを最適化する。このときのＣＮＮのパラメータの初期値として、認識器記憶部１０８に保存されているＣＮＮのパラメータを用いる。

【００４４】

ステップＳ１７では、学習部１０７は、認識器記憶部１０８に格納されているＣＮＮのパラメータを、ステップＳ１６における学習（最適化）で得られたＣＮＮのパラメータに更新する。

【００４５】

図２（ａ）に戻って次に、ステップＳ４では、画像取得部１０１は、撮像装置１から送出された撮像画像を取得する。ステップＳ５では、認識部１０２は先ず、認識器記憶部１０８に保存されているＣＮＮ（認識器）のモデル構造とＣＮＮのパラメータとを用いてＣ
１０
ＮＮを構成する。そして認識部１０２は、ステップＳ４で画像取得部１０１が取得した撮像画像に対して該構成したＣＮＮを適用することで、該撮像画像におけるオブジェクトの位置を認識位置として認識する。ステップＳ５において認識されたオブジェクトの認識位置の出力先は特定の出力先に限らない。例えば、図４の画面の表示領域４２にステップＳ４で取得した撮像画像を表示すると共に、ステップＳ５で認識した認識位置若しくはその近傍位置にマーカ等の情報を配置しても良い。

【００４６】

このように、本実施形態によれば、画像に対してラベル付けする際に、ユーザは現在の認識器で認識される代表位置を参考にラベル付けができるため、ユーザのラベル付けの手間が低減される。
２０

【００４７】

< 第１の実施形態の変形例 >

第１の実施形態では、ステップＳ１３において、クラスタリングにＫｍｅａｎｓアルゴリズムを用いた。しかし、クラスタリングのアルゴリズムは特定のアルゴリズムに限らない。例えば、Ｍｅａｎ ＳｈｉｆｔやＮｅａｒｅｓｔ Ｎｅｉｇｈｂｏｒなどといった他のクラスタリングアルゴリズムを用いてもよいし、ユーザがひとつのクラスとしてまとめた領域を、マウスで画像中をドラッグして指定してもよい。このようにすることで、まとめて処理する認識結果の領域を自由に調整することができる。

【００４８】

また第１の実施形態では、ステップＳ１３において、クラスに対応する代表位置として該クラスに属する認識位置の平均位置を用いたが、他の方法を用いて代表位置を算出しても良い。例えば、クラスに属する認識位置の中央値を該クラスの代表位置としても良いし、クラスに属する認識位置をガウス分布に当てはめた際のピーク位置を該クラスの代表位置としても良い。このようにすることで、クラスに属する認識位置にはずれ値などのノイズがあった場合でも、安定した代表位置を算出することができる。
３０

【００４９】

また第１の実施形態では、ステップＳ１３において、撮像画像上に代表位置を示すマーカを配置したが、表示する情報はこれらの情報に限らず、例えば、代表位置を示すマーカだけでなく、認識位置を示すマーカも撮像画像上に合成して表示しても良いし、一部の対象物体の位置を撮像画像上に合成して表示しても良い。このようにすることで、ユーザは
４０
現状の認識器でどのような認識結果が得られているかが分かる。

【００５０】

また第１の実施形態では、ステップＳ１４ａにおいて、ラベル付けはユーザのマウス操作によって行った。しかし、他の方法でもってラベル付けを行っても良い。例えば、キーボードを用いた操作入力によってラベル付けを実施しても良いし、タッチパネルを用いた操作入力によってラベル付けを行っても良い。

【００５１】

また第１の実施形態では、ステップＳ１４ｂでは、分類部１０３によって分類された代表位置を初期値とし、該初期値に対してラベル付けを行っていたが、これに限る必要はない。例えば、初期値は設定せず、代表位置を示す情報を合成した撮像画像を図４の画面と
５０

は別画面として表示することで、ユーザは代表位置を参考にラベル付けをしてもよい。また、図4の画面上に代表位置を半透明のマーカとして描画することで、ユーザは代表位置を参考にラベル付けをしてもよい。

【0052】

[第2の実施形態]

本実施形態を含む以降の各実施形態や各変形例では、第1の実施形態との差分について説明し、以下で特に触れない限りは第1の実施形態と同様であるものとする。本実施形態に係る画像処理装置は、第1の実施形態に係る画像処理装置の機能に加え、各クラスに対する信頼度を算出し、該信頼度に応じてユーザによるラベル付けを行うのか否かを切り替える機能を有する。ユーザは信頼度が低い場合のみ、ラベルを確認、修正すればよいため、ラベル付けの手間が低減される。

10

【0053】

先ず本実施形態に係る物体認識装置の機能構成例について、図6のブロック図を用いて説明する。図6において図1に示した機能部と同じ機能部には同じ参照番号を付しており、該機能部に係る説明は省略する。本実施形態に係る物体認識装置2000は、画像処理装置2100の構成が図1の画像処理装置1100と異なる。

【0054】

分類部203は、分類部103の機能に加え、上記の代表位置リストと、認識部102が認識した認識位置の集合と、該認識位置がどのクラスに分類されたのかを示す情報と、を含む分類情報を生成して信頼度算出部209に送出する機能を有する。

20

【0055】

信頼度算出部209は、クラスごとに、該クラスに属する認識位置の数に応じた信頼度を算出する。そして信頼度算出部209は、信頼度が閾値未満となるクラスの数が規定数（例えば1）以上である場合には、ユーザによるラベル付けが必要と判断し、その旨を入力部105に通知する。これにより入力部105は第1の実施形態と同様に、ユーザによるラベル付けの為の処理を行う。一方、信頼度算出部209は、信頼度が閾値未満となるクラスの数が規定数未満である場合、ユーザによるラベル付けは不必要と判断し、代表位置リストと、ステップS11で画像取得部101が取得した撮像画像と、を学習データとして記憶部106に格納する。

【0056】

30

本実施形態に係るステップS2及びS3における処理の詳細について、図7のフローチャートに従って説明する。なお、図7において図2(b)に示した処理ステップと同じ処理ステップには同じステップ番号を付しており、該処理ステップに係る説明は省略する。図7においてステップS11～S15が上記のステップS2において実行され、ステップS16、S17が上記のステップS3において実行される。

【0057】

ステップS23で分類部203は、上記のステップS13の処理を行うと共に、代表位置リストと、認識部102が認識した認識位置の集合と、該認識位置がどのクラスに分類されたのかを示す情報と、を含む分類情報を生成して信頼度算出部209に送出する。

【0058】

40

ステップS24aでは、信頼度算出部209は、クラスごとに、該クラスに属する認識位置の数に応じた信頼度を算出する。着目クラスの信頼度は、該着目クラスに属する認識位置の数が多いほど高く、少ないほど低くなるように算出する。例えば、着目クラスに属する認識位置の数を該着目クラスの信頼度として使用しても良い。例えば、図3(d)の代表位置マーカ37に対応するクラスの信頼度は、認識位置マーカ群35の認識位置の数である。

【0059】

対象物体が確かに存在する位置の近傍画素では、対象物体近辺の画像が入力されるため、対象物体があるという判定がなされる確率が高い。一方、対象物体ではない位置は、たまたま対象物体と似たような特徴が抽出されてしまった場合に誤認識が起こる。この場合

50

、前者よりも対象物体であると判定される確率は比較的に低い。以上の理由により、対象物体として認識された位置の数が多いほど信頼度が高いと考えられる。

【 0 0 6 0 】

そしてステップ S 2 4 b では、信頼度算出部 2 0 9 は、信頼度が閾値未満となるクラスの数を選択し、該指標が規定数以上であるか否かを判断する。なお指標に適用可能なものは他にも考えられ、例えば、全クラス数に対する「信頼度が閾値未満となるクラスの数」の割合を指標として用いても良い。

【 0 0 6 1 】

この判断の結果、指標が規定数以上である場合には、ユーザによるラベル付けが必要と判断し、処理はステップ S 1 4 a に進む。一方、この判断の結果、指標が規定数未満である場合、信頼度算出部 2 0 9 は、ユーザによるラベル付けは不必要と判断し、処理はステップ S 1 5 に進む。この場合、ステップ S 1 5 では、信頼度算出部 2 0 9 は、代表位置リストと、ステップ S 1 1 で画像取得部 1 0 1 が取得した撮像画像と、を学習データとして記憶部 1 0 6 に格納する。

10

【 0 0 6 2 】

このように、本実施形態によれば、新たな画像に対してラベル付けする際に、ユーザは現在の認識器で認識される代表位置の信頼度が低い場合のみラベル付けをすればよい。そのため、ラベル付けをしなければならぬ画像数が減少し、ユーザのラベル付けの手間が低減される。

【 0 0 6 3 】

20

< 第 2 の実施形態の変形例 >

第 2 の実施形態では、着目クラスの信頼度は、該着目クラスに属する認識位置の数が多いほど高く、少ないほど低くなるように算出したが、着目クラスの信頼度の求め方は特定の求め方に限らない。例えば、認識部 1 0 2 の CNN は、対象物体かどうかの確率として出力されるため、例えば、着目クラスに属する認識位置に対応する CNN の出力（確率）の合計値が高いほど該着目クラスの信頼度を高くし、該合計値が低いほど該着目クラスの信頼度を低くするようにしても良い。また、クラスに属する認識位置の数をそのまま該クラスの信頼度とするのではなく、対象物体かどうかの確率を重みとした線形和として信頼度を定義してもよいし、各認識位置の代表位置からの距離の逆数を重みとした線形和で信頼度を定義してもよい。

30

【 0 0 6 4 】

また、第 2 の実施形態では、ユーザによるラベル入力（代表位置の編集操作）が必要かどうかを判断するために信頼度と比較する閾値は予め定められた規定値であるものとしたが、これに限る必要はない。例えば、図 4 の画面においてユーザ操作に応じてこの閾値の値を設定できるようにしても良い。閾値はスライダを用いて設定してもよいし、テキスト入力によって設定してもよい。また、ラベルが変更された代表位置とラベルが変更されなかった代表位置のそれぞれの信頼度の統計をとることで閾値を求めてもよい。たとえば、ラベル変更が行われた代表位置の信頼度の平均値が 1 0、ラベルが変更されなかった代表位置の信頼度の平均値が 5 0 であれば、その間の 3 0 を閾値とすれば良い。このようにすることで、ユーザが閾値を決める手間が低減される。

40

【 0 0 6 5 】

また第 2 の実施形態では、ユーザによるラベル入力が必要ない場合、信頼度算出部 2 0 9 は、代表位置リストと、ステップ S 1 1 で画像取得部 1 0 1 が取得した撮像画像と、を学習データとして記憶部 1 0 6 に格納していたが、これに限る必要はない。ユーザによるラベル入力が必要ない場合とは、現状の認識器で十分に認識できているため、例えば、学習データを記憶部 1 0 6 に格納することなく破棄してもよいし、ユーザに学習データとして追加するかどうかの判断を委ねてもよい。このようにすることで、一部の処理（主に学習部）が省略されるため、より早く一連の処理を終えられる。

【 0 0 6 6 】

また第 2 の実施形態では、ユーザによるラベル入力が必要な場合は、入力部 1 0 5 は、

50

代表位置リストに登録されている全ての代表位置を編集候補として画面に表示して編集操作を受け付けていたが、これに限る必要はない。信頼度が閾値以上のクラスについてはラベル入力が必要はないため、入力部 105 は信頼度が閾値未満となるクラスの代表位置のみを編集候補として画面に表示して編集操作を受け付けるようにしてもよい。逆に、信頼度が極端に低い場合は誤認識と考えられるため、下限の閾値も設け、信頼度が下限の閾値以上で且つ上限の閾値（ステップ S 24 b で使用した閾値）を下回るクラスの代表位置のみを編集候補として画面に表示して編集操作を受け付けるようにしてもよい。このようにすることで、ユーザが修正するラベル量が少なくなり、ユーザのラベル付けの手間が低減される。また、図 4 の画面に、各代表位置に対する信頼度を表す情報を表示するようにしても良い。信頼度を表す情報の表示としては、例えば、信頼度をテキスト表示しても良いし、信頼度に応じた色や形を有する規定のパターン画像を表示しても良いし、信頼度ごとに予め作成されたパターン画像を表示しても良い。これにより、ユーザがラベル付けする際に、代表位置をどの程度信用するかの参考になる。

【0067】

〔第 3 の実施形態〕

本実施形態に係る画像処理装置は、第 1 の実施形態に係る画像処理装置の機能に加え、学習後の CNN の性能の評価値を求め、該求めた評価値に応じて該 CNN を更新するか否かを判断する機能を有する。評価値が高くなる場合のみ CNN を更新することで、CNN の性能向上は加速し、より良い結果が得られるようになり、ユーザによるラベル修正量が少なくなるため、ラベル付けの手間が低減される。

【0068】

先ず本実施形態に係る物体認識装置の機能構成例について、図 8 のブロック図を用いて説明する。図 8 において図 1 に示した機能部と同じ機能部には同じ参照番号を付しており、該機能部に係る説明は省略する。本実施形態に係る物体認識装置 3000 は、画像処理装置 3100 の構成が図 1 の画像処理装置 1100 と異なる。

【0069】

学習部 307 は、上記の学習部 107 の機能に加え、学習処理によって得られた CNN のパラメータを、評価部 310 にも送出する。評価部 310 は、認識器記憶部 108 に保存されている CNN のモデル構造及びパラメータに基づく CNN（旧 CNN）の性能の評価値を、記憶部 311 に格納されているデータを用いて求める。更に評価部 310 は、学習部 307 から送出されたパラメータと認識器記憶部 108 に保存されている CNN のモデル構造とに基づく CNN（新 CNN）の性能を評価するための評価値を、記憶部 311 に格納されているデータを用いて求める。記憶部 311 には、試験画像と、該試験画像における代表位置と、が試験データとして格納されている。

【0070】

更新判定部 312 は、旧 CNN の評価値と新 CNN の評価値とを比較し、該比較の結果に基づいて、新 CNN のパラメータを認識器記憶部 108 に格納するか否かを判断する。そして更新判定部 312 は、旧 CNN の評価値よりも新 CNN の評価値が高い場合には、新 CNN のパラメータを認識器記憶部 108 に格納することで、認識器記憶部 108 が保持する旧 CNN を新 CNN に更新する。

【0071】

本実施形態に係るステップ S 2 及び S 3 における処理の詳細について、図 9 のフローチャートに従って説明する。図 9 においてステップ S 11 ~ S 15 が上記のステップ S 2 において実行され、ステップ S 36 ~ S 39 が上記のステップ S 3 において実行される。

【0072】

ステップ S 36 では、学習部 307 は、上記のステップ S 16 の処理を行うと共に、学習処理によって得られた CNN のパラメータを、評価部 310 にも送出する。

【0073】

ステップ S 37 では、評価部 310 は、記憶部 311 に格納されているデータを用いて、旧 CNN の評価値及び新 CNN の評価値を求める。旧 CNN 及び新 CNN のそれぞれの

10

20

30

40

50

評価値を求める方法には様々な方法があり、特定の方法に限らない。以下にCNN（旧CNN及び新CNN）の評価値を求める方法の一例を説明する。

【0074】

先ず、評価部310は、記憶部311から試験データを取得する。試験データには、1つのオブジェクトが写っている試験画像と、該試験画像における該オブジェクトの位置である代表位置と、のセットが複数セット含まれているものとする。そして評価部310は、試験データに含まれている複数セットのうち未選択のセットを選択セットとして選択し、上記のステップS12と同様に、該選択セットに含まれている試験画像に対してCNNを適用することで該試験画像における認識位置を取得する。そして評価部310は、選択セットの試験画像から認識した全ての認識位置が、該選択セットに含まれている代表位置から規定距離以内にあれば成功と判断する。一方、評価部310は、選択セットの試験画像から認識した認識位置のうち1つでも、該選択セットに含まれている代表位置から規定距離以内になければ失敗と判断する。このような選択セットに対する成功／失敗の判断処理を、試験データに含まれている全てのセットについて行うことで、それぞれのセットについて成功／失敗を判断する。そして評価部310は、（（成功と判断したセットの数）／全セット数）を評価値として求める。このような評価値を旧CNN及び新CNNのそれぞれについて求める。なお、上記の通り、このような評価値の算出方法は一例であり、他の算出方法を用いて評価値を算出しても良い。

10

【0075】

ステップS38では、更新判定部312は、旧CNNの評価値と新CNNの評価値との大小比較を行う。この大小比較の結果、旧CNNの評価値<新CNNの評価値であれば、処理はステップS39に進み、旧CNNの評価値 新CNNの評価値であれば、認識器記憶部108に格納されているパラメータの更新を行うことなく、図9のフローチャートに従った処理は終了する。ステップS39では、更新判定部312は、新CNNのパラメータを認識器記憶部108に格納することで、認識器記憶部108が保持するCNNを新CNNに更新する。

20

【0076】

このように、本実施形態によれば、新たな画像に対してラベル付けする際に、参考としてユーザに提示される代表位置の精度が向上していくため、ラベル修正量が少なくなり、ユーザのラベル付けの手間が低減される。

30

【0077】

<第3の実施形態の変形例>

第3の実施形態では、上記のステップS1の処理は省いても良い。すなわち、評価部310によって、新CNNの評価値が算出されるため、評価値用閾値を設け、新CNNの評価値が評価値用閾値以上になるまで繰り返しCNNの学習を続けるようにしてもよい。また、何度も学習を行う中で、評価値の変化が規定値未満となった場合に、学習を終了するようにしてもよい。たとえば、評価値の変化が0.01%未満で5回続き、性能が飽和したと考えられるような場合に、学習を終了する。このようにすることで、ユーザが学習が必要かどうかを判断する手間が削減される。

【0078】

40

また、第3の実施形態では、入力部105は、撮像画像と代表位置とを常に記憶部106に格納していた。しかし、学習データだけでなく、試験データを増やす必要もあるため、画像と代表位置とのセットを記憶部106だけでなく記憶部311にも格納するようにしても良い。なお、画像と代表位置のセットは記憶部106と記憶部311の両方に格納するようにしても良いし、記憶部106と記憶部311のうちランダムに決まる一方に格納するようにしても良い。また、画像と代表位置のセットを記憶部106及び記憶部311の何れに格納するのかをユーザが操作装置3を操作して指定しても良い。

【0079】

また、第2の実施形態で説明した信頼度算出部209を画像処理装置3100に組み込んで、信頼度算出部209による信頼度に応じて画像と代表位置のセットを記憶部106

50

及び記憶部 3 1 1 の何れに格納するのかを決定しても良い。たとえば、信頼度が閾値未満である代表位置と、対応する画像と、を記憶部 1 0 6 に格納し、信頼度が閾値以上である代表位置と、対応する画像と、を記憶部 3 1 1 に格納する。このようにすることで、学習データと試験データを効率よく収集することができる。

【 0 0 8 0 】

また、ステップ S 3 8 において、CNNの更新を行うか否かの判断だけでなく、新たに登録された学習データ（以下、新学習データ）を保持するかどうかの判断をしてもよい。CNNの性能が悪化したということは、新学習データのラベルに誤りが含まれている可能性が高いため、新学習データを破棄するようにしてもよい。または、新学習データのラベルが本当に正しいかを確認するため、ユーザに再度ラベル入力を要求してもよい。このようにすることで、ラベル誤りのある学習データを記憶する頻度を減らすことができる。

10

【 0 0 8 1 】

また、第 3 の実施形態では、旧 CNN 及び新 CNN の 2 つを評価していたが、これに限る必要はない。旧 CNN は複数あってもよいし、新 CNN も複数あってもよい。この場合、たとえば旧 CNN は、旧 CNN と新 CNN のうち評価値が上位 3 つの CNN を認識器記憶部 1 0 8 に記憶するようにすればよい。また、たとえば新 CNN は、初期値をランダムや旧 CNN のいずれかにしたり、学習データをランダムに選択するなどして、複数種の学習結果が得られるようにすればよい。

【 0 0 8 2 】

[第 4 の実施形態]

上記の各実施形態や各変形例では、新たに学習データが追加されるごとに学習を行っていたが、これに限る必要はない。例えば、複数セット（画像と代表位置のセット）が溜まってから学習を行ってもよいし、ユーザが学習を行うタイミングを操作装置 3 を操作して指定してもよい。

20

【 0 0 8 3 】

また、上記の各実施形態や各変形例では、ディープラーニングの一種である CNN を学習して認識器を構築したが、これに限る必要はない。特徴としては、LBP (Local Binary Pattern) を用いてもよいし、BoF (Bag of Features) などを用いてもよい。また、識別器としては、決定木を用いてもよいし、SVM (Support Vector Machine) などを用いてもよい。

30

【 0 0 8 4 】

また、上記の各実施形態や各変形例では、画像中の対象物体の位置のみを認識したが、これに限る必要はない。例えば、CNNの出力を対象物体かどうかの 2 クラス判別ではなく、姿勢の分類もするように出力を増やすことで、画像中の対象物体の位置姿勢を認識してもよい。以下、図 1 0 を用いて位置だけでなく姿勢も認識する場合の処理を説明する。

【 0 0 8 5 】

図 1 0 (a) の x マーカ 5 0 1 は認識すべき対象物体の画像中の位置 (x , y) を表しており、白抜き矢印 5 0 2 が該認識すべき対象物体の画像中の姿勢、黒矢印 5 0 3 は対象物体の姿勢を 8 姿勢に量子化した方向を示している。図 1 0 (a) のようにラベルを対象物体の位置姿勢 (x , y ,) として設定し、CNNを「 8 姿勢のうちどの姿勢の対象物体か」+「対象物体ではない」の 9 クラス判別するものとして設計する。

40

【 0 0 8 6 】

このような CNN を作成して、第 1 の実施形態におけるステップ S 1 2 と同様に認識を行うと、図 1 0 (b) のように、複数の位置姿勢の認識結果（認識位置姿勢）が得られる。図 1 0 (b) の x マーカ群 5 0 4 が認識された物体の位置群、白抜き矢印群 5 0 5 が認識された姿勢群を表している。

【 0 0 8 7 】

この認識結果を第 1 の実施形態におけるステップ S 1 3 と同様に、クラスタリングを行って代表位置姿勢を求めてユーザに提示する。代表位置姿勢を求める際は、各対象物体の位置姿勢 (x , y ,) を特徴とし、Kmeans アルゴリズムを用いてクラスタリング

50

してもよいし、位置 (x, y) をクラスタリングした後に同クラス内における姿勢 をクラスタリングしてもよい。このようにすることで、図 10 (c) の マーカ 506 (代表位置を示す) と黒矢印 507 (代表姿勢を示す) で示されるような代表位置姿勢を求めることができる。

【0088】

入力部 105 では、位置だけでなく姿勢も編集できるようにする。姿勢の修正は、例えば、対象物体の位置からマウスのカーソル位置への方向ベクトルとして設定できるようにすればよい。また、対象物体の位置を決定した後に姿勢を決定するようにしてもよいし、位置編集モードと姿勢編集モードとを切り換えられるようにしてもよい。

【0089】

尚、上記の説明は二次元画像中の位置姿勢の場合について説明したが、三次元空間中の位置姿勢のように次元が増えたとしても、ラベルを (x, y, z, , ,) とすればよく (, , はそれぞれ x、y、z 軸周りの角度)、同様の処理が可能である。

【0090】

[第 5 の実施形態]

図 1, 6, 8 に示したそれぞれの画像処理装置 1100, 2100, 3100 の各機能部はハードウェアで実装しても良いが、一部をソフトウェア (コンピュータプログラム) で実装しても良い。後者の場合、記憶部 106、認識器記憶部 108、記憶部 311 以外の各機能部の機能を実現するコンピュータプログラムを実行可能なコンピュータ装置は画像処理装置 1100, 2100, 3100 に適用可能である。画像処理装置 1100, 2100, 3100 に適用可能なコンピュータ装置のハードウェア構成例について、図 11 のブロック図を用いて説明する。

【0091】

CPU 1101 は、RAM 1102 や ROM 1103 に格納されているコンピュータプログラムやデータを用いて処理を実行する。これにより CPU 1101 は、コンピュータ装置全体の動作制御を行うと共に、画像処理装置 1100, 2100, 3100 のそれぞれが行うものとして上述した各処理を実行若しくは制御する。

【0092】

RAM 1102 は、I/F (インターフェース) 1105 を介して撮像装置 1 から取得した撮像画像、ROM 1103 や外部記憶装置 1104 からロードされたコンピュータプログラムやデータ、を格納するためのエリアを有する。更に RAM 1102 は、CPU 1101 が各種の処理を実行する際に用いるワークエリアを有する。このように RAM 1102 は、各種のエリアを適宜提供することができる。ROM 1103 には、書換不要のコンピュータプログラムやデータ、例えば BIOS のコンピュータプログラムやデータが格納されている。

【0093】

外部記憶装置 1104 は、ハードディスクドライブ装置に代表される大容量情報記憶装置である。外部記憶装置 1104 には、OS (オペレーティングシステム) や、画像処理装置 1100, 2100, 3100 が行うものとして上述した各処理を CPU 1101 に実行若しくは制御させるためのコンピュータプログラムやデータが保存されている。外部記憶装置 1104 に保存されているコンピュータプログラムには、図 1, 6, 8 において記憶部 106、認識器記憶部 108、記憶部 311 以外の各機能部の機能を CPU 1101 に実現させるためのコンピュータプログラムが含まれている。また、外部記憶装置 1104 に保存されているデータには、上記の説明において既知の情報として説明したものが含まれている。外部記憶装置 1104 に保存されているコンピュータプログラムやデータは、CPU 1101 による制御に従って適宜 RAM 1102 にロードされ、CPU 1101 による処理対象となる。なお、上記の記憶部 106、認識器記憶部 108、記憶部 311 は、RAM 1102 や外部記憶装置 1104 を用いて実装可能である。なお、外部記憶装置 1104 には、USB メモリや SD カードなどの、様々なメモリ装置を適用することができる。

10

20

30

40

50

【 0 0 9 4 】

I / F 1 1 0 5 は、上記の撮像装置 1、表示装置 2、操作装置 3 を接続するためのインターフェース群により構成されている。C P U 1 1 0 1、R A M 1 1 0 2、R O M 1 1 0 3、外部記憶装置 1 1 0 4、I / F 1 1 0 5 は何れもバス 1 1 0 6 に接続されている。

【 0 0 9 5 】

なお、上記の各実施形態や各変形例の一部若しくは全部を適宜組み合わせ使用しても構わない。また、上記の各実施形態や各変形例の一部若しくは全部を選択的に使用しても構わない。

【 0 0 9 6 】

(その他の実施例)

10

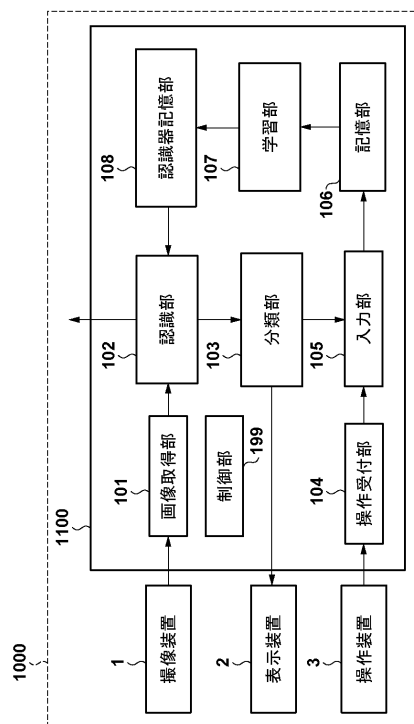
本発明は、上述の実施形態の 1 以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける 1 つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、1 以上の機能を実現する回路（例えば、A S I C）によっても実現可能である。

【符号の説明】

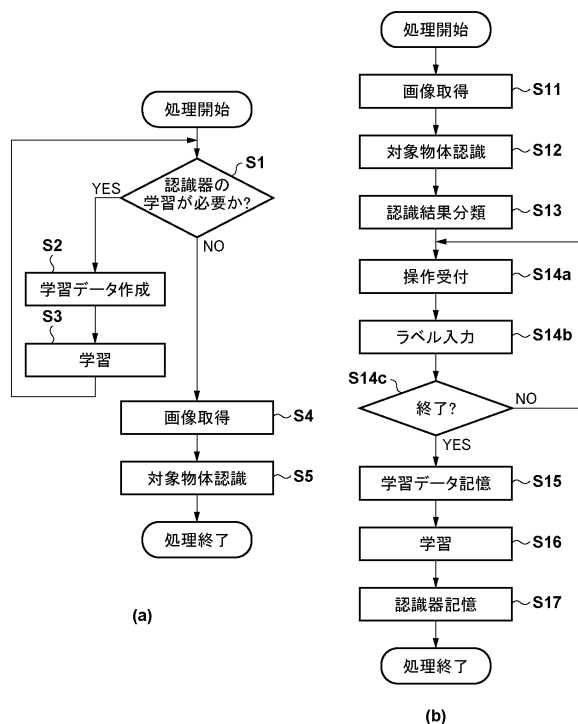
【 0 0 9 7 】

1 0 1 : 画像取得部 1 0 2 : 認識部 1 0 3 : 分類部 1 0 4 : 操作受付部 1 0 5 : 入力部
1 0 6 : 記憶部 1 0 7 : 学習部 1 0 8 : 認識器記憶部

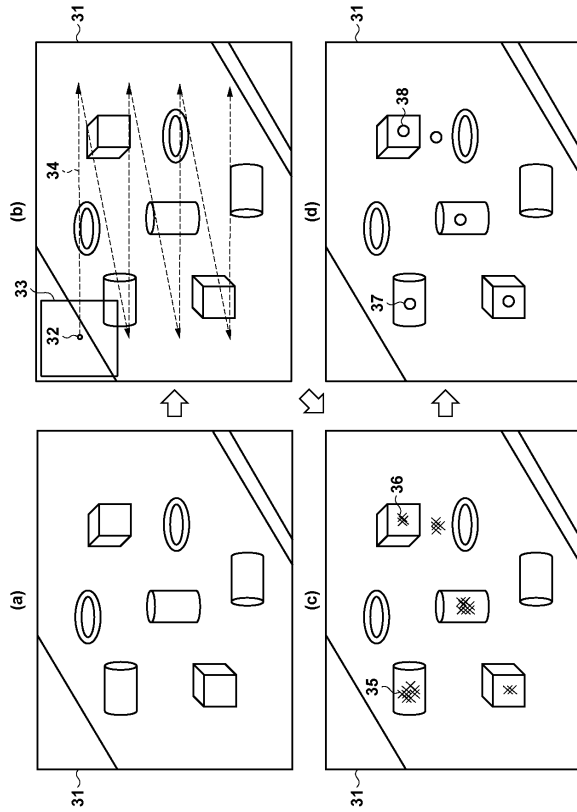
【図 1】



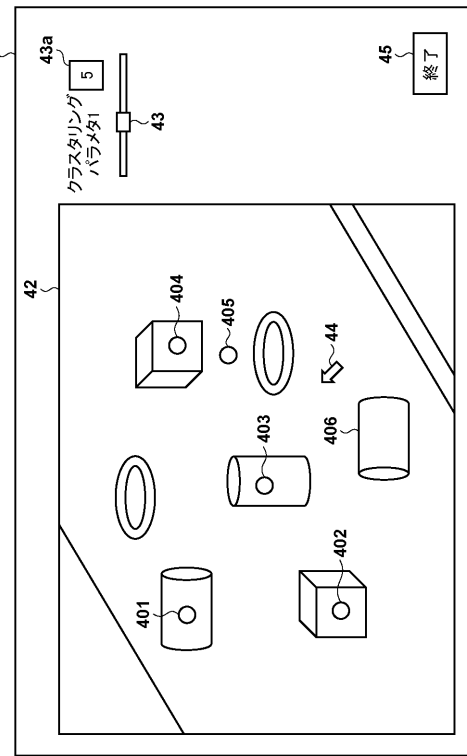
【図 2】



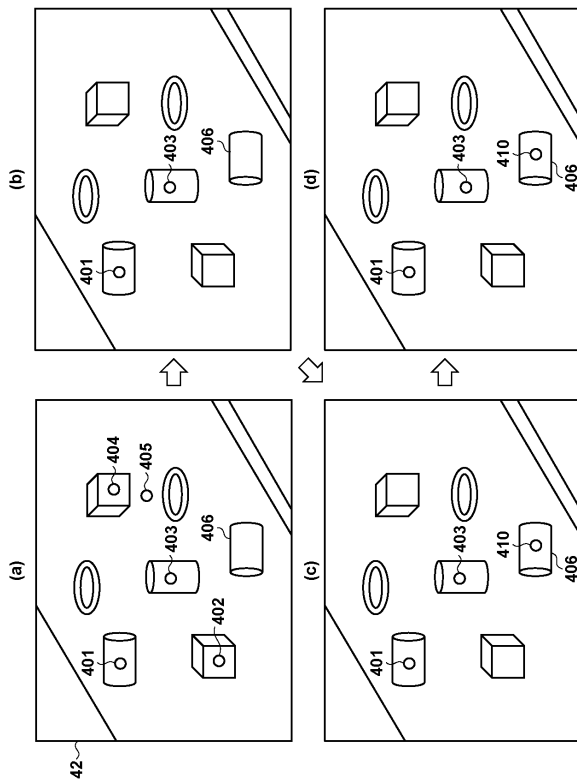
【図 3】



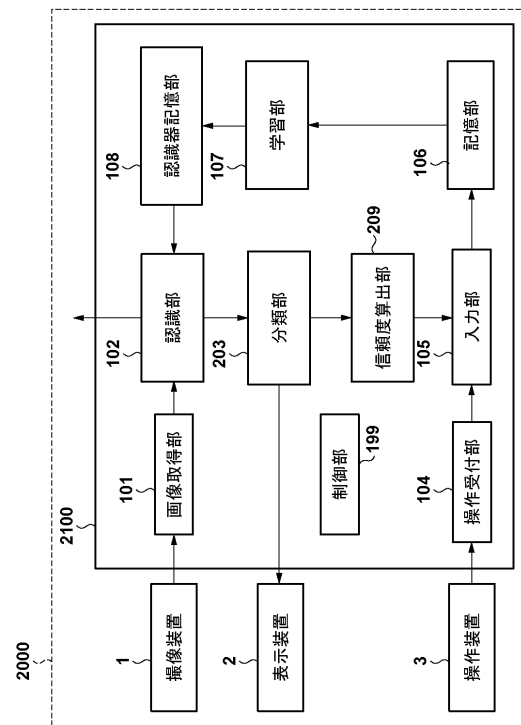
【図 4】



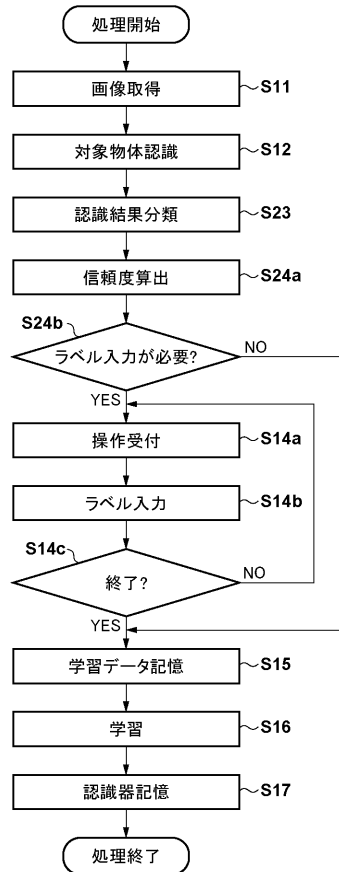
【図 5】



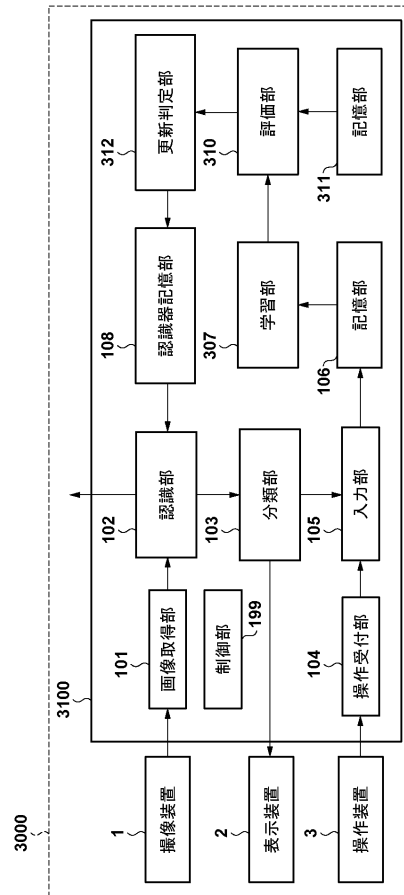
【図 6】



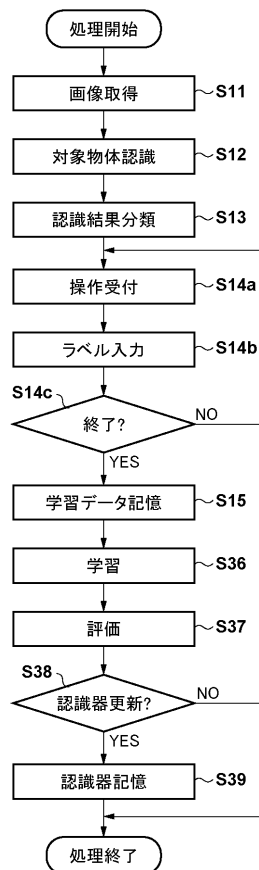
【図 7】



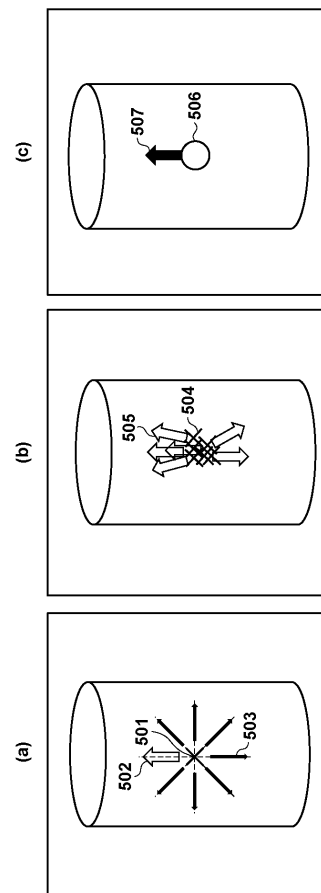
【図 8】



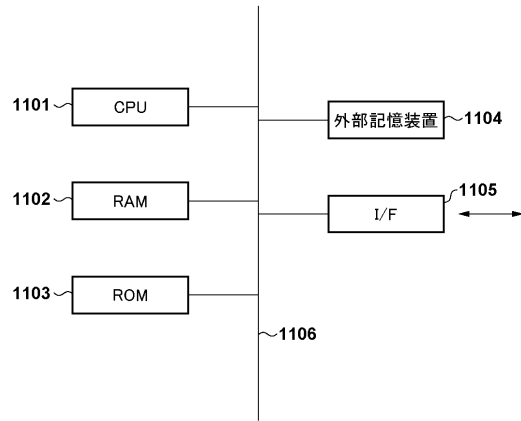
【図 9】



【図 10】



【図 11】



フロントページの続き

(56)参考文献 特開 2011-145791 (JP, A)
特開 2011-060227 (JP, A)
特開 2013-052490 (JP, A)
特開 2011-060221 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06T 7/00 - 7/90
G06N 20/00
G06N 3/04
G06N 3/08
IEEE Xplore