(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2018-126798
(P2018-126798A)
(43)公開日　平成30年8月16日(2018.8.16)

(51)Int.Cl.		F I			テーマコード（参考）
B25J 13/08	(2006.01)	B25J	13/08	Z	3C707

審査請求 未請求　請求項の数 13　ＯＬ　（全 47 頁）

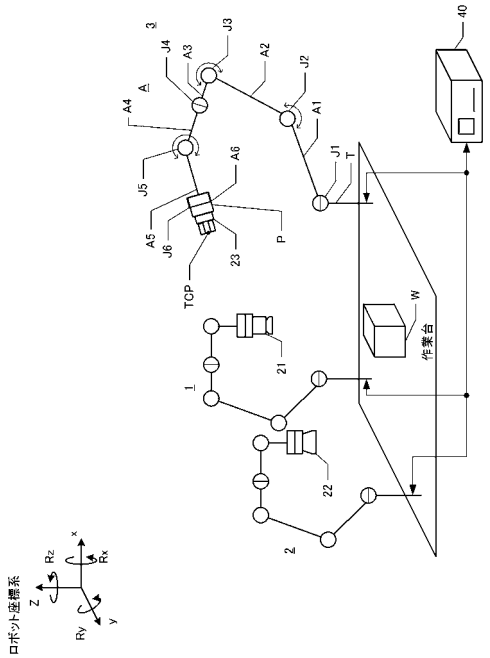
(21)出願番号　特願2017-19313(P2017-19313)
(22)出願日　平成29年2月6日(2017.2.6)

(71)出願人　000002369
セイコーエプソン株式会社
東京都新宿区新宿四丁目１番６号
(74)代理人　100116665
弁理士　渡辺　和昭
(74)代理人　100164633
弁理士　西田　圭介
(74)代理人　100179475
弁理士　仲井　智至
(72)発明者　長谷川　浩
長野県諏訪市大和３丁目３番５号　セイコーエプソン株式会社内

Ｆターム(参考)　3C707 KS03 KS04 KS23 KS24 KS33 KT01 KT05 KT11 LU02 LU09 LV24 LW08 LW12 LW15 MT06

(54)【発明の名称】制御装置、ロボットおよびロボットシステム

(57)【要約】
【課題】ロボットの性能を充分に引き出せる力制御パラメーターを設定することはやはり困難であった。
【解決手段】機械学習を用いて、ロボットの力制御に関する力制御パラメーターを算出する算出部と、算出された前記力制御パラメーターに基づいて前記ロボットを制御する制御部と、を備える制御装置が構成される。
【選択図】図１



【特許請求の範囲】
【請求項１】
機械学習を用いて、ロボットの力制御に関する力制御パラメーターを算出する算出部と、
算出された前記力制御パラメーターに基づいて前記ロボットを制御する制御部と、
を備える制御装置。
【請求項２】
前記力制御パラメーターは、
前記ロボットがインピーダンス制御で動作する際のインピーダンスパラメーターを含む、
請求項１に記載の制御装置。
【請求項３】
前記力制御パラメーターは、
前記ロボットの動作の始点と終点との少なくとも一方を含む、
請求項１または請求項２のいずれかに記載の制御装置。
【請求項４】
前記力制御パラメーターは、
前記ロボットのツールセンターポイントからのオフセット点の位置を含む、
請求項１～請求項３のいずれかに記載の制御装置。
【請求項５】
前記算出部は、
状態変数として、少なくとも前記ロボットの位置情報を観測する状態観測部と、
前記状態変数に基づいて前記力制御パラメーターを学習する学習部と、を含む、
請求項１～請求項４のいずれかに記載の制御装置。
【請求項６】
前記位置情報は、
前記ロボットが備える慣性センサーの出力と、前記ロボットの外部に配置された位置検出部の出力と、の少なくとも一方に基づいて算出される、
請求項５に記載の制御装置。
【請求項７】
前記学習部は、
前記状態変数に基づいて前記力制御パラメーターを変化させる行動を決定し、前記力制御パラメーターを最適化する、
請求項５または請求項６のいずれかに記載の制御装置。
【請求項８】
前記学習部は、
前記ロボットが行った作業の良否に基づいて、前記行動による報酬を評価する、
請求項７に記載の制御装置。
【請求項９】
前記学習部は、
前記作業が正常に完了した場合、前記作業の所要時間が基準よりも短い場合、の少なくとも１つにおいて前記報酬を正と評価する、
請求項８に記載の制御装置。
【請求項１０】
前記学習部は、
前記ロボットが破損した場合、前記ロボットの作業対象である対象物が破損した場合、の少なくとも１つにおいて前記報酬を負と評価する、
請求項８または請求項９のいずれかに記載の制御装置。
【請求項１１】
前記算出部は、

前記状態変数の観測と、当該状態変数に応じた前記行動の決定と、当該行動によって得られる前記報酬の評価とを繰り返すことによって、前記力制御パラメーターを最適化する、
請求項８～請求項１０のいずれかに記載の制御装置。
【請求項１２】
請求項１～請求項１１のいずれかに記載された制御装置によって制御されるロボット。
【請求項１３】
請求項１～請求項１１のいずれかに記載された制御装置と、前記制御装置によって制御される前記ロボットと、
を備えるロボットシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、制御装置、ロボットおよびロボットシステムに関する。
【背景技術】
【０００２】
ロボットに作業を行わせるためには、各種の設定が必要であり、従来、各種の設定は人為的に行われている。しかし、当該設定を行うためには高度なノウハウが必要であり、難易度が高い。そこで、従来、予め決められた手順に従ってインピーダンス制御系の慣性パラメーターと粘性パラメーターを調整する技術が開発されている（例えば、特許文献１）。
【０００３】
また、従来、工作機械の工具補正の頻度を最適化するために機械学習を利用した技術が知られている（特許文献２）。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特許第４９６２５５１号公報
【特許文献２】特許第５９６９６７６号公報
【発明の概要】
【発明が解決しようとする課題】
【０００５】
従来の技術を利用しても、パラメーターの設定には高度な知見が必須であった。例えば、特許文献１においては、振動回数と閾値とを比較する必要がある。しかし、理想的な閾値を予め決めることは実際には難しい。また、予め決められた手順での調整によってパラメーターが理想的な値になるとは限らない（他の手順であればより理想化できる可能性がある）。従って、ロボットの性能を充分に引き出せる力制御パラメーターを設定することはやはり困難であった。
【課題を解決するための手段】
【０００６】
上記課題の少なくとも一つを解決するために、制御装置は、機械学習を用いて、ロボットの力制御に関する力制御パラメーターを算出する算出部と、算出された力制御パラメーターに基づいてロボットを制御する制御部と、を備える。この構成によれば、人為的に決められた力制御パラメーターよりも高性能に力制御を行う力制御パラメーターを高い確率で算出することができる。
【０００７】
さらに、力制御パラメーターは、ロボットがインピーダンス制御で動作する際のインピーダンスパラメーターを含む構成であっても良い。この構成によれば、人為的な調整によって適切な設定を行うことが困難な、インピーダンスパラメーターを自動的に調整することができる。

【０００８】
さらに、力制御パラメーターは、ロボットの動作の始点と終点との少なくとも一方を含む構成であっても良い。この構成によれば、人為的に設定された始点や終点を、より高性能に力制御を行うように自動的に調整することができる。
【０００９】
さらに、力制御パラメーターは、ロボットのツールセンターポイントからのオフセット点の位置を含む構成であっても良い。この構成によれば、人為的な調整によって適切な設定を行うことが困難な、ツールセンターポイントからのオフセット点の位置を自動的に調整することができる。
【００１０】
さらに、算出部は、状態変数として、少なくともロボットの位置情報を観測する状態観測部と、状態変数に基づいて力制御パラメーターを学習する学習部と、を含む構成であっても良い。この構成によれば、高性能な力制御を行う力制御パラメーターを容易に算出することができる。
【００１１】
さらに、位置情報は、ロボットが備える慣性センサーの出力と、ロボットの外部に配置された位置検出部の出力と、の少なくとも一方に基づいて算出される構成であっても良い。慣性センサーによれば、ロボットで汎用的に使用されるセンサーに基づいて位置情報を算出することができる。ロボットの外部に配置された位置検出部は、ロボットの動作に影響されることなく位置情報を算出することができる。
【００１２】
さらに、学習部は、状態変数に基づいて力制御パラメーターを変化させる行動を決定し、力制御パラメーターを最適化する構成であっても良い。この構成によれば、ロボットの使用環境に応じた力制御パラメーターとなるように最適化することができる。
【００１３】
さらに、学習部は、ロボットが行った作業の良否に基づいて、行動による報酬を評価する構成であっても良い。この構成によれば、ロボットの作業の質を高めるように力制御パラメーターを最適化することができる。
【００１４】
さらに、学習部は、作業が正常に完了した場合、作業の所要時間が基準よりも短い場合、の少なくとも１つにおいて報酬を正と評価する構成であっても良い。作業が正常に完了した場合に報酬を正と評価する構成によれば、ロボットの作業を成功させる力制御パラメーターを容易に算出することができる。作業の所要時間が基準よりも短い場合に報酬を正と評価する構成によれば、ロボットを短い時間で作業させる力制御パラメーターを容易に算出することができる。
【００１５】
さらに、学習部は、ロボットが破損した場合、ロボットの作業対象である対象物が破損した場合、の少なくとも１つにおいて報酬を負と評価する構成であっても良い。ロボットが破損した場合に報酬を負と評価する構成によれば、ロボットを破損させる可能性が低い力制御パラメーターを容易に算出することができる。ロボットの作業対象である対象物が破損した場合に報酬を負と評価する構成によれば、対象物を破損させる可能性が低い力制御パラメーターを容易に算出することができる。
【００１６】
さらに、算出部は、状態変数の観測と、当該状態変数に応じた行動の決定と、当該行動によって得られる報酬の評価とを繰り返すことによって、力制御パラメーターを最適化する構成であっても良い。この構成によれば、力制御パラメーターを自動的に最適化することができる。
【図面の簡単な説明】
【００１７】
【図１】ロボットシステムの斜視図である。

【図２】制御装置の機能ブロック図である。
【図３】パラメーターを示す図である。
【図４】加減速特性を示す図である。
【図５】ピックアップ処理のフローチャートである。
【図６】算出部に関連する構成のブロック図である。
【図７】光学パラメーターを学習する際の例を示す図である。
【図８】多層ニューラルネットワークの例を示す図である。
【図９】学習処理のフローチャートである。
【図１０】動作パラメーターを学習する際の例を示す図である。
【図１１】力制御パラメーターを学習する際の例を示す図である。
【発明を実施するための形態】
【００１８】
以下、本発明の実施形態について添付図面を参照しながら以下の順に説明する。なお、各図において対応する構成要素には同一の符号が付され、重複する説明は省略される。
（１）ロボットシステムの構成：
（２）ロボットの制御：
（３）ピックアップ処理：
（４）学習処理：
（４－１）光学パラメーターの学習：
（４－２）光学パラメーターの学習例：
（４－３）動作パラメーターの学習：
（４－４）動作パラメーターの学習例：
（４－５）力制御パラメーターの学習：
（４－６）力制御パラメーターの学習例：
（５）他の実施形態：
【００１９】
（１）ロボットシステムの構成：
図１は本発明の一実施形態にかかる制御装置で制御されるロボットを示す斜視図である。本発明の一実施例としてのロボットシステムは、図１に示すように、ロボット１～３を備えている。ロボット１～３はエンドエフェクターを備える６軸ロボットであり、ロボット１～３には異なるエンドエフェクターが取り付けられている。すなわち、ロボット１には、撮像部２１が取り付けられ、ロボット２には照明部２２が取り付けられ、ロボット３にはグリッパー２３が取り付けられている。なお、ここでは、撮像部２１および照明部２２を光学系と呼ぶ。
【００２０】
ロボット１～３は、制御装置４０によって制御される。制御装置４０はケーブルによりロボット１～３と通信可能に接続される。なお、制御装置４０の構成要素がロボット１に備えられていても良い。また、制御装置４０は複数の装置によって構成されても良い（例えば、後述する学習部と制御部とが異なる装置に備えられる等）。また、制御装置４０は、図示しない教示装置をケーブル、または無線通信によって接続可能である。教示装置は、専用のコンピューターであってもよいし、ロボット１を教示するためのプログラムがインストールされた汎用のコンピューターであってもよい。さらに、制御装置４０と教示装置とは、一体に構成されていてもよい。
【００２１】
ロボット１～３は、アームに各種のエンドエフェクターを装着して使用される単腕ロボットであり、本実施形態において、ロボット１～３においてアームや軸の構成は同等である。図１においてはロボット３においてアームや軸の構成を説明する符号が付されている。ロボット３において示されたように、ロボット１～３は、基台Ｔと、６個のアーム部材Ａ１～Ａ６と、６個の関節Ｊ１～Ｊ６を備える。基台Ｔは作業台に固定されている。基台Ｔと６個のアーム部材Ａ１～Ａ６は関節Ｊ１～Ｊ６によって連結される。アーム部材Ａ１

～Ａ６とエンドエフェクターは可動部であり、これらの可動部が動作することによってロボット１～３は各種の作業を行うことができる。

【００２２】

本実施形態において、関節Ｊ２、Ｊ３、Ｊ５は曲げ関節であり、関節Ｊ１、Ｊ４、Ｊ６はねじり関節である。アームＡのうち最も先端側のアーム部材Ａ６には、力覚センサーＰとエンドエフェクターとが装着される。ロボット１～３は、６軸のアームを駆動させることによって、可動範囲内においてエンドエフェクターを任意の位置に配置し、任意の姿勢（角度）とすることができる。

【００２３】

ロボット３が備えるエンドエフェクターはグリッパー２３であり、対象物Ｗを把持することができる。ロボット２が備えるエンドエフェクターは照明部２２であり、照射範囲に光を照射することができる。ロボット１が備えるエンドエフェクターは撮像部２１であり、視野内の画像を撮像することができる。本実施形態においては、ロボット１～３が備えるエンドエフェクターに対して相対的に固定された位置がツールセンターポイント（ＴＣＰ）として定義される。ＴＣＰの位置はエンドエフェクターの基準の位置となり、ＴＣＰが原点となり、エンドエフェクターに対して相対的に固定された３次元直交座標系であるＴＣＰ座標系が定義される。

【００２４】

力覚センサーＰは、６軸の力検出器である。力覚センサーＰは、力覚センサー上の点を原点とした３次元直交座標系であるセンサー座標系において互いに直交する３個の検出軸と平行な力の大きさと、当該３個の検出軸まわりのトルクの大きさとを検出する。なお、本実施例では６軸ロボットを例にしているが、ロボットの態様は種々の態様であっても良いし、ロボット１～３の態様が異なっていてもよい。また、関節Ｊ６以外の関節Ｊ１～Ｊ５のいずれか１つ以上に力検出器としての力覚センサーを備えても良い。

【００２５】

ロボット１～３が設置された空間を規定する座標系をロボット座標系というとき、ロボット座標系は、水平面上において互いに直交するｘ軸とｙ軸と、鉛直上向きを正方向とするｚ軸とによって規定される３次元の直交座標系である（図１参照）。ｚ軸における負の方向は概ね重力方向と一致する。またｘ軸周りの回転角をＲｘで表し、ｙ軸周りの回転角をＲｙで表し、ｚ軸周りの回転角をＲｚで表す。ｘ，ｙ，ｚ方向の位置により３次元空間における任意の位置を表現でき、Ｒｘ，Ｒｙ，Ｒｚ方向の回転角により３次元空間における任意の姿勢を表現できる。以下、位置と表記した場合、姿勢も意味し得ることとする。また、力と表記した場合、トルクも意味し得ることとする。

【００２６】

なお、本実施形態においてはロボットに作用する力を制御する力制御が実行可能であり、力制御においては、任意の点に作用する当該作用力が目標力になるように制御される。各種の部位に作用する力は、３次元直交座標系である力制御座標系において定義される。目標力（トルクを含む）は、力制御座標系で表現された力の作用点を起点としたベクトルで表現可能であり、後述する学習が行われる以前において、目標力ベクトルの起点は力制御座標系の原点であり、作用力の方向は力制御座標系の１軸方向と一致している。ただし、後述する学習が行われた場合、目標力ベクトルの起点は力制御座標系の原点と異なり得るし、目標力ベクトルの方向は力制御座標系の軸方向と異なり得る。

【００２７】

本実施形態において各種の座標系の関係は予め定義されており、各種の座標系での座標値は互いに変換可能である。すなわち、ＴＣＰ座標系、センサー座標系、ロボット座標系、力制御座標系における位置やベクトルは互いに変換可能である。ここでは簡単のため、制御装置４０がＴＣＰの位置およびＴＣＰに作用する作用力をロボット座標系で制御する説明をするが、ロボット１～３の位置やロボット１～３に作用する力は、各種の座標系で定義でき、互いに変換可能であるため、位置や力がどの座標系で定義され、制御されても良い。むろん、ここで述べた座標系以外にも他の座標系（例えば対象物に固定されたオブ

ジェクト座標系等）が定義され、変換可能であっても良い。
【００２８】
（２）ロボットの制御：
ロボット１は、教示を行うことにより各種作業が可能となる汎用ロボットであり、図２に示すようにアクチュエーターとしてのモーターＭ１～Ｍ６と、センサーとしてのエンコーダーＥ１～Ｅ６とを備える。アームを制御することはモーターＭ１～Ｍ６を制御することを意味する。モーターＭ１～Ｍ６とエンコーダーＥ１～Ｅ６とは、関節Ｊ１～Ｊ６のそれぞれに対応して備えられており、エンコーダーＥ１～Ｅ６はモーターＭ１～Ｍ６の回転角度を検出する。また、各モーターＭ１～Ｍ６には電力を供給する電源線が接続されており、各電源線には電流計が備えられている。従って、制御装置４０は、各モーターＭ１～Ｍ６に供給された電流を計測することができる。
【００２９】
制御装置４０は、コンピューター等のハードウェア資源と記憶部４４に記憶された各種のソフトウェア資源を備え、プログラムを実行可能である。本実施形態において制御装置４０は、算出部４１、検出部４２、制御部４３として機能する。なお、ハードウェア資源は、ＣＰＵ，ＲＡＭ，ＲＯＭ等からなる構成であっても良いし、ＡＳＩＣ等によって構成されても良く、種々の構成を採用可能である。
【００３０】
本実施形態において検出部４２は対象物を検出する処理を実行することが可能であり、制御部４３はロボット１～３のアームを駆動することが可能である。検出部４２は、光学系２０を構成する撮像部２１と照明部２２とに接続されている。検出部４２は、撮像部２１を制御し、撮像部２１が備える撮像センサーによって撮像された画像を取得することができる。また、検出部４２は、照明部２２を制御し、出力光の明るさを変化させることができる。
【００３１】
撮像部２１から画像が出力されると、検出部４２は、撮像画像に基づいてテンプレートマッチング処理を行い、対象物の位置（位置姿勢）を検出する処理を行う。すなわち、検出部４２は、記憶部４４に記憶されたテンプレートデータ４４ｃに基づいてテンプレートマッチング処理を実行する。テンプレートデータ４４ｃは複数の位置姿勢毎のテンプレートである。従って、テンプレートデータ４４ｃに対して位置姿勢をＩＤ等で対応づけておけば、適合したテンプレートデータ４４ｃの種類によって検出部４２から見た対象物の位置姿勢を特定することができる。
【００３２】
具体的には、検出部４２は、複数の位置姿勢毎のテンプレートデータ４４ｃを順次処理対象とし、テンプレートデータ４４ｃの大きさを変化させながら、撮像された画像と比較する。そして、検出部４２は、テンプレートデータ４４ｃと画像との差分が閾値以下の像を対象物の像として検出する。
【００３３】
対象物の像が検出されると、検出部４２は、予め決められた座標系の関係と適合したテンプレートデータ４４ｃの大きさに基づいて対象物の位置姿勢を特定する。すなわち、テンプレートデータ４４ｃの大きさから撮像部２１と対象物との光軸方向の距離が判明し、画像内で検出された対象物の位置から光軸に垂直な方向の位置が判明する。
【００３４】
そこで、例えば、撮像部２１の撮像センサーの光軸と撮像平面上の２軸とがＴＣＰ座標系の各軸に平行に定義されている場合であれば、検出部４２は、テンプレートデータ４４ｃの大きさと、テンプレートデータ４４ｃが画像と適合した位置とに基づいて、ＴＣＰ座標系において対象物の位置を特定することができる。また、検出部４２は、適合したテンプレートデータ４４ｃのＩＤに基づいて、ＴＣＰ座標系における対象物の姿勢を特定することができる。このため、検出部４２は、上述の座標系の対応関係を利用し、任意の座標系、例えば、ロボット座標系における対象物の位置姿勢を特定することができる。

【００３５】
なお、テンプレートマッチング処理は、対象物の位置姿勢を特定するための処理であれば良く、種々の処理を採用可能である。例えば、テンプレートデータ４４ｃと画像との差分は、階調値の差分によって評価されても良いし、画像の特徴（例えば、画像の勾配等）の差分によって評価されても良い。
【００３６】
検出部４２は、パラメーターを参照して当該テンプレートマッチング処理を行う。すなわち、記憶部４４には、各種のパラメーター４４ａが記憶されており、当該パラメーター４４ａには、検出部４２の検出に関するパラメーターが含まれている。図３は、パラメーター４４ａの例を示す図である。図３に示す例において、パラメーター４４ａは、光学パラメーターと動作パラメーターと力制御パラメーターとを含んでいる。
【００３７】
光学パラメーターは、検出部４２の検出に関するパラメーターである。動作パラメーターと力制御パラメーターとはロボット１～３を制御する際のパラメーターであり、詳細は後述する。光学パラメーターは、撮像部２１に関する撮像部パラメーターと、照明部２２に関する照明部パラメーターと、撮像部２１によって撮像された対象物の画像に対する画像処理に関する画像処理パラメーターとが含まれる。
【００３８】
図３においては、これらのパラメーターの例が示されている。すなわち、対象物を撮像する際に撮像部２１が配置される位置が撮像部の位置として定義され、撮像部パラメーターに含まれている。また、撮像部２１は、露光時間と絞りを調整可能な機構を備えており、対象物を撮像する際の露光時間および絞りの値が撮像部パラメーターに含まれている。なお、撮像部の位置は、種々の手法で記述されて良く、例えば、撮像部２１のＴＣＰの位置がロボット座標系で記述される構成等を採用可能である。
【００３９】
検出部４２は、撮像部パラメーターを参照し、撮像部２１の位置を後述する位置制御部４３ａに受け渡す。この結果、位置制御部４３ａは、目標位置Ｌｔを生成し、当該目標位置Ｌｔに基づいてロボット１を制御する。また、検出部４２は、撮像部パラメーターを参照し、撮像部２１の露光時間と絞りを設定する。この結果、撮像部２１においては当該露光時間と絞りによって撮像が行われる状態となる。
【００４０】
また、対象物を撮像する際に照明部２２が配置される位置が照明部の位置として定義され、照明部パラメーターに含まれている。また、照明部２２は、明るさを調整可能な機構を備えており、対象物を撮像する際の明るさの値が照明部パラメーターに含まれている。照明部の位置も、種々の手法で記述されて良く、例えば、照明部２２のＴＣＰの位置がロボット座標系で記述される構成等を採用可能である。
【００４１】
検出部４２は、照明部パラメーターを参照し、照明部２２の位置を後述する位置制御部４３ａに受け渡す。この結果、位置制御部４３ａは、目標位置Ｌｔを生成し、当該目標位置Ｌｔに基づいてロボット２を制御する。また、検出部４２は、照明部パラメーターを参照し、照明部２２の明るさを設定する。この結果、照明部２２においては当該明るさの光が出力される状態となる。
【００４２】
検出部４２は、撮像部２１によって撮像された画像に対してテンプレートマッチング処理を適用する際に、画像処理パラメーターを参照する。すなわち、画像処理パラメーターには、テンプレートマッチング処理を実行する際の処理順序を示す画像処理シーケンスが含まれている。また、本実施形態において、テンプレートマッチング処理における閾値が可変であり、現在のテンプレートマッチングの閾値が画像処理パラメーターに含まれている。さらに、検出部４２は、テンプレートデータ４４ｃと画像とを比較する前に各種の処理を実行可能である。図３においては、各種の処理として平滑化処理と鮮鋭化処理が例示

されており、それぞれの強度が画像処理パラメーターに含まれている。

【００４３】

撮像部２１から画像が出力されると、検出部４２は、画像処理シーケンスに基づいて、画像処理の順序（実行するか否かを含む）を決定し、当該順序で平滑化処理や鮮鋭化処理等の画像処理を実行する。このとき、検出部４２は、画像処理パラメーターに記述された強度で平滑化処理や鮮鋭化処理等の画像処理を実行する。また、画像処理シーケンスに含まれる比較（テンプレートデータ４４ｃと画像との比較）を実行する際には、画像処理パラメーターが示す閾値に基づいて比較を行う。

【００４４】

なお、以上のように検出部４２は、光学パラメーターに基づいて撮像部２１や照明部２２の位置を特定し、ロボット１、ロボット２を動作させることが可能であるが、ロボット１およびロボット２を駆動する際の位置は、後述する動作パラメーターや力制御パラメーターによって与えられてもよい。

【００４５】

本実施形態において、制御部４３は、位置制御部４３ａ、力制御部４３ｂ、接触判定部４３ｃ、サーボ４３ｄを備えている。また、制御部４３においては、モーターＭ１～Ｍ６の回転角度の組み合わせと、ロボット座標系におけるＴＣＰの位置との対応関係Ｕ１が図示しない記憶媒体に記憶され、座標系の対応関係Ｕ２が定義され、図示しない記憶媒体に記憶されている。従って、制御部４３や後述する算出部４１は、対応関係Ｕ２に基づいて、任意の座標系におけるベクトルを他の座標系におけるベクトルに変換することができる。例えば、制御部４３、算出部４１は、力覚センサーＰの出力に基づいてセンサー座標系でのロボット１～３への作用力を取得し、ロボット座標系におけるＴＣＰの位置に作用する力に変換することができる。また、制御部４３、算出部４１は、力制御座標系で表現された目標力をロボット座標系におけるＴＣＰの位置における目標力に変換することができる。むろん、対応関係Ｕ１，Ｕ２は記憶部４４に記憶されていても良い。

【００４６】

制御部４３は、アームを駆動することによって、ロボット１～３とともに移動する各種の部位の位置や各種の部位に作用する力を制御することができ、位置の制御は主に位置制御部４３ａ、力の制御は主に力制御部４３ｂによって実行される。サーボ４３ｄは、サーボ制御を実行することが可能であり、エンコーダーＥ１～Ｅ６の出力が示すモーターＭ１～Ｍ６の回転角度D_aと、制御目標である目標角度Ｄｔとを一致させるフィードバック制御を実行する。すなわち、サーボ４３ｄは、回転角度D_aと目標角度Ｄｔとの偏差、当該偏差の積分、当該偏差の微分にサーボゲインＫｐｐ，Ｋｐｉ，Ｋｐｄを作用させたＰＩＤ制御を実行することができる。

【００４７】

さらに、サーボ４３ｄは、当該サーボゲインＫｐｐ，Ｋｐｉ，Ｋｐｄが作用した出力と、回転角度D_aの微分との偏差、当該偏差の積分、当該偏差の微分にサーボゲインＫｖｐ，Ｋｖｉ，Ｋｖｄを作用させたＰＩＤ制御を実行することができる。当該サーボ４３ｄによる制御は、モーターＭ１～Ｍ６のそれぞれに対して実行可能である。従って、各サーボゲインはロボット１～３が備える６軸のそれぞれについて実行可能である。なお、本実施形態において、制御部４３は、サーボ４３ｄに制御信号を出力し、サーボゲインＫｐｐ，Ｋｐｉ，Ｋｐｄ，Ｋｖｐ，Ｋｖｉ，Ｋｖｄを変化させることができる。

【００４８】

記憶部４４には、上述のパラメーター４４ａに加え、ロボット１～３を制御するためのロボットプログラム４４ｂが記憶される。本実施形態において、パラメーター４４ａおよびロボットプログラム４４ｂは、教示によって生成され、記憶部４４に記憶されるが、後述する算出部４１によって修正され得る。なお、ロボットプログラム４４ｂは、主に、ロボット１～３が実施する作業のシーケンス（工程の順序）を示し、予め定義されたコマンドの組み合わせによって記述される。また、パラメーター４４ａは、主に、各工程を実現するために必要とされる具体的な値であり、各コマンドの引数として記述される。

【００４９】
ロボット１～３を制御するためのパラメーター４４ａには、上述の光学パラメーターの他に、動作パラメーターと力制御パラメーターが含まれる。動作パラメーターは、ロボット１～３の動作に関するパラメーターであり、本実施形態においては、位置制御の際に参照されるパラメーターである。すなわち、本実施形態において、一連の作業は複数の工程に分けられ、各工程を実施する際のパラメーター４４ａが教示によって生成される。動作パラメーターには、当該複数の工程における始点と終点を示すパラメーターが含まれている。当該始点と終点は、種々の座標系で定義されて良く、本実施形態においては制御対象のロボットのＴＣＰの始点および終点がロボット座標系で定義される。すなわち、ロボット座標系の各軸についての並進位置と回転位置とが定義される。

【００５０】
また、動作パラメーターには、複数の工程におけるＴＣＰの加減速特性が含まれている。加減速特性は、ロボット１～３のＴＣＰが各工程の始点から終点まで移動する際の期間と当該期間内の各時刻におけるＴＣＰの速度を示している。図４は、当該加減速特性の例を示す図であり、始点におけるＴＣＰの移動開始時刻t_1からＴＣＰが終点に到達する時刻t_4までの期間内の各時刻においてＴＣＰの速度Ｖが定義されている。また、本実施形態において加減速特性には定速期間が含まれる。

【００５１】
定速期間は時刻t_2～t_3の期間であり、この期間内に置いて速度は一定である。また、この期間の前後においてＴＣＰは加速し、また、減速する。すなわち、時刻t_1～t_2までの期間においてＴＣＰは加速し、時刻t_3～t_4までの期間においてＴＣＰは減速する。当該加減速特性も種々の座標系で定義されて良く、本実施形態においては制御対象のロボットのＴＣＰについての速度であり、ロボット座標系で定義される。すなわち、ロボット座標系の各軸についての並進速度と回転速度（角速度）とが定義される。

【００５２】
さらに、動作パラメーターには、サーボゲインKpp，Kpi，Kpd，Kvp，Kvi，Kvdが含まれている。すなわち、制御部４３は、動作パラメーターとして記述された値になるようにサーボ４３ｄに制御信号を出力し、サーボゲインKpp，Kpi，Kpd，Kvp，Kvi，Kvdを調整することができる。本実施形態において当該サーボゲインは、上述の工程毎の値であるが、後述の学習等によってより短い期間毎の値とされても良い。

【００５３】
力制御パラメーターは、ロボット１～３の力制御に関するパラメーターであり、本実施形態においては、力制御の際に参照されるパラメーターである。始点、終点、加減速特性、サーボゲインは、動作パラメーターと同様のパラメーターであり、始点、終点、加減速特性はロボット座標系の３軸の並進と回転について定義される。また、サーボゲインはモーターＭ１～Ｍ６のそれぞれについて定義される。ただし、力制御の場合、始点および終点の中の少なくとも一部は定義されない場合（任意とされる場合）もある。例えば、ある方向に作用する力が０になるように衝突回避や倣い制御が行われる場合、当該方向における始点および終点は定義されず、当該方向の力を０にするように位置が任意に変化し得る状態が定義される場合もある。

【００５４】
また、力制御パラメーターには、力制御座標系を示す情報が含まれている。力制御座標系は、力制御の目標力を定義するための座標系であり、後述の学習が行われる前においては目標力ベクトルの起点が原点であり、目標力ベクトルの方向に１軸が向いている。すなわち、教示において力制御における各種の目標力が定義される際に、各作業の各工程における目標力の作用点が教示される。例えば、対象物の一点を他の物体に当て、両者の接触点で対象物から他の物体に一定の目標力を作用させた状態で対象物の向きを変化させる場合において、対象物が他の物体と接触する点が目標力の作用点となり、当該作用点を原点とした力制御座標系が定義される。そこで、力制御パラメーターにおいては、力制御の目標力が作用する点を原点とし、目標力の方向に１軸が向いている座標系、すなわち、力制

御座標系を特定するための情報を、パラメーターに含んでいる。なお、当該パラメーターは種々の定義が可能であるが、例えば、力制御座標系と他の座標系（ロボット座標系等）との関係を示すデータによって定義可能である。

【００５５】

さらに、力制御パラメーターには、目標力が含まれている。目標力は、各種の作業において、任意の点に作用すべき力として教示される力であり、力制御座標系において定義される。すなわち、目標力を示す目標力ベクトルが、目標力ベクトルの起点と、起点からの６軸成分（３軸の並進力、３軸のトルク）として定義され、力制御座標系で表現されている。なお、力制御座標系と他の座標系との関係を利用すれば、当該目標力を任意の座標系、例えば、ロボット座標系におけるベクトルに変換することが可能である。

【００５６】

さらに、力制御パラメーターには、インピーダンスパラメーターが含まれている。すなわち、本実施形態において力制御部４３ｂが実施する力制御は、インピーダンス制御である。インピーダンス制御は、仮想の機械的インピーダンスをモーターＭ１～Ｍ６によって実現する制御である。この際、ＴＣＰが仮想的に有する質量が仮想慣性係数ｍとして定義され、ＴＣＰが仮想的に受ける粘性抵抗が仮想粘性係数ｄとして定義され、ＴＣＰが仮想的に受ける弾性力のバネ定数が仮想弾性係数ｋとして定義される。インピーダンスパラメーターはこれらのｍ，ｄ，ｋであり、ロボット座標系の各軸に対する並進と回転について定義される。本実施形態において当該力制御座標系、目標力、インピーダンスパラメーターは、上述の工程毎の値であるが、後述の学習等によってより短い期間毎の値とされても良い。

【００５７】

本実施形態において、一連の作業は複数の工程に分けられ、各工程を実施するロボットプログラム４４ｂが教示によって生成されるが、位置制御部４３ａは、ロボットプログラム４４ｂが示す各工程をさらに微小時間ΔＴ毎の微小工程に細分化する。そして、位置制御部４３ａは、パラメーター４４ａに基づいて微小工程毎の目標位置Ｌｔを生成する。力制御部４３ｂは、パラメーター４４ａに基づいて一連の作業の各工程における目標力ｆLtを取得する。

【００５８】

すなわち、位置制御部４３ａは、動作パラメーターまたは力制御パラメーターが示す始点、終点、加減速特性を参照し、始点から終点まで当該加減速特性で移動する場合（姿勢の場合は姿勢が変化する場合）の微小工程毎のＴＣＰの位置を目標位置Ｌｔとして生成する。力制御部４３ｂは、各工程についての力制御パラメーターが示す目標力を参照し、力制御座標系とロボット座標系との対応関係Ｕ２に基づいて当該目標力をロボット座標系における目標力ｆLtに変換する。当該目標力ｆLtは、任意の点に作用する力として変換され得るが、ここでは、後述の作用力がＴＣＰに作用している力として表現されるため、当該作用力と目標力ｆLtとを運動方程式で解析するため、目標力ｆLtがＴＣＰの位置における力に変換されるとして説明を行う。むろん、工程によっては、目標力ｆLtが定義されない場合もあり、この場合、力制御を伴わない位置制御が行われる。

【００５９】

なお、ここでＬの文字は、ロボット座標系を規定する軸の方向（ｘ，ｙ，ｚ，Ｒｘ，Ｒｙ，Ｒｚ）のなかのいずれか１個の方向を表すこととする。また、Ｌは、Ｌ方向の位置も表すこととする。例えば、Ｌ＝ｘの場合、ロボット座標系にて設定された目標位置のｘ方向成分がＬｔ＝ｘｔと表記され、目標力のｘ方向成分がｆLt＝ｆxtと表記される。

【００６０】

位置制御や力制御を実行するため、制御部４３は、ロボット１～３の状態を取得することができる。すなわち、制御部４３は、モーターＭ１～Ｍ６の回転角度Ｄaを取得し、対応関係Ｕ１に基づいて、当該回転角度Ｄaをロボット座標系におけるＴＣＰの位置Ｌ（ｘ，ｙ，ｚ，Ｒｘ，Ｒｙ，Ｒｚ）に変換することができる。また制御部４３は、対応関係Ｕ２を参照し、ＴＣＰの位置Ｌと、力覚センサーＰの検出値および位置とに基づいて、力覚

センサーＰに現実に作用している力をＴＣＰに作用している作用力f_Lに変換してロボット座標系において特定することができる。

【００６１】

すなわち、力覚センサーＰに作用している力は、センサー座標系で定義される。そこで、制御部４３は、ロボット座標系におけるＴＣＰの位置Ｌと対応関係Ｕ２と力覚センサーＰの検出値に基づいて、ロボット座標系においてＴＣＰに作用する作用力f_Lを特定する。また、ロボットに作用するトルクは、作用力f_Lと、ツール接触点（エンドエフェクターとワークの接触点）から力覚センサーＰまでの距離とから算出することができ、図示されないf_Lトルク成分として特定される。なお、制御部４３は、作用力f_Lに対して重力補償を行う。重力補償とは、作用力f_Lから重力成分を除去する処理である。重力補償は、例えば、ＴＣＰの姿勢ごとにＴＣＰに作用する作用力f_Lの重力成分を予め調査しておき、作用力f_Lから当該重力成分を減算するなどして実現可能である。

【００６２】

ＴＣＰに作用する重力以外の作用力f_Lと、ＴＣＰに作用すべき目標力ｆLtとが特定されると、力制御部４３ｂは、対象物等の物体がＴＣＰに存在し、当該ＴＣＰに力が作用し得る状態において、インピーダンス制御による補正量ΔＬ（以後、力由来補正量ΔＬと呼ぶ。）を取得する。すなわち、力制御部４３ｂはパラメーター４４ａを参照して目標力ｆLtとインピーダンスパラメーターｍ，ｄ，ｋを取得し、運動方程式（１）に代入して力由来補正量ΔＬを取得する。なお、当該力由来補正量ΔＬは、ＴＣＰが機械的インピーダンスを受けた場合に、目標力ｆLtと作用力f_Lとの力偏差Δf_L（ｔ）を解消するために、ＴＣＰが移動すべき位置Ｌの大きさを意味する。

【数１】

$$m\Delta\ddot{S}(t) + d\Delta\dot{S}(t) + k\Delta S(t) = \Delta f_S(t) \quad \cdots (1)$$

【００６３】

（１）式の左辺は、ＴＣＰの位置Ｌの２階微分値に仮想慣性係数ｍを乗算した第１項と、ＴＣＰの位置Ｌの微分値に仮想粘性係数ｄを乗算した第２項と、ＴＣＰの位置Ｌに仮想弾性係数ｋを乗算した第３項とによって構成される。（１）式の右辺は、目標力ｆLtから現実の作用力f_Lを減算した力偏差Δf_L（ｔ）によって構成される。（１）式における微分とは、時間による微分を意味する。

【００６４】

力由来補正量ΔＬが得られると、制御部４３は、対応関係Ｕ１に基づいて、ロボット座標系を規定する各軸の方向の動作位置を、各モーターＭ１～Ｍ６の目標の回転角度である目標角度Ｄｔに変換する。サーボ４３ｄは、目標角度ＤｔからモーターＭ１～Ｍ６の現実の回転角度であるエンコーダーＥ１～Ｅ６の出力（回転角度Ｄａ）を減算することにより、駆動位置偏差Ｄｅ（＝Ｄｔ－Ｄａ）を算出する。サーボ４３ｄは、パラメーター４４ａを参照してサーボゲインＫｐｐ，Ｋｐｉ，Ｋｐｄ，Ｋｖｐ，Ｋｖｉ，Ｋｖｄを取得し、駆動位置偏差ＤｅにサーボゲインＫｐｐ，Ｋｐｉ，Ｋｐｄを乗算した値と、現実の回転角度Ｄａの時間微分値である駆動速度との差である駆動速度偏差に、サーボゲインＫｖｐ，Ｋｖｉ，Ｋｖｄを乗算した値とを加算することにより、制御量Ｄｃを導出する。制御量Ｄｃは、モーターＭ１～Ｍ６のそれぞれについて特定され、各モーターＭ１～Ｍ６の制御量ＤｃでモーターＭ１～Ｍ６のそれぞれが制御される。制御部４３がモーターＭ１～Ｍ６を制御する信号は、ＰＷＭ（Pulse Width Modulation）変調された信号である。

【００６５】

以上のように、運動方程式に基づいて目標力ｆLtから制御量Ｄｃを導出してモーターＭ１～Ｍ６を制御するモードを力制御モードというものとする。また制御部４３は、エンドエフェクター等の構成要素が対象物Ｗから力を受けない非接触状態の工程では、力制御を行わず、目標位置から線形演算で導出する回転角度でモーターＭ１～Ｍ６を制御する。目標位置から線形演算で導出する回転角度でモーターＭ１～Ｍ６を制御するモードを位置制御モードというものとする。さらに、制御部４３は、目標位置から線形演算で導出する回

転角度と目標力を運動方程式に代入して導出する回転角度とを例えば線型結合によって統合し、統合した回転角度でモーターＭ１～Ｍ６を制御するハイブリッドモードでもロボット１を制御することができる。これらのモードはロボットプログラム４４ｂによって予め決められる。

【００６６】

位置制御モードまたはハイブリッドモードで制御を行う場合、位置制御部４３ａは、微小工程毎の目標位置Ｌｔを取得する。微小工程毎の目標位置Ｌｔが得られると、制御部４３は、対応関係Ｕ１に基づいて、ロボット座標系を規定する各軸の方向の動作位置を、各モーターＭ１～Ｍ６の目標の回転角度である目標角度Ｄｔに変換する。サーボ４３ｄは、パラメーター４４ａを参照してサーボゲインKpp，Kpi，Kpd，Kvp，Kvi，Kvdを取得し、目標角度Ｄｔに基づいて、制御量Ｄｃを導出する。制御量Ｄｃは、モーターＭ１～Ｍ６のそれぞれについて特定され、各モーターＭ１～Ｍ６の制御量Ｄｃでモーターｍ１～Ｍ６のそれぞれが制御される。この結果、各工程において、ＴＣＰは、微小工程毎の目標位置Ｌｔを経由し、加減速特性に従って始点から終点まで移動する。

【００６７】

なお、ハイブリッドモードでは、制御部４３は、微小工程毎の目標位置Ｌｔに、力由来補正量ΔＬを加算することにより動作位置（Ｌｔ＋ΔＬ）を特定し、当該動作位置に基づいて目標角度Ｄｔを取得し、制御量Ｄｃを取得する。

【００６８】

接触判定部４３ｃは、ロボット１～３が作業において想定されていない物体と接触したか否かを判定する機能を実行する。本実施形態において、接触判定部４３ｃは、ロボット１～３のそれぞれが備える力覚センサーＰの出力を取得し、出力が予め決められた基準値を超えた場合にロボット１～３が作業において想定されていない物体と接触したと判定する。この場合において、種々の処理が行われて良いが、本実施形態において接触判定部４３ｃは、ロボット１～３の制御量Ｄｃを０としてロボット１～３を停止させる。なお、停止させる際の制御量は、種々の制御量であって良く、直前の制御量Ｄｃをキャンセルする制御量でロボット１～３を動作させる構成等であっても良い。

【００６９】

（３）ピックアップ処理：

次に、以上の構成におけるロボット１～３の動作を説明する。ここでは、ロボット２の照明部２２で照明され、ロボット１の撮像部２１で撮像された対象物Ｗをロボット３のグリッパー２３でピックアップする作業を例にして説明する。むろん、ロボット１～３による作業は、ピックアップ作業に限定されず、他にも種々の作業（例えば、ネジ締め作業、挿入作業、ドリルによる穴あけ作業、バリ取り作業、研磨作業、組み立て作業、製品チェック作業等）に適用可能である。ピックアップ処理は、上述のコマンドによって記述されたロボット制御プログラムによって検出部４２および制御部４３が実行する処理によって実現される。本実施形態においてピックアップ処理は、作業台に対象物Ｗが配置した状態で実行される。

【００７０】

図５は、ピックアップ処理のフローチャートの例を示す図である。ピックアップ処理が開始されると、検出部４２は、撮像部２１が撮像した画像を取得する（ステップＳ１００）。すなわち、検出部４２は、パラメーター４４ａを参照して照明部２２の位置を特定し、当該位置を位置制御部４３ａに対して受け渡す。この結果、位置制御部４３ａは、現在の照明部２２の位置を始点、パラメーター４４ａが示す照明部２２の位置を終点とした位置制御を実行し、パラメーター４４ａが示す照明部の位置に照明部２２を移動させる。次に、検出部４２はパラメーター４４ａを参照して照明部２２の明るさを特定し、照明部２２を制御して照明の明るさを当該明るさに設定する。

【００７１】

さらに、検出部４２は、パラメーター４４ａを参照して撮像部２１の位置を特定し、当該位置を位置制御部４３ａに対して受け渡す。この結果、位置制御部４３ａは、現在の撮

像部２１の位置を始点、パラメーター４４ａが示す撮像部２１の位置を終点とした位置制御を実行し、パラメーター４４ａが示す照明部の位置に撮像部２１を移動させる。次に、検出部４２はパラメーター４４ａを参照して撮像部２１の露光時間および絞りを特定し、撮像部２１を制御して露光時間および絞りを当該露光時間および絞りに設定する。露光時間および絞りの設定が完了すると、撮像部２１は、画像を撮像し、検出部４２に対して出力する。検出部４２は、当該画像を取得する。

【００７２】

次に、検出部４２は、画像に基づいて、対象物の検出が成功したか否かを判定する（ステップＳ１０５）。すなわち、検出部４２は、パラメーター４４ａを参照して画像処理シーケンスを特定し、当該画像処理シーケンスが示す各処理をパラメーター４４ａが示す強度で実行する。また、検出部４２は、テンプレートデータ４４ｃを参照し、テンプレートデータ４４ｃと画像との差分を閾値と比較し、差分が閾値以下である場合に、対象物の検出が成功したと判定する。

【００７３】

ステップＳ１０５において、対象物の検出が成功したと判定されない場合、検出部４２は、テンプレートデータ４４ｃと画像の相対位置、またはテンプレートデータ４４ｃの大きさ、の少なくとも一方を変化させ、ステップＳ１００以降の処理を繰り返す。一方、ステップＳ１０５において、対象物の検出が成功したと判定された場合、制御部４３は、制御目標を特定する（ステップＳ１１０）。

【００７４】

本例におけるピックアップ処理は、検出部４２が検出した対象物Ｗの位置姿勢に合わせてロボット３のグリッパー２３を移動させ、姿勢を変化させ、グリッパー２３で対象物Ｗをピックアップし、所定の位置まで対象物Ｗを運んでグリッパー２３から対象物Ｗを離す作業である。そこで、位置制御部４３ａおよび力制御部４３ｂは、ロボットプログラム４４ｂに基づいて一連の作業を構成する複数の工程を特定する。

【００７５】

制御目標の特定対象となる工程は、各工程の中で未処理かつ時系列で先に存在する工程である。制御目標の特定対象となる工程が力制御モードの工程である場合、力制御部４３ｂは、パラメーター４４ａの力制御パラメーターを参照し、力制御座標系、目標力を取得する。力制御部４３ｂは、力制御座標系に基づいて、当該目標力をロボット座標系の目標力ｆLtに変換する。また、力制御部４３ｂは、力覚センサーＰの出力をＴＣＰに作用している作用力f_Lに変換する。さらに、力制御部４３ｂは、パラメーター４４ａの力制御パラメーターを参照し、インピーダンスパラメーターｍ，ｄ，ｋに基づいて、力由来補正量ΔＬを制御目標として取得する。

【００７６】

制御目標の特定対象となる工程が位置制御モードである場合、位置制御部４３ａは、当該工程を微小工程に細分化する。そして、位置制御部４３ａは、パラメーター４４ａの動作パラメーターを参照し、始点、終点、および加減速特性に基づいて、微小工程毎の目標位置Ｌtを制御目標として取得する。制御目標の特定対象となる工程がハイブリッドモードである場合、位置制御部４３ａは、当該工程を微小工程に細分化し、パラメーター４４ａの力制御パラメーターを参照し、始点、終点、および加減速特性に基づいて、微小工程毎の目標位置Ｌtを取得し、力制御座標系、目標力ｆLt、インピーダンスパラメーター、作用力f_Lに基づいて力由来補正量ΔＬを取得する。これらの目標位置Ｌtおよび力由来補正量ΔＬが制御目標である。

【００７７】

制御目標が特定されると、サーボ４３ｄは、現在の制御目標でロボット３を制御する（ステップＳ１１５）。すなわち、現在の工程が力制御モードまたはハイブリッドモードの工程である場合、サーボ４３ｄは、パラメーター４４ａの力制御パラメーターを参照し、サーボゲインに基づいて、制御目標に対応する制御量Ｄcを特定し、モーターＭ１～Ｍ６のそれぞれを制御する。現在の工程が位置制御モードの工程である場合、サーボ４３ｄは

、パラメーター４４ａの動作パラメーターを参照し、サーボゲインに基づいて、制御目標に対応する制御量Ｄｃを特定し、モーターＭ１～Ｍ６のそれぞれを制御する。

【００７８】

次に、制御部４３は、現在の工程が終了したか否かを判定する（ステップＳ１２０）。当該判定は、種々の終了判定条件によって実行されてよく、位置制御であれば、例えば、ＴＣＰが目標位置に達したことや目標位置においてＴＣＰが整定したこと等が挙げられる。力制御であれば、例えば、作用力が目標力に一致した状態から作用力が指定の大きさ以上、または指定の大きさ以下に変化したことや、ＴＣＰが指定の範囲外になったこと等が挙げられる。前者は、例えば、ピックアップ作業における対象物の把持動作の完了や、把持解除動作の完了等が挙げられる。後者は、例えば、ドリルによる対象物の貫通作業においてドリルが貫通した場合等が挙げられる。

【００７９】

むろん、他にも各工程が失敗したと推定される場合において、工程が終了したと判定されて良い。ただし、この場合には、作業の中止や中断が行われることが好ましい。工程の失敗を判定するための終了判定条件としては、例えば、ＴＣＰの移動速度や加速度が上限値を超えた場合やタイムアウトが発生した場合等が挙げられる。終了判定条件を充足したか否かは、各種のセンサー、力覚センサーＰや撮像部２１、他のセンサー等が利用されて良い。

【００８０】

ステップＳ１２０において、現在の工程が終了したと判定されない場合、制御部４３は、微小時間ΔＴ後に、次の微小工程についてステップＳ１１５以降の処理を実行する。すなわち、現在の工程が位置制御モードまたはハイブリッドモードである場合、位置制御部４３ａは、次の微小工程における目標位置Ｌｔを制御目標としてロボット３を制御する。また、現在の工程が力制御モードまたはハイブリッドモードである場合、力制御部４３ｂは、再度力覚センサーＰの出力に基づいて作用力f_Lを取得し、最新の作用力f_Lに基づいて特定される力由来補正量ΔＬを制御目標としてロボット３を制御する。

【００８１】

ステップＳ１２０において、現在の工程が終了したと判定された場合、制御部４３は、作業が終了したか否かを判定する（ステップＳ１２５）。すなわち、ステップＳ１２０で終了したと判定された工程が最終工程であった場合、制御部４３は、作業が終了したと判定する。ステップＳ１２５で作業が終了したと判定されなかった場合、制御部４３は、作業シーケンスの次の工程を現在の工程に変更し（ステップＳ１３０）、ステップＳ１１０以降の処理を実行する。ステップＳ１２５で作業が終了したと判定された場合、制御部４３は、作業が終了したと判定し、ピックアップ処理を終了する。

【００８２】

（４）学習処理：

本実施形態にかかる制御装置４０は、以上のように、パラメーター４４ａに基づいてロボット１～３を制御することができる。上述の実施形態において、パラメーター４４ａは教示によって生成されたが、人為的な教示によってパラメーター４４ａを最適化することは困難である。

【００８３】

例えば、検出部４２による対象物Ｗの検出において、同じ対象物Ｗであっても、光学パラメーターが異なれば対象物Ｗの位置、画像内での対象物の像や位置、対象物Ｗに生じる影など、様々な要素が変化し得る。従って、光学パラメーターを変化させると、検出部４２による対象物Ｗの検出精度が変化し得る。そして、光学パラメーターを変化させた場合に対象物Ｗの検出精度がどのように変化するのかは、必ずしも明らかではない。

【００８４】

また、動作パラメーターや力制御パラメーターはロボット１～３の制御に利用されるが、ロボット１～３のように複数の自由度（可動軸）を有するロボットは極めて多数のパターンで動作することが可能である。そして、ロボット１～３においては、振動や異音、オ

ーバーシュート等の好ましくない動作が発生しないようにパターンが決められている必要がある。さらに、エンドエフェクターとして各種の装置が取り付けられる場合、ロボット１～３の重心が変化し得るため、最適な動作パラメーター、力制御パラメーターも変化し得る。そして、動作パラメーターや力制御パラメーターを変化させた場合に、ロボット１～３の動作がどのように変化するのかは、必ずしも明らかではない。

【００８５】

さらに、力制御パラメーターは、ロボット１～３において力制御が行われる場合に利用されるが、ロボット１～３において実施される各作業において、力制御パラメーターを変化させた場合に、ロボット１～３の動作がどのように変化するのかは、必ずしも明らかではない。例えば、どのような方向においてどのようなインピーダンスパラメーターが最適であるのか、全ての作業工程において推定することは困難である。このため、検出部４２の検出精度を高めたり、ロボット１～３の潜在的な性能を引き出したりするためには極めて多数の試行錯誤を行う必要がある。

【００８６】

しかし、人為的に極めて多数の試行錯誤を行うことは困難であるため、対象物Ｗの検出精度が充分に高く、当該検出精度がほぼ上限に達していると推定される状態や、ロボット１～３の潜在的な性能が引き出されている状態（所要時間や消費電力等のパフォーマンスのさらなる向上が困難な状態）を人為的に実現することは困難である。また、パラメーター４４ａの調整を行うためには、パラメーター４４ａの変化による検出精度の変化やロボット１～３の動作の変化を熟知しているオペレーターが必要になり、熟知していないオペレーターがパラメーター４４ａの調整を行うことは困難である。また、常に熟練のオペレーターを必要とするシステムはとても不便である。

【００８７】

そこで、本実施形態においては、人為的なパラメーター４４ａの決定作業を行うことなく、自動的にパラメーター４４ａを決定するための構成を備えている。なお、本実施形態によれば、多少のパラメーター４４ａの変更によって検出精度がより向上しないと推定される（検出精度が極大であると推定される）状態や、多少のパラメーター４４ａの変更によってロボット１～３の性能が高性能化することはないと推定される（性能が極大であると推定される）状態を実現することができる。本実施形態においては、これらの状態を最適化された状態と呼ぶ。

【００８８】

本実施形態において制御装置４０は、パラメーター４４ａの自動的な決定のために算出部４１を備えている。本実施形態において、算出部４１は、機械学習を用いて、光学パラメーターと動作パラメーターと力制御パラメーターとを算出することができる。図６は、算出部４１の構成を示す図であり、図２に示す構成の一部を省略し、算出部４１の詳細を示した図である。なお、図６に示す記憶部４４は、図２に示す記憶部４４と同一の記憶媒体であり、各図においては記憶された情報の一部の図示が省略されている。

【００８９】

算出部４１は、状態変数を観測する状態観測部４１ａと、観測された状態変数に基づいてパラメーター４４ａを学習する学習部４１ｂとを備えている。本実施形態において、状態観測部４１ａは、パラメーター４４ａを変化させたことによって生じた結果を状態変数として観測する。このため、状態観測部４１ａは、サーボ４３ｄの制御結果と、エンコーダーＥ１～Ｅ６の値と、力覚センサーＰの出力と、検出部４２が取得する画像とを状態変数として取得可能である。

【００９０】

具体的には、状態観測部４１ａは、サーボ４３ｄの制御結果として、モーターＭ１～Ｍ６に供給される電流値を観測する。当該電流値は、モーターＭ１～Ｍ６で出力されるトルクに相当する。エンコーダーＥ１～Ｅ６の出力は、対応関係Ｕ１に基づいてロボット座標系におけるＴＣＰの位置に変換される。従って、状態観測部４１ａは、ロボット１であれば撮像部２１の位置、ロボット２であれば照明部２２の位置、ロボット３であればグリッ

パー２３の位置を観測することができる。

【００９１】

力覚センサーＰの出力は、対応関係Ｕ２に基づいてロボット座標系におけるＴＣＰに作用する作用力に変換される。従って、状態観測部４１ａは、ロボット１～３への作用力を状態変数として観測することができる。検出部４２が取得する画像は、撮像部２１で撮像された画像であり、状態観測部４１ａは、当該画像を状態変数として観測することができる。状態観測部４１ａは、学習対象のパラメーター４４ａに応じて観測対象の状態変数を適宜選択することができる。

【００９２】

学習部４１ｂは、学習によってパラメーター４４ａを最適化することができればよく、本実施形態においては、強化学習によってパラメーター４４ａを最適化する。具体的には、学習部４１ｂは、状態変数に基づいてパラメーター４４ａを変化させる行動を決定し、当該行動を実行する。当該行動後の状態に応じて報酬を評価すれば、当該行動の行動価値が判明する。そこで、算出部４１は、状態変数の観測と、当該状態変数に応じた行動の決定と、当該行動によって得られる報酬の評価とを繰り返すことによって、パラメーター４４ａを最適化する。

【００９３】

本実施形態において、算出部４１は、パラメーター４４ａの中から学習対象のパラメーターを選択して学習を行うことができる。本実施形態においては、光学パラメーターの学習と、動作パラメーターの学習と、力制御パラメーターの学習とのそれぞれを独立して実行することができる。

【００９４】

（４－１）光学パラメーターの学習：

図７はエージェントと環境とからなる強化学習のモデルに沿って光学パラメーターの学習例を説明する図である。図７に示すエージェントは、予め決められた方策に応じて行動ａを選択する機能に相当し、学習部４１ｂによって実現される。環境は、エージェントが選択した行動ａと現在の状態ｓとに基づいて次の状態ｓ'を決定し、行動ａと状態ｓと状態ｓ'とに基づいて即時報酬ｒを決定する機能に相当し、状態観測部４１ａおよび学習部４１ｂによって実現される。

【００９５】

本実施形態においては、予め決められた方策によって学習部４１ｂが行動ａを選択し、状態観測部４１ａが状態の更新を行う処理を繰り返すことにより、ある状態ｓにおけるある行動ａの行動価値関数Ｑ（ｓ，ａ）を算出するＱ学習が採用される。すなわち、本例においては、下記の式（２）によって行動価値関数を更新する。そして、行動価値関数Ｑ（ｓ，ａ）が適正に収束した場合には、当該行動価値関数Ｑ（ｓ，ａ）を最大化する行動ａが最適な行動であると見なされ、当該行動ａを示すパラメーター４４ａが最適化されたパラメーターであると見なされる。

【数２】

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)) \cdots (2)$$

【００９６】

ここで、行動価値関数Ｑ（ｓ，ａ）は、状態ｓにおいて行動ａを取った場合において将来にわたって得られる収益（本例では割引報酬総和）の期待値である。報酬はｒであり、状態ｓ、行動ａ、報酬ｒの添え字ｔは、時系列で繰り返す試行過程における１回分のステップを示す番号（試行番号と呼ぶ）であり、行動決定後に状態が変化すると試行番号がインクリメントされる。従って、式（２）内の報酬r_{t+1}は状態s_tで行動a_tが選択され、状態がs_{t+1}になった場合に得られる報酬である。αは学習率、γは割引率である。また、ａ'は、状態s_{t+1}で取り得る行動a_{t+1}の中で行動価値関数Ｑ（s_{t+1}，a_{t+1}）を最大化する行動であり、$\max_{a'}Q(s_{t+1}, a')$は、行動ａ'が選択されたことによって最大化された行動価値関数である。

【００９７】
光学パラメーターの学習においては、光学パラメーターを変化させることが行動の決定に相当しており、学習対象のパラメーターと取り得る行動とを示す行動情報４４ｄが記憶部４４に予め記録される。すなわち、当該行動情報４４ｄに学習対象として記述された光学パラメーターが学習対象となる。図７においては、光学パラメーターの中の撮像部パラメーターと、照明部パラメーターと、画像処理パラメーターとの一部が学習対象となっている例を示している。
【００９８】
具体的には、撮像部パラメーターの中で撮像部２１のｘ座標、ｙ座標が学習対象となっている。従って、この例においてｚ座標やｘｙｚ軸に対する回転（姿勢）は学習対象となっておらず、撮像部２１は、対象物Ｗが置かれる作業台に向いている状態であるとともに、撮像部２１のｘ－ｙ平面内での移動が学習対象である。むろん、他の撮像部パラメーター、例えば、撮像部２１の姿勢やｚ座標、露光時間や絞りが学習対象であっても良い。
【００９９】
また、図７に示す例においては、照明部パラメーターの中で、照明部２２のｘ座標、ｙ座標および照明部の明るさが学習対象となっている。従って、この例においてｚ座標やｘｙｚ軸に対する回転（姿勢）は学習対象となっておらず、照明部２２は、対象物Ｗが置かれる作業台に向いている状態であるとともに、照明部２２のｘ－ｙ平面内での移動が学習対象である。むろん、他の照明部パラメーター、例えば、照明部２２の姿勢やｚ座標が学習対象であっても良い。
【０１００】
さらに、図７に示す例においては、画像処理パラメーターの中で、平滑化処理の強度と鮮鋭化処理の強度とテンプレートマッチングの閾値が学習対象となっている。従って、この例において、画像処理シーケンスは学習対象となっておらず、撮像部２１で撮像された画像に対する画像処理の順序は変化しない（むろん、画像処理シーケンスが学習対象である実施形態も採用可能である）。
【０１０１】
図７に示す例において行動には値を一定値増加させる行動と、値を一定値減少させる行動とが存在する。従って、図７に示す全８個のパラメーターにおいて取り得る行動は全１６個である（行動ａ１～行動ａ１６）。行動情報４４ｄは、学習対象のパラメーターと取り得る行動とを示しているため、図７に示す例であれば、図示した８個のパラメーターが行動情報４４ｄに学習対象として記述される。また、各行動を特定するための情報（行動のＩＤ、各行動での増減量等）が行動情報４４ｄに記述される。
【０１０２】
図７に示す例において、報酬は対象物Ｗの検出の成否に基づいて特定される。すなわち、学習部４１ｂは、行動ａとして光学パラメーターを変化させた後、当該光学パラメーターによってロボット１，２を動作させ、検出部４２によって撮像部２１が撮像した画像を取得する。そして、学習部４１ｂは、当該光学パラメーターに基づいてテンプレートマッチング処理を実行し、対象物Ｗの検出が成功したか否かを判定する。さらに、学習部４１ｂは、検出の成否によって行動ａ、状態ｓ、ｓ'の報酬を決定する。当該報酬は、対象物Ｗの検出の成否に基づいて決定されれば良く、例えば、検出の成功に正（例えば＋１）、検出の失敗に負（例えば－１）の報酬を与える構成等を採用可能である。この構成によれば、対象物の検出精度を高めるように最適化を行うことができる。
【０１０３】
現在の状態ｓにおいて行動ａが採用された場合における次の状態ｓ'は、行動ａとしてのパラメーターの変化が行われた後にロボット１，２を動作させ、状態観測部４１ａが状態を観測することによって特定可能である。なお、本例にかかる光学パラメーターの学習においてロボット３は動作しない。図７に示す例において、状態変数には、撮像部２１のｘ座標、ｙ座標と、照明部２２のｘ座用、ｙ座標、照明部２２の明るさと、平滑化処理の強度、鮮鋭化処理の強度、テンプレートマッチングの閾値と、撮像部２１で撮像された画

像とが含まれている。
【０１０４】
従って、この例において、状態観測部４１ａは、行動ａが実行された後に、ロボット１のエンコーダーＥ１～Ｅ６の出力をＵ１に基づいて変換して撮像部２１のｘ座標およびｙ座標を観測する。また、状態観測部４１ａは、行動ａが実行された後に、ロボット２のエンコーダーＥ１～Ｅ６の出力をＵ１に基づいて変換して照明部２２のｘ座標およびｙ座標を観測する。
【０１０５】
本実施形態において、照明部２２の明るさは、パラメーター４４ａによって誤差無く調整可能であると見なされており（または誤差が影響ないと見なされており）、状態観測部４１ａは、パラメーター４４ａに含まれる照明部の明るさを取得して状態変数が観測されたと見なす。むろん、照明部２２の明るさは、センサー等によって実測されても良いし、撮像部２１が撮像した画像に基づいて（例えば、平均階調値等により）観測されても良い。状態観測部４１ａは、平滑化処理の強度、鮮鋭化処理の強度、テンプレートマッチングの閾値についても、パラメーター４４ａを参照して現在の値を取得し、状態変数が観測されたと見なす。
【０１０６】
さらに、状態観測部４１ａにおいては、撮像部２１が撮像し、検出部４２が取得した画像を状態変数として取得する（図７に示す太枠）。すなわち、状態観測部４１ａは、撮像部２１が撮像した画像（対象物が存在し得る注目領域等の画像であっても良い）の画素毎の階調値を状態変数として観測する。撮像部のｘ座標等は、行動であるとともに観測対象としての状態であるが、撮像部２１が撮像した画像は行動ではない。従って、この意味で、撮像された画像は、光学パラメーターの変化から直接的に推定することが困難な変化をし得る状態変数である。また、検出部４２は、当該画像に基づいて対象物を検出するため、当該画像は検出の成否に直接的に影響を与え得る状態変数である。従って、状態変数として、当該画像を観測することにより、人為的に改善することが困難なパラメーターの改善を行い、効果的に検出部４２の検出精度を高めるように光学パラメーターを最適化することが可能になる。
【０１０７】
（４－２）光学パラメーターの学習例：
次に、光学パラメーターの学習例を説明する。学習の過程で参照される変数や関数を示す情報は、学習情報４４ｅとして記憶部４４に記憶される。すなわち、算出部４１は、状態変数の観測と、当該状態変数に応じた行動の決定と、当該行動によって得られる報酬の評価とを繰り返すことによって行動価値関数Ｑ（ｓ，ａ）を収束させる構成が採用されている。そこで、本例において、学習の過程で状態変数と行動と報酬との時系列の値が、順次、学習情報４４ｅに記録されていく。
【０１０８】
行動価値関数Ｑ（ｓ，ａ）は、種々の手法で算出されて良く、多数回の試行に基づいて算出されても良いが、本実施形態においては、行動価値関数Ｑ（ｓ，ａ）を近似的に算出する一手法であるＤＱＮ（Ｄｅｅｐ　Ｑ－Ｎｅｔｗｏｒｋ）が採用されている。ＤＱＮにおいては、多層ニューラルネットワークを用いて行動価値関数Ｑ（ｓ，ａ）を推定する。本例においては、状態ｓを入力とし、選択し得る行動の数Ｎ個の行動価値関数Ｑ（ｓ，ａ）の値を出力とする多層ニューラルネットワークが採用されている。
【０１０９】
図８は、本例において採用されている多層ニューラルネットワークを模式的に示す図である。図８において、多層ニューラルネットワークは、Ｍ個（Ｍは２以上の整数）の状態変数を入力とし、Ｎ個（Ｎは２以上の整数）個の行動価値関数Ｑの値を出力としている。例えば、図７に示す例であれば、撮像部のｘ座標～テンプレートマッチングの閾値までの８個の状態変数と撮像された画像の画素数との和がＭ個であり、Ｍ個の状態変数の値が多層ニューラルネットワークに入力される。図８においては、試行番号ｔにおけるＭ個の状

態をs_{1t}～s_{Mt}として示している。

【０１１０】

Ｎ個は選択し得る行動ａの数であり、多層ニューラルネットワークの出力は、入力された状態ｓにおいて特定の行動ａが選択された場合の行動価値関数Ｑの値である。図８においては、試行番号ｔにおいて選択し得る行動a_{1t}～a_{Nt}のそれぞれにおける行動価値関数ＱをＱ（s_t，a_{1t}）～Ｑ（s_t，a_{Nt}）として示している。当該Ｑに含まれるs_tは入力された状態s_{1t}～s_{Mt}を代表して示す文字である。図７に示す例であれば、１６個の行動が選択可能であるためＮ＝１６である。むろん、行動ａの内容や数（Ｎの値）、状態ｓの内容や数（Ｍの値）は試行番号ｔに応じて変化しても良い。

【０１１１】

図８に示す多層ニューラルネットワークは、各層の各ノードにおいて直前の層の入力（１層目においては状態ｓ）に対する重みｗの乗算とバイアスｂの加算とを実行し、必要に応じて活性化関数を経た出力を得る（次の層の入力になる）演算を実行するモデルである。本例においては、層ＤＬがＰ個（Ｐは１以上の整数）存在し、各層において複数のノードが存在する。

【０１１２】

図８に示す多層ニューラルネットワークは各層における重み、とバイアスｂ、活性化関数、層の順序等によって特定される。そこで、本実施形態においては、当該多層ニューラルネットワークを特定するためのパラメーター（入力から出力を得るために必要な情報）が学習情報４４ｅとして記憶部４４に記録される。なお、学習の際には、多層ニューラルネットワークを特定するためのパラメーターの中で可変の値（例えば，重みｗとバイアスｂ）を更新していくことになる。ここでは、学習の過程で変化し得る多層ニューラルネットワークのパラメーターをθと表記する。当該θを使用すると、上述の行動価値関数Ｑ（s_t，a_{1t}）～Ｑ（s_t，a_{Nt}）は、Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）とも表記できる。

【０１１３】

次に、図９に示すフローチャートに沿って学習処理の手順を説明する。光学パラメーターの学習処理は、ロボット１，２の運用過程において実施されても良いし、実運用の前に事前に学習処理が実行されてもよい。ここでは、実運用の前に事前に学習処理が実行される構成（多層ニューラルネットワークを示すθが最適化されると、その情報が保存され、次回以降の運用で利用される構成）に従って学習処理を説明する。

【０１１４】

学習処理が開始されると、算出部４１は、学習情報４４ｅを初期化する（ステップＳ２００）。すなわち、算出部４１は、学習を開始する際に参照されるθの初期値を特定する。初期値は、種々の手法によって決められて良く、過去に学習が行われていない場合においては、任意の値やランダム値等がθの初期値となっても良いし、ロボット１，２や撮像部２１，照明部２２の光学特性を模擬するシミュレーション環境を準備し、当該環境に基づいて学習または推定したθを初期値としてもよい。

【０１１５】

過去に学習が行われた場合は、当該学習済のθが初期値として採用される。また、過去に類似の対象についての学習が行われた場合は、当該学習におけるθが初期値とされても良い。過去の学習は、ロボット１，２を用いてユーザーが行ってもよいし、ロボット１，２の製造者がロボット１，２の販売前に行ってもよい。この場合、製造者は、対象物や作業の種類に応じて複数の初期値のセットを用意しておき、ユーザーが学習する際に初期値を選択する構成であっても良い。θの初期値が決定されると、当該初期値が現在のθの値として学習情報４４ｅに記憶される。

【０１１６】

次に、算出部４１は、パラメーターを初期化する（ステップＳ２０５）。ここでは、光学パラメーターが学習対象であるため、算出部４１は、光学パラメーターを初期化する。すなわち、算出部４１は、ロボット１のエンコーダーＥ１～Ｅ６の出力を対応関係Ｕ１で

変換し、撮像部２１の位置を初期値として設定する。また、算出部４１は、予め決められた初期の露光時間（過去に学習が行われた場合には最新の露光時間）を撮像部２１の露光時間の初期値として設定する。さらに、算出部４１は、撮像部２１に制御信号を出力し、現在の絞りの値を初期値として設定する。

【０１１７】

さらに、算出部４１は、ロボット２のエンコーダーＥ１～Ｅ６の出力を対応関係Ｕ１で変換し、照明部２２の位置を初期値として設定する。また、算出部４１は、予め決められた初期の明るさ（過去に学習が行われた場合には最新の明るさ）を照明部２２の明るさの初期値として設定する。さらに、算出部４１は、平滑化処理の強度、鮮鋭化処理の強度、テンプレートマッチングの閾値、画像処理シーケンスについて予め決められた初期値（過去に学習が行われた場合には最新の値）を設定する。初期化されたパラメーターは記憶部４４に現在のパラメーター４４ａとして記憶される。

【０１１８】

次に、状態観測部４１ａは、状態変数を観測する（ステップＳ２１０）。すなわち、制御部４３は、パラメーター４４ａおよびロボットプログラム４４ｂを参照してロボット１，２を制御する。検出部４２は、制御後の状態で撮像部２１が撮像した画像に基づいて対象物Ｗの検出処理（上述のステップＳ１００，Ｓ１０５に相当）を実行する。この後、状態観測部４１ａは、ロボット１のエンコーダーＥ１～Ｅ６の出力をＵ１に基づいて変換して撮像部２１のｘ座標およびｙ座標を観測する。また、状態観測部４１ａは、ロボット２のエンコーダーＥ１～Ｅ６の出力をＵ１に基づいて変換して照明部２２のｘ座標およびｙ座標を観測する。さらに、状態観測部４１ａは、パラメーター４４ａを参照して照明部２２に設定されるべき明るさを取得して状態変数が観測されたと見なす。

【０１１９】

さらに、状態観測部４１ａは、平滑化処理の強度、鮮鋭化処理の強度、テンプレートマッチングの閾値についても、パラメーター４４ａを参照して現在の値を取得し、状態変数が観測されたと見なす。さらに、状態観測部４１ａにおいては、撮像部２１が撮像し、検出部４２が取得した画像を取得し、各画素の階調値を状態変数として取得する。

【０１２０】

次に、学習部４１ｂは、行動価値を算出する（ステップＳ２１５）。すなわち、学習部４１ｂは、学習情報４４ｅを参照してθを取得し、学習情報４４ｅが示す多層ニューラルネットワークに最新の状態変数を入力し、Ｎ個の行動価値関数Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）を算出する。

【０１２１】

なお、最新の状態変数は、初回の実行時においてステップＳ２１０、２回目以降の実行時においてステップＳ２２５の観測結果である。また、試行番号ｔは初回の実行時において０、２回目以降の実行時において１以上の値となる。学習処理が過去に実施されていない場合、学習情報４４ｅが示すθは最適化されていないため、行動価値関数Ｑの値としては不正確な値となり得るが、ステップＳ２１５以後の処理の繰り返しにより、行動価値関数Ｑは徐々に最適化していく。また、ステップＳ２１５以後の処理の繰り返しにおいて、状態ｓ、行動ａ、報酬ｒは、各試行番号ｔに対応づけられて記憶部４４に記憶され、任意のタイミングで参照可能である。

【０１２２】

次に、学習部４１ｂは、行動を選択し、実行する（ステップＳ２２０）。本実施形態においては、行動価値関数Ｑ（ｓ，ａ）を最大化する行動ａが最適な行動であると見なされる処理が行われる。そこで、学習部４１ｂは、ステップＳ２１５において算出されたＮ個の行動価値関数Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）の値の中で最大の値を特定する。そして、学習部４１ｂは、最大の値を与えた行動を選択する。例えば、Ｎ個の行動価値関数Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）の中でＱ（s_t，a_{Nt}；θ_t）が最大値であれば、学習部４１ｂは、行動a_{Nt}を選択する。

【０１２３】

行動が選択されると、学習部４１ｂは、当該行動に対応するパラメーター４４ａを変化させる。例えば、図７に示す例において、撮像部のｘ座標を一定値増加させる行動ａ１が選択された場合、学習部４１ｂは、光学パラメーターの撮像部パラメーターが示す撮像部の位置においてｘ座標を一定値増加させる。パラメーター４４ａの変化が行われると、制御部４３は、当該パラメーター４４ａを参照してロボット１，２を制御する。検出部４２は、制御後の状態で撮像部２１が撮像した画像に基づいて対象物Ｗの検出処理を実行する。

【０１２４】

次に、状態観測部４１ａは、状態変数を観測する（ステップＳ２２５）。すなわち、状態観測部４１ａは、ステップＳ２１０における状態変数の観測と同様の処理を行って、状態変数として、撮像部２１のｘ座標およびｙ座標、照明部２２のｘ座標およびｙ座標、照明部２２に設定されるべき明るさ、平滑化処理の強度、鮮鋭化処理の強度、テンプレートマッチングの閾値、撮像部２１が撮像した画像の各画素の階調値を取得する。なお、現在の試行番号がｔである場合（選択された行動がa_tである場合）、ステップＳ２２５で取得される状態ｓはs_{t+1}である。

【０１２５】

次に、学習部４１ｂは、報酬を評価する（ステップＳ２３０）。本例において、報酬は、対象物Ｗの検出の成否に基づいて決定される。そこで、学習部４１ｂは、検出部４２から対象物の検出結果の成否（ステップＳ１０５の成否）を取得し、検出成功であれば既定量の正の報酬、検出失敗であれば既定量の負の報酬を取得する。なお、現在の試行番号がｔである場合、ステップＳ２３０で取得される報酬ｒはr_{t+1}である。

【０１２６】

本実施形態においては式（２）に示す行動価値関数Ｑの更新を目指しているが、行動価値関数Ｑを適切に更新していくためには、行動価値関数Ｑを示す多層ニューラルネットワークを最適化（θを最適化）していかなくてはならない。図８に示す多層ニューラルネットワークによって行動価値関数Ｑを適正に出力させるためには、当該出力のターゲットとなる教師データが必要になる。すなわち、多層ニューラルネットワークの出力と、ターゲットとの誤差を最小化するようにθを改善することによって、多層ニューラルネットワークが最適化されることが期待される。

【０１２７】

しかし、本実施形態において、学習が完了していない段階では行動価値関数Ｑの知見がなく、ターゲットを特定することは困難である。そこで、本実施形態においては、式（２）の第２項、いわゆるＴＤ誤差（Ｔｅｍｐｏｒａｌ　Ｄｉｆｆｅｒｅｎｃｅ）を最小化する目的関数によって多層ニューラルネットワークを示すθの改善を実施する。すなわち、$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta_t))$をターゲットとし、ターゲットと$Q(s_t, a_t; \theta_t)$との誤差が最小化するようにθを学習する。ただし、ターゲット$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta_t))$は、学習対象のθを含んでいるため、本実施形態においては、ある程度の試行回数にわたりターゲットを固定する（例えば、最後に学習したθ（初回学習時はθの初期値）で固定する）。本実施形態においては、ターゲットを固定する試行回数である既定回数が予め決められている。

【０１２８】

このような前提で学習を行うため、ステップＳ２３０で報酬が評価されると、学習部４１ｂは目的関数を算出する（ステップＳ２３５）。すなわち、学習部４１ｂは、試行のそれぞれにおけるＴＤ誤差を評価するための目的関数（例えば、ＴＤ誤差の２乗の期待値に比例する関数やＴＤ誤差の２乗の総和等）を算出する。なお、ＴＤ誤差は、ターゲットが固定された状態で算出されるため、固定されたターゲットを$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta^-))$と表記すると、ＴＤ誤差は$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) - Q(s_t, a_t; \theta_t))$である。当該ＴＤ誤差の式において報酬$r_{t+1}$は、行動$a_t$によってステップＳ２３０で得られた報酬である。

【０１２９】

また、$max_{a'}Q(s_{t+1}, a'; \theta_-)$は、行動$a_t$によってステップＳ２２５で算出される状態$s_{t+1}$を、固定された$\theta_-$で特定される多層ニューラルネットワークの入力とした場合に得られる出力の中の最大値である。$Q(s_t, a_t; \theta_t)$は、行動a_tが選択される前の状態s_tを、試行番号ｔの段階のθ_tで特定される多層ニューラルネットワークの入力とした場合に得られる出力の中で、行動a_tに対応した出力の値である。

【０１３０】

目的関数が算出されると、学習部４１ｂは、学習が終了したか否か判定する（ステップＳ２４０）。本実施形態においては、ＴＤ誤差が充分に小さいか否かを判定するための閾値が予め決められており、目的関数が閾値以下である場合、学習部４１ｂは、学習が終了したと判定する。

【０１３１】

ステップＳ２４０において学習が終了したと判定されない場合、学習部４１ｂは、行動価値を更新する（ステップＳ２４５）。すなわち、学習部４１ｂは、ＴＤ誤差のθによる偏微分に基づいて目的関数を小さくするためのθの変化を特定し、θを変化させる。むろん、ここでは、各種の手法でθを変化させることが可能であり、例えば、ＲＭＳＰｒｏｐ等の勾配降下法を採用可能である。また、学習率等による調整も適宜実施されて良い。以上の処理によれば、行動価値関数Ｑがターゲットに近づくようにθを変化させることができる。

【０１３２】

ただし、本実施形態においては、上述のようにターゲットが固定されているため、学習部４１ｂは、さらに、ターゲットを更新するか否かの判定を行う。具体的には学習部４１ｂは、既定回数の試行が行われたか否かを判定し（ステップＳ２５０）、ステップＳ２５０において、既定回数の試行が行われたと判定された場合に、学習部４１ｂは、ターゲットを更新する（ステップＳ２５５）。すなわち、学習部４１ｂは、ターゲットを算出する際に参照されるθを最新のθに更新する。この後、学習部４１ｂは、ステップＳ２１５以降の処理を繰り返す。一方、ステップＳ２５０において、既定回数の試行が行われたと判定されなければ、学習部４１ｂは、ステップＳ２５５をスキップしてステップＳ２１５以降の処理を繰り返す。

【０１３３】

ステップＳ２４０において学習が終了したと判定された場合、学習部４１ｂは、学習情報４４ｅを更新する（ステップＳ２６０）。すなわち、学習部４１ｂは、学習によって得られたθを、ロボット１，２による作業や検出部４２による検出の際に参照されるべきθとして学習情報４４ｅに記録する。当該θを含む学習情報４４ｅが記録されている場合、ステップ１００～Ｓ１０５のようにロボット１，２による作業が行われる際に、検出部４２はパラメーター４４ａに基づいて対象物の検出処理を行う。そして、検出部４２による検出が成功するまで、撮像部２１による撮像が繰り返される工程においては、状態観測部４１ａによる現在の状態の観測と、学習部４１ｂによる行動の選択が繰り返される。むろん、この際、学習部４１ｂは、状態を入力として算出された出力Ｑ（ｓ，ａ）の中で最大値を与える行動ａを選択する。そして、行動ａが選択された場合、行動ａが行われた状態に相当する値となるようにパラメーター４４ａが更新される。

【０１３４】

以上の構成によれば、検出部４２は、行動価値関数Ｑが最大化される行動ａを選択しながら対象物の検出処理を実行することができる。当該行動価値関数Ｑは、上述の処理により、多数の試行が繰り返された結果、最適化されている。そして、当該試行は、算出部４１によって自動で行われ、人為的に実施不可能な程度の多数の試行を容易に実行することができる。従って、本実施形態によれば、人為的に決められた光学パラメーターよりも高い確率で対象物を高精度に検出することができる。

【０１３５】

さらに、本実施形態において検出部４２は、対象物の位置姿勢を検出する構成であるため、本実施形態によれば高精度に対象物の位置姿勢を検出することができる。さらに、本

実施形態によれば、最適化された行動価値関数Ｑに基づいて、光学パラメーターである撮像部パラメーターを算出することができる。従って、対象物の検出精度を高めるように撮像部２１を調整することができる。さらに、本実施形態によれば、最適化された行動価値関数Ｑに基づいて、光学パラメーターである照明部パラメーターを算出することができる。従って、対象物の検出精度を高めるように照明部２２を調整することができる。

【０１３６】

さらに、本実施形態によれば、最適化された行動価値関数Ｑに基づいて、光学パラメーターである画像処理パラメーターを算出することができる。従って、対象物の検出精度を高める画像処理を実行することが可能になる。さらに、本実施形態によれば、自動で行動価値関数Ｑが最適化されるため、高精度に対象物を検出する光学パラメーターを容易に算出することができる。また、行動価値関数Ｑの最適化は自動的に行われるため、最適な光学パラメーターの算出も自動的に行うことができる。

【０１３７】

さらに、本実施形態において学習部４１ｂは、状態変数としての画像に基づいて光学パラメーターを変化させる行動を決定し、光学パラメーターを最適化する。従って、照明部２２によって照明が行われている実環境下において撮像部２１で実際に撮像した画像に基づいて光学パラメーターを最適化することができる。従って、ロボット１，２の使用環境に応じた光学パラメーターとなるように最適化することができる。

【０１３８】

本実施形態においては、撮像部２１の位置および照明部２２の位置が行動に含まれており、当該行動に基づいて行動価値関数Ｑを最適化することで撮像部２１の位置および照明部２２の位置に関するパラメーター４４ａを最適化することができる。従って、学習後においては、少なくとも、撮像部２１と照明部２２の相対位置関係が理想化される。また、対象物Ｗが作業台の固定位置またはほぼ固定された位置に置かれるのならば、学習後において、撮像部２１と照明部２２のロボット座標系における位置が理想化されると考えることもできる。さらに、本実施形態においては、撮像部２１によって撮像された画像が状態として観測される。従って、本実施形態によれば、各種の画像の状態に対応した撮像部２１の位置や照明部２２の位置が理想化される。

【０１３９】

（４－３）動作パラメーターの学習：

動作パラメーターの学習においても、学習対象のパラメーターを選択することが可能であり、ここでは、その一例を説明する。図１０は、動作パラメーターの学習例を図７と同様のモデルで説明した図である。本例も式（２）に基づいて行動価値関数Ｑ（ｓ，ａ）を最適化する。従って、最適化後の行動価値関数Ｑ（ｓ，ａ）を最大化する行動ａが最適な行動であると見なされ、当該行動ａを示すパラメーター４４ａが最適化されたパラメーターであると見なされる。

【０１４０】

動作パラメーターの学習においても、動作パラメーターを変化させることが行動の決定に相当しており、学習対象のパラメーターと取り得る行動とを示す行動情報４４ｄが記憶部４４に予め記録される。すなわち、当該行動情報４４ｄに学習対象として記述された動作パラメーターが学習対象となる。図１０においては、ロボット３における動作パラメーターの中のサーボゲインと加減速特性が学習対象であり、動作の始点および終点は学習対象となっていない。なお、動作の始点および終点は教示位置であるが、本実施形態においては他の位置は教示されない。従って、本実施形態においては、ロボット３に対して教示された教示位置を含まない構成である。

【０１４１】

具体的には、動作パラメーターの中のサーボゲインＫpp，Ｋpi，Ｋpd，Ｋvp，Ｋvi，Ｋvdは、モーターＭ１～Ｍ６のそれぞれについて定義され、６軸のそれぞれについて増減可能である。従って、本実施形態においては、１軸あたり６個のサーボゲインのそれぞれを増加または減少させることが可能であり、増加について３６個の行動、減少についても３

６個の行動、計７２個の行動（行動ａ１～ａ７２）を選択し得る。

【０１４２】

一方、動作パラメーターの中の加減速特性は図４に示すような特性であり、モーターＭ１～Ｍ６のそれぞれについて（６軸について）定義される。本実施形態において加減速特性は、加速域における加速度、減速域における加速度、車速が０より大きい期間の長さ（図４に示すt_4）を変化させることができる。なお、本実施形態において加速域や減速域におけるカーブは加速度の増減によって定義され、例えば、増減後の加速度がカーブ中央の傾きを示し、当該中央の周囲のカーブは予め決められた規則に従って変化する。むろん、加減速特性の調整法は他にも種々の手法が採用可能である。

【０１４３】

いずれにしても、本実施形態においては、１軸あたり３個の要素（加速域、減速域、期間）で加減速特性を調整可能であり、各要素に応じた数値（加速度、期間長）を増加または減少させることが可能である。従って、増加について１８個の行動、減少についても１８個の行動、計３６個の行動（行動ａ７３～ａ１０８）を選択し得る。本実施形態においては、以上のようにして予め定義された行動の選択肢に対応するパラメーターが、行動情報４４ｄに学習対象として記述される。また、各行動を特定するための情報（行動のＩＤ、各行動での増減量等）が行動情報４４ｄに記述される。

【０１４４】

図１０に示す例において、報酬はロボット３が行った作業の良否に基づいて評価される。すなわち、学習部４１ｂは、行動ａとして動作パラメーターを変化させた後、当該動作パラメーターによってロボット３を動作させ、検出部４２によって検出された対象物をピックアップする作業を実行する。さらに、学習部４１ｂは、作業の良否を観測し、作業の良否を評価する。そして、学習部４１ｂは、作業の良否によって行動ａ、状態ｓ、ｓ'の報酬を決定する。

【０１４５】

なお、作業の良否は作業の成否（ピックアップ成否等）のみならず、作業の質を含む。具体的には、学習部４１ｂは、図示しない計時回路に基づいて作業の開始から終了まで（ステップＳ１１０の開始からステップＳ１２５で終了と判定されるまで）の所要時間を取得する。そして、学習部４１ｂは、作業の所要時間が基準よりも短い場合に正（例えば＋１）、作業の所要時間が基準よりも長い場合に負（例えば－１）の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回の作業の所要時間であっても良いし、過去の最短所要時間であっても良いし、予め決められた時間であっても良い。

【０１４６】

さらに、学習部４１ｂは、作業の各工程において、ロボット３のエンコーダーＥ１～Ｅ６の出力をＵ１に基づいて変換してグリッパー２３の位置を取得する。そして、学習部４１ｂは、各工程の目標位置（終点）と、工程終了の際のグリッパー２３の位置とのずれ量を取得し、グリッパー２３の位置と目標位置とのずれ量が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回のずれ量であっても良いし、過去の最短のずれ量であっても良いし、予め決められたずれ量であっても良い。

【０１４７】

さらに、学習部４１ｂは、作業の各工程において取得したグリッパー２３の位置を、整定する以前の所定期間にわたって取得し、当該期間における振動強度を取得する。そして、学習部４１ｂは、当該振動強度の程度が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回の振動強度の程度であっても良いし、過去の最小の振動強度の程度であっても良いし、予め決められた振動強度の程度であっても良い。振動強度の程度は、種々の手法で特定されて良く、目標位置からの乖離の積分値や閾値以上の振動が生じている期間など、種々の手法を採用可能である。なお、所定期間は、種々の期間とすることが可能であり、工程の始点から終点にわたる期間であれば、動作中の振動強度による報酬が評価され、工程の終

期が所定期間とされれば、残留振動の強度による報酬が評価される。

【０１４８】

さらに、学習部４１ｂは、作業の各工程の終期において取得したグリッパー２３の位置を、整定する以前の所定期間にわたって取得し、当該期間における目標位置からの乖離の最大値をオーバーシュート量として取得する。そして、学習部４１ｂは、当該オーバーシュート量が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回のオーバーシュート量の程度であっても良いし、過去の最小のオーバーシュート量であっても良いし、予め決められたオーバーシュート量であっても良い。

【０１４９】

さらに、本実施形態においては、制御装置４０、ロボット１～３、作業台等の少なくとも１カ所に集音装置が取り付けられており、学習部４１ｂは、作業中に集音装置が取得した音を示す情報を取得する。そして、学習部４１ｂは、作業中の発生音の大きさが基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回の作業または工程の発生音の大きさの程度であっても良いし、過去の発生音の大きさの最小値であっても良いし、予め決められた大きさであっても良い。また、発生音の大きさは、音圧の最大値で評価されても良いし、所定期間内の音圧の統計値（平均値等）で評価されても良く、種々の構成を採用可能である。

【０１５０】

現在の状態ｓにおいて行動ａが採用された場合における次の状態ｓ'は、行動ａとしてのパラメーターの変化が行われた後にロボット３を動作させ、状態観測部４１ａが状態を観測することによって特定可能である。なお、本例にかかる動作パラメーターの学習は、ロボット１，２による対象物の検出完了後に、ロボット３に関して実行される。

【０１５１】

図１０に示す例において、状態変数には、モーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力が含まれている。従って、状態観測部４１ａは、サーボ４３ｄの制御結果として、モーターＭ１～Ｍ６に供給される電流値を観測することができる。当該電流値は、モーターＭ１～Ｍ６で出力されるトルクに相当する。また、エンコーダーＥ１～Ｅ６の出力は、対応関係Ｕ１に基づいてロボット座標系におけるＴＣＰの位置に変換される。従って、状態観測部４１ａは、ロボット３が備えるグリッパー２３の位置情報を観測することになる。

【０１５２】

力覚センサーＰの出力は積分することによってロボットの位置に変換することができる。すなわち、状態観測部４１ａは、対応関係Ｕ２に基づいてロボット座標系においてＴＣＰへの作用力を積分することでＴＣＰの位置を取得する。従って、本実施形態において状態観測部４１ａは、力覚センサーＰの出力も利用してロボット３が備えるグリッパー２３の位置情報を観測する。なお、状態は、各種の手法で観測されて良く、上述の変換が行われない値（電流値やエンコーダー、力覚センサーの出力値）が状態として観測されても良い。

【０１５３】

状態観測部４１ａは、行動であるサーボゲインや加減速特性の調整結果を直接的に観測しているのではなく、調整の結果、ロボット３で得られた変化をモーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力として観測している。従って、行動による影響を間接的に観測していることになり、この意味で、本実施形態の状態変数は、動作パラメーターの変化から直接的に推定することが困難な変化をし得る状態変数である。

【０１５４】

また、モーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力は、ロボット３の動作を直接的に示しており、当該動作は作業の良否を直接的に示している。従って、状態変数として、モーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の

値、力覚センサーＰの出力を観測することにより、人為的に改善することが困難なパラメーターの改善を行い、効果的に作業の質を高めるように動作パラメーターを最適化することが可能になる。この結果、人為的に決められた動作パラメーターよりも高性能な動作を行う動作パラメーターを高い確率で算出することができる。

【０１５５】

（４－４）動作パラメーターの学習例：

次に、動作パラメーターの学習例を説明する。学習の過程で参照される変数や関数を示す情報は、学習情報４４ｅとして記憶部４４に記憶される。すなわち、算出部４１は、状態変数の観測と、当該状態変数に応じた行動の決定と、当該行動によって得られる報酬の評価とを繰り返すことによって行動価値関数Ｑ（ｓ，ａ）を収束させる構成が採用されている。そこで、本例において、学習の過程で状態変数と行動と報酬との時系列の値が、順次、学習情報４４ｅに記録されていく。

【０１５６】

なお、本実施形態において、動作パラメーターの学習は位置制御モードで実行される。位置制御モードでの学習を実行するためには、位置制御モードのみで構成される作業がロボット３のロボットプログラム４４ｂとして生成されても良いし、任意のモードを含む作業がロボット３のロボットプログラム４４ｂとして生成されている状況において、その中の位置制御モードのみを用いて学習してもよい。

【０１５７】

行動価値関数Ｑ（ｓ，ａ）は、種々の手法で算出されて良く、多数回の試行に基づいて算出されても良いが、ここでは、ＤＱＮによって行動価値関数Ｑを最適化する例を説明する。行動価値関数Ｑの最適化に利用される多層ニューラルネットワークは、上述の図８において模式的に示される。図１０に示すような状態が観測される本例であれば、ロボット３におけるモーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力（６軸の出力）が状態であるため、状態ｓの数Ｍ＝１８である。また、図１０に示す行動が選択され得る本例であれば、１０８個の行動が選択可能であるためＮ＝１０８である。むろん、行動ａの内容や数（Ｎの値）、状態ｓの内容や数（Ｍの値）は試行番号ｔに応じて変化しても良い。

【０１５８】

本実施形態においても、当該多層ニューラルネットワークを特定するためのパラメーター（入力から出力を得るために必要な情報）が学習情報４４ｅとして記憶部４４に記録される。ここでも学習の過程で変化し得る多層ニューラルネットワークのパラメーターをθと表記する。当該θを使用すると、上述の行動価値関数Ｑ（s_t，a_{1t}）～Ｑ（s_t，a_{Nt}）は、Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）とも表記できる。

【０１５９】

次に、図９に示すフローチャートに沿って学習処理の手順を説明する。動作パラメーターの学習処理は、ロボット３の運用過程において実施されても良いし、実運用の前に事前に学習処理が実行されてもよい。ここでは、実運用の前に事前に学習処理が実行される構成（多層ニューラルネットワークを示すθが最適化されると、その情報が保存され、次回以降の運用で利用される構成）に従って学習処理を説明する。

【０１６０】

学習処理が開始されると、算出部４１は、学習情報４４ｅを初期化する（ステップＳ２００）。すなわち、算出部４１は、学習を開始する際に参照されるθの初期値を特定する。初期値は、種々の手法によって決められて良く、過去に学習が行われていない場合においては、任意の値やランダム値等がθの初期値となっても良いし、ロボット３や対象物を模擬するシミュレーション環境を準備し、当該環境に基づいて学習または推定したθを初期値としてもよい。

【０１６１】

過去に学習が行われた場合は、当該学習済のθが初期値として採用される。また、過去に類似の対象についての学習が行われた場合は、当該学習におけるθが初期値とされても

良い。過去の学習は、ロボット３を用いてユーザーが行ってもよいし、ロボット３の製造者がロボット３の販売前に行ってもよい。この場合、製造者は、対象物や作業の種類に応じて複数の初期値のセットを用意しておき、ユーザーが学習する際に初期値を選択する構成であっても良い。θの初期値が決定されると、当該初期値が現在のθの値として学習情報４４ｅに記憶される。

【０１６２】

次に、算出部４１は、パラメーターを初期化する（ステップＳ２０５）。ここでは、動作パラメーターが学習対象であるため、算出部４１は、動作パラメーターを初期化する。すなわち、学習が行われていない状態であれば、算出部４１は、教示によって生成されたパラメーター４４ａに含まれる動作パラメーターを初期値として設定する。過去に何らかの学習が行われた状態であれば、算出部４１は、学習の際に最後に利用されていたパラメーター４４ａに含まれる動作パラメーターを初期値として設定する。

【０１６３】

次に、状態観測部４１ａは、状態変数を観測する（ステップＳ２１０）。すなわち、制御部４３は、パラメーター４４ａおよびロボットプログラム４４ｂを参照してロボット３を制御する（上述のステップＳ１１０～Ｓ１３０に相当）。この後、状態観測部４１ａは、モーターＭ１～Ｍ６に供給される電流値を観測する。また、状態観測部４１ａは、エンコーダーＥ１～Ｅ６の出力を取得し、対応関係Ｕ１に基づいてロボット座標系におけるＴＣＰの位置に変換する。さらに、状態観測部４１ａは、力覚センサーＰの出力を積分し、ＴＣＰの位置を取得する。

【０１６４】

次に、学習部４１ｂは、行動価値を算出する（ステップＳ２１５）。すなわち、学習部４１ｂは、学習情報４４ｅを参照してθを取得し、学習情報４４ｅが示す多層ニューラルネットワークに最新の状態変数を入力し、Ｎ個の行動価値関数$Q(s_t, a_{1t}; \theta_t)$～$Q(s_t, a_{Nt}; \theta_t)$を算出する。

【０１６５】

なお、最新の状態変数は、初回の実行時においてステップＳ２１０、２回目以降の実行時においてステップＳ２２５の観測結果である。また、試行番号ｔは初回の実行時において０、２回目以降の実行時において１以上の値となる。学習処理が過去に実施されていない場合、学習情報４４ｅが示すθは最適化されていないため、行動価値関数Ｑの値としては不正確な値となり得るが、ステップＳ２１５以後の処理の繰り返しにより、行動価値関数Ｑは徐々に最適化していく。また、ステップＳ２１５以後の処理の繰り返しにおいて、状態ｓ、行動ａ、報酬ｒは、各試行番号ｔに対応づけられて記憶部４４に記憶され、任意のタイミングで参照可能である。

【０１６６】

次に、学習部４１ｂは、行動を選択し、実行する（ステップＳ２２０）。本実施形態においては、行動価値関数Ｑ（ｓ，ａ）を最大化する行動ａが最適な行動であると見なされる処理が行われる。そこで、学習部４１ｂは、ステップＳ２１５において算出されたＮ個の行動価値関数$Q(s_t, a_{1t}; \theta_t)$～$Q(s_t, a_{Nt}; \theta_t)$の値の中で最大の値を特定する。そして、学習部４１ｂは、最大の値を与えた行動を選択する。例えば、Ｎ個の行動価値関数$Q(s_t, a_{1t}; \theta_t)$～$Q(s_t, a_{Nt}; \theta_t)$の中で$Q(s_t, a_{Nt}; \theta_t)$が最大値であれば、学習部４１ｂは、行動a_{Nt}を選択する。

【０１６７】

行動が選択されると、学習部４１ｂは、当該行動に対応するパラメーター４４ａを変化させる。例えば、図１０に示す例において、モーターＭ１のサーボゲインKppを一定値増加させる行動ａ１が選択された場合、学習部４１ｂは、動作パラメーターが示すモーターＭ１のサーボゲインKppの値を一定値増加させる。パラメーター４４ａの変化が行われると、制御部４３は、当該パラメーター４４ａを参照してロボット３を制御し、一連の作業を実行させる。なお、本実施形態においては、行動選択のたびに一連の作業が実行されるが、行動選択のたびに一連の作業の一部が実行される構成（一連の作業を構成する複数の

工程の少なくとも１工程が実行される構成）であっても良い。

【０１６８】

次に、状態観測部４１ａは、状態変数を観測する（ステップＳ２２５）。すなわち、状態観測部４１ａは、ステップＳ２１０における状態変数の観測と同様の処理を行って、状態変数として、モーターＭ１～Ｍ６に供給される電流値、エンコーダーＥ１～Ｅ６の出力に基づいて特定されるＴＣＰの位置、力覚センサーＰの出力に基づいて特定されるＴＣＰの位置を取得する。なお、現在の試行番号がｔである場合（選択された行動がa_tである場合）、ステップＳ２２５で取得される状態ｓはs_{t+1}である。

【０１６９】

次に、学習部４１ｂは、報酬を評価する（ステップＳ２３０）。すなわち、学習部４１ｂは、図示しない計時回路に基づいて作業の開始から終了までの所要時間を取得し、作業の所要時間が基準よりも短い場合に正、作業の所要時間が基準よりも長い場合に負の報酬を取得する。さらに、学習部４１ｂは、作業の各工程の終了段階におけるグリッパー２３の位置を取得し、各工程の目標位置とのずれ量を取得する。そして、学習部４１ｂは、グリッパー２３の位置と目標位置とのずれ量が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する。一連の作業が複数の工程で構成される場合、各工程の報酬の和が取得されても良いし、統計値（平均値等）が取得されても良い。

【０１７０】

さらに、学習部４１ｂは、作業の各工程において取得したグリッパー２３の位置に基づいて振動強度を取得する。そして、学習部４１ｂは、当該振動強度の程度が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する。一連の作業が複数の工程で構成される場合、各工程の報酬の和が取得されても良いし、統計値（平均値等）が取得されても良い。

【０１７１】

さらに、学習部４１ｂは、作業の各工程の終期において取得したグリッパー２３の位置に基づいてオーバーシュート量を取得する。そして、学習部４１ｂは、当該オーバーシュート量が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する一連の作業が複数の工程で構成される場合、各工程の報酬の和が取得されても良いし、統計値（平均値等）が取得されても良い。

【０１７２】

さらに、学習部４１ｂは、作業中に集音装置が取得した音を示す情報を取得する。そして、学習部４１ｂは、作業中の発生音の大きさが基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する。なお、現在の試行番号がｔである場合、ステップＳ２３０で取得される報酬ｒはr_{t+1}である。

【０１７３】

本例においても、式（２）に示す行動価値関数Ｑの更新を目指しているが、行動価値関数Ｑを適切に更新していくためには、行動価値関数Ｑを示す多層ニューラルネットワークを最適化（θを最適化）していかなくてはならない。そして、図８に示す多層ニューラルネットワークによって行動価値関数Ｑを適正に出力させるためには、当該出力のターゲットとなる教師データが必要になる。すなわち、多層ニューラルネットワークの出力と、ターゲットとの誤差を最小化するようにθを改善すると、多層ニューラルネットワークが最適化されることが期待される。

【０１７４】

しかし、本実施形態において、学習が完了していない段階では行動価値関数Ｑの知見がなく、ターゲットを特定することは困難である。そこで、本実施形態においては、式（２）の第２項、いわゆるＴＤ誤差を最小化する目的関数によって多層ニューラルネットワークを示すθの改善を実施する。すなわち、$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta_t))$をターゲットとし、ターゲットと$Q(s_t, a_t; \theta_t)$との誤差が最小化するようにθを学習する。ただし、ターゲット$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta_t))$は、学習対象のθを含んでいるため、本実施形態においては、ある程度の試行回数にわたりター

ゲットを固定する（例えば、最後に学習したθ（初回学習時はθの初期値）で固定する）。本実施形態においては、ターゲットを固定する試行回数である既定回数が予め決められている。

【０１７５】

このような前提で学習を行うため、ステップＳ２３０で報酬が評価されると、学習部４１ｂは目的関数を算出する（ステップＳ２３５）。すなわち、学習部４１ｂは、試行のそれぞれにおけるＴＤ誤差を評価するための目的関数（例えば、ＴＤ誤差の２乗の期待値に比例する関数やＴＤ誤差の２乗の総和等）を算出する。なお、ＴＤ誤差は、ターゲットが固定された状態で算出されるため、固定されたターゲットを（$r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$）と表記すると、ＴＤ誤差は（$r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) - Q(s_t, a_t; \theta_t)$）である。当該ＴＤ誤差の式において報酬$r_{t+1}$は、行動$a_t$によってステップＳ２３０で得られた報酬である。

【０１７６】

また、$\max_{a'} Q(s_{t+1}, a'; \theta^-)$は、行動$a_t$によってステップＳ２２５で算出される状態$s_{t+1}$を、固定された$\theta^-$で特定される多層ニューラルネットワークの入力とした場合に得られる出力の中の最大値である。$Q(s_t, a_t; \theta_t)$は、行動a_tが選択される前の状態s_tを、試行番号ｔの段階のθ_tで特定される多層ニューラルネットワークの入力とした場合に得られる出力の中で、行動a_tに対応した出力の値である。

【０１７７】

目的関数が算出されると、学習部４１ｂは、学習が終了したか否か判定する（ステップＳ２４０）。本実施形態においては、ＴＤ誤差が充分に小さいか否かを判定するための閾値が予め決められており、目的関数が閾値以下である場合、学習部４１ｂは、学習が終了したと判定する。

【０１７８】

ステップＳ２４０において学習が終了したと判定されない場合、学習部４１ｂは、行動価値を更新する（ステップＳ２４５）。すなわち、学習部４１ｂは、ＴＤ誤差のθによる偏微分に基づいて目的関数を小さくするためのθの変化を特定し、θを変化させる。むろん、ここでは、各種の手法でθを変化させることが可能であり、例えば、ＲＭＳＰｒｏｐ等の勾配降下法を採用可能である。また、学習率等による調整も適宜実施されて良い。以上の処理によれば、行動価値関数Ｑがターゲットに近づくようにθを変化させることができる。

【０１７９】

ただし、本実施形態においては、上述のようにターゲットが固定されているため、学習部４１ｂは、さらに、ターゲットを更新するか否かの判定を行う。具体的には学習部４１ｂは、既定回数の試行が行われたか否かを判定し（ステップＳ２５０）、ステップＳ２５０において、既定回数の試行が行われたと判定された場合に、学習部４１ｂは、ターゲットを更新する（ステップＳ２５５）。すなわち、学習部４１ｂは、ターゲットを算出する際に参照されるθを最新のθに更新する。この後、学習部４１ｂは、ステップＳ２１５以降の処理を繰り返す。一方、ステップＳ２５０において、既定回数の試行が行われたと判定されなければ、学習部４１ｂは、ステップＳ２５５をスキップしてステップＳ２１５以降の処理を繰り返す。

【０１８０】

ステップＳ２４０において学習が終了したと判定された場合、学習部４１ｂは、学習情報４４ｅを更新する（ステップＳ２６０）。すなわち、学習部４１ｂは、学習によって得られたθを、ロボット３による作業の際に参照されるべきθとして学習情報４４ｅに記録する。当該θを含む学習情報４４ｅが記録されている場合、ステップ１１０～Ｓ１３０のようにロボット３による作業が行われる際に、制御部４３はパラメーター４４ａに基づいてロボット３を制御する。そして、当該作業の過程においては、状態観測部４１ａによる現在の状態の観測と、学習部４１ｂによる行動の選択が繰り返される。むろん、この際、学習部４１ｂは、状態を入力として算出された出力Ｑ（ｓ，ａ）の中で最大値を与える行

動ａを選択する。そして、行動ａが選択された場合、行動ａが行われた状態に相当する値となるようにパラメーター４４ａが更新される。

【０１８１】

以上の構成によれば、制御部４３は、行動価値関数Ｑが最大化される行動ａを選択しながら作業を実行することができる。当該行動価値関数Ｑは、上述の処理により、多数の試行が繰り返された結果、最適化されている。そして、当該試行は、算出部４１によって自動で行われ、人為的に実施不可能な程度の多数の試行を容易に実行することができる。従って、本実施形態によれば、人為的に決められた動作パラメーターよりも高い確率でロボット３の作業の質を高めることができる。

【０１８２】

さらに、本実施形態においては、行動によってパラメーター４４ａとしてのサーボゲインが変化する。従って、人為的な調整によって適切な設定を行うことが困難な、モーターを制御するためのサーボゲインを自動的に調整することができる。さらに、本実施形態においては、行動によってパラメーター４４ａとしての加減速特性が変化する。従って、人為的な調整によって適切な設定を行うことが困難な加減速特性を自動的に調整することができる。

【０１８３】

さらに、本実施形態においては、行動によってロボットの動作の始点および終点が変化しない。従って、本実施形態においては、ロボット３が予定された始点および終点から外れ、利用者の意図しない動作が行われることを防止することができる。さらに、本実施形態においては、行動によってロボットに対する教示位置である始点および終点は変化しない。従って、本実施形態においては、ロボット３が教示された位置から外れ、利用者の意図しない動作が行われることを防止することができる。なお、本実施形態において、教示位置は始点および終点であるが、他の位置が教示位置となってもよい。例えば、始点と終点との間で通過すべき位置や取るべき姿勢がある場合、これらが教示位置（教示姿勢）であっても良い。

【０１８４】

さらに、本実施形態においては、ロボット３が行った作業の良否に基づいて行動による報酬を評価するため、ロボット３の作業を成功させるようにパラメーターを最適化することができる。さらに、本実施形態においては、作業の所要時間が基準よりも短い場合に報酬を正と評価するため、ロボット３を短い時間で作業させる動作パラメーターを容易に算出することができる。さらに、本実施形態においては、ロボット３の位置と目標位置とのずれ量が基準よりも小さい場合に報酬を正と評価するため、ロボット３を目標位置に正確に移動させる動作パラメーターを容易に算出することができる。

【０１８５】

さらに、本実施形態においては、振動強度が基準よりも小さい場合に報酬を正と評価するため、ロボット３の動作による振動を発生させる可能性が低い動作パラメーターを容易に算出することができる。さらに、本実施形態においては、ロボット３の位置のオーバーシュートが基準よりも小さい場合に報酬を正と評価するため、ロボット３がオーバーシュートする可能性が低い動作パラメーターを容易に算出することができる。さらに、本実施形態においては、発生音が基準よりも小さい場合に報酬を正と評価するため、ロボット３に異常を発生させる可能性が低い動作パラメーターを容易に算出することができる。

【０１８６】

さらに、本実施形態によれば、自動で行動価値関数Ｑが最適化されるため、高性能な動作を行う動作パラメーターを容易に算出することができる。また、行動価値関数Ｑの最適化は自動的に行われるため、最適な動作パラメーターの算出も自動的に行うことができる。

【０１８７】

さらに、本実施形態においては、ロボット３において汎用的に使用される力覚センサーＰによってロボット３の位置情報を取得するため、ロボット３で汎用的に使用されるセン

サーに基づいて位置情報を算出することができる。

【０１８８】

さらに、本実施形態において学習部４１ｂは、状態変数としてのロボット３の動作結果を実測し、動作パラメーターを最適化する。従って、ロボット３によって作業が行われている実環境下において合わせて動作パラメーターを最適化することができる。従って、ロボット３の使用環境に応じた動作パラメーターとなるように最適化することができる。

【０１８９】

さらに、本実施形態において状態観測部４１ａは、ロボット３にエンドエフェクターとしてのグリッパー２３が設けられた状態で状態変数を観測する。また、学習部４１ｂは、ロボット３にエンドエフェクターとしてのグリッパー２３が設けられた状態で行動としてのパラメーター４４ａの変更が実行される。この構成によれば、エンドエフェクターとしてのグリッパー２３を用いた動作を行うロボット３に適した動作パラメーターを容易に算出することができる。

【０１９０】

さらに、本実施形態において状態観測部４１ａは、エンドエフェクターとしてのグリッパー２３が対象物を把持した状態で状態変数を観測する。また、学習部４１ｂは、エンドエフェクターとしてのグリッパー２３が対象物を把持した状態で行動としてのパラメーター４４ａの変更が実行される。この構成によれば、エンドエフェクターとしてのグリッパー２３で対象物を把持して動作を行うロボット３に適した動作パラメーターを容易に算出することができる。

【０１９１】

（４－５）力制御パラメーターの学習：

力制御パラメーターの学習においても、学習対象のパラメーターを選択することが可能であり、ここでは、その一例を説明する。図１１は、力制御パラメーターの学習例を図７と同様のモデルで説明した図である。本例も式（２）に基づいて行動価値関数Ｑ（ｓ，ａ）を最適化する。従って、最適化後の行動価値関数Ｑ（ｓ，ａ）を最大化する行動ａが最適な行動であると見なされ、当該行動ａを示すパラメーター４４ａが最適化されたパラメーターであると見なされる。

【０１９２】

力制御パラメーターの学習においても、力制御パラメーターを変化させることが行動の決定に相当しており、学習対象のパラメーターと取り得る行動とを示す行動情報４４ｄが記憶部４４に予め記録される。すなわち、当該行動情報４４ｄに学習対象として記述された力制御パラメーターが学習対象となる。図１１においては、ロボット３における力制御パラメーターの中のインピーダンスパラメーターと、力制御座標系と、目標力と、ロボット３の動作の始点および終点が学習対象である。なお、力制御における動作の始点および終点は教示位置であるが、力制御パラメーターの学習によって変動し得る。また、力制御座標系の原点は、ロボット３のＴＣＰ（ツールセンターポイント）からのオフセット点であり、学習前においては目標力が作用する作用点である。従って、力制御座標系（原点座標と軸回転角）と目標力とが変化すると、ＴＣＰからのオフセット点の位置が変化することになり、目標力の作用点が力制御座標系の原点ではない場合も生じ得る。

【０１９３】

力制御パラメーターの中のインピーダンスパラメーターｍ，ｋ，ｄは、ロボット座標系の各軸に対する並進と回転について定義される。従って、本実施形態においては、１軸あたり３個のインピーダンスパラメーターｍ，ｄ，ｋのそれぞれを増加または減少させることが可能であり、増加について１８個の行動、減少についても１８個の行動、計３６個の行動（行動ａ１～ａ３６）を選択し得る。

【０１９４】

一方、力制御座標系は当該座標系の原点座標と、力制御座標系の軸の回転角度と、がロボット座標系を基準として表現されることによって定義される。従って、本実施形態においては、原点座標の３軸方向への増減と、３軸の軸回転角の増減とが可能であり、原点座

標の増加について３個、減少について３個、軸回転角の増加について３個、減少について３個の行動が可能であり、計１２個の行動（行動ａ３７～ａ４８）を選択し得る。目標力は、目標力ベクトルで表現され、目標力の作用点と、力制御座標系の６軸それぞれの成分の大きさ（３軸の並進力、３軸のトルク）によって定義される。従って、本実施形態においては、目標力の作用点の３軸方向への増減について６個、６軸それぞれの成分の増加について６個、減少について６個の行動が可能であり、計１８個の行動（行動ａ４９～ａ６６）を選択し得る。

【０１９５】

ロボット３の動作の始点および終点は、ロボット座標系の各軸方向に沿って座標の増減が可能であり、始点の増減について６個、終点の増減について６個の計１２個の行動（行動ａ６７～ａ７８）を選択し得る。本実施形態においては、以上のようにして予め定義された行動の選択肢に対応するパラメーターが、行動情報４４ｄに学習対象として記述される。また、各行動を特定するための情報（行動のＩＤ、各行動での増減量等）が行動情報４４ｄに記述される。

【０１９６】

図１１に示す例において、報酬はロボット３が行った作業の良否に基づいて評価される。すなわち、学習部４１ｂは、行動ａとして力制御パラメーターを変化させた後、当該力制御パラメーターによってロボット３を動作させ、検出部４２によって検出された対象物をピックアップする作業を実行する。さらに、学習部４１ｂは、作業の良否を観測し、作業の良否を評価する。そして、学習部４１ｂは、作業の良否によって行動ａ、状態ｓ、ｓ'の報酬を決定する。

【０１９７】

なお、作業の良否は作業の成否（ピックアップ成否等）のみならず、作業の質を含む。具体的には、学習部４１ｂは、図示しない計時回路に基づいて作業の開始から終了まで（ステップＳ１１０の開始からステップＳ１２５で終了と判定されるまで）の所要時間を取得する。そして、学習部４１ｂは、作業の所要時間が基準よりも短い場合に正（例えば＋１）、作業の所要時間が基準よりも長い場合に負（例えば－１）の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回の作業の所要時間であっても良いし、過去の最短所要時間であっても良いし、予め決められた時間であっても良い。

【０１９８】

さらに、学習部４１ｂは、作業の各工程において、ロボット３のエンコーダーＥ１～Ｅ６の出力をＵ１に基づいて変換してグリッパー２３の位置を取得する。そして、学習部４１ｂは、作業の各工程において取得したグリッパー２３の位置を、整定する以前の所定期間にわたって取得し、当該期間における振動強度を取得する。そして、学習部４１ｂは、当該振動強度の程度が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回の振動強度の程度であっても良いし、過去の最小の振動強度の程度であっても良いし、予め決められた振動強度の程度であっても良い。

【０１９９】

振動強度の程度は、種々の手法で特定されて良く、目標位置からの乖離の積分値や閾値以上の振動が生じている期間など、種々の手法を採用可能である。なお、所定期間は、種々の期間とすることが可能であり、工程の始点から終点にわたる期間であれば、動作中の振動強度による報酬が評価され、工程の終期が所定期間とされれば、残留振動の強度による報酬が評価される。なお、力制御においては、前者の振動強度による報酬の方が重要である場合が多い。前者の振動強度による報酬の方が重要であれば、後者の残留振動の強度による報酬は評価されない構成とされても良い。

【０２００】

さらに、学習部４１ｂは、作業の各工程の終期において取得したグリッパー２３の位置を、整定する以前の所定期間にわたって取得し、当該期間における目標位置からの乖離の最大値をオーバーシュート量として取得する。そして、学習部４１ｂは、当該オーバーシ

ュート量が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回のオーバーシュート量の程度であっても良いし、過去の最小のオーバーシュート量であっても良いし、予め決められたオーバーシュート量であっても良い。

【０２０１】

さらに、本実施形態においては、制御装置４０、ロボット１～３、作業台等の少なくとも１カ所に集音装置が取り付けられており、学習部４１ｂは、作業中に集音装置が取得した音を示す情報を取得する。そして、学習部４１ｂは、作業中の発生音の大きさが基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を与える。なお、基準は種々の要素によって特定されて良く、例えば、前回の作業または工程の発生音の大きさの程度であっても良いし、過去の発生音の大きさの最小値であっても良いし、予め決められた大きさであっても良い。また、発生音の大きさは、音圧の最大値で評価されても良いし、所定期間内の音圧の統計値（平均値等）で評価されても良く、種々の構成を採用可能である。

【０２０２】

なお、力制御パラメーターの学習においては、動作パラメーターの学習において報酬とされていた、目標位置からの乖離は報酬に含まれない。すなわち、力制御パラメーターの学習においては、工程の始点や終点が学習に応じて変動し得るため、報酬には含まれていない。

【０２０３】

現在の状態ｓにおいて行動ａが採用された場合における次の状態ｓ'は、行動ａとしてのパラメーターの変化が行われた後にロボット３を動作させ、状態観測部４１ａが状態を観測することによって特定可能である。なお、本例にかかる力制御パラメーターの学習は、ロボット１，２による対象物の検出完了後に、ロボット３に関して実行される。

【０２０４】

図１１に示す例において、状態変数には、モーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力が含まれている。従って、状態観測部４１ａは、サーボ４３ｄの制御結果として、モーターＭ１～Ｍ６に供給される電流値を観測する。当該電流値は、モーターＭ１～Ｍ６で出力されるトルクに相当する。また、エンコーダーＥ１～Ｅ６の出力は、対応関係Ｕ１に基づいてロボット座標系におけるＴＣＰの位置に変換される。従って、状態観測部４１ａは、ロボット３が備えるグリッパー２３の位置情報を観測することになる。

【０２０５】

本実施形態においては、ロボットの運動中に力覚センサーＰによって検出された出力を積分することによってロボットの位置を算出することができる。すなわち、状態観測部４１ａは、対応関係Ｕ２に基づいてロボット座標系において運動中のＴＣＰへの作用力を積分することでＴＣＰの位置を取得する。従って、本実施形態において状態観測部４１ａは、力覚センサーＰの出力も利用してロボット３が備えるグリッパー２３の位置情報を観測する。なお、状態は、各種の手法で観測されて良く、上述の変換が行われない値（電流値やエンコーダー、力覚センサーの出力値）が状態として観測されても良い。

【０２０６】

状態観測部４１ａは、行動であるインピーダンスパラメーターや力制御座標系、工程の始点および終点の調整結果を直接的に観測しているのではなく、調整の結果、ロボット３で得られた変化をモーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力として観測している。従って、行動による影響を間接的に観測していることになり、この意味で、本実施形態の状態変数は、力制御パラメーターの変化から直接的に推定することが困難な変化をし得る状態変数である。

【０２０７】

また、モーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力は、ロボット３の動作を直接的に示しており、当該動作は作業の良否を直接的に示している。従って、状態変数として、モーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の

値、力覚センサーＰの出力を観測することにより、人為的に改善することが困難なパラメーターの改善を行い、効果的に作業の質を高めるように力制御パラメーターを最適化することが可能になる。この結果、人為的に決められた力制御パラメーターよりも高性能な動作を行う力制御パラメーターを高い確率で算出することができる。

【０２０８】

（４－６）力制御パラメーターの学習例：

次に、力制御パラメーターの学習例を説明する。学習の過程で参照される変数や関数を示す情報は、学習情報４４ｅとして記憶部４４に記憶される。すなわち、算出部４１は、状態変数の観測と、当該状態変数に応じた行動の決定と、当該行動によって得られる報酬の評価とを繰り返すことによって行動価値関数Ｑ（ｓ，ａ）を収束させる構成が採用されている。そこで、本例において、学習の過程で状態変数と行動と報酬との時系列の値が、順次、学習情報４４ｅに記録されていく。

【０２０９】

なお、本実施形態において、力制御パラメーターの学習は力制御モードで実行される（位置制御のみが行われる位置制御モードでは力制御パラメーターの学習は行われない）。力制御モードでの学習を実行するためには、力制御モードのみで構成される作業がロボット３のロボットプログラム４４ｂとして生成されても良いし、任意のモードを含む作業がロボット３のロボットプログラム４４ｂとして生成されている状況において、その中の力制御モードのみを用いて学習してもよい。

【０２１０】

行動価値関数Ｑ（ｓ，ａ）は、種々の手法で算出されて良く、多数回の試行に基づいて算出されても良いが、ここでは、ＤＱＮによって行動価値関数Ｑを最適化する例を説明する。行動価値関数Ｑの最適化に利用される多層ニューラルネットワークは、上述の図８において模式的に示される。図１１に示すような状態が観測される本例であれば、ロボット３におけるモーターＭ１～Ｍ６の電流、エンコーダーＥ１～Ｅ６の値、力覚センサーＰの出力（６軸の出力）が状態であるため、状態ｓの数Ｍ＝１８である。図１１に示すような行動が選択され得る本例であれば、６０個の行動が選択可能であるためＮ＝７８である。むろん、行動ａの内容や数（Ｎの値）、状態ｓの内容や数（Ｍの値）は試行番号ｔに応じて変化しても良い。

【０２１１】

本実施形態においても、当該多層ニューラルネットワークを特定するためのパラメーター（入力から出力を得るために必要な情報）が学習情報４４ｅとして記憶部４４に記録される。ここでも学習の過程で変化し得る多層ニューラルネットワークのパラメーターをθと表記する。当該θを使用すると、上述の行動価値関数Ｑ（s_t，a_{1t}）～Ｑ（s_t，a_{Nt}）は、Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）とも表記できる。

【０２１２】

次に、図９に示すフローチャートに沿って学習処理の手順を説明する。力制御パラメーターの学習処理は、ロボット３の運用過程において実施されても良いし、実運用の前に事前に学習処理が実行されてもよい。ここでは、実運用の前に事前に学習処理が実行される構成（多層ニューラルネットワークを示すθが最適化されると、その情報が保存され、次回以降の運用で利用される構成）に従って学習処理を説明する。

【０２１３】

学習処理が開始されると、算出部４１は、学習情報４４ｅを初期化する（ステップＳ２００）。すなわち、算出部４１は、学習を開始する際に参照されるθの初期値を特定する。初期値は、種々の手法によって決められて良く、過去に学習が行われていない場合においては、任意の値やランダム値等がθの初期値となっても良いし、ロボット３や対象物を模擬するシミュレーション環境を準備し、当該環境に基づいて学習または推定したθを初期値としてもよい。

【０２１４】

過去に学習が行われた場合は、当該学習済のθが初期値として採用される。また、過去

に類似の対象についての学習が行われた場合は、当該学習におけるθが初期値とされても良い。過去の学習は、ロボット３を用いてユーザーが行ってもよいし、ロボット３の製造者がロボット３の販売前に行ってもよい。この場合、製造者は、対象物や作業の種類に応じて複数の初期値のセットを用意しておき、ユーザーが学習する際に初期値を選択する構成であっても良い。θの初期値が決定されると、当該初期値が現在のθの値として学習情報４４ｅに記憶される。

【０２１５】

次に、算出部４１は、パラメーターを初期化する（ステップＳ２０５）。ここでは、力制御パラメーターが学習対象であるため、算出部４１は、力制御パラメーターを初期化する。すなわち、学習が行われていない状態であれば、算出部４１は、教示によって生成されたパラメーター４４ａに含まれる力制御パラメーターを初期値として設定する。過去に何らかの学習が行われた状態であれば、算出部４１は、学習の際に最後に利用されていたパラメーター４４ａに含まれる力制御パラメーターを初期値として設定する。

【０２１６】

次に、状態観測部４１ａは、状態変数を観測する（ステップＳ２１０）。すなわち、制御部４３は、パラメーター４４ａおよびロボットプログラム４４ｂを参照してロボット３を制御する（上述のステップＳ１１０～Ｓ１３０に相当）。この後、状態観測部４１ａは、モーターＭ１～Ｍ６に供給される電流値を観測する。また、状態観測部４１ａは、エンコーダーＥ１～Ｅ６の出力を取得し、対応関係Ｕ１に基づいてロボット座標系におけるＴＣＰの位置に変換する。さらに、状態観測部４１ａは、力覚センサーＰの出力を積分し、ＴＣＰの位置を取得する。

【０２１７】

次に、学習部４１ｂは、行動価値を算出する（ステップＳ２１５）。すなわち、学習部４１ｂは、学習情報４４ｅを参照してθを取得し、学習情報４４ｅが示す多層ニューラルネットワークに最新の状態変数を入力し、Ｎ個の行動価値関数Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）を算出する。

【０２１８】

なお、最新の状態変数は、初回の実行時においてステップＳ２１０、２回目以降の実行時においてステップＳ２２５の観測結果である。また、試行番号ｔは初回の実行時において０、２回目以降の実行時において１以上の値となる。学習処理が過去に実施されていない場合、学習情報４４ｅが示すθは最適化されていないため、行動価値関数Ｑの値としては不正確な値となり得るが、ステップＳ２１５以後の処理の繰り返しにより、行動価値関数Ｑは徐々に最適化していく。また、ステップＳ２１５以後の処理の繰り返しにおいて、状態ｓ、行動ａ、報酬ｒは、各試行番号ｔに対応づけられて記憶部４４に記憶され、任意のタイミングで参照可能である。

【０２１９】

次に、学習部４１ｂは、行動を選択し、実行する（ステップＳ２２０）。本実施形態においては、行動価値関数Ｑ（ｓ，ａ）を最大化する行動ａが最適な行動であると見なされる処理が行われる。そこで、学習部４１ｂは、ステップＳ２１５において算出されたＮ個の行動価値関数Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）の値の中で最大の値を特定する。そして、学習部４１ｂは、最大の値を与えた行動を選択する。例えば、Ｎ個の行動価値関数Ｑ（s_t，a_{1t}；θ_t）～Ｑ（s_t，a_{Nt}；θ_t）の中でＱ（s_t，a_{Nt}；θ_t）が最大値であれば、学習部４１ｂは、行動a_{Nt}を選択する。

【０２２０】

行動が選択されると、学習部４１ｂは、当該行動に対応するパラメーター４４ａを変化させる。例えば、図１１に示す例において、ロボット座標系のｘ軸に関するインピーダンスパラメーターｍを一定値増加させる行動ａ１が選択された場合、学習部４１ｂは、力制御パラメーターが示すｘ軸に関するインピーダンスパラメーターｍを一定値増加させる。パラメーター４４ａの変化が行われると、制御部４３は、当該パラメーター４４ａを参照してロボット３を制御し、一連の作業を実行させる。なお、本実施形態においては、行動

選択のたびに一連の作業が実行されるが、行動選択のたびに一連の作業の一部が実行される構成（一連の作業を構成する複数の工程の少なくとも１工程が実行される構成）であっても良い。

【０２２１】

次に、状態観測部４１ａは、状態変数を観測する（ステップＳ２２５）。すなわち、状態観測部４１ａは、ステップＳ２１０における状態変数の観測と同様の処理を行って、状態変数として、モーターＭ１～Ｍ６に供給される電流値、エンコーダーＥ１～Ｅ６の出力に基づいて特定されるＴＣＰの位置、力覚センサーＰの出力に基づいて特定されるＴＣＰの位置を取得する。なお、現在の試行番号がｔである場合（選択された行動がa_tである場合）、ステップＳ２２５で取得される状態ｓはs_{t+1}である。

【０２２２】

次に、学習部４１ｂは、報酬を評価する（ステップＳ２３０）。すなわち、学習部４１ｂは、図示しない計時回路に基づいて作業の開始から終了までの所要時間を取得し、作業の所要時間が基準よりも短い場合に正、作業の所要時間が基準よりも長い場合に負の報酬を取得する。さらに、学習部４１ｂは、作業の各工程におけるグリッパー２３の位置を取得し、作業の各工程において取得したグリッパー２３の位置に基づいて振動強度を取得する。そして、学習部４１ｂは、当該振動強度の程度が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する。一連の作業が複数の工程で構成される場合、各工程の報酬の和が取得されても良いし、統計値（平均値等）が取得されても良い。

【０２２３】

さらに、学習部４１ｂは、作業の各工程の終期において取得したグリッパー２３の位置に基づいてオーバーシュート量を取得する。そして、学習部４１ｂは、当該オーバーシュート量が基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する一連の作業が複数の工程で構成される場合、各工程の報酬の和が取得されても良いし、統計値（平均値等）が取得されても良い。

【０２２４】

さらに、学習部４１ｂは、作業中に集音装置が取得した音を示す情報を取得する。そして、学習部４１ｂは、作業中の発生音の大きさが基準よりも小さい場合に正、基準よりも大きい場合の負の報酬を取得する。なお、現在の試行番号がｔである場合、ステップＳ２３０で取得される報酬ｒはr_{t+1}である。

【０２２５】

本例においても、式（２）に示す行動価値関数Ｑの更新を目指しているが、行動価値関数Ｑを適切に更新していくためには、行動価値関数Ｑを示す多層ニューラルネットワークを最適化（θを最適化）していかなくてはならない。そして、図８に示す多層ニューラルネットワークによって行動価値関数Ｑを適正に出力させるためには、当該出力のターゲットとなる教師データが必要になる。すなわち、多層ニューラルネットワークの出力と、ターゲットとの誤差を最小化するようにθを改善すると、多層ニューラルネットワークが最適化されることが期待される。

【０２２６】

しかし、本実施形態において、学習が完了していない段階では行動価値関数Ｑの知見がなく、ターゲットを特定することは困難である。そこで、本実施形態においては、式（２）の第２項、いわゆるＴＤ誤差を最小化する目的関数によって多層ニューラルネットワークを示すθの改善を実施する。すなわち、$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta_t))$をターゲットとし、ターゲットと$Q(s_t, a_t; \theta_t)$との誤差が最小化するようにθを学習する。ただし、ターゲット$(r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta_t))$は、学習対象のθを含んでいるため、本実施形態においては、ある程度の試行回数にわたりターゲットを固定する（例えば、最後に学習したθ（初回学習時はθの初期値）で固定する）。本実施形態においては、ターゲットを固定する試行回数である既定回数が予め決められている。

【０２２７】

このような前提で学習を行うため、ステップＳ２３０で報酬が評価されると、学習部４１ｂは目的関数を算出する（ステップＳ２３５）。すなわち、学習部４１ｂは、試行のそれぞれにおけるＴＤ誤差を評価するための目的関数（例えば、ＴＤ誤差の２乗の期待値に比例する関数やＴＤ誤差の２乗の総和等）を算出する。なお、ＴＤ誤差は、ターゲットが固定された状態で算出されるため、固定されたターゲットを（$r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$）と表記すると、ＴＤ誤差は（$r_{t+1}+\gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)-Q(s_t, a_t; \theta_t)$）である。当該ＴＤ誤差の式において報酬$r_{t+1}$は、行動$a_t$によってステップＳ２３０で得られた報酬である。

【０２２８】

また、$\max_{a'} Q(s_{t+1}, a'; \theta^-)$は、行動$a_t$によってステップＳ２２５で算出される状態$s_{t+1}$を、固定された$\theta^-$で特定される多層ニューラルネットワークの入力とした場合に得られる出力の中の最大値である。$Q(s_t, a_t; \theta_t)$は、行動a_tが選択される前の状態s_tを、試行番号ｔの段階のθ_tで特定される多層ニューラルネットワークの入力とした場合に得られる出力の中で、行動a_tに対応した出力の値である。

【０２２９】

目的関数が算出されると、学習部４１ｂは、学習が終了したか否か判定する（ステップＳ２４０）。本実施形態においては、ＴＤ誤差が充分に小さいか否かを判定するための閾値が予め決められており、目的関数が閾値以下である場合、学習部４１ｂは、学習が終了したと判定する。

【０２３０】

ステップＳ２４０において学習が終了したと判定されない場合、学習部４１ｂは、行動価値を更新する（ステップＳ２４５）。すなわち、学習部４１ｂは、ＴＤ誤差のθによる偏微分に基づいて目的関数を小さくするためのθの変化を特定し、θを変化させる。むろん、ここでは、各種の手法でθを変化させることが可能であり、例えば、ＲＭＳＰｒｏｐ等の勾配降下法を採用可能である。また、学習率等による調整も適宜実施されて良い。以上の処理によれば、行動価値関数Ｑがターゲットに近づくようにθを変化させることができる。

【０２３１】

ただし、本実施形態においては、上述のようにターゲットが固定されているため、学習部４１ｂは、さらに、ターゲットを更新するか否かの判定を行う。具体的には学習部４１ｂは、既定回数の試行が行われたか否かを判定し（ステップＳ２５０）、ステップＳ２５０において、既定回数の試行が行われたと判定された場合に、学習部４１ｂは、ターゲットを更新する（ステップＳ２５５）。すなわち、学習部４１ｂは、ターゲットを算出する際に参照されるθを最新のθに更新する。この後、学習部４１ｂは、ステップＳ２１５以降の処理を繰り返す。一方、ステップＳ２５０において、既定回数の試行が行われたと判定されなければ、学習部４１ｂは、ステップＳ２５５をスキップしてステップＳ２１５以降の処理を繰り返す。

【０２３２】

ステップＳ２４０において学習が終了したと判定された場合、学習部４１ｂは、学習情報４４ｅを更新する（ステップＳ２６０）。すなわち、学習部４１ｂは、学習によって得られたθを、ロボット３による作業の際に参照されるべきθとして学習情報４４ｅに記録する。当該θを含む学習情報４４ｅが記録されている場合、ステップ１１０～Ｓ１３０のようにロボット３による作業が行われる際に、制御部４３はパラメーター４４ａに基づいてロボット３を制御する。そして、当該作業の過程においては、状態観測部４１ａによる現在の状態の観測と、学習部４１ｂによる行動の選択が繰り返される。むろん、この際、学習部４１ｂは、状態を入力として算出された出力Ｑ（ｓ，ａ）の中で最大値を与える行動ａを選択する。そして、行動ａが選択された場合、行動ａが行われた状態に相当する値となるようにパラメーター４４ａが更新される。

【０２３３】

以上の構成によれば、制御部４３は、行動価値関数Ｑが最大化される行動ａを選択しな

がら作業を実行することができる。当該行動価値関数Ｑは、上述の処理により、多数の試行が繰り返された結果、最適化されている。そして、当該試行は、算出部４１によって自動で行われ、人為的に実施不可能な程度の多数の試行を容易に実行することができる。従って、本実施形態によれば、人為的に決められた力制御パラメーターよりも高い確率でロボット３の作業の質を高めることができる。

【０２３４】

さらに、本実施形態においては、行動によってパラメーター４４ａとしてのインピーダンスパラメーターが変化する。従って、人為的な調整によって適切な設定を行うことが困難な、インピーダンスパラメーターを自動的に調整することができる。さらに、本実施形態においては、行動によってパラメーター４４ａとしての始点と終点が変化する。従って、人為的に設定された始点や終点を、より高性能に力制御を行うように自動的に調整することができる。

【０２３５】

さらに、本実施形態においては、行動によってパラメーター４４ａとしての力制御座標系が変化する。この結果、ロボット３のＴＣＰからのオフセット点の位置が変化する。従って、人為的な調整によって適切な設定を行うことが困難な、ＴＣＰからのオフセット点の位置を自動的に調整することができる。さらに、本実施形態においては、行動によってパラメーター４４ａとしての目標力が変化し得る。従って、人為的な調整によって適切な設定を行うことが困難な、目標力を自動的に調整することができる。特に、力制御座標系と目標力との組み合わせを人為的に理想化することは困難であるため、これらの組が自動的に調整される構成は、有用である。

【０２３６】

さらに、本実施形態においては、ロボット３が行った作業の良否に基づいて行動による報酬を評価するため、ロボット３の作業を成功させるようにパラメーターを最適化することができる。さらに、本実施形態においては、作業の所要時間が基準よりも短い場合に報酬を正と評価するため、ロボット３を短い時間で作業させる力制御パラメーターを容易に算出することができる。

【０２３７】

さらに、本実施形態においては、振動強度が基準よりも小さい場合に報酬を正と評価するため、ロボット３の動作による振動を発生させる可能性が低い力制御パラメーターを容易に算出することができる。さらに、本実施形態においては、ロボット３の位置のオーバーシュートが基準よりも小さい場合に報酬を正と評価するため、ロボット３がオーバーシュートする可能性が低い力制御パラメーターを容易に算出することができる。さらに、本実施形態においては、発生音が基準よりも小さい場合に報酬を正と評価するため、ロボット３に異常を発生させる可能性が低い力制御パラメーターを容易に算出することができる。

【０２３８】

さらに、本実施形態によれば、自動で行動価値関数Ｑが最適化されるため、高性能な力制御を行う力制御パラメーターを容易に算出することができる。また、行動価値関数Ｑの最適化は自動的に行われるため、最適な力制御パラメーターの算出も自動的に行うことができる。

【０２３９】

さらに、本実施形態においては、ロボット３において汎用的に使用される力覚センサーＰによってロボット３の位置情報を取得するため、ロボット３で汎用的に使用されるセンサーに基づいて位置情報を算出することができる。

【０２４０】

さらに、本実施形態において学習部４１ｂは、状態変数としてのロボット３の動作結果を実測し、力制御パラメーターを最適化する。従って、ロボット３によって作業が行われている実環境下において合わせて力制御パラメーターを最適化することができる。従って、ロボット３の使用環境に応じた力制御パラメーターとなるように最適化することができ

る。
【０２４１】
さらに、本実施形態において状態観測部４１ａは、ロボット３にエンドエフェクターとしてのグリッパー２３が設けられた状態で状態変数を観測する。また、学習部４１ｂは、ロボット３にエンドエフェクターとしてのグリッパー２３が設けられた状態で行動としてのパラメーター４４ａの変更が実行される。この構成によれば、エンドエフェクターとしてのグリッパー２３を用いた動作を行うロボット３に適した力制御パラメーターを容易に算出することができる。
【０２４２】
さらに、本実施形態において状態観測部４１ａは、エンドエフェクターとしてのグリッパー２３が対象物を把持した状態で状態変数を観測する。また、学習部４１ｂは、エンドエフェクターとしてのグリッパー２３が対象物を把持した状態で行動としてのパラメーター４４ａの変更が実行される。この構成によれば、エンドエフェクターとしてのグリッパー２３で対象物を把持して動作を行うロボット３に適した力制御パラメーターを容易に算出することができる。
【０２４３】
（５）他の実施形態：
以上の実施形態は本発明を実施するための一例であり、他にも種々の実施形態を採用可能である。例えば、制御装置は、ロボットに内蔵されていても良いし、ロボットの設置場所と異なる場所、例えば外部のサーバー等に備えられていても良い。また、制御装置は、複数の装置で構成されていても良く、制御部４３と算出部４１とが異なる装置で構成されても良い。また、制御装置は、ロボットコントローラー、ティーチングペンダント、ＰＣ、ネットワークにつながるサーバー等であっても良いし、これらが含まれていても良い。さらに、上述の実施形態の一部の構成が省略されてもよいし、処理の順序が変動または省略されてもよい。さらに、上述の実施形態においては、ＴＣＰについて目標位置や目標力の初期ベクトルが設定されたが、他の位置、例えば力覚センサーＰについてのセンサー座標系の原点やネジの先端等について目標位置や目標力の初期ベクトルが設定されても良い。
【０２４４】
ロボットは、任意の態様の可動部で任意の作業を実施できれば良い。エンドエフェクターは、対象物に関する作業に利用される部位であり、任意のツールが取り付けられて良い。対象物は、ロボットによる作業対象となる物体であれば良く、エンドエフェクターによって把持された物体であっても良いし、エンドエフェクターが備えるツールで扱われる物体であっても良く、種々の物体が対象物となり得る。
【０２４５】
ロボットに作用させる目標力は、当該ロボットを力制御によって駆動する際にロボットに作用させる目標力であれば良く、例えば、力覚センサー等の力検出部によって検出される力（または当該力から算出される力）を特定の力に制御する際に、当該力が目標力となる。また、力覚センサー以外のセンサー、例えば加速度センサーで検出される力（または当該力から算出される力）が目標力になるように制御されても良いし、加速度や角速度が特定の値になるように制御されても良い。
【０２４６】
さらに、上述の学習処理においては、試行のたびにθの更新によって行動価値を更新し、既定回数の試行が行われるまでターゲットを固定したが、複数回の試行が行われてからθの更新が行われてもよい。例えば、第１既定回数の試行が行われるまでターゲットが固定され、第２既定回数（＜第１既定回数）の試行が行われるまでθを固定する構成が挙げられる。この場合、第２既定回数の試行後に第２既定回数分のサンプルに基づいてθを更新し、さらに試行回数が第１既定回数を超えた場合に最新のθでターゲットを更新する構成となる。
【０２４７】

さらに、学習処理においては、公知の種々の手法が採用されてよく、例えば、体験再生や報酬のＣｌｉｐｐｉｎｇ等が行われてもよい。さらに、図８においては、層ＤＬがＰ個（Ｐは１以上の整数）存在し、各層において複数のノードが存在するが、各層の構造は、種々の構造を採用可能である。例えば、層の数やノードの数は種々の数を採用可能であるし、活性化関数としても種々の関数を採用可能であるし、ネットワーク構造が畳み込みニューラルネットワーク構造等になっていても良い。また、入力や出力の態様も図８に示す例に限定されず、例えば、状態ｓと行動ａとが入力される構成や、行動価値関数Ｑを最大化する行動ａがｏｎｅ－ｈｏｔベクトルとして出力される構成が少なくとも利用される例が採用されても良い。

【０２４８】

上述の実施形態においては、行動価値関数に基づいてｇｒｅｅｄｙ方策で行動を行って試行しながら、行動価値関数を最適化することにより、最適化された行動価値関数に対するｇｒｅｅｄｙ方策が最適方策であると見なしている。この処理は、いわゆる価値反復法であるが、他の手法、例えば、方策反復法によって学習が行われてもよい。さらに、状態ｓ、行動ａ、報酬ｒ等の各種変数においては、各種の正規化が行われてよい。

【０２４９】

機械学習の手法としては、種々の手法を採用であり、行動価値関数Ｑに基づいたε－ｇｒｅｅｄｙ方策によって試行が行われてもよい。また、強化学習の手法としても上述のようなＱ学習に限定されず、ＳＡＲＳＡ等の手法が用いられても良い。また、方策のモデルと行動価値関数のモデルを別々にモデル化した手法、例えば、Ａｃｔｏｒ－Ｃｒｉｔｉｃアルゴリズムが利用されても良い。Ａｃｔｏｒ－Ｃｒｉｔｉｃアルゴリズムを利用するのであれば、方策を示すａｃｔｏｒであるμ（ｓ；θ）と、行動価値関数を示すｃｒｉｔｉｃであるＱ（ｓ，ａ；θ）とを定義し、μ（ｓ；θ）にノイズを加えた方策に従って行動を生成して試行し、試行結果に基づいてａｃｔｏｒとｃｒｉｔｉｃを更新することで方策と行動価値関数とを学習する構成であっても良い。

【０２５０】

算出部は、機械学習を用いて、学習対象のパラメーターを算出することができればよく、パラメーターとしては、光学パラメーター、画像処理パラメーター、動作パラメーター、力制御パラメーターの少なくとも１個であれば良い。機械学習は、サンプルデータを用いてよりよいパラメーターを学習する処理であれば良く、上述の強化学習以外にも、教師あり学習やクラスタリングなど種々の手法によって各パラメーターを学習する構成を採用可能である。

【０２５１】

光学系は、対象物を撮像することができる。すなわち、対象物が含まれる領域を視野にした画像を取得する構成を備える。光学系の構成要素としては上述のように、撮像部や照明部を含むことが好ましく、他にも種々の構成要素が含まれていて良い。また、上述のように、撮像部や照明部はロボットのアームによって移動可能であっても良いし、２次元的な移動機構によって移動可能であっても良いし、固定的であっても良い。むろん、撮像部や照明部は交換可能であっても良い。また、光学系で用いる光（撮像部による検出光や照明部の出力光）の帯域は可視光帯域に限定されず、赤外線や紫外線、Ｘ線等の任意の電磁波が用いられる構成が採用可能である。

【０２５２】

光学パラメーターは、光学系の状態を変化させ得る値であれば良く、撮像部や照明部等で構成される光学系において状態を直接的または間接的に特定するための数値等が光学パラメーターとなる。例えば、撮像部や照明部等の位置や角度等を示す値のみならず、撮像部や照明部の種類を示す数値（ＩＤや型番等）が光学パラメーターとなり得る。

【０２５３】

検出部は、算出された光学パラメーターによる光学系での撮像結果に基づいて、対象物を検出することができる。すなわち、検出部は、学習された光学パラメーターによって光学系を動作させて対象物を撮像し、撮像結果に基づいて対象物の検出処理を実行する構成

を備える。

【０２５４】

検出部は対象物を検出することができればよく、上述の実施形態のように、対象物の位置姿勢が検出される構成の他、対象物の有無が検出される構成であっても良く、種々の構成を採用可能である。なお、対象物の位置姿勢は、例えば、３軸における位置と３軸に対する回転角とによる６個のパラメーターによって定義可能であるが、むろん、必要に応じて任意の数のパラメーターが考慮されなくても良い。例えば、平面上に設置された対象物であれば、少なくとも１個の位置に関するパラメーターが既知であるとして検出対象から除外されても良い。また、平面に固定的な向きで設置された対象物であれば、姿勢に関するパラメーターが検出対象から除外されても良い。

【０２５５】

対象物は、光学系で撮像され、検出される対象となる物体であればよく、ロボットの作業対象となるワークや、ワークの周辺の物体、ロボットの一部など、種々の物体が想定可能である。また、撮像結果に基づいて対象物を検出する手法としても種々の手法を採用可能であり、画像の特徴量抽出によって対象物が検出されても良いし、対象物の動作（人等の可動物体等の検出）によって対象物が検出されても良く、種々の手法が採用されてよい。

【０２５６】

制御部は、対象物の検出結果に基づいてロボットを制御することができる。すなわち、制御部は、対象物の検出結果に応じてロボットの制御内容を決定する構成を備える。従って、ロボットの制御は、上述のような対象物をつかむための制御の他にも種々の制御が行われてよい。例えば、対象物に基づいてロボットの位置決めをする制御や、対象物に基づいてロボットの動作を開始または終了させる制御など、種々の制御が想定される。

【０２５７】

ロボットの態様は、種々の態様であって良く、上述の実施形態のような垂直多関節ロボット以外にも直交ロボット、水平多関節ロボット、双腕ロボット等であって良い。また、種々の態様のロボットが組み合わされても良い。むろん、軸の数やアームの数、エンドエフェクターの態様等は種々の態様を採用可能である。例えば、撮像部２１や照明部２２がロボット３の上方に存在する平面に取り付けられ、当該平面上で撮像部２１や照明部２２が移動可能であっても良い。

【０２５８】

状態観測部は、行動等の試行に応じて変化した結果を観測することができればよく、各種のセンサー等によって状態が観測されても良いし、ある状態から他の状態に変化させる制御が行われ、制御の失敗（エラー等）が観測されなければ当該他の状態が観測されたと見なされる構成であっても良い。前者のセンサーによる観測は、位置等の検出の他にも撮像センサーによる画像の取得も含まれる。

【０２５９】

さらに、上述の実施形態における行動や状態、報酬は例であり、他の行動や状態、報酬を含む構成や任意の行動や状態が省略された構成であっても良い。例えば、撮像部２１や照明部２２が交換可能であるロボット１，２において、撮像部２１や照明部２２の種類の変更を行動として選択可能であり、状態として種類を観測可能であっても良い。接触判定部４３ｃによる判定結果に基づいて報酬が決定されても良い。すなわち、学習部４１ｂにおける学習過程において、接触判定部４３ｃが作業において想定されていない物体とロボットとが接触したと判定した場合、当該直前の行動による報酬を負に設定する構成を採用可能である。この構成によれば、ロボットが想定外の物体に接触しないようにパラメーター４４ａを最適化することができる。

【０２６０】

また、例えば、光学パラメーターの最適化に際して、ロボット１～３によって対象物の検出結果に基づいた作業（例えば、上述のピックアップ作業等）を行い、学習部４１ｂが、対象物の検出結果に基づいてロボット１～３が行った作業の良否に基づいて、行動によ

る報酬を評価する構成であってもよい。この構成は、例えば、図７に示す報酬の中で、対象物の検出の替わりに、または、対象物の検出に加えて作業の成否（例えば、ピックアップの成否）を報酬とする構成が挙げられる。

【０２６１】

作業の成否は、例えば、作業の成否を判定可能な工程（ピックアップの工程等）におけるステップＳ１２０の判定結果等で定義可能である。この場合、行動や状態において、ロボット１～３の動作に関する行動や状態が含まれても良い。さらに、この構成においては、ロボット１～３の作業対象である対象物を撮像部２１および照明部２２を備える光学系で撮像した画像を状態とすることが好ましい。この構成によれば、ロボットの作業を成功させるように光学パラメーターを最適化することができる。なお、光学パラメーターや動作パラメーター、力制御パラメーターを学習するために観測される状態としての画像は、撮像部２１で撮像された画像そのものであっても良いし、撮像部２１で撮像された画像に対して画像処理（例えば、上述の平滑化処理や鮮鋭化処理等）が行われた後の画像であっても良い。

【０２６２】

さらに、光学パラメーター、動作パラメーター、力制御パラメーターのそれぞれを別個に最適化するのではなく、これらのパラメーターの中の２種以上を最適化する構成が採用されてもよい。例えば、図７に示す例において、動作パラメーターや力制御パラメーターを変化させる行動が含まれる構成であれば、光学パラメーターとともに、動作パラメーターや力制御パラメーターを最適化することが可能である。この場合、最適化された動作パラメーターや力制御パラメーターに基づいてロボット１～３が制御される。この構成によれば、対象物の検出を伴う作業を行うパラメーターを最適化することができ、対象物の検出精度を高める学習を実行することができる。

【０２６３】

画像処理パラメーターは、対象物の撮像結果としての画像を変化させ得る値であれば良く、図３に示す例に限定されず、追加または削除されてよい。例えば、画像処理の有無や画像処理の強度、画像処理の順序など、実行される画像処理アルゴリズムを特定するための数値（処理順序等を示すフラグ等を含む）等が画像処理パラメーターとなり得る。より具体的には、画像処理としては、二値化処理、直線検出処理、円検出処理、色検出処理、ＯＣＲ処理等があげられる。

【０２６４】

さらに、画像処理は、複数の種類の画像処理を組み合わせた処理であってもよい。例えば、円検出処理とＯＣＲ処理を組み合わせて、「円内の文字を認識する処理」という処理が行われてもよい。いずれにしても、各画像処理の有無や強度を示すパラメーターが画像処理パラメーターとなり得る。また、これらの画像処理パラメーターの変化が行動となり得る。

【０２６５】

動作パラメーターは、上述の実施形態に挙げられたパラメーターに限定されない。例えば、学習対象となる動作パラメーターに、ロボット１～３が備える慣性センサーに基づいて制御を行うためのサーボゲインが含まれていても良い。すなわち、慣性センサーの出力に基づいた制御ループでモーターＭ１～Ｍ６が制御される構成において、当該制御ループにおけるサーボゲインが行動によって変化する構成であっても良い。例えば、ロボット１～３に取り付けられたエンコーダーＥ１～Ｅ６に基づいてロボット１～３の特定の部位の角速度を算出し、慣性センサーの一種であるジャイロセンサーによって当該特定の部位の角速度を検出し、両者の差分にジャイロサーボゲインを乗じてフィードバック制御を行う構成において、当該ジャイロサーボゲインが行動によって変化する構成が挙げられる。この構成であれば、ロボットの特定の部位に生じる角速度の振動成分を抑制する制御を行うことができる。むろん、慣性センサーはジャイロセンサーに限定されず、加速度センサー等において同様のフィードバック制御が行われる構成において加速度ゲインが行動によって変化する構成であっても良い。以上の構成によれば、人為的な調整によって適切な設定

を行うことが困難な、慣性センサーに基づいて制御を行うためのサーボゲインを自動的に調整することができる。なお、加速度センサーはロボットの運動によって生じる加速度を検知するセンサーであり、上述の力覚センサーはロボットに作用する力を検知するセンサーである。通常、加速度センサーと力覚センサーとは異なるセンサーであるが、一方が他方の機能を代替できる場合には、一方が他方として機能しても良い。

【０２６６】

むろん、力制御パラメーターも上述の実施形態に挙げられたパラメーターに限定されないし、学習対象となるパラメーターも適宜選択されてよい。例えば、目標力に関し、６軸中の全成分または一部の成分が行動として選択し得ない（すなわち固定である）構成であっても良い。この構成は、固定された固定対象物（細い筒等）に、ロボットが把持した対象物を挿入する作業において、目標力は固定対象物のある点に対して固定的な成分を有するが、ロボットの挿入作業に応じて力制御座標系が変化するように学習する構成等を想定する事ができる。

【０２６７】

学習部４１ｂは、ロボット３が把持した対象物を作業完了前に落とした場合、ロボット３の作業対象である対象物の一部が作業完了前に分離した場合、ロボット３が破損した場合、ロボット３の作業対象である対象物が破損した場合、の少なくとも１つにおいて報酬を負と評価する構成であっても良い。ロボット３が把持した対象物を作業完了前に落とした場合に報酬を負と評価する構成によれば、対象物を落とさずに作業を完了させる可能性が高い動作パラメーターや力制御パラメーターを容易に算出することができる。

【０２６８】

ロボット３の作業対象である対象物の一部が作業完了前に分離した場合に報酬を負と評価する構成によれば、対象物を分離させることなく作業を完了させる可能性が高い動作パラメーターや力制御パラメーターを容易に算出することができる。ロボット３が破損した場合に報酬を負と評価する構成によれば、ロボット３を破損させる可能性が低い動作パラメーターや力制御パラメーターを容易に算出することができる。

【０２６９】

ロボット３の作業対象である対象物が破損した場合に報酬を負と評価する構成によれば、対象物を破損させる可能性が低い動作パラメーターや力制御パラメーターを容易に算出することができる。なお、ロボット３が把持した対象物を作業完了前に落としたか否か、ロボット３の作業対象である対象物の一部が作業完了前に分離したか否か、ロボット３が破損したか否か、ロボット３の作業対象である対象物が破損したか否かは、各種のセンサー、例えば撮像部２１等によって検出される構成を採用可能である。

【０２７０】

さらに、学習部４１ｂは、ロボット３による作業が正常に完了した場合において報酬を正と評価する構成であっても良い。ロボット３による作業が正常に完了した場合に報酬を正と評価する構成によれば、ロボット３の作業を成功させる動作パラメーターや力制御パラメーターを容易に算出することができる。

【０２７１】

さらに、ロボット３の位置を検出するための位置検出部は、上述の実施形態のようなエンコーダー、力覚センサーに限定されず、他のセンサー、専用の慣性センサーや撮像部２１等の光学センサー、距離センサー等であっても良い。また、センサーはロボットに内蔵されていても良いが、ロボットの外部に配置されても良い。ロボットの外部に配置された位置検出部を利用すれば、ロボットの動作に影響されることなく位置情報を算出することができる。

【０２７２】

さらに、算出部４１は、ロボットの異なる複数の動作に基づいて、複数の動作に共通の動作パラメーターや力制御パラメーターを算出する構成であっても良い。複数の動作は、最適化された動作パラメーターを利用して実行される動作を含んでいれば良い。従って、複数の動作は、異なる種類の複数の作業（ピックアップ作業、研磨作業、ネジ締め作業な

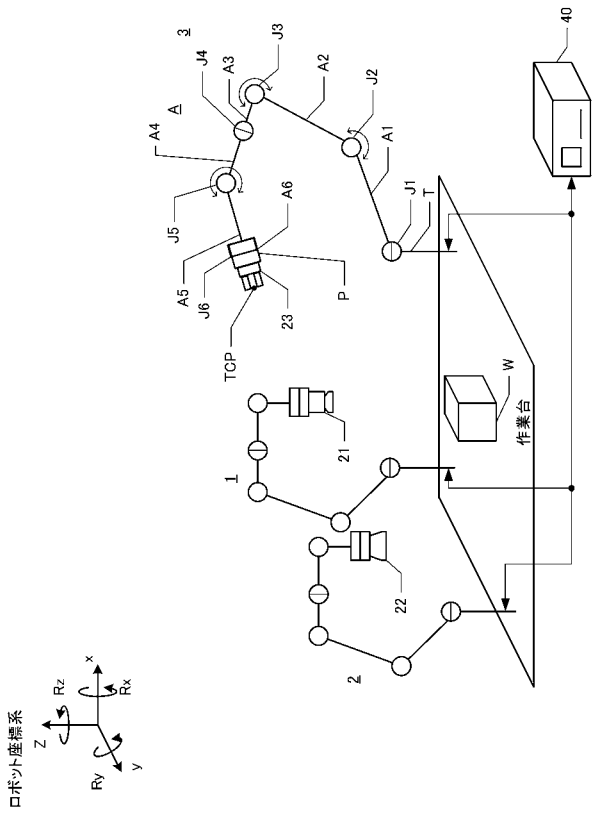
ど）である構成や、同種の作業（ネジの大きさが異なる複数のネジ締め作業等）である構成等が挙げられる。この構成によれば、各種の動作に適用可能な汎用的な動作パラメーターや力制御パラメーターを容易に算出することができる。

【符号の説明】

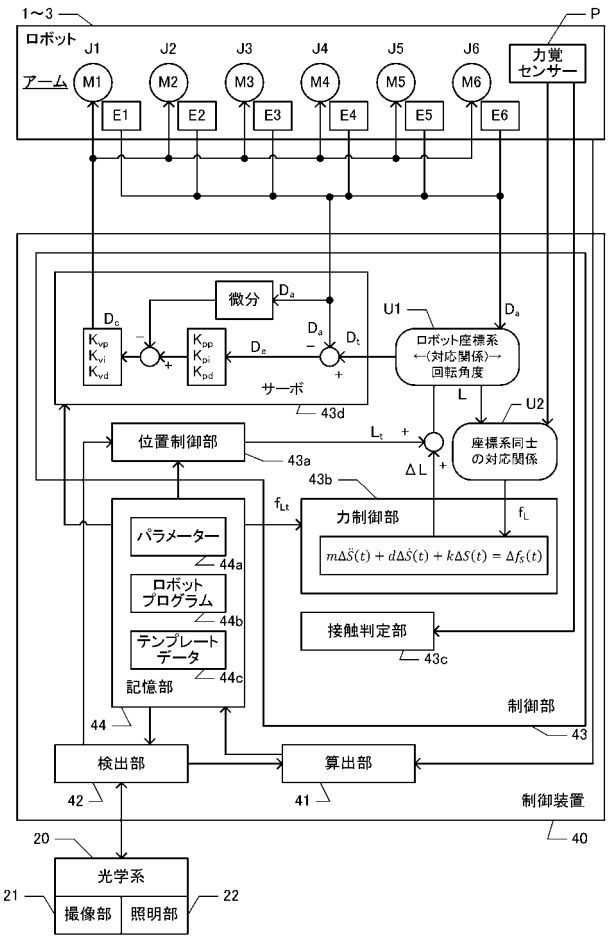
【０２７３】

１～３…ロボット、２０…光学系、２１…撮像部、２２…照明部、２３…グリッパー、４０…制御装置、４１…算出部、４１ａ…状態観測部、４１ｂ…学習部、４２…検出部、４３…制御部、４３ａ…位置制御部、４３ｂ…力制御部、４３ｃ…接触判定部、４３ｄ…サーボ、４４…記憶部、４４ａ…パラメーター、４４ｂ…ロボットプログラム、４４ｃ…テンプレートデータ、４４ｄ…行動情報、４４ｅ…学習情報

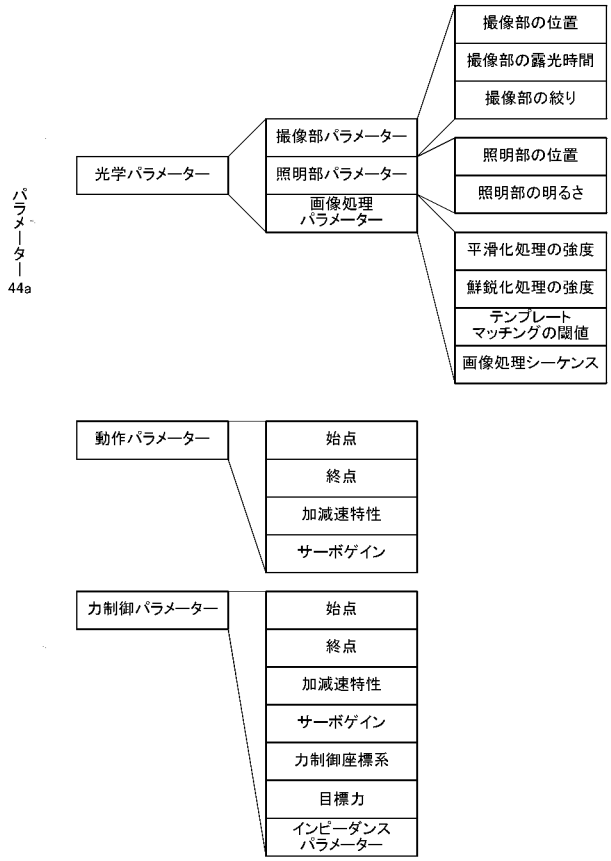
【図１】



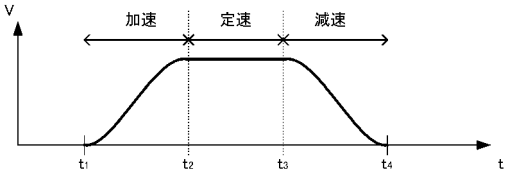
【図２】



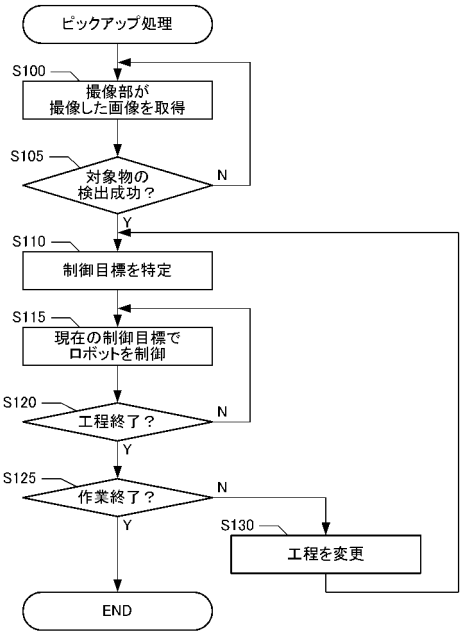
【図３】


パラメーター
44a
光学パラメーター
撮像部パラメーター
照明部パラメーター
画像処理パラメーター
撮像部の位置
撮像部の露光時間
撮像部の絞り
照明部の位置
照明部の明るさ
平滑化処理の強度
鮮鋭化処理の強度
テンプレートマッチングの閾値
画像処理シーケンス
動作パラメーター
始点
終点
加減速特性
サーボゲイン
力制御パラメーター
始点
終点
加減速特性
サーボゲイン
力制御座標系
目標力
インピーダンスパラメーター

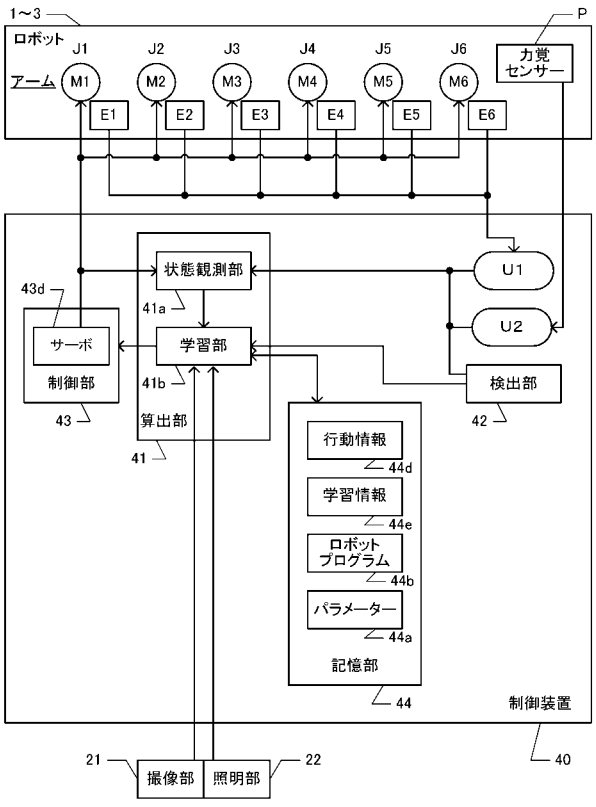
【図４】


V
加速
定速
減速
t1
t2
t3
t4
t

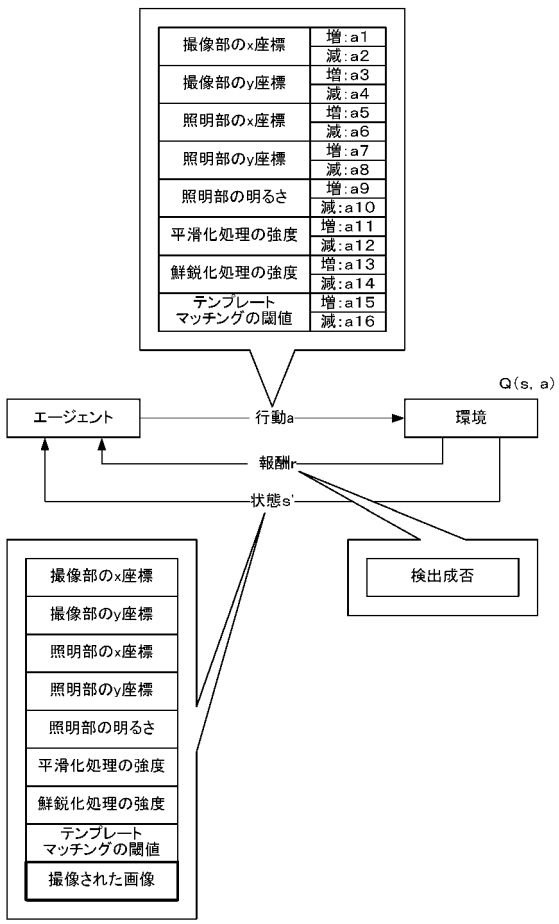
【図５】


ピックアップ処理
S100
撮像部が撮像した画像を取得
S105
対象物の検出成功？
N
Y
S110
制御目標を特定
S115
現在の制御目標でロボットを制御
S120
工程終了？
N
Y
S125
作業終了？
N
Y
S130
工程を変更
END

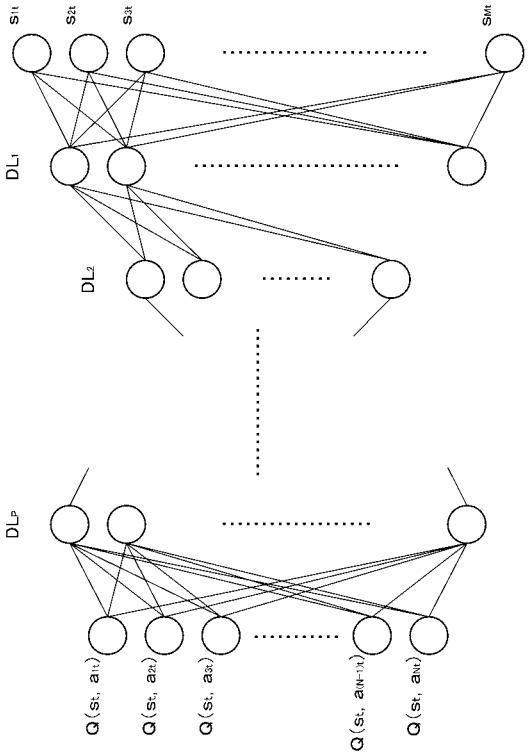
【図６】


1～3
ロボット
アーム
J1
J2
J3
J4
J5
J6
M1
M2
M3
M4
M5
M6
E1
E2
E3
E4
E5
E6
P
力覚センサー
状態観測部
41a
学習部
41b
算出部
41
43d
サーボ
制御部
43
U1
U2
検出部
42
行動情報
44d
学習情報
44e
ロボットプログラム
44b
パラメーター
44a
記憶部
44
制御装置
40
21
撮像部
照明部
22

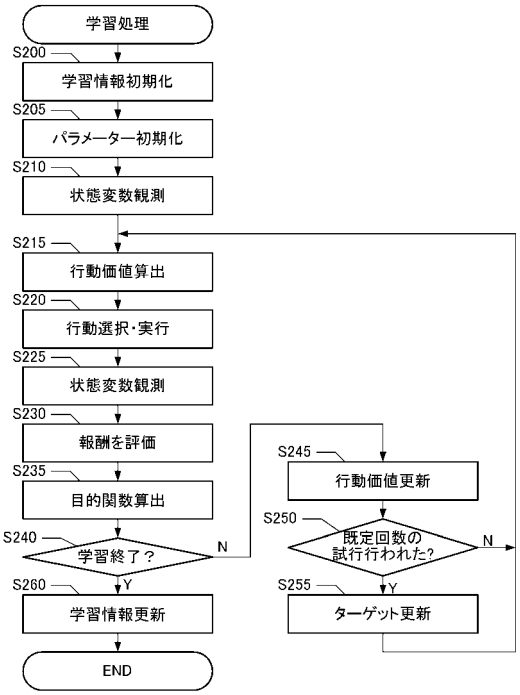
【図７】


撮像部のx座標 増:a1 減:a2
撮像部のy座標 増:a3 減:a4
照明部のx座標 増:a5 減:a6
照明部のy座標 増:a7 減:a8
照明部の明るさ 増:a9 減:a10
平滑化処理の強度 増:a11 減:a12
鮮鋭化処理の強度 増:a13 減:a14
テンプレートマッチングの閾値 増:a15 減:a16
Q(s, a)
エージェント
行動a
環境
報酬r
状態s'
検出成否
撮像部のx座標
撮像部のy座標
照明部のx座標
照明部のy座標
照明部の明るさ
平滑化処理の強度
鮮鋭化処理の強度
テンプレートマッチングの閾値
撮像された画像

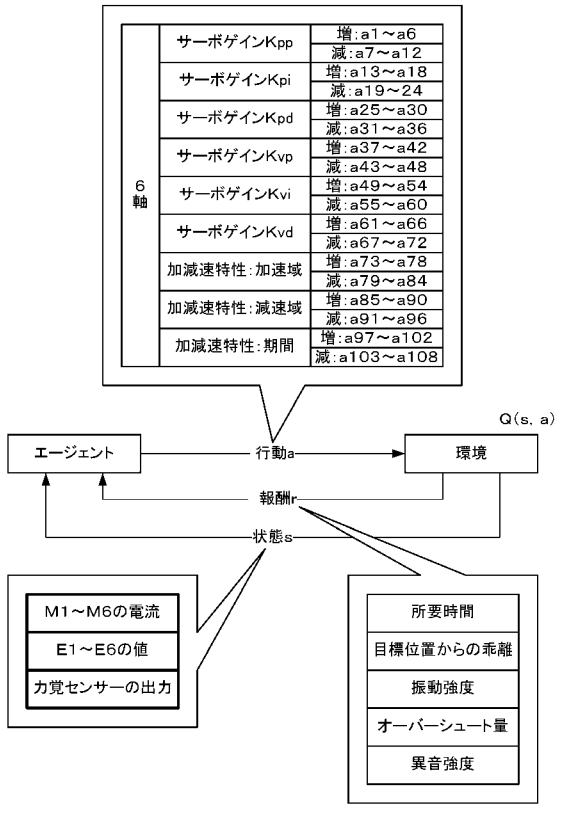
【図８】


DL1
DL2
DLP
Q(St, at1)
Q(St, at2)
Q(St, at3)
Q(St, atN-1)
Q(St, atN)

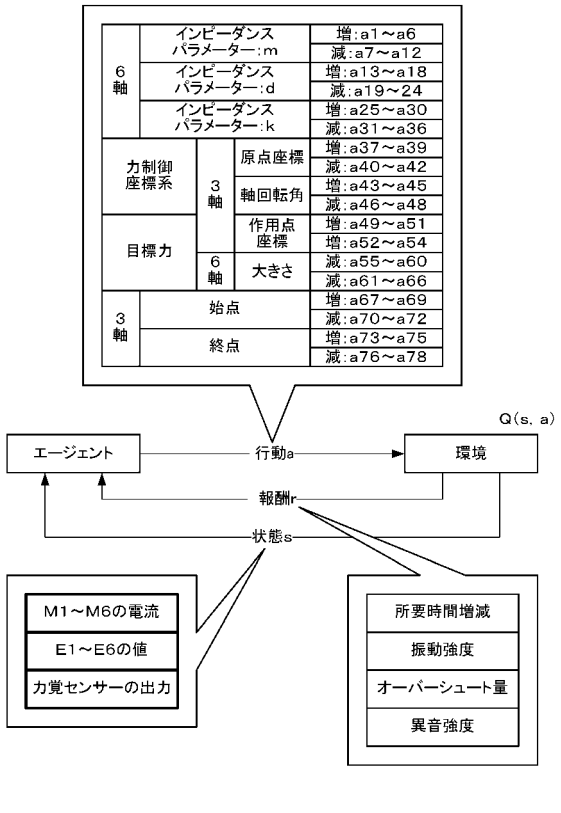
【図９】


学習処理
S200
学習情報初期化
S205
パラメーター初期化
S210
状態変数観測
S215
行動価値算出
S220
行動選択・実行
S225
状態変数観測
S230
報酬を評価
S235
目的関数算出
S240
学習終了？
N
Y
S245
行動価値更新
S250
既定回数の試行行われた?
N
Y
S255
ターゲット更新
S260
学習情報更新
END

【図１０】


6軸
サーボゲインKpp 増:a1～a6 減:a7～a12
サーボゲインKpi 増:a13～a18 減:a19～24
サーボゲインKpd 増:a25～a30 減:a31～a36
サーボゲインKvp 増:a37～a42 減:a43～a48
サーボゲインKvi 増:a49～a54 減:a55～a60
サーボゲインKvd 増:a61～a66 減:a67～a72
加減速特性:加速域 増:a73～a78 減:a79～a84
加減速特性:減速域 増:a85～a90 減:a91～a96
加減速特性:期間 増:a97～a102 減:a103～a108
Q(s, a)
エージェント
行動a
環境
報酬r
状態s
M1～M6の電流
E1～E6の値
力覚センサーの出力
所要時間
目標位置からの乖離
振動強度
オーバーシュート量
異音強度

【図１１】


6軸
インピーダンスパラメーター:m 増:a1～a6 減:a7～a12
インピーダンスパラメーター:d 増:a13～a18 減:a19～24
インピーダンスパラメーター:k 増:a25～a30 減:a31～a36
力制御座標系
3軸
原点座標 増:a37～a39 減:a40～a42
軸回転角 増:a43～a45 減:a46～a48
目標力
作用点座標 増:a49～a51 増:a52～a54
6軸
大きさ 減:a55～a60 減:a61～a66
3軸
始点 増:a67～a69 減:a70～a72
終点 増:a73～a75 減:a76～a78
Q(s, a)
エージェント
行動a
環境
報酬r
状態s
M1～M6の電流
E1～E6の値
力覚センサーの出力
所要時間増減
振動強度
オーバーシュート量
異音強度