

(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2019-192022  
(P2019-192022A)

(43) 公開日 令和1年10月31日(2019.10.31)

(51) Int.Cl.		F I		テーマコード (参考)
<b>G06T 7/00</b>	<b>(2017.01)</b>	G06T 7/00	350B	5B057
<b>G06N 20/00</b>	<b>(2019.01)</b>	G06N 99/00	153	5L096
<b>G06T 1/00</b>	<b>(2006.01)</b>	G06T 1/00	300	

審査請求 未請求 請求項の数 22 O L (全 27 頁)

(21) 出願番号 特願2018-85520 (P2018-85520)  
(22) 出願日 平成30年4月26日 (2018.4.26)

(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 100126240  
弁理士 阿部 琢磨  
(74) 代理人 100124442  
弁理士 黒岩 創吾  
(72) 発明者 矢澤 友則  
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) 5B057 AA05 BA02 CA12 CA16 CB12  
CB16 CE08 DA07 DB02 DC03  
DC08 DC14 DC16 DC22 DC33  
DC40  
5L096 BA05 CA02 FA06 FA62 FA64  
FA67 FA69 GA34 KA04

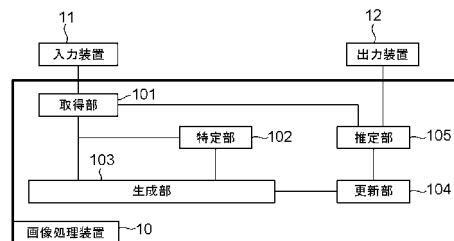
(54) 【発明の名称】 画像処理装置、画像処理方法及びプログラム

(57) 【要約】

【課題】 重なり合った物体に対するピックアップ位置の学習に用いる教師データを生成する技術を提供する。

【解決手段】 画像を用いた、重なり合った物体群に対するピックアップ位置の学習において、ピックアップ位置として正解または不正解を示す教師データの画像を生成する画像処理装置であって、少なくとも1つの物体が写った第1の画像と、重なり合った物体群が写った第2の画像とを取得する取得手段と、前記第1の画像から1つの前記物体の領域と、前記領域におけるピックアップ位置の正解に対応する第1の位置を特定する特定手段と、前記領域の画像と前記第2の画像とを合成した第3の画像のうち、前記第1の位置に前記第1の位置であることを示す第1のラベルを付与し、前記第3の画像のうち、前記領域の境界周辺に前記第1の位置とは異なることを示す第2のラベルを付与した前記教師データである第1の教師データを生成する生成手段とを有する。

【選択図】 図2



## 【特許請求の範囲】

## 【請求項 1】

画像を用いた、重なり合った物体群に対するピッキング位置の学習において、ピッキング位置として正解または不正解を示す教師データの画像を生成する画像処理装置であって、

少なくとも 1 つの物体が写った第 1 の画像と、重なり合った物体群が写った第 2 の画像とを取得する取得手段と、

前記第 1 の画像から 1 つの前記物体の領域と、前記領域におけるピッキング位置の正解に対応する第 1 の位置を特定する特定手段と、

前記領域の画像と前記第 2 の画像とを合成した第 3 の画像のうち、前記第 1 の位置に前記第 1 の位置であることを示す第 1 のラベルを付与し、前記第 3 の画像のうち、前記領域の境界周辺に前記第 1 の位置とは異なることを示す第 2 のラベルを付与した前記教師データである第 1 の教師データを生成する生成手段とを有することを特徴とする画像処理装置。

10

## 【請求項 2】

前記第 1 の位置は、前記物体の領域の中心であ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】

前記第 1 のラベルは、前記第 3 の画像のうち、前記第 1 の位置を基点とした所定の第 1 の幅より内側の領域に付与されることを特徴とする請求項 1 または 2 に記載の画像処理装置。

20

## 【請求項 4】

前記境界周辺の幅は、前記物体の中心から該物体の境界の点までの距離より小さい幅であることを特徴とする請求項 1 乃至 3 のいずれか 1 項に記載の画像処理装置。

## 【請求項 5】

前記第 2 のラベルは、前記第 3 の画像のうち、前記第 1 の位置を基点とした所定の第 2 の幅より外側に対して付与されることを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載の画像処理装置。

## 【請求項 6】

前記第 2 のラベルは、前記第 3 の画像のうち、前記境界周辺であって前記領域の外側の領域に付与されることを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の画像処理装置。

30

## 【請求項 7】

前記生成手段は、前記第 3 の画像のうち、前記領域に対して前記第 2 のラベルを付与した後に、前記第 1 の位置を基点とした第 3 の幅より小さい領域から該第 2 のラベルを削除し、前記第 1 の位置を基点とした第 4 の幅より小さい領域に前記第 1 のラベルを付与した前記第 1 の教師データを生成し、

前記第 3 の幅は前記第 4 の幅より大きいことを特徴とする請求項 1 乃至 6 のいずれか 1 項に画像処理装置。

## 【請求項 8】

前記第 3 の画像は、前記領域の一部と前記第 2 の画像とを合成した画像であり、

前記生成手段は、前記領域の境界周辺であって該境界の内側の領域に、前記第 2 のラベルを付与した前記第 1 の教師データを生成することを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の画像処理装置。

40

## 【請求項 9】

前記生成手段は、前記第 3 の画像の部分画像であって所定の大きさである第 4 の画像から第 2 の教師データを生成し、

前記第 2 の教師データは、前記第 4 の画像の中心にピッキング位置があるか否かを示す画像であることを特徴とする請求項 1 乃至 8 の何れか 1 項に記載の画像処理装置。

## 【請求項 10】

50

前記生成手段は、前記第 3 の画像の部分画像であって所定の大きさである第 4 の画像の中心に前記第 1 の位置が含まれた場合に、前記第 1 のラベルを該第 4 の画像の中心に付与した前記第 2 の教師データを生成す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】

前記生成手段は、前記第 3 の画像の部分画像であって所定の大きさである第 4 の画像の中心が、前記領域でありかつ前記第 1 の位置ではない場合または前記領域の境界周辺に含まれる場合に、前記第 2 のラベルを該第 4 の画像の中心に付与した前記第 2 の教師データを生成することを特徴とする請求項 9 または 10 に記載の画像処理装置。

【請求項 12】

前記第 2 の教師データは、入力された画像に対してピッキング位置があるか否かの正解または不正解を出力する学習モデルの教師データであって、

前記第 4 の画像は、前記学習モデルに入力する画像と同じ大きさであることを特徴とする請求項 10 または 11 に記載の画像処理装置。

【請求項 13】

前記第 1 の画像は、前記物体と異なる色である背景に置かれた前記物体が撮像された画像であって、

前記特定手段は、該画像の輝度の差を用いて前記物体の領域を特定することを特徴とする請求項 1 乃至 12 いずれか 1 項に記載の画像処理装置。

【請求項 14】

前記特定手段は、複数の物体が写った画像と、該物体のうち重なり合っていない物体を 1 つだけを取り除いた後に撮像された画像との画像特徴の差に基づいて 1 つの前記物体の領域を特定することを特徴とする請求項 1 乃至 13 いずれか 1 項に記載の画像処理装置。

【請求項 15】

前記第 2 の画像の画像特徴に基づいて前記物体の領域と背景との境界を検出する検出手段をさらに有することを特徴とする請求項 1 乃至 14 のいずれか 1 項に記載の画像処理装置。

【請求項 16】

前記特定手段は、前記検出手段で検出された前記境界の長さに合わせて前記第 1 の画像から前記物体の領域の部分画像を特定し、

前記生成手段は、前記部分画像を前記第 2 の画像の前記境界に沿って合成した画像のうち、前記境界の周辺でかつ前記領域に含まれる領域に対して、前記第 2 のラベルを付与した前記第 1 の教師データを生成することを特徴とする請求項 15 に記載の画像処理装置。

【請求項 17】

前記生成手段が生成した教師データに基づいて重なり合った物体群を撮像した入力画像に対する吸着位置を出力する学習モデルのパラメータを更新する更新手段を更に有することを特徴とする請求項 1 乃至 16 のいずれか 1 項に記載の画像処理装置。

【請求項 18】

前記物体の 3 次元形状モデルと前記第 2 の画像における前記物体とを照合する照合手段をさらに有し、

任意の姿勢である前記 3 次元形状モデルと前記第 2 の画像における物体の姿勢が一致した場合、

前記特定手段は、前記 3 次元形状モデルを前記第 2 の画像に投影して前記 3 次元形状モデルの前記第 1 の位置を特定することを特徴とする請求項 1 乃至 17 いずれか 1 項に記載の画像処理装置。

【請求項 19】

画像を用いた、重なり合った物体群に対するピッキング位置の学習において、ピッキング位置としての正解または不正解を示す教師データの画像を生成する画像処理装置であって、

少なくとも 1 つの物体が写った第 1 の画像と、重なり合った物体群が写った第 2 の画像とを取得する取得手段と、

10

20

30

40

50

前記第 1 の画像から 1 つの前記物体の領域と、前記領域におけるピッキング位置の正解に対応する第 1 の位置とを特定する特定手段と、

前記物体の領域の画像と前記第 2 の画像とを合成した第 3 の画像のうち、前記物体の領域の境界周辺に前記第 1 の位置とは異なることを示すラベルを付与した前記教師データを生成する生成手段とを有することを特徴とする画像処理装置。

【請求項 20】

重なり合った物体群に対するピッキング位置の学習において入力画像に対してピッキング位置の正解または不正解を示す教師データを生成する画像処理装置であって、

少なくとも 1 つの物体が写った第 1 の画像と、重なり合った物体群が写った第 2 の画像とを取得する取得手段と、

前記第 2 の画像の画像特徴に基づいて前記物体の領域と背景との境界を検出する検出手段と、

前記検出手段で検出された前記境界の長さに合わせて前記第 1 の画像から前記物体の領域の部分画像を特定する特定手段と、

前記部分画像を前記第 2 の画像の前記境界に沿って合成した画像のうち、前記境界の周辺でかつ前記領域に含まれる領域に対して前記第 1 の位置とは異なることを示すラベルを付与した前記教師データを生成する生成手段とを有することを特徴とする画像処理装置。

【請求項 21】

コンピュータを請求項 1 乃至 20 の何れか 1 項に記載の画像処理装置が有する各手段として機能させるためのプログラム。

【請求項 22】

画像を用いた、重なり合った物体群に対するピッキング位置の学習において、ピッキング位置として正解または不正解を示す教師データの画像を生成する画像処理方法であって、

少なくとも 1 つの物体が写った第 1 の画像と、重なり合った物体群が写った第 2 の画像とを取得する取得工程と、

前記第 1 の画像から 1 つの前記物体の領域と、前記領域におけるピッキング位置の正解に対応する第 1 の位置とを特定する特定工程と、

前記領域の画像と前記第 2 の画像とを合成した第 3 の画像のうち、前記第 1 の位置に前記第 1 の位置であることを示す第 1 のラベルを付与し、前記第 3 の画像のうち、前記領域の境界周辺に前記第 1 の位置とは異なることを示す第 2 のラベルを付与した前記教師データである第 1 の教師データを生成する生成工程とを有することを特徴とする画像処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像を処理して、ピッキング位置の学習に用いる教師データを生成する技術に関する。

【背景技術】

【0002】

ディープラーニングを行うには、推定する環境を再現した大量の教師データが必要である。ディープラーニングといった機械学習の教師データを生成する方法として、群衆に対して人物領域の画像を合成して教師データを得る、特許文献 1 に開示された方法が知られている。

【0003】

近年、産業用ロボットが物体をピッキングする技術において、複数の物体が重なり合った状態である物体群の画像から、物体の位置や姿勢を推定する学習モデルを使ったディープラーニングの活用が提案されている。産業用ロボットがピッキングするときは、重なり合った物体群からピッキングに適した位置や姿勢の物体を見つける必要がある。

【先行技術文献】

10

20

30

40

50

## 【特許文献】

【0004】

【特許文献1】特許第6008045号

## 【非特許文献】

【0005】

【非特許文献1】J. Hesch and S. Roumeliotis. A direct least-squares (DLS) method for PnP. Proc. ICCV, pages 383-390, 2011.

【非特許文献2】Nobuyuki Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics, 1979.

【非特許文献3】Kingma, D. P., & Ba, J. L. Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, 1-13, 2015

【非特許文献4】立野圭祐, 小竹大輔, 内山晋二. ピンピッキングのための距離・濃淡画像を最ゆうに統合する高精度高安定なモデルフィッティング手法. 電子情報通信学会論文誌D, Vol. 94, No. 8, pp. 1410-1422, 2011. 8.

## 【発明の概要】

【発明が解決しようとする課題】

【0006】

重なり合った物体群から1つの物体をピッキングする場合、重なり合った物体群のうちピッキング装置の手前の方にある物体をピッキングする必要がある。すなわち、全体が見えている物体をピッキングすることが望まれる。従来、このような重なり合った物体群の手前にある物体に注目させつつ、ディープラーニングに用いる教師データを生成する技術は、確立されていなかった。

【0007】

本発明は上記課題に鑑みてなされたものであり、重なり合った物体群に対するピッキング位置の機械学習に用いる教師データを生成する技術を提供することを目的とする。

【課題を解決するための手段】

【0008】

上記課題を解決する本発明にかかる画像処理装置は、画像を用いた、重なり合った物体群に対するピッキング位置の学習において、ピッキング位置として正解または不正解を示す教師データの画像を生成する画像処理装置であって、少なくとも1つの物体が写った第1の画像と、重なり合った物体群が写った第2の画像とを取得する取得手段と、前記第1の画像から1つの前記物体の領域と、前記領域におけるピッキング位置の正解に対応する第1の位置を特定する特定手段と、前記領域の画像と前記第2の画像とを合成した第3の画像のうち、前記第1の位置に前記第1の位置であることを示す第1のラベルを付与し、前記第3の画像のうち、前記領域の境界周辺に前記第1の位置とは異なることを示す第2のラベルを付与した前記教師データである第1の教師データを生成する生成手段とを有する。

【発明の効果】

【0009】

重なり合った物体群に対するピッキング位置の機械学習に用いる教師データを生成できる。

【図面の簡単な説明】

【0010】

【図1】ハードウェア構成の一例を示す図

【図2】画像処理装置の機能構成例を示すブロック図

10

20

30

40

50

【図 3】処理対象である物体が撮像された画像の一例を示す図

【図 4】マーカーボードの一例を示す図

【図 5】特定部に係る機能構成例を示すブロック図

【図 6】(a)差分画像(b)前景画像の一例を示す図

【図 7】生成部に係る機能構成例を示すブロック図

【図 8】(a)学習実行時(b)推定処理実行時に画像処理装置が実行する処理の流れを示すフローチャート

【図 9】特定部が実行する処理の流れを示すフローチャート

【図 10】生成部が実行する処理の流れを示すフローチャート

【図 11】(a)合成画像(b)膨張領域の一例を示す図

【図 12】(a)重なり合った物体群が撮像された画像(b)物体の部分画像(c)合成画像の一例を示す図

【図 13】(a)物体群が撮像された画像(b)差分画像の一例を示す図

【図 14】特定部に係る機能構成例を示すブロック図

【図 15】特定部が実行する処理の流れを示すフローチャート

【図 16】特定部に係る機能構成例を示すブロック図

【図 17】特定部が実行する処理の流れを示すフローチャート

【図 18】重なり合った物体群を横から見た様子を示す図

【発明を実施するための形態】

【0011】

(第1の実施形態)

第1の実施形態では、物体が写った画像から教師データの生成を行う画像処理装置について説明する。ここでは、ピッキング位置を推定する画像解析を使ったピッキングマシンに本発明を適用した例について述べる。マシンによるピッキングにおいては、把持式と吸着式が主流であるが、ここでは吸着式を例にあげて説明する。

【0012】

本実施形態における教師データは、重なり合った物体群が撮像された画像から吸着によるピッキングに適したピッキング位置を推定する学習モデルに用いる。以下、吸着式のピッキング装置におけるピッキング位置を吸着位置と記述する。

【0013】

ここで、本実施形態の概要を説明する。安定的に物体をピッキングする為には、物体の重心付近をピッキングすることが求められる。一方で、重なり合った物体群のうち、他の物体(上あるいは手前にある物体)によって遮蔽される物体(下あるいは奥の物体)は、その物体の重心付近を吸着して持ち上げることが難しい。そこで、本実施形態における画像処理装置は、『他の物体によって遮蔽された物体(下あるいは奥にある物体)の領域は、ピッキング位置にある物体ではない』ことを学習させる教師データを生成する。具体的には、重なり合った物体群が撮像された画像を背景として、1つの物体が撮像された画像を合成する。また、1つの物体が撮像された画像に対して、物体の重心付近の領域と、物体とその境界周辺の領域を特定する。前者の領域に対しては、吸着位置であることを示す教師ラベル(後述する第1のラベル)を付与した正解の教師データを作成する。後者の領域に対しては、吸着位置ではないことを示す教師ラベル(後述する第2のラベル)を付与し、不正解の教師データを作成する。物体同士が少しずつ重なっている場合、上の物体の境界周辺に吸着位置でないことを示すラベルが付与されている。そのため、重なり合った物体群が撮像された画像から吸着によるピッキングに適した吸着位置を推定する推定処理実行時に下の物体の吸着位置は検出されない。また、物体同士が重なっている領域が小さい場合、重なっている境界周辺に一定の幅のラベルが付与され、それ以外の下物体の領域には吸着位置にはラベルが付与されない。遮蔽されている物体全体に不正解を表すラベルを付けてしまうと、吸着可能な位置であっても不正解であると推定する学習を行う可能性がある。本発明を適用した画像処理装置は、物体の中心に他の物体が重なっているような判断が難しい部分を再現した教師データが多く必要な場面で、画像における不正解と

10

20

30

40

50

すべき領域を効率的に生成できる。また、これらの教師データを学習モデルに与えることによって、遮蔽された物体の領域には吸着位置が出力されないように学習モデルのパラメータを更新する。これによって、推定処理実行時に、遮蔽された物体を検出することを抑制する。以下、図面を用いて詳細に説明する。なお、以下の説明において、複数の物体のことは物体群と記載し、1つの物体は物体と記載する。物体は、工業用部品を物体とするが、ピッキング装置でピッキング可能なものであれば何でも良い。また、物体の数は、部品の数、部品を組み合わせたユニット数、袋や箱に詰められたセット数等であって、例えば、1回のピッキング動作でピッキングしたい単位でカウントする。

#### 【0014】

図1を用いて、本実施形態のハードウェアの構成例を示す。CPU(H01)は、RAM(H03)をワークメモリとして、ROM(H02)や記憶装置H04に格納されたOSやその他プログラムを読みだして実行し、システムバスH00に接続された各構成を制御して、各種処理の演算や論理判断などを行う。CPU(H01)が実行する処理には、実施形態の画像処理が含まれる。記憶装置(H04)は、ハードディスクドライブや外部記憶装置などであり、実施形態の処理にかかるプログラムや各種データを記憶する。入力部(H05)は、カメラなどの撮像装置、ユーザー指示を入力するためのボタン、キーボード、タッチパネルなどの入力デバイスである。なお、記憶装置(H04)は例えばSATAなどのインタフェースを介して、入力部(H05)は例えばUSBなどのシリアルバスを介して、それぞれシステムバス(H00)に接続されるが、それらの詳細は省略する。通信I/F(H06)は無線通信で外部の機器と通信を行う。表示部(H07)はディスプレイである。ハードウェア構成はこれに限ったものではない。

#### 【0015】

図2に、第1の実施形態における画像処理装置10の機能構成例を示す。画像処理装置10は、画像を用いた、重なり合った物体群に対するピッキング位置の学習において、ピッキング位置として正解または不正解を示す教師データの画像を生成する。画像処理装置10は入力装置11から取得した画像から、推定部105が保持する学習モデルを使って画像に映っている特定の物体の位置または姿勢を出力装置12へ出力する。学習モデルとは、入力画像から入力画像に対応する吸着位置を出力するニューラルネットワークに基づくネットワーク構造とそのパラメータとする。入力装置11は、画像処理装置10に画像を入力する。ここでは、入力装置11は撮像装置(カラーカメラ)であるとする。撮像装置は、例えばグレースケールカメラや、赤外線カメラや、広角レンズカメラや、パノラマカメラ、ステレオカメラであっても良い。また、入力装置は、距離画像を取得する距離センサを有する装置でも良い。例えば、LiDAR、TOF方式等のアクティブ距離センサである。出力装置12は、ロボットアーム等のピッキング装置である。このピッキング装置は、例えば画像処理装置10が推定した結果を用いて、吸着部を物体の吸着位置まで移動させて物体を吸着し、搬送するタスクを行う。吸着式以外でも把持するハンドや吊り下げるフックを持つ装置でも良い。画像処理装置10から推定結果を出力する表示装置でも良い。画像処理装置10は、取得部101、特定部102、生成部103、更新部104、推定部105から構成される。画像処理装置10の機能構成はこれに限ったものではない。例えば、更新部104や推定部105は外部にあって、無線または有線の通信によって接続されていても良い。例えば、学習モデルを保持する記憶部を有しても良い。また、入力装置11または出力装置12が有する機能構成を含んでも良い。

#### 【0016】

取得部101は、学習実行時においては、図3に示すように、物体1が写った画像1130及び重なり合った物体群が写った画像1131を取得する。また、推定処理実行時においては、重なり合った物体群が写った画像を取得する。これらの画像は入力装置11もしくは記憶装置H04から取得する。画像1130は、図3(a)に示すように、マーカーボード111にある姿勢で置かれた物体1を撮像した画像である。ここでは物体は1つだけ置かれるものとする。また、画像1130は、物体の姿勢ごとに分類されている。詳細は後述するが、マーカーボード111は、撮像された2次元画像から3次元情報を得る

10

20

30

40

50

為に利用する。画像 1 1 3 0 は、学習を実行する前に前もって準備した画像である。学習の準備とは、教師データを生成する作業を指す。画像 1 1 3 0 を準備する方法は後述する。なお、学習実行とは、教師データを学習モデルに与え、学習モデルのそれぞれの層で行う計算の重みづけ係数などのパラメータを更新する動作を指す。画像 1 1 3 1 は、図 3 ( b ) に示すように、重なり合った物体群が写った画像である。重なり合った物体群が写った画像を背景に使うことによって、様々な姿勢の物体が写った画像の教師データを作成でき、効果的に学習できる。なお、これらの画像は、入力装置 1 1 もしくは記憶装置 H 0 4 から取得する。本実施形態において、画像はカラー画像であるものとして説明する。

#### 【 0 0 1 7 】

図 4 を使って、マーカーボード 1 1 1 について説明する。マーカーボード 1 1 1 は、マーカー領域 1 1 1 1 と背景用途領域 1 1 1 2 を有する。マーカー領域 1 1 1 1 とは、規則的に配列された白と黒マーカーがある領域を指す。なお、マーカーの色は白と黒に限定されず、2 種類のマーカーの色の差が区別出来ればよい。マーカー領域 1 1 1 1 に記されているマーカーの色、マーカー同士の間隔、配置位置のマーカー情報はあらかじめ計測しておく。計測された情報は記憶装置 H 0 4 で保持され、特定部 1 0 2 が取得する。また、画像 1 1 3 0 ごとに撮像装置の内部パラメータ（焦点距離、画像中心位置、レンズ歪み等）はあらかじめキャリブレーションしておく。このマーカー情報を用いて、マーカーボードに対する撮像装置の相対的な位置または姿勢を求めることが出来る。撮像装置の位置や姿勢を求める方法は S 1 2 0 1 で述べる。背景用途領域 1 1 1 2 は、物体の色とは異なる単一色の領域である。物体の差分画像を特定する際に、この色の差（輝度）を用いる。差分画像を求める方法は S 1 2 0 2 で述べる。

10

20

#### 【 0 0 1 8 】

ここで、画像 1 1 3 0 を準備する方法をの概略を述べる。まず、領域 1 1 1 2 に物体を 1 つ配置する。物体を配置した位置は計測しておき、マーカー情報として保持する。このとき、画像 1 1 3 0 は物体の姿勢グループ毎に分類する。姿勢グループとは、物体を見え方で分類したもので、例えば、表と裏といったふうに分類された平均的な姿勢を指す。すなわち、画像 1 1 3 0 は物体の姿勢に関する情報と紐づけられているものとする。次に、撮像装置を動かして、様々な角度から物体を撮像する。撮像装置は、デジタルカメラ等の一般的なカメラを用いてよい。後述する S 1 2 0 3 の手順を実施すると、マーカーボード 1 1 1 を使うことによって、吸着位置の推定精度を上げることができる。また、簡単なセッティングで教師データとなる画像を準備出来る為、工場等の現場でも使いやすい。

30

#### 【 0 0 1 9 】

特定部 1 0 2 では、画像 1 1 3 0 から 1 つの物体の領域と、物体の領域におけるピックアップ位置の正解に対応する中心位置を特定する。図 5 に、第 1 の実施形態における特定部 1 0 2 の機能構成例を示す。各機能構成の概略を説明する。位置姿勢特定部 1 0 2 1 では、画像 1 1 3 0 が撮像されたときの撮像装置の位置または姿勢を特定する。画像 1 1 3 0 に映っているマーカーボード 1 1 1 と、その画像を撮像した撮像装置との相対的な位置と姿勢を推定する。これによって、多様な姿勢をとる物体の画像を取得できる。領域特定部 1 0 2 2 は、画像 1 1 3 0 から背景との差分画像 1 1 4 を特定し、物体 1 つ分の画像である前景画像 1 1 3 2 を特定する。中心位置特定部 1 0 2 3 では、位置姿勢特定部 1 0 2 1 で求めた相対位置姿勢と、領域特定部 1 0 2 2 で求めた前景画像 1 1 3 2 とに基づいて、前景画像 1 1 3 2 の中心位置を特定する。この中心位置が物体の重心近傍であると仮定したとき、中心位置は吸着に適した位置であるといえる。具体的な方法は後述する。

40

#### 【 0 0 2 0 】

図 6 を用いて、領域特定部 1 0 2 2 で特定される画像を説明する。差分画像 1 1 4 とは、図 6 ( a ) に示すような 1 つの物体の領域を示す画像である。ここでは検出したい物体以外を隠した画像を指す。差分画像 1 1 4 は、画像 1 1 3 0 の部分画像から輝度の差分に基づいて特定する。具体的な処理は S 1 2 0 2 で述べる。前景画像 1 1 3 2 とは、図 6 ( b ) に示すように、物体を 1 つだけ取得した画像である。すなわち、画像 1 1 3 0 における物体の領域より外側が削除された画像である。差分画像 1 1 4 を画像 1 1 3 0 に照合す

50

ることによって特定する。

#### 【0021】

生成部103では、取得部101で取得した画像1131と、特定部102で特定された前景画像1132とを合成した画像と物体の領域の中心位置に基づいて、教師データ(第1の教師データ)を生成する。第1の教師データとは、画像の各画素または領域に対してピッキング位置であるか否かを示す正解値または不正解値といった教師ラベルを付与した画像である。図7に、生成部103の機能構成の一例を示す。生成部103は、画像生成部1031、第1のラベル生成部1032、第2のラベル生成部1033から構成される。以下概要を説明する。画像生成部1031では、画像1131と、前景画像1132を利用して、合成画像を生成する。合成画像とは、重なり合った物体群を背景として、1つの物体を前景として重畳した画像である。第1のラベル生成部1032では、画像生成部1031で生成された合成画像のうち、物体の中心位置に対して第1のラベルを付与した第1の教師データを生成する。本実施形態における第1のラベルとは、吸着位置であることを示すラベルである。第2のラベル生成部1033では、画像生成部1031で生成された合成画像のうち物体の領域とその境界周辺に対して第2のラベルを付与した第1の教師データを生成する。本実施形態における第2のラベルとは、吸着位置ではないことを示すラベルである。第2の教師データは、画像1131の部分画像を用いて生成される。生成する方法は後述する。さらに第1の教師データから、各画素を中心として一定の大きさをスライドさせて生成した第2の教師データを生成しても良い。第2の教師データは後述の推定部105が保持する学習モデルのパラメータの更新に用いる。

10

20

#### 【0022】

更新部104では、生成部103で生成された教師データを用いて、推定部105が保持する学習モデルのパラメータの更新を行う。学習モデルについては推定部105で述べる。パラメータの更新とは、学習モデルの入力側の層に画像を設定し、出力画像の層に画像に対する正解値を設定し、ニューラルネットワークを経由して算出される出力が設定した正解値に近づくようにニューラルネットワークのパラメータを調整する処理を指す。なお、更新部104は、画像処理装置10の外部にあっても良い。つまり、学習を実行する別の装置に画像処理装置10で生成した教師データを入力しても良い。

#### 【0023】

推定部105では、推定処理実行時に、保持している学習モデルを使って入力画像に映っている吸着可能な物体の位置を推定し、出力装置12へ出力する。本実施形態の学習モデルにおいては、物体の吸着位置が正解値として設定されており、学習モデルは吸着位置を推定結果として出力する。ここでは、学習モデルはニューラルネットワークであり、大きさが定数の正方形の画像(縦が128画素であり、横が128画素の画像など)を入力として受け付ける。入力画像の大きさは既知であるものとする。また、学習モデルは、その画像の中心が吸着位置であるか吸着位置でないかをクラス分類するニューラルネットワークで構成する。このクラス分類器は、入力した画像の中心が吸着位置か否かを判定する。さらに吸着位置である場合、物体の姿勢の判定も行う。なお、学習モデルは、入力画像から物体の位置または姿勢、物体のクラス(種類)、物体のテクスチャを推定するものでも良い。また、物体に印字されている文字や記号を認識する学習モデルでも良い。推定部105は、画像処理装置10とは別の装置に具備されても良い。例えば、出力装置12に推定部が具備されていても良い。

30

40

#### 【0024】

図8のフローチャートを用いて、教師データの準備段階を含む学習実行時と推定処理実行時における画像処理装置10が実行する処理の流れを説明する。以下、フローチャートは、CPUが制御プログラムを実行することにより実現されるものとする。以下の説明では、各工程(ステップ)について先頭にSを付けて表記することで、工程(ステップ)の表記を省略する。図8(a)のフローチャートは、学習実行時に画像処理装置10が実行する処理手順を示す。図8(b)のフローチャートは、吸着可能な物体の位置または物体の姿勢を推定する学習モデルを用いた推定処理実行時に、画像処理装置10が実行する処

50

理手順を示す。はじめに、準備段階を含む学習を実行するとき、画像処理装置 10 が実行する処理手順について図 8 ( a ) を参照して説明する。

【 0 0 2 5 】

S 1 0 0 1 では、取得部 1 0 1 が、少なくとも 1 つの物体が写った第 1 の画像と、重なり合った物体が写った第 2 の画像とを取得する。図 3 を参照すると、処理対象である物体が写った画像 1 1 3 0 ( 第 1 の画像 ) と山積みされた物体群が写った画像 1 1 3 1 ( 第 2 の画像 ) を取得する。

【 0 0 2 6 】

S 1 0 0 2 では、特定部 1 0 2 が、画像 1 1 3 0 ( 第 1 の画像 ) から物体の領域を特定し、さらに物体の領域から中心位置 ( 第 1 の位置 ) を特定する。図 9 のフローチャートに基づき、S 1 0 0 2 で特定部 1 0 2 が実行する処理の流れを説明する。

10

【 0 0 2 7 】

S 1 2 0 1 では、位置姿勢特定部 1 0 2 1 が、画像 1 1 3 0 に写ったマーカーボード 1 1 1 と、撮像した撮像装置との相対的な位置と姿勢を特定する。マーカーボードと撮像装置との相対的な位置と姿勢 ( 以下、相対位置姿勢と記載する ) は、3次元空間上での姿勢の差分を表す 3 次正方行列と、位置の差分を表す 3 次元ベクトルで表される情報である。まず、画像 1 1 3 0 に写っているマーカー領域 1 1 1 1 のマーカーを、ハフ変換による円検出で取得する。次に、円検出で取得したマーカーのうち白のマーカーの成すベクトルと、既知であるマーカー同士の間隔を利用して、相対位置姿勢を最小二乗法で取得する。最小二乗法の適用方法としては、例えば非特許文献 1 に示す公知の方法を利用する。非特許文献 1 では、画像上の点とモデルとして保持している 3 次元点を、位置姿勢をパラメータとした方程式に表し、その両辺の差の二乗誤差の和の最小化をする。最小化する式は非特許文献 1 の表現方法では 4 次式になるので、姿勢パラメータで偏微分した式が 0 になるような連立方程式をコンパニオンマトリクス由来の方法で解く。ここで、位置と姿勢はマーカーボード 1 1 1 上に定めた任意の 1 点を原点とした座標系で表されるものとする。なお、マーカーボード 1 1 1 を使った方法以外でも、撮像装置の位置と姿勢が取得できる方法であれば良い。特定された相対位置姿勢は画像 1 1 3 0 に紐づけられる。

20

【 0 0 2 8 】

S 1 2 0 2 では、領域特定部 1 0 2 2 が、物体と異なる色である背景に置かれた物体が撮像された画像の輝度の差を用いて物体の領域を特定する。具体的には、画像 1 1 3 0 に写った物体の領域を表す差分画像 1 1 4 を特定する。さらに、画像 1 1 3 0 に対して差分画像 1 1 4 を使って前景画像 1 1 3 2 を特定する。手順を説明する。まず画像 1 1 3 0 から、背景用途領域 1 1 1 2 の領域だけ取得した部分画像を作る。なお、部分画像は、物体を含み、かつ背景用途領域の外側は含まないようにすればどのように取得しても良い。次に、部分画像から背景用途領域 1 1 1 2 の色と類似度が低い色の領域を物体の領域として特定する。類似度は、あらかじめ取得しておいた背景用途領域 1 1 1 2 の色と、物体の平均の色の RGB 空間で距離の近さによって判定する。背景用途領域 1 1 1 2 は、物体の色とは異なる単一色である領域であるため、画像特徴として輝度を用いて物体の領域を特定できる。さらに、差分画像 1 1 4 を用いて画像 1 1 3 0 から前景画像 1 1 3 2 を特定する。例えば、図 3 ( a ) の画像 1 1 3 0 の部分画像から特定した差分画像 1 1 4 が図 6 ( a ) である。更に、特定した差分画像 1 1 4 を使って画像 1 1 3 0 の物体の領域を特定したものが前景画像 1 1 3 2 である。前景画像 1 1 3 2 は、図 6 ( b ) に示すような物体を 1 つだけ取得した画像である。

30

40

【 0 0 2 9 】

S 1 2 0 3 では、中心位置特定部 1 0 2 3 が、S 1 2 0 1 で求めた相対位置姿勢を使って、S 1 2 0 2 で求めた前景画像 1 1 3 2 ( 物体の領域 ) の中心位置 ( 第 1 の位置 ) を特定する。物体を安定的に吸着するには物体の重心付近を吸着すると良い為、画像上の物体の領域における中心を物体の重心として仮定する。すなわち、この工程では前景画像 1 1 3 2 に対して、吸着に適した位置を特定する。まず、撮像装置の相対位置姿勢から、画像 1 1 3 0 の姿勢グループごとに、撮像装置が基準の位置姿勢である前景画像 1 1 3 2 を選

50

択する。すなわち、相対位置姿勢が撮像装置の視線方向とマーカボード111の法線方向の内積の絶対値が最も大きくなる画像を、代表の前景画像1132として取得する。次に、代表の前景画像1132から中心位置を特定する為に、代表の前景画像1132における中心座標を画像座標系で求める。例えば、代表の前景画像1132に対応する差分画像114の輝度が一定以上の領域（白い領域）の重心を特定する。さらに、代表の前景画像1132における中心座標を、画像座標系からマーカボード座標系に変換する。これによって、2次元から3次元の情報に変換される。このとき、奥行情報（重力方向）の座標は0を与える。なお、部品の厚みに合わせて一定の値を与えても良い。奥行情報にも値を与えることによって、実際の重心に近い場所を吸着位置として学習できる為、より適切な吸着位置を推定できる。以下、この代表の前景画像1132における中心位置を代表中心位置と記載する。代表の前景画像1132における以外の画像1130に対しては、姿勢グループ毎にその姿勢における代表中心位置を、S1201で求めた相対位置姿勢を使ってマーカボード座標系からそれぞれの画像座標系に変換する。その画像座標系を中心位置として特定する。物体の取り得る姿勢のうち出現頻度の高い姿勢を代表画像として準備することによって、推定処理実行時に吸着位置を推定しやすくする。または、物体の姿勢により吸着や把持位置が異なる場合、ある程度姿勢を分類することで明確に物体の姿勢を区別して正しく吸着できるようになる。

10

20

30

40

50

#### 【0030】

図18を使って、ある一定の姿勢で置かれた物体を、撮像装置を様々な位置や姿勢に動かして撮像することによるメリットを説明する。図18における太線86は重なり合った物体群によって形成された山の様子を表している。前提として、ピッキング装置には、重なり合った物体群を上から撮像する撮像装置が備えられていることが多い。また、重なり合った物体群の頂上は平らではなく、図18(a)(b)(c)のように様々な形状を取り得る。そのため、吸着に適した物体は以下に説明するケースがありうる。例えば、図18(a)は、重なり合った物体群が放物線の山の形を成しており、その上に物体82がある。物体82は、重なり合った物体群を矢印81の方向から見て、最も手前にあって、撮像装置に対して平行な姿勢で置かれているため吸着の成功率が高い。図18(b)は、重なり合った物体が放物線の山の形を成しており、その谷底にある物体84が遮蔽されていない様子を示す。物体84は、物体85よりも下方に位置しているが、他の物体によって遮蔽されていないため、吸着に適しているものと考えられる。そのため、山の1番上にある物体が最もピッキングしやすいとは限らない。また、図18(c)は、いくつかの頂点のある山の形に物体が載っている様子を示す。破線80より上にある物体は、矢印81の方向から観察したときに、遮蔽されている部分が少ないと考えられる。そのため複数のピッキングの候補として残しても良い。これらのピッキングに適した物体の候補は、各図における矢印81の方向から観察したときに、それぞれ異なる姿勢の為、物体の領域の見え方が異なる。本実施形態のS1002において、様々な角度から物体を撮像し、実際のピッキングを行う物体の見え方に近い角度で物体を撮像した画像から、図6(b)に示す1つの物体の領域を用意することができる。また、物体の3次元重心は、その物体を2次元の画像に投影した物体の中心と一致するとは限らない。つまり、3次元における物体の重心とそれを投影した画像における物体の領域の中心にはずれが発生し得る。本実施形態では、相対位置姿勢を使うことで、このずれをある程度修正する効果がある。その為、2次元画像から物体の重心、すなわち吸着に適した位置を求めた場合でも吸着位置を精度良く学習できる。なお、本実施形態は撮像装置を一定の位置と姿勢（先に述べた基準の位置姿勢）に固定して撮像された画像があれば実施できる。その際は、S1201では、予め撮像装置の位置と姿勢をキャリブレーションによって取得する。

#### 【0031】

S1003では、物体の領域の画像と重なり合った物体群が写った画像とを合成した画像のうち、物体の領域の中心に吸着位置であることを示す第1のラベルを付与する。また合成画像のうち、物体の領域とその領域の境界周辺に吸着位置とは異なることを示す第2のラベルを付与する。これによって、第1の教師データを生成する。なお、吸着位置は、

物体の領域の中心であるとする。物体の中心は、物体の重心に近い為、吸着に適している。図10のフローチャートに基づき、生成部が実行するS1003の処理の流れを説明する。

#### 【0032】

S10031では、画像生成部1031が、前景画像1132（1つの物体の領域の画像）と画像1131（第2の画像）を合成して合成画像11330（第3の画像）を生成する。画像11330は図11（a）で一例を示すように、画像1131の部分画像の上に前景画像1132を合成した画像である。画像11330は、特定部102で求めた吸着位置の座標（図11の点1100）と、物体の領域にそれぞれ所定の値を対応させた領域マップを紐づけられている。領域マップとは、例えば、吸着位置には1、物体の領域と後述する膨張させた領域には2、その他の背景部分には0といった値を与え、画素に対応させて2次元配列したものである。

10

#### 【0033】

ここでは更に、学習モデルに与える第2の教師データを作成するために、学習モデルを構成するニューラルネットワークにおける入力画像と同じ大きさである部分画像11331（第4の画像）を合成画像11330から生成する。部分画像11331は、第2の教師データの基となる画像である。第2の教師データは、入力された画像に対してピックアップ位置があるか否かの正解または不正解を出力する学習モデルの画像である。すなわち、第2の教師データは、画像11331（第4の画像）の中心にピックアップ位置があるか否かを示す画像である。第2の教師データは、画像11331（第4の画像）の中心に吸着位置（第1の位置）が含まれた場合に、第1のラベルを画像11331の中心に付与される。また、画像11331（第4の画像）の中心が、物体の領域でありかつ中心位置ではない場合または物体の領域の境界周辺に含まれる場合には、第2のラベルを画像11331の中心に付与した第2の教師データが生成される。以下、ニューラルネットワークの入力画像の大きさはN画素×N画素（Nは2の冪）であると仮定する。ニューラルネットワークの入力画像と同じ大きさの部分画像11331を生成する場合、リサイズ等の手間が省ける上、リサイズをした際に発生しうる画像の劣化を抑制することが出来る。なお、部分画像11331の大きさは一定の大きさであれば良い。ニューラルネットワークの入力画像と異なる大きさの部分画像11331を生成する場合は、後述するS1004で学習モデルに教師データを与える際にリサイズする。まず、画像1131から、物体1つが十分に収まり2つ以上収まらない程度の大きさの長方形である部分画像を取得する。なお、この長方形の大きさは、前景画像の物体の端から一定の幅を持たせるようにする。一定の幅とは、ニューラルネットワークの入力画像と同じ大きさのスライディングウィンドウを画像11330に用いた時に、物体の輪郭を中心としたスライディングウィンドウがはみ出さない程度の大きさであるとする。ここでは、画像11330がT画素×S画素（T, Sはそれぞれ自然数で、 $T > S > N$ ）であると仮定する。1枚の画像1131から複数の部分画像を取得することによって効率的に教師データを生成できる。次に、画像1131の上に、前景画像1132を合成して、合成画像11330を生成する。このとき、前景画像1132における物体の中心位置に対応する画像上の座標と、画像1131の部分画像における中心の座標が一致するように、重畳して合成する。更に、合成画像11330の所定の領域を中心に一定の大きさで取得された部分画像11331を生成する。このとき、部分画像11331は、中心座標を、画像11330における領域（ $N/2 < x < (T - N/2)$ ）かつ $N/2 < y < S - N/2$ ）の各画素に移動させて取得する。つまり、N画素×N画素の枠を画像11330の端から端まで1画素ずつスライディングさせて部分画像11331を取得する。これによって、画像1131に写っていた物体が必ず遮蔽されるように画像を生成出来る為、正解と間違いやすい不正解の教師データを精度良く効率的に生成できる。なお、先に画像1131に前景画像1132を重畳した合成画像1133からラベリングに適した大きさにして取得しても良い。

20

30

40

#### 【0034】

S10032では、第1のラベル生成部1032が、前景画像1132の中心位置を用

50

いて、合成画像 1 1 3 3 0 に対して第 1 のラベルを付与した教師データ（第 1 の教師データ）を生成する。この第 1 のラベルは、吸着位置（ピックアップ位置）であることを示す情報である。領域マップを参照して、中心 1 1 0 0 を基点とした幅  $D_0$ （第 1 の幅）の範囲（面積  $S$  の値の平方根の 1 0 パーセントである幅  $D_0$ ）に対して第 1 のラベルを付与する。つまり、合成画像 1 1 3 3 0 の中心の近傍の領域に第 1 のラベルを生成する。

#### 【 0 0 3 5 】

第 2 の教師データは、部分画像 1 1 3 3 1 をスライドさせたときに、部分画像 1 1 3 3 1 における画像中心と、前景画像 1 1 3 2 の中心 1 1 0 0 との距離が所定の値より小さい場合に、部分画像 1 1 3 3 1 の中心に対して第 1 のラベルを付与して生成される。つまり、第 1 のラベルは、合成画像のうち、物体の中心を基点とした所定の第 1 の幅より内側の領域に付与される。図 1 1 を用いて説明する。まず、予め前景画像 1 1 3 2 の物体の領域の面積  $S$  を取得する。面積  $S$  の値に応じた幅  $D_0$ （面積  $S$  の値の平方根の 1 0 パーセントなど）を取得する。 $D_0$  は物体の領域を超えないように第 1 のラベル生成部 1 0 3 2 が決定する（例えば幅  $D_0$  は面積  $S$  の平方根より小さい）。物体の大きさや形状に合わせてユーザーが幅  $D_0$  を設定しても良い。まず、部分画像 1 1 3 3 1 の中心と、前景画像 1 1 3 2 の中心位置（図 1 1 ( a ) の  $x$  印）との距離  $d$  を取得する。距離  $d$  と幅  $D_0$  を比較して、距離  $d$  が幅  $D_0$  よりも小さい場合、その部分画像 1 1 3 3 1 の中心の画素に対して第 1 のラベルを生成する。図 1 1 ( b ) の位置 1 1 0 0 に示される  $x$  印から周囲  $D_0$  の範囲内の領域に部分画像 1 1 3 3 1 の中心が当てはまったときに、部分画像 1 1 3 3 1 の中心の画素に対して第 1 のラベルを付与される。

10

20

#### 【 0 0 3 6 】

なお、第 1 のラベルは、吸着位置である場合は 1、その領域が吸着位置でない場合は 0 をとる 2 値情報である。なお、第 1 のラベルは 2 値情報以外でも、例えば確率を表す数値をもつ情報であっても良い。また、部分画像 1 1 3 3 1 に対して第 1 のラベルを付与したデータを、正解を示す第 2 の教師データとする。

#### 【 0 0 3 7 】

$S_1 0 0 3 3$  では、第 2 のラベル生成部 1 0 3 3 が、合成画像 1 1 3 3 0 と前景画像 1 1 3 2 に基づいて、物体の領域と物体を膨張させた領域（境界周辺）に、吸着位置（ピックアップ位置）ではないことを表す第 2 のラベルを付与した第 1 の教師データを生成する。図 1 1 ( b ) を使って説明する。まず、前景画像 1 1 3 2 の領域でかつ、 $S_1 0 0 3 2$  で第 1 のラベルを付与されていない領域に対して第 2 のラベルを付与する。すなわち、第 2 のラベルは、合成画像のうち、物体領域の中心位置を基点とした所定の第 2 の幅より外側に対して付与される。判断する方法は、例えば、 $S_1 0 0 3 1$  で合成画像 1 1 3 3 0 の各領域に対応させて物体の領域を表す領域マップを生成しておき、そのマップを参照する。 $S_1 0 0 3 2$  において、前景画像 1 1 3 2 と画像 1 1 3 3 0 の中心が一致するように合成しているため、前景画像 1 1 3 2 の領域は画像 1 1 3 3 0 の画像座標系で表現できる。物体の中心以外の領域に「吸着位置ではない」ことを示すラベルを付与することによって、吸着位置をより正確に学習させることが出来る。

30

#### 【 0 0 3 8 】

第 2 の教師データは、部分画像 1 1 3 3 1 をスライドさせたときに、前述の距離  $d$  が幅  $D_0$  よりも大きく、かつ部分画像 1 1 3 3 1 の中心が前景画像 1 1 3 2 の領域である場合に、部分画像 1 1 3 3 1 に第 2 のラベルを付与して作成される。部分画像 1 1 3 3 1 の中心座標も、画像 1 1 3 3 0 の画像座標系で表現できるため、前景画像 1 1 3 2 の領域に部分画像 1 1 3 3 1 の中心座標が含まれているか否かを求めればよい。

40

#### 【 0 0 3 9 】

なお、第 1 のラベルと第 2 のラベルが近接しないようにラベルを付与しても良い。つまり、物体の中心 1 1 0 0 を基点とした幅  $D_1$ （第 2 の幅）だけ離れた領域に対して第 2 のラベルを付与する。ただし、 $D_0 < D_1$  である。例えば、幅  $D_1$  は面積  $S$  の値の平方根の 1 5 パーセントなどと設定する。このような領域を設定することによって、中心位置の周辺の間違いやすい領域に正解が存在しないことを学習できるため、推定精度が向上する効

50

果がある。

【 0 0 4 0 】

次に、前景画像 1 1 3 2 を物体の中心から見て外側に膨張した境界周辺の幅に対して第 2 のラベルを付与した第 1 の教師データを生成する。境界周辺の幅は、物体の中心から物体の境界の点までの距離より小さい幅である。つまり、第 2 のラベルは、合成画像のうち、物体の境界周辺であって物体の領域の外側の領域に付与される。図 1 1 b の外側の斜線領域が境界周辺の幅である。この第 2 のラベルは、吸着位置ではないことを示す情報である。幅  $D_2$  (境界周辺の幅) だけ膨張させた物体の領域は画像 1 1 3 3 0 の画像座標系で表現できる。画像 1 1 3 3 0 の画像座標系で表現された部分画像 1 1 3 3 1 の中心座標が、幅  $D_2$  だけ膨張させた物体の領域に含まれているか否かを求める。すなわち、合成画像 1 1 3 3 0 のうち、物体の領域 1 1 3 2 を基点として膨張させた幅  $D_2$  である周辺領域に含まれ、かつ物体の領域には含まれない合成画像 1 1 3 3 0 の領域に対して、第 2 のラベルを生成する。第 2 の教師データは、部分画像 1 1 3 3 1 をスライドさせたときに中心が前述の領域マップにおける膨張させた領域である場合に、部分画像 1 1 3 3 1 に第 2 のラベルを付与して作成する。また、物体の領域の境界周辺に部分画像 1 1 3 3 1 の中心が含まれる場合に、部分画像 1 1 3 3 1 に対して第 2 のラベルを付与されたデータを不正解の第 2 の教師データとして生成する。

10

【 0 0 4 1 】

図 1 1 に示す幅  $D_2$  (膨張させた幅) は、2 つの物体を外接させたときに 2 つの物体の中心を結んだ距離の半分より小さくなるようにする。これは、本来正解とすべき中心が存在する領域に不正解を示すラベルを生成しないようにするためである。例えば、物体が半径  $r$  の円柱形である場合は、 $D_2 < r$  とする。上記条件を満たす幅にすることによって、正解の吸着位置に不正解のラベルを付与することを抑制できる。例えば、中心が遮蔽されていないが遮蔽された領域が大きいため、ピックアップできない物体があるケースを考える。上にある物体を A、下にある物体 (遮蔽された物体) を B とする。このケースは、物体 B の遮蔽された領域がある程度大きい物体 B の中心は物体 A によって遮蔽されていない場合に起こり得る。物体の中心をすべからく推定するような学習モデルは、遮蔽された物体の吸着位置 (中心) を推定するので、物体 B の中心が露出していればそこに中心があるものだと推定する。しかし、ピックアップする場合は、物体 B が物体 A に遮蔽されている領域が大きいくらい吸着に失敗しやすくなる。そこで、物体 A が物体 B の上にあることを情報として利用して、物体 B の中心に吸着位置ではないことを示すラベルを付与する。そうすることによって、学習モデルは遮蔽された物体には吸着位置がないことを学習できる。また、遮蔽がある部分を広くとりすぎたためにピックアップ可能な物体を検出しないケースを考える。物体 A と物体 B の中心同士が、物体 1 つぶんほど、十分に離れている場合は、物体 A の下にある物体 B の中心も吸着可能である可能性が高い。そのため、膨張する幅  $D_2$  はあまり大きすぎると学習に悪影響を与える。そこで、幅  $D_2$  の目安として、物体の外周を基点とした幅において、隣接した 2 つ物体の中心を結んだ幅に物体 1 つが入らない大きさにする。例えば、幅  $D_2$  の大きさは、前景画像 1 1 3 2 の物体の領域の面積  $S$  の値に応じた値 (面積の値の平方根の 10 パーセントなど) にする。以上のように、適切な幅  $D_2$  だけ膨張させた領域に吸着位置でないことを示す第 2 のラベルを設定することによって、吸着位置をより精度良く学習できるようになる。なお、第 2 のラベルは中心付近 (例えば物体の中心から幅  $D_0$  または幅  $D_1$  の範囲) には付与しないようにする。これによって、中心付近に吸着位置がないことを誤って学習することを抑制する。なお、幅  $D_2$  は物体の形状によって設定する。例えば、円形の物体であれば円の半径に応じて幅  $D_2$  を決定する。直方体の物体であれば、上面の長方形の短辺に基づいて幅  $D_2$  を決定する。なお、物体の輪郭の任意の一点からの距離が幅  $D_2$  より小さい領域に対して物体の周辺領域を設定しても良い。

20

30

40

【 0 0 4 2 】

なお、第 2 のラベルは、吸着位置でない場合は 2、その領域が物体の領域または物体を膨張させた領域でない場合は 0 をとる 2 値情報である。なお、第 1 のラベルは 2 値情報以

50

外でも、例えば確率を表す数値をもつ情報であっても良い。

【0043】

第1のラベルまたは第2のラベルが付与されなかった領域については、特にラベルを与えない、または0というクラス値を付与する。取得した背景画像はランダムに選択するため、重なり合った物体群の1番上にくる物体や遮蔽された領域の少ない物体が含まれる可能性がある。それらに対して吸着位置でないことを示すラベルを付与すると、推定処理時に実際の吸着位置を推定できなくなる可能性がある。前景となる物体の周囲以外にラベルを付与しないことによって、遮蔽された領域が少ない物体を検出するようにでき、効率的に教師データを作成できる。

【0044】

ここで説明した第1の教師データは、合成画像11330の各画素に対応させてクラス値を2次元配列にした画像である。第1の教師データは、例えば、画像のセグメンテーションを行う学習モデルの場合に有効である。第2の教師データは、部分画像11331の中心の画素に対して、第1のラベルまたは第2のラベルを付与した画像である。この例以外にも、教師データのサイズは、学習モデルに合わせて調整して良い。第2の教師データは、第1の教師データに基づいて作成しても良い。つまり、部分画像11331に対して合成画像11330の各画素のクラスを参照して第1または第2のラベルを付与することで、第2の教師データを生成する。

【0045】

なお、S1003の手順は説明した順番でなくても実行できる。例えば、先に吸着位置でないことを示す第2のラベルを付与しても良い。また、所定の領域に教師ラベルを付与しないようにしても良い。例えば、吸着位置を表す第1のラベルと、吸着位置でないことを表す第2のラベルとの境界部分に対応する領域に教師ラベルに付与することを抑制する。つまり、中心位置の近傍にラベルを付与しない領域をつくる。具体的に手順を説明する。まず、前景画像1132の領域とその領域を幅D2だけ膨張させた周辺に第2のラベルを付与する。次に、前景画像1132の中心1100を基点とした一定の幅D1（第3の幅）の領域に対して、第2のラベルを削除する。このとき、幅D0 < 幅D1であり、幅D1も幅D0と同様に面積Sの平方根の20パーセントと決める。更に、前景画像1141の中心1100を基点とした一定の幅D0（第4の幅）の領域に対して、改めて第1のラベルを付与する。ある正解の周りに、ラベルを与えない処理は、点あるいは小さい領域に対して付与された正解の教師ラベルが学習過程において再現されにくくなるというニューラルネットワークの性質に対して有効な処理である。

【0046】

S1004では、更新部104が、教師データに基づいて重なり合った物体群を撮像した入力画像に対する吸着位置を出力する学習モデルのパラメータを更新する。つまり、生成部103で生成した第2の教師データに基づいて推定部105が保持する学習モデルのパラメータの更新をする。このパラメータは、推定処理実行時に学習モデルのクラス分類処理の閾値を表現するために用いる。まず、教師データとして、吸着位置であることを表す第1のラベルが付与された第2の教師データと、吸着位置でないことを表す第2のラベルが付与された第2の教師データを同じ数だけ選択する。それらに対応した第2の教師データをランダムに選択する。ラベルが2種類の場合、その出現数が等しい場合が最もラベルの違いを比較および修正する場面が多くなるため、学習に適している。なお、データの数は必ずしも同数である必要はない。これらの教師データを用いて、ニューラルネットワークのパラメータを更新する。

【0047】

以上の工程を行うことにより、学習時に必要な教師データを効率的に自動的に生成することが可能になる。なお、必ずしも、説明した順序のみで実行することを限定するものではない。例えば、S1201とS1202は並列的に処理しても良い。続いて、学習モデルを使った推定処理を実行するとき、画像処理装置10が実行する処理手順について図8(b)を参照して説明する。

10

20

30

40

50

## 【 0 0 4 8 】

S 1 1 0 1では、取得部 1 0 1が、重なり合った物体群を映した被推定画像を取得する。S 1 1 0 2では、推定部 1 0 5が、被推定画像にある物体の吸着位置を推定する。被推定画像の物体の吸着位置を推定する領域のすべての画素に対して、ニューラルネットワークの入力画像と同じ大きさの被推定画像の部分画像を生成する。被推定画像のすべての部分画像に対して、推定部 1 0 5が保持するニューラルネットワークで推定を行い、推定結果（吸着位置であるか否か、物体の姿勢のクラス）を元の被推定画像と同じ座標にマッピングする。吸着位置をマッピングした結果が推定結果である。推定結果は出力装置 1 2に出力される。なお、被推定画像の部分画像をニューラルネットワークの大きさに合わせる以外の方法でも良い。例えば、被推定画像を拡大縮小して大きさを調整しても良い。

10

## 【 0 0 4 9 】

重なり合った物体群を撮像した画像を教師データ生成に用いることで、物体の姿勢が多様に再現される。その結果、多様に遮蔽された物体の不正解の教師データを表現できるため、遮蔽された対象の誤検出率を下げるができる。

## 【 0 0 5 0 】

画像は、カラー画像である例を示したが、物体の位置を推定するための情報が十分に得られるものであれば何でも良い。例えば、グレー画像や距離画像でも良い。歪補正やガンマ補正などの画像処理を行った画像を用いても良い。赤外画像などの画像でも良い。

## 【 0 0 5 1 】

領域特定部 1 0 2 2で画像 1 1 3 0から差分画像 1 1 4を生成する方法は、RGB空間での距離によって類似度を判定する例を示したが、物体の領域を算出できる適切な類似度なら何でも良い。例えば、HSV空間（Hue、Saturation、Lightness空間）の距離を利用しても良い。輝度の差から類似度を求めても良い。また、例えば、背景差分を用いて領域を特定しても良い。

20

## 【 0 0 5 2 】

中心位置特定部 1 0 2 3で中心位置を特定する方法は、カメラの視線方向とマーカードの法線方向の一致度が高い代表画像から算出する例を示したが、吸着に適した位置または領域を求める方法であれば他の方法でも良い。マーカードに複数回部品を乗せる場合、それぞれを異なる姿勢として中心位置を求めても良い。姿勢のグループはユーザーが予め準備する以外にも、例えば、画像 1 1 3 0における物体の姿勢の平均に最も近いもので分類するようにしても良い。姿勢検出の分解能を上げるために複数の代表画像を決定しても良い。より多様なパリエーションの物体の姿勢において、それぞれ中心位置を特定することによって、推定精度が向上する。

30

## 【 0 0 5 3 】

中心位置特定部 1 0 2 3で中心位置を特定する方法は、前景画像 1 1 3 2の画像重心に対応する3次元座標である例を示したが、物体に対して幾何的な処理を行う時に用いる原点を決めることができれば何でも良い。例えば、あらかじめマーカード 1 1 1上の特定の座標を決めておき画像に投影した点を用いても良い。

## 【 0 0 5 4 】

画像生成部 1 0 3 1で、合成画像 1 1 3 3 0から生成する部分画像 1 1 3 3 1は、学習モデルを構成するニューラルネットワークにおける入力画像と同じ大きさであるという説明をしたが、必ずしも同じ大きさでなくても良い。同じ大きさである場合、処理の手間が省ける為、効率的に教師データを生成できる。一方で、部分画像 1 1 3 3 1は一定の大きさの画像を揃えることが出来ればその大きさは何でもよい。他の大きさの部分画像 1 1 3 3 1を用いる場合は、更新部 1 0 4に教師ラベルを付与した画像 1 1 3 3 1を入力する際に、画像 1 1 3 3 1をニューラルネットワークにおける入力画像の大きさにリサイズする。リサイズする方法は、拡大、縮小、変形等の公知の画像処理方法を用いればよい。

40

## 【 0 0 5 5 】

第1のラベル生成部 1 0 3 2は、吸着位置か否かの2値のラベルを生成する例を示したが、物体の位置または姿勢を求めることができれば3値以上のラベルを生成しても良い。

50

例えば、中心位置特定部 1023 で複数の姿勢から代表の前景画像を選択し、中心位置であることに加え、その代表の前景画像における物体の姿勢を示すラベルを生成しても良い。物体の表と裏のそれぞれの画像 1130 を取得しそれぞれを区別する教師ラベルを生成しても良い。

#### 【0056】

第1のラベル生成部 1032 の第1のラベルの付与は、部分画像 11331 の中心を基点とした所定の幅に対して生成する例を示したが、物体の位置または姿勢の推定に必要な情報を得られる教師データを生成できれば何でも良い。例えば、幅  $D_0$  を距離画像の値に基づいて設定しても良い。これは、例えば、重なり合った物体群の上方に距離センサが設置されている場合に、距離センサから見て奥にある物体に対しては、幅  $D_0$  を小さく与える。また、距離センサから見て手前にある物体に対しては、幅  $D_0$  を大きく与える。距離画像が得られない場合でも、同じ大きさの物体が撮像された画像から、大きく映っている物体には大きめの幅  $D_0$  を与えるようにしても良い。これによって、より安定した教師データを作成できる。差分画像 114 の画像重心を中心位置としても良い。領域の先端などの特徴部位に特徴であることを示すラベルを付与しても良い。領域を収縮して得られる領域に対して物体の内部であることを表すラベルを生成しても良い。例えば、直方体の物体は、重心とその両端の見え方に変化がない場合、重心が含まれる画像とそうでない画像を見分けるのが難しい。そこで、長方形の物体の領域に対して、吸着位置の領域を物体の領域を内側に  $W\%$  だけ縮小させる。この縮小した領域には重心が含まれる可能性が高い。この方法は、特に細長い形状の物体に有効である。これによって、より安定した教師データを作成できる。

#### 【0057】

第2のラベル生成部 1033 は、差分画像 114 の領域の境界部分の外側に、面積  $S$  に応じた幅だけ膨張した領域に吸着位置でないことを表す第2のラベルを付与する例を示したが、遮蔽された物体を表現できるように第2のラベルが設定されれば何でも良い。例えば、画像 11330 の値の局所的な周波数によってその近傍の教師ラベルの生成する領域の幅を調整しても良い。周波数が高い、すなわち、物体が混み合って重なっている領域については、あまり判別が難しくないので、幅  $D_2$  を与えなくても良い。一方で、遮蔽されている物体側における画像特徴の変化が乏しい領域には幅  $D_2$  を拡大して与えると良い。遮蔽された物体でも重なりから離れた領域に正解があるという不安定な教師データを作成することを抑制できる。また、物体が影をつくる領域に応じて、第2のラベルを与える幅を設定しても良い。遮蔽する領域が大きい形状の物体である場合は、その近傍の教師ラベルの生成する領域の幅を大きめに調整しても良い。遮蔽する領域が小さい形状の物体である場合は、その近傍の教師ラベルの生成する領域の幅を小さめに調整しても良い。

#### 【0058】

更新部 104 は、推定部 105 の保持するパラメータを更新できる方法であればどの方法を用いても良い。例えば、非特許文献3のような、学習係数を自動的に決定する方法を用いてパラメータを更新しても良い。非特許文献3では、パラメータ更新のときに微分で決定した勾配方向の1次モーメントと2次モーメントを用いる。この比を使って更新幅を決定する。更新幅を自動で決定できるため、効率的な学習を行える。

#### 【0059】

推定部 105 が保持する学習モデルであるクラス分類器の入力画像の大きさは、定数の大きさの正方形の画像である例を示したが、長方形や三角形等でも良い。例えば、物体が画像上で現れるときの縦横比があらかじめ分かっている場合などはその縦横比の長方形にしても良い。また、推定部 105 が保持する学習モデル(クラス分類器)は、ニューラルネットワークによるものである例を示したが、教師データを用いて物体の位置の検出が行えるものであれば何でも良い。例えば、サポートベクターマシンによるものでも良い。ランダムイズドフォレストによるものでも良い。入力画像ごとに中心位置とのオフセットを推定する回帰推定を行っても良い。

#### 【0060】

10

20

30

40

50

推定部 105 における位置推定方法は、クラス分類結果の面積の固定閾値である例を示したが、確からしい吸着位置を求められる方法であれば何でも良い。例えば、吸着位置であることを表す領域の形状により計算しても良い。吸着位置であることを表す領域の周囲が吸着位置でないことを確認して計算しても良い。クラス分類結果の尤度を反映しても良い。輝度やスケールなどの複数の条件で実行してそれらを組み合わせた結果を用いて推定しても良い。

#### 【0061】

物体の画像を準備するときにマーカーボードを使う方法を述べたが、マーカーボードを使わないで物体の位置を取得できる方法であれば何でも良い。マーカーボードを使う場合は、人が手でカメラを持って撮像する為、簡単な設備で画像を用意出来る。他の方法として、例えば、ロボットアームに撮像装置を装着したのを使う方法がある。予めロボットと撮像装置の位置と姿勢をキャリブレーションしておく。この撮像装置を使って物体を撮像すればカメラパラメータを使って物体の位置をカメラ座標から世界座標に変換して求めることができる。これによって実際の環境に近い画像を使って教師データを生成できるため、効率的に画像を用意出来る。

#### 【0062】

##### (変形例1)

第1の実施形態では、重なり合った物体群を映した画像を背景として教師データを生成する例を述べた。しかし、重なり合った物体が写った画像における物体が前景にあるような画像を作成して、遮蔽された物体の学習を行っても良い。つまり、前景画像に映っている物体が、重なり合った物体が写った画像における物体によって遮蔽されているように画像を合成する。重なり合った物体群物体画像に映っている様々な姿勢の物体を利用できるので、効率的に間違いやすい不正解を示す教師データを揃えることが出来る。ここでは、合成画像(第3の画像)は、前景画像の部分画像(物体の領域の一部)と(第2の画像)とを合成した画像である。そして、生成部103は、物体の領域の境界周辺であって境界の内側の領域に、第2のラベルを付与した第1の教師データを生成する。そのために、第1の実施形態の学習実行時において、画像処理装置が実行するS10031、S10032、S10033に、以下の処理を加える。

#### 【0063】

S10031において、画像生成部1031が、画像11310の上に、前景画像1132の一部である部分画像11321を重畳して合成画像1135を生成する。図12を使って説明する。まず、図12(a)で図示されるような画像11310の画像に対してエッジ検出を行う。つまり、重なり合った物体群が写った画像(第2の画像)の画像特徴に基づいて物体の領域と背景との境界を検出する。エッジ検出では、輝度や距離値などの画像特徴が非連続な部分を検出することで、物体の境界部分を検出する。エッジ検出結果のうち、エッジが連続して現れている部分を選択し、その部分を背景境界1134とする。背景境界1134は枝分かれしない曲線または直線となるように選択する。背景境界1134は、画像11310において、物体の下にある背景が表出している境界部分を指す。画像11310は、物体がある程度まばらに配置された状態である画像が望ましい。ここでは、図12(a)の物体2が重なり合った物体のうち、一番上にあるものとして考える。次に、前景画像1132が背景境界1134によって2つの領域に分割されるように取得する。前景画像1132から背景境界1134によって分割された2つの領域のうち、片方を選択したものを物体の部分画像11321とする。物体の部分画像11321は、図12(b)で示すように、画像11310における物体2と重なる部分が少ない方を選択する。選択する方法は、例えば、距離画像を取得できる場合は、距離センサから見て奥に存在する領域に対応する部分画像を選択すれば良い。選択された物体の部分画像11321と画像11310は、背景境界1134に合わせるようにして合成する。合成画像1135は、物体の部分画像11321を前景として合成しているが、画像11310における物体2が部分画像11321の上に置かれているものとして考える。

#### 【0064】

10

20

30

40

50

変形例 1 においては、S 1 0 0 3 2 はスキップする。すなわち、合成画像 1 1 3 5 に対しては第 1 のラベル生成部 1 0 3 2 が第 1 のラベルを生成しない。変形例 1 においては、明らかに吸着に適していると判断できる領域が画像 1 1 3 5 に存在しないためである。

#### 【 0 0 6 5 】

S 1 0 0 3 3 において、第 2 のラベル生成部 1 0 3 3 が、部分画像 1 1 3 2 1 の領域の境界周辺であって部分画像 1 1 3 2 1 の内側の領域に第 2 のラベルを付与した教師データを生成する。さらに、合成画像 1 1 3 5 の背景境界 1 1 3 4 から所定の値 D 3 より近傍の領域でかつ部分画像 1 1 3 2 1 を重畳した領域に、吸着位置でないことを表す第 2 のラベルを付与した教師データを生成する。領域 D 3 は、図 1 2 ( c ) における斜線部の領域で示す。但し、領域 D 3 は、物体の面積 S の平方根の 1 0 % といったふうに決定する。領域 D 3 は、ユーザーが任意の値を決定しても良い。この処理によって、不正解の教師データを多く生成できる。なお、部分画像 1 1 3 2 1 の境界内側にラベルを付与せず、背景境界 1 1 3 4 の部分に対してのみラベルを付与しても良い。また、部分画像 1 1 3 2 1 がかなり遮蔽されている場合は、部分画像 1 1 3 2 1 の領域全体に第 2 のラベルを付与しても良い。また、部分画像 1 1 3 2 1 があまり遮蔽されていない場合は、部分画像 1 1 3 2 1 の境界周辺であって内側の領域に対してのみ第 2 のラベルを付与しても良い。部分画像 1 1 3 2 1 が、例えば、物体の領域の 8 0 % 以上の大きさである場合は、吸着位置が遮蔽されていない可能性が高いため、第 2 のラベルを付与することを抑制する。この方法によってより精度良く学習できる。

10

#### 【 0 0 6 6 】

この機能構成を有する画像処理装置では、ディープラーニングにおける不正解とすべき教師データを効率的に生成できる。

20

#### 【 0 0 6 7 】

( 変形例 2 )

変形例 2 では、物体の領域を画像の差分により特定した差分画像を用いて、教師データの生成を行う画像処理装置について説明する。ここで用いる画像の例を図 1 3 に示す。画像 1 1 3 0 1 と画像 1 1 3 0 2 は、整列して置かれた物体を 1 つずつ人またはピッキング装置が取っていくタスクを実行する過程を撮像した画像である。図 1 3 ( a ) の画像 1 1 3 0 1 は物体 1 を取る前の画像である。図 1 3 ( a ) の画像 1 1 3 0 2 は物体 1 を取った後の画像である。画像 1 1 3 0 1 は物体 1 を取る前の画像である。図 1 3 の画像 1 1 3 0 2 は物体 1 を取った後の画像である。本変形例では、所定の動作の前後で撮像した 2 枚以上の画像から、輝度等の画像特徴の変化のあった部分のみを表した画像を用いる。これによって、マーカーボード等の道具を用いなくても、簡易な準備で効率的に教師データを生成できる。

30

#### 【 0 0 6 8 】

図 1 4 に基づいて本変形例における特定部 1 0 2 の機能構成例の差分を説明する。特定部 1 0 2 は、複数の物体群が写った画像と、物体群のうち重なり合っていない物体を 1 つだけを取り除いた後に撮像された画像との画像特徴の差に基づいて 1 つの物体の領域を特定する。特定部 1 0 2 は、差分特定部 1 0 2 3 と中心位置特定部 1 0 2 4 から構成される。差分特定部 1 0 2 3 は、複数の画像 1 1 3 1 の差分を特定し、この差分から差分画像を特定する。差分の算出方法は、物体の領域を算出するために必要な情報が得られれば何でも良い。例えば、ノイズの影響を避けるためにガウスフィルタを適用してから差分を算出しても良い。複数の画像 1 1 3 1 を取得して平均した画像から画像 1 1 3 0 を引いても良い。差分画像を特定するときは、画像 1 1 7 の値の絶対値を持つ列に対し 2 値化処理を行う。中心位置特定部 1 0 2 4 が、画像 1 1 7 から物体の領域の中心位置を特定する。

40

#### 【 0 0 6 9 】

前述の実施形態と差がある部分について、図 8 のフローチャートに基づき画像処理装置 1 0 が実行する処理の流れを説明する。S 1 0 0 1 では、取得部 1 0 1 が、物体が写っている画像 1 1 3 0 1 と画像 1 1 3 0 2 を取得する。続いて、図 1 5 のフローチャートに基づき特定部 1 0 2 が実行する S 1 0 0 2 の処理の流れを説明する。

50

## 【 0 0 7 0 】

S 2 2 0 1では、差分特定部 1 0 2 3が、複数の物体が写った画像と、物体のうち重なり合っていない物体を1つだけを取り除いた後に撮像された画像との画像特徴の差に基づいて物体の領域を特定する。図 1 3に基づいて説明すると、画像 1 1 3 0 1と画像 1 1 3 0 2の差分を特定し、物体 1の領域の差分画像を特定する。すなわち、図 1 3 ( b )に示す差分画像 1 1 7は、画像 1 1 3 0 1の輝度値から画像 1 1 3 0 2の輝度値を引いた値から特定される。差分画像 1 1 7は、画像特徴として前述の差分の値(輝度等)を保持する2次元のデータ列である。このように物体 1の領域部分の輝度のみが変化するため、画像 1 1 7では物体 1の領域が浮き上がる。画像 1 1 7を特定するときは、画像 1 1 7の値の絶対値を持つ列に対し2値化処理を行う。2値化処理の方法は非特許文献 2などを用いる。非特許文献 2の方法では、値のヒストグラムから2値化処理に適切な閾値を決定して2値化する。2値化処理の結果、撮像条件における想定される物体の面積と同程度の領域それぞれに対して、その領域を画像 1 1 7として生成する。なお、2値化処理の方法は、物体の領域を適切に推定できる方法なら何でも良い。例えば、事前に他の画像で2値化処理を試しておき適切な閾値を算出して固定閾値として利用しても良いし、画像の平均の輝度の値によって閾値をスケールしても良い。

10

## 【 0 0 7 1 】

S 2 2 0 2では、中心位置特定部 1 0 2 4が、画像 1 1 7における物体の領域の中心位置を特定する。画像 1 1 7の輝度が一定の値より大きい部分が物体の領域にあたるので、物体の領域の重心を中心位置とする。画像 1 1 7において画像座標から世界座標系に変換した位置を物体 1の中心位置とする。

20

## 【 0 0 7 2 】

この画像処理装置によって、マーカーボード等の道具を用いなくても、簡易な準備で効率的に教師データを生成できる。

## 【 0 0 7 3 】

( 変形例 3 )

変形例 3では、物体の3次元形状モデルが利用可能な場合において、重なり合った物体が写った画像から教師データの生成を行う画像処理装置について説明する。3次元形状モデルとは、物体の3次元形状を数値化して表現したデータで、例えば物体の3D-CADデータである。3次元形状モデルが入手可能である場合は、物体の姿勢についての情報をより精度よく捉える事が出来る。そのため、より適切な吸着位置を推定するためのデータを効率的に生成することが出来る。また、物体の姿勢によって吸着に適した位置が重心から少しずれる場合があっても、安定して吸着位置を特定することが出来る。図 1 6に基づいて特定部 1 0 2の機能構成例を説明する。

30

## 【 0 0 7 4 】

マッチング部 1 0 2 5は、物体の3次元形状モデルと画像 1 1 3 1 ( 第 2 の画像 )における前記物体とを照合する。そして、物体画像における遮蔽されていない物体の位置及び姿勢を特定する。さらに、特定された物体と3次元形状モデルを利用して物体の領域を示す差分画像を特定する。具体的な方法については後述する。マッチング部 1 0 2 5は、あらかじめ対象となる物体の3次元形状モデルを視点ごとに分割して保持する。3次元形状モデルは、複数の視点から見た物体の表面点の3次元座標とエッジとなる点の3次元座標を有している。視点は、例えば、正二十面体の重心が原点と一致するように配置されているとき、各面の中心を通るベクトルに倣う視点の集合を選ぶ。記憶装置等から取得しても良い。また、あらかじめ、画像 1 1 3 1を取得したカメラの位置と姿勢を示すカメラパラメータを取得しておく。なお、3次元形状モデルは、3次元形状を表すボクセルであっても良い。この場合、マッチング時に投影結果を算出しても良い。3次元形状モデルは、平面や球面などのプリミティブ形状に分割して保持されても良い。

40

## 【 0 0 7 5 】

中心位置特定部 1 0 2 6は、任意の姿勢である3次元形状モデルと画像 1 1 3 1における物体の姿勢が一致した場合、3次元形状モデルを画像 1 1 3 1に投影して3次元形状モ

50

デルの中心位置を特定する。つまり、差分画像の中心位置を特定する。S 3 2 0 1 で用いた 3 次元形状モデルの重心位置を中心位置として特定する。

【 0 0 7 6 】

図 7 に基づいて、前述までの生成部 1 0 3 の機能構成例と差がある部分のみ説明する。画像生成部 1 0 3 1 は、画像 1 1 3 1 から、マッチング部 1 0 2 5 で特定した物体を含むように部分画像を生成する。

【 0 0 7 7 】

図 8 ( a ) のフローチャートと図 1 7 のフローチャートに基づいて画像処理装置が行う処理の流れについて説明する。図 8 ( a ) の処理については差がある部分についてのみ説明する。

【 0 0 7 8 】

S 3 2 0 1 では、マッチング部 1 0 2 5 が、物体の 3 次元形状モデルがうまくあてはまる物体を画像 1 1 3 1 から特定する。そして、特定された物体と 3 次元形状モデルを利用して物体の領域の差分画像を特定する。画像 1 1 3 1 から特定された物体について、3 次元形状モデルに基づいてその物体の位置及び姿勢を特定する。予め、画像 1 1 3 1 はエッジ検出を行っておく。これによって、画像 1 1 3 1 に写っている物体の輪郭が検出される。まず、この 3 次元形状モデルを用いて、画像 1 1 3 1 から輪郭が一致する物体を 1 つ特定する。非特許文献 4 の方法によりマッチングを行う。非特許文献 4 の方法では、カラー画像上のエッジと距離画像上の点と 3 次元モデルとして保持している輪郭及び 3 次元点を、既知の初期位置姿勢を基に対応を決定し、初期位置姿勢からの勾配法による位置姿勢の推定を行う。初期位置姿勢付近の濃淡画像と距離画像と 3 次元形状モデルの間で尤もらしさが高くなる 3 次元形状モデルの位置姿勢を推定する。初期位置姿勢は、位置は奥行を示す距離画像における平均値と、画像座標における物体が存在する領域のうちランダムな座標とを与え、姿勢はランダムな値を与える。複数回マッチング（最近傍探索）を行い、推定した位置姿勢から 3 次元形状モデルを画像 1 1 3 1 の画像に投影したときの各画像特徴の誤差が適切な閾値以下になる画像 1 1 3 1 に写っている物体をマッチング結果として取得する。閾値は、事前にマッチングを行い、マッチングがうまくいった場合とマッチングがうまくいかなかった場合をサンプリングしておき、それぞれの場合の再投影誤差の平均値の中間の値を利用する。閾値には任意の値をユーザーが設定しても良い。次に、取得した物体に最も姿勢が近い 3 次元形状モデルの輪郭部分の 3 次元座標を画像 1 1 3 1 に投影する。投影した点が囲む領域を差分画像として取得する。ここで、投影した 3 次元形状モデルの重心位置も取得する。なお、初期位置姿勢は、ランダムな値を与える例を示したが、適切なマッチング結果を得られる方法であれば何でも良い。例えば、あらかじめユーザーが初期位置姿勢を手動で指定しても良いし、テンプレートマッチングなどで自動的な検出を行っても良い。この方法は、2 次元と 3 次元の両方の情報を用いる為、位置姿勢を決定しにくい場合にも対応できる。3 次元形状モデルを適切に画像 1 1 3 1 に投影し、その領域を求められるものであればどのような方法でも良い。例えば、輪郭の情報を用いてエッジの画像上の距離を最小化するように位置及び姿勢を算出しても良い。複数の平面部分が距離画像とあてはまるように位置及び姿勢を算出しても良い。

【 0 0 7 9 】

S 3 2 0 2 では、中心位置特定部 1 0 2 6 が、差分画像の中心位置を特定する。差分画像の中心位置の座標は、S 3 2 0 1 で用いた 3 次元形状モデルの重心位置を中心位置として特定する。中心位置の座標は、3 次元形状モデルを投影した 2 次元画像に対応する重心座標を画像座標系に変換しても良い。他にも、S 3 2 0 1 でマッチングするとき、3 次元物体が 1 つだけ収まる 3 次元のバウンディングボックスを生成し、このバウンディングボックスの重心を物体の重心として中心位置を特定しても良い。

【 0 0 8 0 】

S 1 0 0 3 1 では、画像生成部 1 0 3 1 が、画像 1 1 3 1 から、S 3 2 0 1 で特定した物体を含むように部分画像を生成する。画像 1 1 3 1 から、対応する差分画像の中心位置を中心として、物体の領域が 1 つ分だけ十分に収まる程度の長方形の領域を取得すること

10

20

30

40

50

で部分画像を生成する。長方形の領域の大きさについての条件は前述の S 1 0 0 3 1 と同様である。すなわち、この部分画像は、第 1 の実施形態における合成画像 1 1 3 3 0 に相当する。

【 0 0 8 1 】

この変形例によって、重なり合った物体群された画像から遮蔽されていない物体を特定する為、効率的に教師データを作成することができる。

【 0 0 8 2 】

本発明により、ディープラーニングにおいて不正解とすべき教師データを効率的に生成できる。

【 0 0 8 3 】

(他の変形例)

以上の各実施形態において、ピックアップ装置は吸着式でなくても良い。例えば、吊り下げ式のピックアップの場合は、物体の中心や穴のある位置を推定する学習モデルに第 1 の実施形態の方法を適用できる。吊り下げ式のピックアップ装置の場合は、ハンドの種類や大きさやピックアップ対象の物体の情報と組み合わせて、ハンドの向きや物体の位置姿勢を推定する学習モデルを用いればよい。ハンド部分がフックになっているピックアップ装置に第 1 の実施形態の方法を適用した例を述べる。フックの大きさや向きは予め計測されている。また、ピックアップ対象である物体には、穴または輪状の構造があり、画像処理装置はそれぞれの画像特徴をデータとして保持しているものとする。図 8 ( a ) のフローチャートを用いて具体的な処理を説明する。第 1 の実施形態と同様の処理については説明を省略し、  
20  
差がある部分について説明する。S 1 2 0 3 では、中心位置特定部 1 0 2 3 が、S 1 2 0 2 で求めた前景画像 1 1 3 2 (物体の領域) のピックアップ位置を特定する。まず、物体の穴をフックでひっかけて吊り下げる場合について説明する。このとき、物体の穴の位置がピックアップ位置である。物体の穴の位置は、前景画像 1 1 3 2 において、輝度の差がある領域を穴がある位置として特定する。次に、輪状の構造を有する物体については、輪状の部分の画像特徴を予め用意しておく。前景画像 1 1 3 2 でエッジ検出を行い、予め用意された画像特徴とマッチングして、最も近いエッジを持つ領域をピックアップ位置として特定する。S 1 0 0 3 2 では、第 1 のラベル生成部 1 0 3 2 が、部分画像 1 1 3 3 1 と前景画像 1 1 3 2 に基づいて、合成画像 1 1 3 3 0 に対して第 1 のラベルを付与した教師データを生成する。この第 1 のラベルは、ピックアップ位置であることを示すラベルである。部分  
30  
画像 1 1 3 3 1 における画像中心と、S 1 2 0 3 で求めた前景画像 1 1 3 2 の中心位置との距離に基づいて、距離が所定の値より小さい場合には、部分画像 1 1 3 3 1 の中心の画素に対してピックアップ位置であることを示すラベルを生成する。S 1 0 0 3 3 では、第 2 のラベル生成部 1 0 3 3 が、合成画像 1 1 3 3 0 と前景画像 1 1 3 2 に基づいて、S 1 2 0 3 で特定したピックアップ位置の周辺領域に対して、ピックアップ位置ではないことを表す第 2 のラベルを付与した教師データを生成する。小さい穴にフックや針を通すタスクにおいて、ピックアップ位置の周りに不正解を表す領域を設定することで、ピックアップ位置をより正確に学習させる効果がある。

【 0 0 8 4 】

尚、CPU はプログラムを実行することで各種の手段として機能することが可能である。なお、CPU と協調して動作する ASIC などの制御回路がこれらの手段として機能しても良い。また、CPU と画像処理装置の動作を制御する制御回路との協調によってこれらの手段が実現されても良い。また、CPU は単一のものである必要はなく、複数であっても良い。この場合、複数の CPU は分散して処理を実行することが可能である。また、複数の CPU は単一のコンピュータに配置されていても良いし、物理的に異なる複数のコンピュータに配置されていても良い。なお、CPU がプログラムを実行することで実現する手段が専用の回路によって実現されても良い。

【符号の説明】

【 0 0 8 5 】

1 0 画像処理装置

10

20

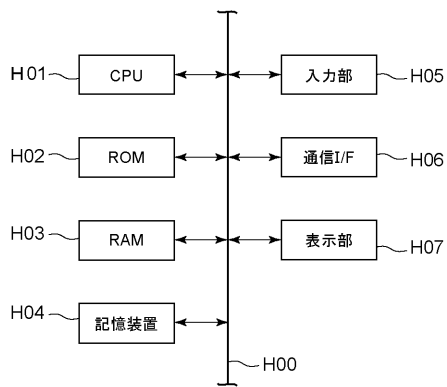
30

40

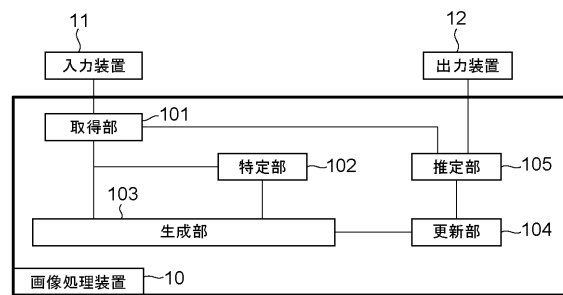
50

- 1 1 入力装置
- 1 2 出力装置
- 1 0 1 取得部
- 1 0 2 特定部
- 1 0 3 生成部
- 1 0 4 更新部
- 1 0 5 推定部

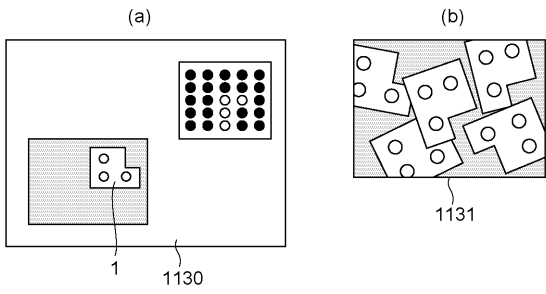
【 図 1 】



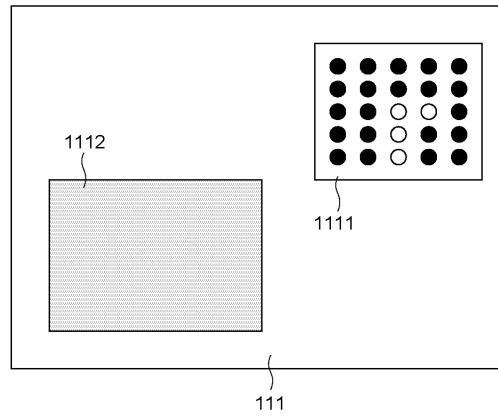
【 図 2 】



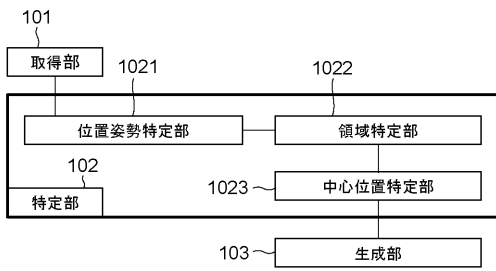
【 図 3 】



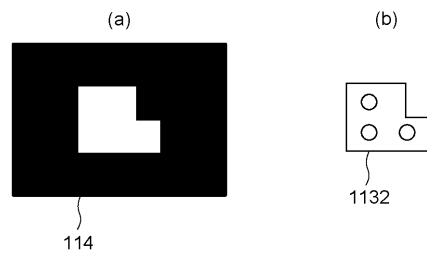
【 図 4 】



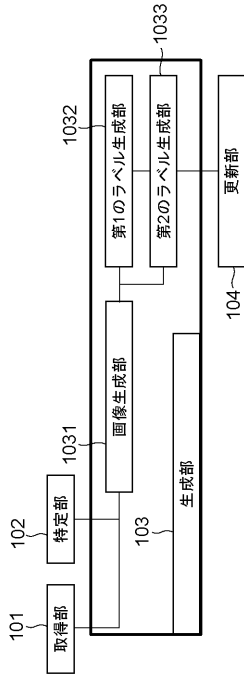
【 図 5 】



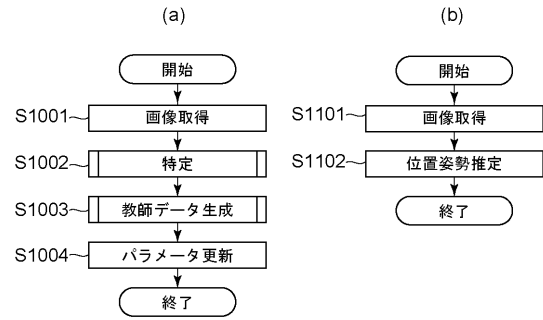
【 図 6 】



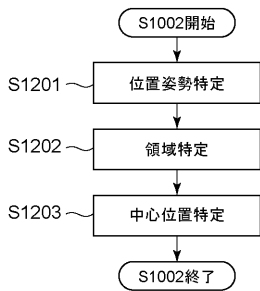
【 図 7 】



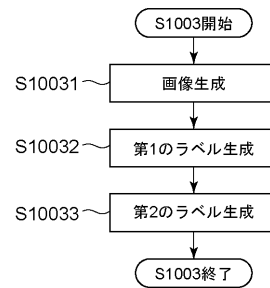
【 図 8 】



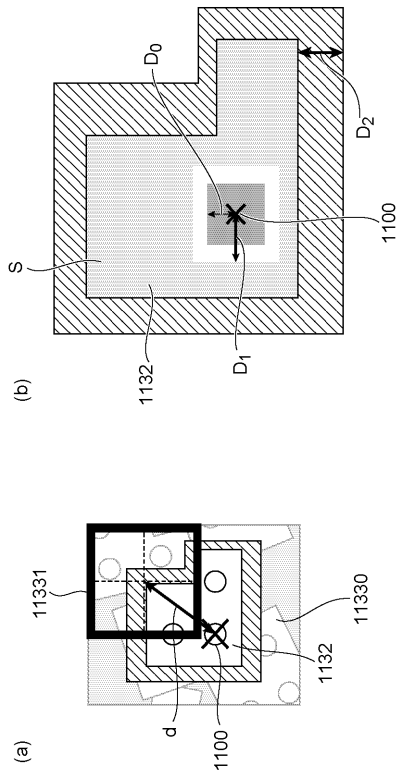
【 図 9 】



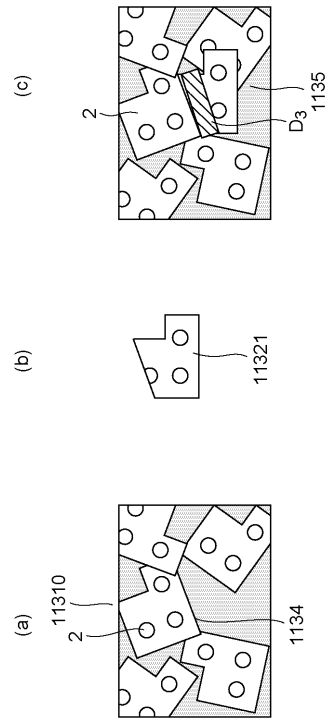
【 図 10 】



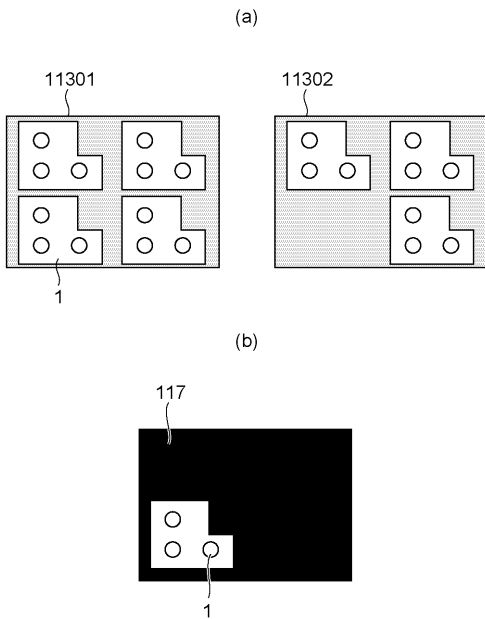
【図 1 1】



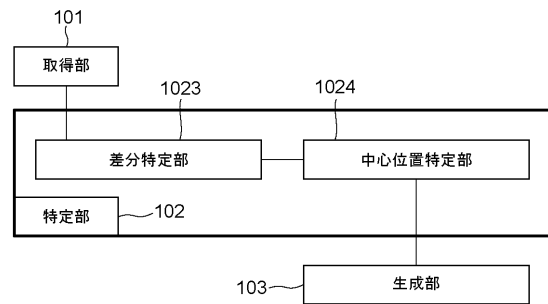
【図 1 2】



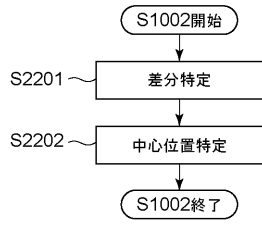
【図 1 3】



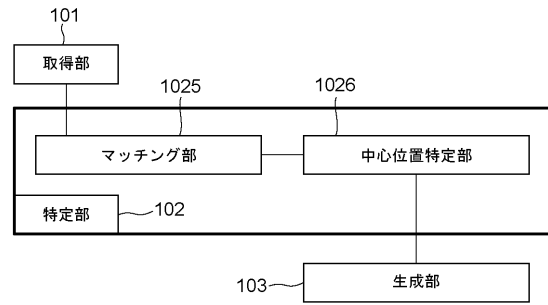
【図 1 4】



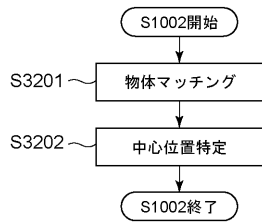
【 図 1 5 】



【 図 1 6 】



【 図 1 7 】



【 図 1 8 】

