

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4877810号
(P4877810)

(45) 発行日 平成24年2月15日(2012.2.15)

(24) 登録日 平成23年12月9日(2011.12.9)

(51) Int. Cl. F 1
G06T 7/00 (2006.01) G06T 7/00 350B
G06T 1/00 (2006.01) G06T 1/00 280

請求項の数 6 外国語出願 (全 22 頁)

(21) 出願番号	特願2007-96733 (P2007-96733)	(73) 特許権者	393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2
(22) 出願日	平成19年4月2日(2007.4.2)	(73) 特許権者	503360115 独立行政法人科学技術振興機構 埼玉県川口市本町四丁目1番8号
(65) 公開番号	特開2008-257353 (P2008-257353A)	(74) 代理人	100099933 弁理士 清水 敏
(43) 公開日	平成20年10月23日(2008.10.23)	(72) 発明者	アレス・ウデ 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
審査請求日	平成22年4月1日(2010.4.1)	(72) 発明者	ゴードン・チェン 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 物体の視覚的表現を学習するための学習システム及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

ビデオカメラと、物体を保持し、回転させ、平行移動させることができる制御可能なマニピュレータとを用いて物体の視覚表現を学習するための学習システムであって、

前記ビデオカメラの出力を受けるとして接続されたフレームグラバと、

前記マニピュレータが前記ビデオカメラの光軸に沿って前記マニピュレータによって保持された前記物体を平行移動させるように、前記マニピュレータを駆動するための平行移動手段と、

前記フレームグラバによって取込まれたフレーム内での前記物体の画像があらかじめ定められた大きさの境界の範囲内に位置し、かつ前記境界の範囲内でできるだけ大きくなるように、前記マニピュレータによって保持された前記物体の位置を決定するための手段と、

前記マニピュレータによって保持された前記物体が前記カメラの光軸と直交する2つの軸の回りで回転するように、前記マニピュレータを駆動するための回転手段と、

前記回転手段によって回転された前記物体の画像を収集するための画像収集手段とを含み、前記画像は前記物体の識別子に関連付けて記憶され、

前記学習システムはさらに、前記画像と前記物体の前記識別子とを用いて、前記物体を識別するように分類器をトレーニングするためのトレーニング手段を含む、学習システム

【請求項2】

前記画像収集手段が

前記フレームグラバによって取込まれた前記フレームの範囲内で前記物体の領域を識別するための識別手段と、

あらかじめ定められた形状を、前記識別手段によって識別された前記物体の前記領域の外縁にフィットするための手段と、

前記あらかじめ定められた形状内で画像を抽出するための手段とを含む、請求項 1 に記載の学習システム。

【請求項 3】

前記識別手段が

前記フレームグラバによって取込まれた前記フレームの範囲内の各画素が、背景、前記マニピュレータの画像の一部、前記物体の画像の一部、及びアウトライアに属す確率を計算するための手段と、

前記各画素について、前記各画素が前記物体の前記画像に属しているかどうかを決定するための手段とを含む、請求項 2 に記載の学習システム。

【請求項 4】

前記回転手段が、前記マニピュレータによって保持された前記物体が前記光軸と直交する 2 つの軸の回りで回転するように、前記マニピュレータを駆動するための手段を含む、請求項 1 に記載の学習システム。

【請求項 5】

前記分類器がサポートベクタマシンベースの分類器である、請求項 1 ~ 請求項 4 のいずれかに記載の学習システム。

【請求項 6】

ビデオカメラと、前記ビデオカメラから出力されたフレームを取込むためのフレームグラバと、物体を保持し、回転させ、平行移動させることができる制御可能なマニピュレータとに接続されたコンピュータ上で実行されるコンピュータプログラムであって、前記コンピュータによって実行されると、前記コンピュータを、

前記マニピュレータが、前記ビデオカメラの光軸に沿って前記マニピュレータによって保持された前記物体を平行移動させるように、前記マニピュレータを駆動するための平行移動手段と、

前記フレームグラバによって取込まれたフレーム内での前記物体の画像が、あらかじめ定められた大きさの境界の範囲内に位置し、かつ前記境界の範囲内でできるだけ大きくなるように、前記マニピュレータによって保持された前記物体の位置を決定するための手段と、

前記マニピュレータによって保持された前記物体が前記ビデオカメラの光軸と直交する 2 つの軸回りで回転するように、前記マニピュレータを駆動するための回転手段と、

前記回転手段によって回転された前記物体の画像を収集するための画像収集手段として機能させ、前記画像は前記物体の識別子に関連づけて記憶され、

前記コンピュータプログラムはさらに、前記コンピュータを、前記画像と前記物体の前記識別子とを用いて、前記物体を識別するように分類器をトレーニングするためのトレーニング手段として機能させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は画像の学習システムに関し、特に、物体についての事前知識なしに学習するために物体の画像を収集するための、ビデオカメラ及びマニピュレータを有する画像学習システムに関する。

【背景技術】

【0002】

物体の認識は自律型ロボットにとって本質的に不可欠なものであるため、過去に多くの注目を集めてきた。現在成功している物体の認識システムの大部分は視野ベースであり、

10

20

30

40

50

物体のスナップショットから好適な表現を構築する。初期のアプローチはさほどの前処理なしに物体について収集されたパターンを使用していたのであるが、現在の研究の大部分は局所的な画像の特徴を使用している。例えば、スケール不変特徴変換 (SIFTキー) (非特許文献5)、及びガボールジェット (非特許文献6) 等である。

【0003】

最近の研究では、ヒューマノイドロボット上での物体の認識が、中心視の仕組み (非特許文献1、非特許文献2) を使用することによって実質的に改良され得ることが示されている。図1は中心視システムを採用したヒューマノイドロボット30を示した図である。図1を参照して、異なるレンズを備えた2つのカメラ40がロボット30の各々の目 (全部で4つのカメラ) に設けられている。一方のカメラは中心視カメラであり、他方は辺縁視カメラである。中心視カメラはロボットの周囲のより狭い部分のより詳細な視界を提供する狭いレンズを備えている (図2(B)を参照)。辺縁視カメラはロボットの周囲をさらに含むが、詳細さでは劣る画像を与えるようなより広いレンズを備えている (図2(A)を参照)。各々の目では、中心視カメラと辺縁視カメラは堅固に平行光軸で接続されている。各々の目のモーターシステムは2つの独立な自由度から成る。

10

【0004】

先行技術のヒューマノイドロボット30は手42を備えた左腕46を有し、それに対応する右側の腕と手との組を備えている。図1に示されるように、ロボット30は物体を認識する能力を利用して、物体44を手42でつかむことができる。これは腕46と手42とが一種のマニピュレータを形成しているからである。

20

【0005】

図1に示されている中心視システムは有用である。なぜなら、第1には、それらによってロボット30は視野がより広く、画素の分布が疎である画像でロボットの周囲を監視したり探索したりすることが可能であり、それによって探索プロセスの能率を上げているからである。第2には、関心のある物体がいったん中心視に現れると、それらは同時に画素の分布が密で、さらなる詳細を含んでいる中心視の領域からさらなる情報を抽出することができるからである。

【0006】

非特許文献3は、図1に示すように、1つの目につき2つのカメラを備えた中心視システムを開示しており、物体の色のテクスチャと形状とについての事前の粗い知識を使用して、人間によって操作された物体を有効に観測することによって認識の表現をどのように学習することができるかを示している。一言で言うと、非特許文献3で開示されているシステムはボトムアップの方法で物体のための表現を学習しようとしているのである。

30

【非特許文献1】A. ウデ、C. G. アトキソン、及びG. チェン、「検出、追跡、認識及び行動のための辺縁視と中心視との組合せ」、インテリジェントロボット及びシステムに関するIEEE/RSJ国際会議予稿集、ラスベガス、ネバダ、2003年、2173-2178ページ。(A. Ude, C. G. Atkeson, and G. Cheng, "Combining peripheral and foveal humanoid vision to detect, pursue, recognize and act," in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, Las Vegas, Nevada, 2003, pp. 2173 - 2178.)

40

【非特許文献2】M. ビヨルクマン及びD. クラギッシュ、「物体の認識とポーズ推定のための中心視と辺縁視との結合」、ロボティクスとオートメーションに関するIEEE会議予稿集、ニューオーリンズ、ルイジアナ、2004年、5135-5140ページ。

(M. Bjorkman and D. Kragic, "Combination of foveal and peripheral vision for object recognition and pose estimation," in Proc. IEEE Conf. Robotics and Automation, New Orleans, Louisiana, 2004, pp.

50

5135 - 5140.)

【非特許文献3】A. ウデ、C. ガスケット、及びG. チェン、「1つの目について2個のカメラを持つ中心視システム」、ロボティクスとオートメーションに関するIEEE国際会議予稿集、オーランド、フロリダ、2006年、3457-3462ページ。(A. Ude, C. Gaskett, and G. Cheng, "Foveated vision systems with two cameras per eye," in Proc. IEEE Int. Conf. Robotics and Automation, Orlando, Florida, 2006, pp. 3457 - 3462.)

【非特許文献4】P. フィッツパトリック、「最初の接触：セグメンテーションのための活性視覚のアプローチ」、インテリジェントロボットとシステムとに関する2003年IEEE/RSJ国際会議予稿集、ラスベガス、ネバダ、2003年、2161-2166ページ。(P. Fitzpatrick, "First contact: an active vision approach to segmentation," in Proc. 2003 IEEE/RSJ Int. Conf. Intelligent Robots and Systems, Las Vegas, Nevada, 2003, pp. 2161 - 2166.)

【非特許文献5】D. G. ロウエ、「3次元の物体認識のための局地的特徴視覚のクラスタリング」、コンピュータビジョンと認識のためのIEEE会議予稿集、カウアイ、ハワイ、2001年、682-688ページ。(D. G. Lowe, "Local feature view clustering for 3D object recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Kauai, Hawaii, 2001, pp. 682 - 688.)

【非特許文献6】L. ウィスコット、J. - M. フェラス、N. クリュージャー、及びC. フォンデルマルスブルグ、「弾性バッチのグラフマッチングによる顔認識」、IEEEトランザクション、パターン分析、機械知能、第19巻、第7号、775-779ページ、1997年。(L. Wiskott, J. - M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," IEEE Trans. Pattern Anal. Machine Intell., vol. 19, no. 7, pp. 775 - 779, 1997.)

【非特許文献7】S. スリニバサン及びK. L. ボイヤー、「視覚ベースの固有空間を使用する頭のポーズ推定」、パターン認識に関する第16番目の国際会議予稿集、ケベック、カナダ、2002年、302-305ページ。(S. Srinivasan and K. L. Boyer, "Head pose estimation using view based eigenspaces," in Proc. 16th Int. Conf. Pattern Recognition, vol. 4, Quebec, Canada, 2002, pp. 302 - 305.)

【非特許文献8】C. ガスケット、A. ウデ、及びG. チェン、「終点の閉ループと学習された終点の開ループの視覚サーボ制御による手と目の調整」、ヒューマノイドロボティクス国際ジャーナル、第2巻、第2号、203-224ページ、2005年。(C. Gaskett, A. Ude, and G. Cheng, "Hand-eye coordination through endpoint closed-loop and learned endpoint open-loop visual servo control," International Journal of Humanoid Robotics, vol. 2, no. 2, pp. 203 - 224, 2005.)

【非特許文献9】D. コマニッシュ、V. ラメッシュ、及びP. メア、「カーネルベースの

10

20

30

40

50

物体のトラッキング」、パターン分析と機械知能に関するIEEEトランザクション、第25巻、第5号、564-577ページ、2003年。(D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Trans. Pattern Anal. Machine Intell., vol. 25, no. 5, pp. 564 - 577, 2003.)

【非特許文献10】S. J. マッケナ、Y. ラージャ、及びS. ゴング、「適応的混合モデルを使用する色の物体のトラッキング」、画像と視覚とのコンピューティング、第17巻、225-231ページ、1999年。(S. J. McKenna, Y. Raja, and S. Gong, "Tracking colour objects using adaptive mixture models," Image and Vision Computing, vol. 17, pp. 225 - 231, 1999.)

10

【非特許文献11】T. ヨアヒム、「大規模サポートベクタマシン学習の実用化」、カーネル法 サポートベクタ学習の進歩、B. シェルコップ、C. J. C. バークス、A. J. スモラ、編、ケンブリッジ、MA: MITプレス、1999年。(T. Joachims, "Making large-scale support vector machine learning practical," in Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press, 1999.)

20

【発明の開示】

【発明が解決しようとする課題】

【0007】

非特許文献3で提案されたシステムは、ヒューマノイド視覚という仕組みでの物体の認識で生じる多くの問題を扱っているが、ゼロから、すなわち物体について何も知らない状態から物体の認識のための完全な表現を学習するためにヒューマノイドロボットの能力を利用することは、ほとんど行なわれていなかった。特に、事前の知識なしに画像の中から物体を発見することは、ロボットにとっては難しい問題であり、純粋なボトムアップの方法で達成することは不可能ではないとしても大変厳しいことである。

30

【0008】

受動的なコンピュータ視覚システムは、トップダウンのプロセスを導入することによってそれを解決しようと試みる。トップダウンのプロセスとは、初期に得られた特徴を関連付け、グループ化してより大きな集合体と組とを得るための助けとなる、物体についての知識を伝えるものである。特徴をグループ化することにより、単なる特徴を使用する場合よりも物体の構成要素を形成しやすいであろうと期待されている。処理階層においてより高いある水準では、早期指標の間の相互作用はシーン分解において最も大きくなり有意義な構成要素(物体)を生成するはずであり、そしてその有意義な構成要素をさらなるシーンの分析と解釈(認識)の目的のために使用することができる。

40

【0009】

残念ながら、完全に一般的な方法で物体の探索を導くようなトップダウンプロセスを定式化することは容易ではない。統計的な学習のアプローチでは、例として与えられた画像からどのようにそのような画像の分解を生成するかを学習するのは困難であろうと考えられる。なぜなら、人間によって行なわれるような画像の分解は、人間が環境と相互に作用するときを得るような経験に依存するからである。この情報は画像から容易に得られるのではなく、むしろ人間の行動が外部の世界にどのように影響するかに関する経験から来るのである。そのような情報を受動的な学習プロセスにどのように持ってくるかができるかは明確ではない。

【0010】

50

ヒューマノイドロボットは、しかし、調べる行為を行ってその応答から学習することによって、因果関係を使用してその世界を探索するという潜在能力を有する。物体をつつくことが物体の境界のための視覚的な証拠を抽出するために使用可能であると示されており、これは分離に適している。目的は、初期の、粗い物体の分離の後のことである。すなわち、ロボットが物体をつかんだ後で物体の表現の学習を容易にするために何をすることができるか、について研究することである。操作なしに物体の表現を構築することは間違いなく可能である。さもなくば、家のような大きな物体をいかにして認識できるだろうか。しかし、操作が学習プロセスを大いに支援し高速化すると考えられることも確かである。

【0011】

したがって、この発明の目的の1つは、物体についていかなる事前知識もなしに物体の視覚表現を学習することができるような、学習システムとコンピュータプログラムとを提供することである。

10

【0012】

この発明の別の目的は、目標となる物体を操作することによって、物体についていかなる事前知識もなしに物体の視覚表現を学習することができるような、学習システムとコンピュータプログラムとを提供することである。

【0013】

この発明のまた別の目的は、物体についていかなる事前知識もなしに物体の分類を学習することができるような、学習システムとコンピュータプログラムとを提供することである。

20

【0014】

この発明のさらなる目的は、目標となる物体を操作することによって、物体についてのいかなる事前知識もなしに物体の分類を学習することができるような、学習システムとコンピュータプログラムとを提供することである。

【課題を解決するための手段】

【0015】

この発明の第1の局面によれば、ビデオカメラと、物体を保持し、回転させ、平行移動させることができる操作可能なマニピュレータとを用いて物体の視覚表現を学習するための学習システムは、ビデオカメラの出力を受けるように接続されたフレームグラバと、マニピュレータがビデオカメラの光軸に沿ってマニピュレータによって保持された物体を平行移動させるように、マニピュレータを駆動するための平行移動手段と、フレームグラバによって取込まれたフレーム内の物体の画像があらかじめ定められた大きさの境界の範囲内に位置し、かつ境界の範囲内でできるだけ大きくなるように、マニピュレータによって保持された物体の位置を決定するための手段と、マニピュレータによって保持された物体がカメラの光軸と直交する2つの軸回りで回転するように、マニピュレータを駆動するための回転手段と、回転手段によって回転された物体の画像を収集するための画像収集手段とを含み、画像は物体の識別子に関連付けて記憶され、学習システムはさらに、画像と物体の識別子とを用いて、物体の識別子を出力するように分類器をトレーニングするためのトレーニング手段を含む。

30

【0016】

好ましくは、画像収集手段は、フレームグラバによって取込まれたフレームの範囲内で物体の領域を識別するための識別手段と、あらかじめ定められた形状を識別手段によって識別された物体の領域の外縁にフィットするための手段と、あらかじめ定められた形状内で画像を抽出するための手段とを含む。

40

【0017】

さらに好ましくは、識別手段は、フレームグラバによって取込まれたフレームの範囲内の各画素が、背景、マニピュレータの画像の一部、物体の画像の一部、及びアウトライアに属す確率を計算するための手段と、各画素について、画素が物体の画像に属しているかどうかを決定するための手段とを含む。

【0018】

50

さらに好ましくは、回転手段はマニピュレータによって保持された物体がビデオカメラの光軸と直交する2つの軸の回りで回転するように、マニピュレータを駆動するための手段を含む。

【0019】

分類器はサポートベクタマシン(SVM)ベースの分類器であってもよい。

【0020】

この発明の第2の局面によれば、コンピュータプログラムは、ビデオカメラと、ビデオカメラから出力されたフレームを取込むためのフレームグラバと、物体を保持し、回転させ、平行移動させることができる制御可能なマニピュレータとに連結されたコンピュータ上で実行される。このコンピュータプログラムは、コンピュータによって実行されると、コンピュータを、マニピュレータがビデオカメラの光軸に沿ってマニピュレータによって保持された物体を平行移動させるように、マニピュレータを駆動させるための平行移動手段と、フレームグラバによって取込まれたフレーム内での物体の画像が、あらかじめ定められた大きさの境界の範囲内に位置し、かつ境界の範囲内でできるだけ大きくなるように、マニピュレータによって保持された物体の位置を決定するための手段と、マニピュレータによって保持された物体がビデオカメラの光軸と直交する2つの軸の回りで回転するように、マニピュレータを駆動するための回転手段と、回転手段によって回転された物体の画像を収集するための画像収集手段として機能させ、画像は物体の識別子に関連付けて記憶され、コンピュータプログラムはさらに、コンピュータを、画像と物体の識別子とを用いて、物体を識別するように分類器をトレーニングするためのトレーニング手段として機能させる。

【発明を実施するための最良の形態】

【0021】

[アプローチとシステムの構成の概要]

図3はヒューマノイドロボットの制御ブロック50の構成を示す図である。ヒューマノイドロボットは、ビデオカメラ52及びそれぞれ手を有している右腕及び左腕を備え、これらはすべてアクチュエータによって操作可能であってその手が物体をつかみ、平行移動させ、回転させることができる。これ以降、ロボットの制御ブロック50を単に「ロボット」と呼ぶ。ロボット50は背景技術の部分で論じられたように、中心視カメラ(全部で4つのカメラ)を備えているが、この実施の形態では、本質的に1つのカメラだけが必要である。したがって、ロボット50はただ1つのカメラ52を有するものとして示されている。

【0022】

なお、ロボット50の外観は本質的には図1に示されているヒューマノイドロボット30のそれと同じである。それは手42を備えた腕46を有し、手42に置かれた物体をつかむ能力を持つ。

【0023】

図3を参照して、ロボット50は、カメラ52からのビデオストリームの各フレームを取込むためのフレームグラバ60と、物体の視覚表現を学習するためにロボット50の種々の部分を制御するためのコントローラ62と、カメラ52、頭、両腕、及び手等を含むロボット50の種々の部分を動作させるためのアクチュエータ66のセットと、コントローラ62からのパラメータに従ってアクチュエータ66を駆動して、望ましい軌道に沿ってロボット50の種々の部分を動作させるためのアクチュエータドライバ64とを含む。

【0024】

コントローラ62は、フレームグラバ60からのフレーム画像を受けて、物体の種々の画像の表現を収集しながら、制御された態様で物体をつかみ、物体を移動(操作)させるためにロボット50の腕と手とを制御するオブジェクト操作プロセス80と、オブジェクト操作プロセス80によって収集された物体の画像を物体の識別子と関連付けて記憶するための物体画像記憶部84と、オブジェクト操作プロセス80によって計算された軌道に従ってアクチュエータ66を駆動するためのパラメータを計算するためのアクチュエータ

パラメータ計算プロセス 82 と、マルチクラスの物体の分類器として働くように、物体画像記憶部 84 に蓄えられた物体の画像を使用してトレーニングするための SVM ベースの分類器 86 とを実現するようにプログラムされたプロセッサで実現される。

【0025】

学習のための次の手順を設計した。これはオブジェクト操作プロセス 80 によって実現され、ロボット 50 が物体を操作する間に物体の外観の画像を抽出する。

【0026】

1) 学習の開始はユーザによって始められ、ユーザはロボット 50 の手に新しい物体を置く。

【0027】

2) ロボット 50 は一旦物体を保持すると、その手を中心視カメラの視界から離れるように移動し、静止した背景の学習を始める。典型的には、オブジェクト操作プロセス 80 は最初に 5 秒間かなり平滑化された画像中の画素の平均値を学習し、さらに続く 5 秒間で各画素の色の分散を学習する。

【0028】

3) ロボット 50 は物体の観察のためにその手を開始位置に移動させる。いったん開始位置に達すると、操作された物体の位置と範囲とを推定するための手順がオブジェクト操作プロセス 80 で始まる。

【0029】

4) ロボット 50 は、物体を中心視の範囲内に維持しようとしながら、予め定められた直線軌道に沿って物体を移動（平行移動）させる。この段階では腕と手との自由度のみを使用する。その軌道に沿ってキャプチャされた各々の画像での物体の位置と範囲とを推定する。手が開始点に戻った後に、物体の外観を学習するために（推定された位置と大きさに関する）軌道上の最適な位置が決定される。

【0030】

5) ロボット 50 は、前のステップで決定された学習のために最適な位置に物体を移動させる。望ましい手の位置と向きに達すると、2 自由度に沿って手を働かせることを始め、物体をデプス方向に回転する。物体を操作する間、物体の外観の画像を収集する。この段階は、手が 2 自由度のためのあらかじめ指定された動きの範囲をカバーし終ると終了する。

【0031】

6) 異なる設定で再びロボットの手に物体を置き、この手順を繰り返す。

【0032】

7) すべての物体がすべての関連のある配置で置かれ、かつすべての外観の画像が物体画像記憶部 84 に収集され記憶されると、分類器 86 は非線形マルチクラス SVM に基づく方法を使用することによってトレーニングされる。マルチクラス SVM については、追って述べる。

【0033】

上記の概要は幾らかの説明を必要とする。ロボットが自分で物体を手にとることが確かに将来のゴールの 1 つであるが、これはまだ実現されていない。物体の存在と位置についての仮説の自動生成のために非特許文献 4 と類似する手順が予想される。これに続いて未知の物体をつかむことが必要となるが、これはそれ自体が困難な仕事である。種々の設定で物体をロボットの手に置くようなユーザの代わりに、自動の様式では、ロボットは自分で物体をつかみなおすことが必要だろう。困難ではあるが、これは物体の姿勢を外観の画像に関連付けるための新しい可能性を開く。なぜならロボットは初期のポーズに関する物体の各々のポーズを推定するために自己受容的な情報を使用することができるからである。その場合にはポーズ推定のための標準的な視覚ベースのアプローチを使用することができる（非特許文献 7 を参照）。

【0034】

ステップ 2) で学習したような背景モデルは、物体の動き及び照明条件の変動のような

10

20

30

40

50

要因のために、度々変化する。これは、しかし、ここではあまり関係がない。なぜなら、物体が十分に制御されているときは、ロボット50は学習の間その環境の中で他に何も動かないことを保証しているからである。さらに、学習された背景モデルは短命であり、物体をつかみなおすたびに新たに学習される。

【0035】

ステップ4)で述べられた、目に向かう方向及び離れる方向への試験的な移動は、物体の外観を学習するために好適な姿勢で物体を置くために必要とされる。この姿勢は物体が画像の中であまり小さく見えない程度に十分に目に近くなくてはならず、物体の投影が画像の外側にはみ出してしまわない程度に十分に遠くなくてはならない。

【0036】

ここでの基準は、物体が中心視の視界の中でできるだけ大きく見えるべきであり、その一方で、画像の大きさが320×240画素である場合に、中心視の境界は画像の境界から少なくとも40画素離れている必要があるというものである。これは図4の(A)~(C)に示されており、破線である内部の長方形130は画像の境界120から40画素だけ離れた境界を示している。

【0037】

図4(A)を参照して、この場合、物体はカメラ52に近すぎ、画像の境界120の中の画像140が大きすぎるので長方形130の範囲内ではない。対照的に物体がカメラ52から遠すぎると、画像144は図4(C)に示されたように小さすぎるだろう。画像142が長方形130の範囲内でできるだけ大きいとき、物体は図4(B)に示されているように理想的な位置にある。

【0038】

この基準に基づいた理想的な位置は、物体を中心視画像の中心に向かう直線に沿って移動させることによって決定される。

【0039】

これを図5を参照して説明する。図5を参照して、学習プロセスにおいて、カメラ52のレンズの軸106は固定された座標系100のx軸と一致している。x軸のほか座標系はy軸とz軸を含む。それらの軸は互いに直交し、座標系の原点で互いに交差する。開始位置はx軸上に定められている。物体102は背景の学習の後、初期位置に置かれる。双方向の矢印104によって示されているように、ロボット50は物体102を座標系のx軸の方向に移動(平行移動)させて、物体102のための理想的な学習位置を見つける。

【0040】

操作プロセスでは、物体102をデプス方向に回転させる。すなわち、それは図5の中の矢印108と110によって示されているように、y軸とz軸との回りで回転されるのであるが、座標系100のx軸の回りでは回転されない。物体102をデプス方向に回転することによって、種々の角度からの物体102の画像をカメラ52から見ることができ、物体102をx軸回りで回転すると物体102の同じ表面ばかりがカメラ52に向くことになる。

【0041】

正確にキャリブレートされたヒューマノイドロボット上では、デカルト座標での直線軌道を簡単に設計することができる。正確なモデルが利用できない場合のために、非特許文献8で述べられている粗い開ループと、より正確な閉ループとの制御システムに基づいた自動の手順を開発した。辺縁視画像からの情報に基づいて物体を中心視の中心に置くために、中心視と辺縁視との間の関係を支配する中心視原則によってシステムを案内する(非特許文献3)。開発された技術によれば、開ループ制御システムによって設計されたように中心視の方向に向かう粗い直線軌道から始めることができ、その後閉ループ制御システムに基づいてその開ループ制御システムを改善することができる。この手順はその軌道上でいくつかの姿勢を生成し、これを補間して正確な関節空間の軌道を生成することで、結果としてデカルト空間内の直線運動が得られる。

10

20

30

40

50

【 0 0 4 2 】

トレーニングの目的のための物体の視界を抽出するために設計された操作の手順は認識にとって理想的な位置を決定するためにも使用することができる。ステップ 1) ~ 5) を実行することによって物体のスナップショットをキャプチャする。このアプローチはスケールリングに対する不変性を容易にする。なぜなら、常にほぼ同じ距離から物体を観るからである。なお、正規化されたスナップショットとなるように変形するためのステップを実行することが依然として必要であるが、異なるスケールで物体の画像上のマッピングをした場合に導入されるような様々なデジタル化のための操作は大いに減少する。わずかに異なる視角から取られる物体のスナップショットに対する分類器の結果を監視することによって、分類器の信頼性をも改善することができる。なぜなら、分類器が正確であるときだけ、その結果が安定しているのが常であるからである。

10

【 0 0 4 3 】

[物体の発見]

この実施の形態の中心は、物体がロボット 50 によって操作される間、画像から物体を識別するための能力である。この目的を達成するために、次のような画像プロセスをモデル化することが必要である。

【 0 0 4 4 】

- (プロセス o によって表される) 未知の物体、
- 背景 (b)、
- 手 (h)、
- 場面の中の何らかの予期せぬ事象をモデル化したアウトライアプロセス (t)。

20

【 0 0 4 5 】

これらのプロセス o 、 b 、 h 及び t は、画素値が与えられるとその画素が物体の部分、背景、ロボット 50 の手及びアウトライアである確率を推定する確率モデルである。その確率に従って、その画素が物体の部分、背景、ロボットの手又はアウトライアとして分類されることになる。

【 0 0 4 6 】

以下の式で表わされるガウシアンプロセスによって、静止した背景中の各画素の色強度をモデル化する。

【 0 0 4 7 】

【 数 1 】

$$\Theta_b = \{ \bar{I}_u, \bar{\Sigma}_u \} \mathbf{u}$$

これは各画素 u において、平均 - \bar{I}_u (ここで「 $\bar{\quad}$ 」の記号は式中文字の上に付されたものを示す) と共分散行列 - $\bar{\Sigma}_u$ と以下の関連の確率分布で特徴付けられる。

【 0 0 4 8 】

【 数 2 】

$$p(\mathbf{I}_u, \mathbf{u} | \Theta_b) = \frac{1}{2\pi \sqrt{\det(\bar{\Sigma}_u)}} \cdot \exp\left(-\frac{1}{2} (\mathbf{I}_u - \bar{\mathbf{I}}_u)^T \bar{\Sigma}_u^{-1} (\mathbf{I}_u - \bar{\mathbf{I}}_u)\right) \quad (1)$$

40

輝度の変化に対してある程度の頑健性を得るために、ここでは色合い及び彩度のどちらか又は以下の式で表わされる正規化された R G B の値によって、色強度を特徴付ける。

【 0 0 4 9 】

【 数 3 】

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B}$$

これら 3 つの値は独立ではないので、この実施の形態では正規化された赤色及び正規化された緑色だけを使用し、それによっていずれの色空間においても色を 2 次元の値にする。平均値と共分散とは、ロボット 50 が物体を中心視に持ってくる直前に、背景の画素の統

50

計量を集めることによって学習される。2つの色空間のどちらを使用しても、大きな違いは観測されなかったが、この点を確認するためにはさらなる実験が必要である。

【0050】

画像中のロボットの手の位置が自己受容的な情報を使用して計算できるとしても、この情報は十分ではない。なぜなら、手のどの部分が目に見えており、どの部分が操作されている物体によって覆われているかを、前もって知ることができないからである。したがって、画像中の手の外観をモデル化することが必要となる。

【0051】

手の外観のモデリングのために、色ヒストグラム（非特許文献9）及びガウス（混合）モデル（非特許文献10）のような物体のトラッキング理論からの標準的なアプローチで実験をした。トラッキングとは異なり、関心は実際に手の位置を測定することにはなく、特定の画素が手に属する確率を推定することだけにある。色ヒストグラムとガウス混合モデルとは共にこの能力を提供する。ガウス混合モデルは次のように定義される。

【0052】

【数4】

$$p(\mathbf{I}_u | \Theta_k) = \sum_{k=1}^K \frac{\omega_k}{2\pi \sqrt{\det(\Sigma_k)}} \cdot \exp\left(-\frac{1}{2}(\mathbf{I}_u - \bar{\mathbf{I}}_k)^T \Sigma_k^{-1} (\mathbf{I}_u - \bar{\mathbf{I}}_k)\right) \quad (2)$$

この実施の形態では、1つの色によって手を特徴付けることができ、したがって手の外観をモデル化するために、ユニモーダルガウシアン（ $K = 1$ ）を使用することができた。

【0053】

動きのキューは確かにロボットの手と背景とから物体を抽出する上で助けにはなるが、そのようなキューだけでは物体の外観の抽出には十分ではない。ロボットが物体を保持しているとき、物体の動きはロボットの手の動きと同じである。したがって、動きのキューだけに基づいて物体と手とを区別することはできない。さらに、動きの推定は通常、差分法によって計算され、そのため比較的ノイズが多くなる。その結果、動きは他のキューのための補助としてだけ使用されるべきであり、画像からの物体のセグメンテーションのための唯一の特徴量として使用されるべきではない。

【0054】

物体についての事前知識がないので、明らかに、実際に学習したいものであるその外観をモデル化することはできない。物体を操作するために使用する開ループ軌道は、しかし、明確に定義されており、その物体が画像中のどこにあるかは大体はわかっている。したがって、前のステップでの物体に属する画素の平均値 - \mathbf{u} と共分散 - Σ を使用することによって、画像の画素が物体の範囲内にある確率をモデル化することができる。これは次の分布となる。

【0055】

【数5】

$$p(\mathbf{u} | \Theta_o) = \frac{1}{2\pi \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{u} - \bar{\mathbf{u}})^T \bar{\Sigma}^{-1} (\mathbf{u} - \bar{\mathbf{u}})\right) \quad (3)$$

図6の(A)～(D)までは物体の外観抽出の例を示す図である。左から右へ向かって、図6は背景の学習のために使用している画像を(A)で、背景の統計量の収集した物のために使用した画像を平滑化したものを(B)で、ロボットの手42によって操作されている間の、楕円250によって表現される画像の物体の推定された範囲を(C)で、及び確率 $P(\mathbf{u} | \Theta_o)$ をしきい値処理し、クローズドなモルフォロジ演算を適用した後の物体の画素の、接続された要素の最大のものを含む2値化画像252を(D)で示す。

【0056】

図6(C)において、楕円250の範囲内にある画像が物体の表現を学習するために使用される。図6(D)において、2値化画像252は説明の目的のためにだけ示したもので、計算には使用しない。この実施の形態のロボット50が、楕円250を抽出するため

10

20

30

40

50

に、簡単な背景を必要とはしないことを示すために、その場面に追加の物体を置いている。

【0057】

ロボットは軌道の中心と画像の中心とを通過する直線に沿って物体を移動させようとするので、物体の位置は常に画像の中心に近く、最初の小さな範囲で画像中で物体が中心合わせされていると仮定して外観の抽出を初期化することができる。その場合、楕円を物体の画像にフィットすることは容易な仕事である。

【0058】

図2に示されているように、中心視の画像は手42に加えて腕46の他の部分を含むことがある。腕46とシーン中に見えるかもしれない他の予期せぬ物体との見かけについて事前情報がないとき、アウトライアプロセスによってその画像中のそのような事象をモデル化することができる。このプロセスには画像中の画素の位置又はこの画素での色強度の値に関わらず、小さい一定の確率 $P(\theta_t)$ が割当てられている。このプロセスとオブジェクトプロセス θ_o との間の相互作用により、次のような結果となる。すなわち、バックグラウンド及び手と異なるテクスチャの領域は、もしもそれが物体の予想位置に近ければ対象物体として分類され、それ以外の場合にはアウトライアに分類される(式(7)を参照)。さらにモデル θ_o 、 θ_b 及び θ_h によって計算された確率がすべて一定の確率 $P(\theta_t)$ より小さい時、画素をアウトライアとして分類する。

【0059】

腕については、それを含む画像の部分は自己受容的な情報を使用して、計算から除外することができる。動的なヒューマノイドロボット上では、自己受容的な情報は画像中の腕の場所について粗い推定を提供するだけである。しかし、腕を含む画像の大部分を計算から除外するためには十分である。実験では、これとアウトライアプロセスとを組み合わせると、画像中の対象物体の範囲を推定するとき、腕をフィルタによって取除くのに十分であることが示された。

【0060】

画像中のどの画素も互いに独立なプロセス $\theta = \{\theta_b, \theta_h, \theta_o, \theta_t\}$ (閉じた世界の仮定)の1つから生じるものと仮定すれば、色 I_u が、ある場所 u で観測される確率を、全確率の法則を用いて次のように書くことができる。

【0061】

【数6】

$$P(I_u, u | \theta) = \omega_b P(I_u, u | \theta_b) + \omega_h P(I_u | \theta_h) + \omega_o P(u | \theta_o) + \omega_t P(\theta_t) \quad (4)$$

ここで ω_x はプロセス θ_x を観測する事前(混合)確率であり、 $\omega_b + \omega_h + \omega_o + \omega_t = 1$ である。

【0062】

未知の物体の現在の位置とその範囲とを推定することが必要である。それらは学習のための外観の画像を提供するだろう。これはプロセス $\theta = \{\theta_b, \theta_h, \theta_o, \theta_t\}$ が与えられたときに画像 I を観測する確率を最大化することによって達成することができる。隣接する画像をプロセスに割当てる相関を無視することによって、画像 I を観測する全体の確率を次のように評価することができる。

【0063】

【数7】

$$P(I) = P(I | \theta) = \prod_u P(I_u, u | \theta) \quad (5)$$

背景と手の色の分布とは静止していると仮定しているため、物体の位置 u と、物体に属する画素の共分散 Σ と、混合確率 ω_b 、 ω_h 、 ω_o 、及び ω_t とに関して(5)を最大化することができる。(5)を最大化する代わりに、以下の式で表わされる負の対数尤度を最小化することのほうが容易である。

10

20

30

40

50

【 0 0 6 4 】

【 数 8 】

$$L(\Theta, \omega) = -\log(P(\mathbf{I}|\Theta)) = -\sum_{\mathbf{u}} \log(P(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta)) \quad (6)$$

ここで $\Theta = (\omega_b, \omega_h, \omega_o, \omega_t)$ である。ラグランジュの乗数理論を使用すると、上記対数尤度がEMアルゴリズムによって最小化され得ることを示すことができる。

【 0 0 6 5 】

【 数 9 】

$$P(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta_x) = \frac{\omega_x p(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta_x)}{\sum_{y \in \{o, h, b, t\}} \omega_y p(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta_y)} \quad (7)$$

10

ただし、 $x = o, h, b, t$ 、と書くと、EMアルゴリズムは画素の確率(7)を推定する期待ステップと、確率 $P(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta_o) = P(\mathbf{I}_{\mathbf{u}}|\Theta_o)$ を用いて以下で示される物体の画素の平均値と共分散を推定する最大化ステップとから成る。

【 0 0 6 6 】

【 数 1 0 】

$$\bar{\mathbf{u}} = \frac{1}{\sum_{\mathbf{u}} P(\mathbf{u}|\Theta_o)} \sum_{\mathbf{u}} P(\mathbf{u}|\Theta_o) \mathbf{u}, \quad (8)$$

20

$$\bar{\Sigma} = \frac{1}{\sum_{\mathbf{u}} P(\mathbf{u}|\Theta_o)} \sum_{\mathbf{u}} P(\mathbf{u}|\Theta_o) (\mathbf{u} - \bar{\mathbf{u}})(\mathbf{u} - \bar{\mathbf{u}})^T \quad (9)$$

確率 $P(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta_b)$ 及び $P(\mathbf{I}_{\mathbf{u}}|\Theta_h)$ はEMプロセスを通して一定のままであり、したがって各々の画像のために一度だけ推定すればよい。このためビデオレート、すなわち30ヘルツで物体の外観の抽出を実現することが容易になった。混合確率は一定と仮定することも、それらを以下のようにEMプロセスの一部として推定することもできる。

【 0 0 6 7 】

【 数 1 1 】

$$\omega_x = \frac{1}{n} \sum_{\mathbf{u}} P(\mathbf{I}_{\mathbf{u}}, \mathbf{u}|\Theta_x) \quad (10)$$

30

ここで n は画素の数であり、 $x = o, h, b, t$ である。

【 0 0 6 8 】

[物体の表現の学習]

「物体の発見」の節で述べた物体の外観の抽出のための技術に関連して、オブジェクト操作プロセス80の出力を使用している認識のための分類器86(図3を参照)を、この実施の形態のロボット50がどのように学習するかを示さなければならない。楕円250(図6(C)を参照)を推定した後、画像は一定の大きさのウィンドウ上に変換される。これはスケーリング及び平面上の回転に対する不変性を保証し、さらに互いに比較することができるような標準的な大きさの画像をも提供している。

40

【 0 0 6 9 】

最大の分類性能を保証するために、分類器86にせよ、他の一般的な何らかの分類器にせよ、供給されたデータにはある種の前処理が必要であり、これは特に高次元の入力データにとって重要である。多くの現代の視覚ベースのアプローチは局地的な特徴の集まりによって視野を特徴付けている。複素ガボールカーネルを使用して画像中の局所的な構造が識別され、画像は最初にグレースケールに変換される。現在のところ、色ヒストグラムのような特徴は確かに有益ではあるけれども、この実施の形態では認識のために色を使用し

50

ない。ガボールカーネルは次のように与えられる。

【 0 0 7 0 】

【 数 1 2 】

$$\Phi_{\mu,\nu}(\mathbf{x}) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} \cdot \exp\left(-\frac{\|k_{\mu,\nu}\|^2 \|\mathbf{x}\|^2}{2\sigma^2}\right) \cdot \left(\exp(ik_{\mu,\nu}^T \mathbf{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right) \quad (11)$$

ここで $k_{\mu,\nu} = k_\nu [\cos(\Phi_\mu), \sin(\Phi_\mu)]^T$ である。画素 \mathbf{x} でのガボールジェットを、この画素で、複数のガボールカーネルで画像をたたみ込むことによって獲得された複素係数 J_j^x の集合として定義する。ガボールカーネルはそれらがいくつかの異なる波長 k_ν と向き Φ_μ とをサンプリングするように選ぶ。

ウィスコットら（非特許文献6）は $k_\nu = 2^{\frac{\nu+2}{\nu}}$, $\nu = 0, \dots, 4$ 及び

$\Phi_\mu = \mu \frac{\pi}{8}$, $\mu = 0, \dots, 7$ を使用することを提案しているが、これは入来する画像の大きさ及

び画像の構造の両方に依存する。彼らはジェット間の類似度を次式によって測定可能であることを示した。

$$S(\{J_i^x\}, \{J_j^y\}) = \frac{a_x^T * a_y}{\|a_x\| \|a_y\|} \quad (12)$$

ここで $a_x = [J_1^x, \dots, J_s^x]^T$ であり s は複素ガボールカーネルの数である。これは複素係数の大きさが画像中のジェットの位置と共にゆっくりと変化するという事実に基づく。

この実施の形態のシステムでは、画素の規則的な格子 X_G 上のガボールジェットをサンプリングすることによって、特徴ベクタを構築する。各々の格子点では、ロボット 50 はガボールジェットを計算し、それを特徴ベクタに加える。当然、格子点はどの画像でも同じ順序で分析されることが必要である。この実施の形態で使用される格子の大きさは 6×6 であり、変形された画像の大きさは、楕円の外の画素は除外して 160×120 であって、各々のガボールジェットの次元は 40 である。この結果、 16080 次元の特徴ベクタが得られる。これらの特徴ベクタはトレーニングのために SVM ベースの分類器 86 に供給される。

【 0 0 7 1 】

[非線形マルチクラスサポートベクタマシン]

さて、今度は非線形マルチクラス SVM を使用する物体認識のための好適な分類器を探すという問題を扱う。非線形マルチクラス SVM に基づく分類は次の決定関数を使用して実行される。

【 0 0 7 2 】

【 数 1 3 】

$$H(\mathbf{x}) = \arg \max_{r \in \Omega} \left\{ \sum_{i=1}^m \tau_{i,r} K(\mathbf{x}_i, \mathbf{x}) + b_r \right\} \quad (13)$$

ここで、 \mathbf{x} は分類されるべき入力特徴ベクタ（この実施の形態でのガボールジェットを収集したもの）であり、 \mathbf{x}_i は SVM のトレーニングのために供給された特徴ベクタであり、 $\tau_{i,r}$, b_r は SVM のトレーニングによって推定された値であり、 $r = \{1, \dots, N\}$ はクラスの識別子（この実施の形態での物体）である。 $\tau_{i,r} = 0$ での特徴ベクタ \mathbf{x}_i はサポートベクタと呼ばれる。SVM のトレーニングは、Mercer の理論

10

20

30

40

50

を満たす全てのカーネル関数 K についてその収束が保証されている、2 次の最適化問題を解くことから成る。

【 0 0 7 3 】

ガボールジェット (1 2) のための類似度尺度はガボールジェットから成る特徴ベクタの分類のためのカーネル関数の設計にとって良い動機付けを提供する。 X_G をガボールジェットが計算される 2 つの正規化された画像の範囲内での全ての格子点の集合とし、 J_{X_G} と L_{X_G} とを 2 つの異なる画像の中ではあるが同じ格子点上で計算されたガボールジェットとする。好適なカーネル関数を次のように定義する。

【 0 0 7 4 】

【 数 1 4 】

$$K(J_{X_G}, L_{X_G}) = \exp\left(-\rho \frac{1}{M} \sum_{x \in X_G} \left(1 - \frac{a_x^T * b_x^T}{\|a_x\| \|b_x\|}\right)\right) \quad (14)$$

10

ここで M は X_G での格子点の数である。この関数は Mercer の条件を満たし、したがって、サポートベクタの学習のために使用することができる。パラメータ ρ は実験的に供給する必要がある。

【 0 0 7 5 】

[ソフトウェアによる実現]

上述のように、この実施の形態のロボット 5 0 のコントローラ 6 2 はコンピュータのハードウェアとその上で実行されるソフトウェアとで実現される。そのコンピュータは汎用プロセッサでもよい。そのソフトウェアは次のような制御の流れを有する。

20

【 0 0 7 6 】

図 7 はコントローラ 6 2 を実現するソフトウェアの制御の流れを示すフロー図である。図 7 を参照して、このプログラムは、ユーザが物体を開始位置でロボットの手に置くまで待つステップ 1 7 0 と、物体を手にしたときに実行される、背景だけが視界に残るようにロボットの視界から離れるように手を動かすステップ 1 7 2 と、ステップ 1 7 2 に続き、背景の確率モデルを学習するステップ 1 7 4 と、ステップ 1 7 4 に続き、物体を持っているロボットの手を開始位置に移動させるステップ 1 7 6 とを含む。

【 0 0 7 7 】

このプログラムはさらに、ロボットの手を直線軌道の中で動かしながら、操作している物体の位置と範囲とを推定するステップ 1 7 8 と、ステップ 1 7 8 に続き、物体の画像を収集するために物体の最適な位置を決定するステップ 1 8 0 と、ステップ 1 8 0 に続き、物体をステップ 1 8 0 で決定された最適な位置に移動させるステップ 1 8 2 とを含む。

30

【 0 0 7 8 】

プログラムはさらに、ステップ 1 8 2 に続き、物体を最適な位置にデプス方向に回転し、抽出し、収集し、物体画像記憶部 8 4 でその画像を記憶するステップ 1 8 4 と、ステップ 1 8 4 に続き、物体のための画像収集プロセスが終了したかどうかを決定するステップ 1 8 6 とを含む。画像収集プロセスを終了すべき場合は、制御はステップ 1 8 8 に進み、そうでなければ、それはステップ 1 7 0 に戻り、物体が異なる姿勢で置かれるまで待つか、別の物体がロボットの手に置かれるまで待つ。

40

【 0 0 7 9 】

ステップ 1 8 8 では、分類器 8 6 のトレーニングプロセスが物体画像記憶部 8 4 に記憶された画像を使用して実行される。トレーニングが完了すると、そのプロセスは終わる。

【 0 0 8 0 】

ステップ 1 7 0 から 1 7 8 までの動作は比較的簡単であり、したがってここでは、詳細に記述しない。

【 0 0 8 1 】

図 8 はステップ 1 8 0 での動作の詳細を示す図である。図 8 を参照して、図 7 のステップ 1 8 0 で実行されるプログラムのルーチンは、図 3 で示されたフレームグラバ 6 0 を利

50

用してカメラ52からのビデオの信号のフレームを取込むステップ200と、確率モデル θ 、 b 、 h 、及び t を利用して各画素の確率を計算するステップ202と、ステップ202に続き、確率に基づいて物体の画像に属する画素の領域を抽出するステップ204と、ステップ204に続き、推定された物体の画像領域に楕円をフィットし正規化することにより物体の画像の大きさを推定するステップ206と、ステップ206に続き、繰返しの終了条件が満たされたか否かを判断するステップ207とを含む。物体の画像の配置と大きさが大きく変化しないとき又は繰返しの最大数に到達したとき、その繰返しは終了する。繰返しの終了条件が満たされない場合、制御はステップ202に戻り、そうでなければ、制御は繰返しを出て次のステップに進む。

【0082】

プログラムは、さらに、現在の腕の姿勢が、前のいかなる腕の姿勢よりも図4に示す最適な大きさに近い物体の画像をもたらす場合、ロボットの関節の設定を記憶するステップ208と、ステップ208に続き、ロボットの手が開始点に戻るか否かを決定するステップ210と、ステップ210で手が開始点にないと決定されたときに実行され、物体を1ステップだけカメラ52から遠くへ、又はカメラ52に近く移動させるステップ212とを含む。ステップ212の後、制御はステップ200に戻る。ステップ210で手が開始点にあると判断されると、制御はこのルーチンを出る。

【0083】

図9は図7のステップ184の詳細なフロー図を示す図である。図9を参照して、このルーチンは、ステップ180で実行されたそれと類似しており、図3で示されるフレームグラバ60を利用してカメラ52からのビデオの信号のフレームを取込むステップ230と、確率モデル θ 、 b 、 h 、及び t を利用して各画素の確率を計算するステップ232と、ステップ232に続き、その確率に基づいて物体の画像に属する画素の領域を抽出するステップ234と、ステップ234に続き、楕円を推定された物体の画像の領域にフィットさせ、正規化することによって物体の画像の大きさを推定するステップ236と、ステップ236に続き、繰返しの終了条件が満たされたか否かを判定するステップ237とを含む。物体の画像の配置と大きさが大きく変化しないか又は繰返しの最大数に到達したとき、その繰返しは終了する。その繰返しの終了条件を満たさなければ、制御はステップ232に戻り、そうでなければ制御はその繰返しを出て次のステップに進む。

【0084】

ルーチンはさらに、ステップ236に続き、物体画像記憶部84内でステップ236でフィットした楕円の範囲内の画像を物体の画像として抽出するステップ238と、ステップ238に続き、学習の段階を終了すべきか否かを決定するステップ240と、学習の段階を終了すべきでないと判定されたときに実行されて、物体をさらに操作するステップ242とを含む。ステップ242の後、制御はステップ230からステップ240の動作を繰返すためにステップ230に戻る。ステップ240で学習の段階を終了すべきと判定されると、制御はこのルーチンを出る。

【0085】

[動作]

ロボット50、特にロボット50のコントローラ62は次のように動作する。最初に、コントローラ62のオブジェクト操作プロセス80はユーザが物体をロボットの手に置くまで待つ。物体がロボットの手の上に置かれると、オブジェクト操作プロセス80は物体をつかみ、ロボットの手をカメラ52の視界から離れて移動させるので、背景だけが視界に残る。所望の軌道でロボットの手を移動するために必要なパラメータがオブジェクト操作プロセス80からの目標軌道に従ってアクチュエータパラメータ計算プロセス82によって計算され、アクチュエータドライバ64に適用される。アクチュエータドライバ64はそのパラメータに従ってアクチュエータ66を駆動し、ロボットの手は望ましい軌道で移動することになる。

【0086】

オブジェクト操作プロセス 80 は、はじめに 5 秒間、背景のかなり平滑化した画像の画素の平均値を学習し、その後の 5 秒間で各画素での色の分散を学習する。

【 0 0 8 7 】

背景を学習した後、ロボット 50 はその手を開始位置に移動させる。手が開始位置に到達すると、ロボット 50 は中心視の範囲内に物体を維持しながら、図 5 中の双方向の矢印によって示される前もって決定された直線軌道に沿って物体を移動させようと試みる。軌道に沿ったキャプチャされた画像の各々の中の物体の位置と範囲とが推定される。手が開始位置に戻った後、物体の外観を学習するための軌道上の最適位置が決定される。

【 0 0 8 8 】

ロボット 50 は学習のための最適位置に物体を移動させる。望ましい手の位置と向きとに到達した後、それは物体をデプス方向に回転し始める。物体を操作している間、物体の外観の画像を物体画像記憶部 84 で収集し、記憶する。手が 2 つの自由度 (y 軸と z 軸との回りでの回転) のための動きのあらかじめ特定された範囲をカバーし終ると、この段階は終了する。

【 0 0 8 9 】

次に、物体が異なる設定で再びロボットの手に置かれ、この手順をこの物体のために繰返す。

【 0 0 9 0 】

上記の繰返しが物体のすべての設定で完了すると、別の物体がロボットの手に置かれ、同様の動作をこの新しい物体のために繰返す。収集された画像はそれぞれの物体の識別子でラベル付けされる。

【 0 0 9 1 】

すべての物体が適切なすべての設定で置かれ、かつすべての外観の画像が収集され、物体画像記憶部 84 に記憶されると、分類器 86 は非線形マルチクラス SVM に基づく方法を用いてトレーニングされる。

【 0 0 9 2 】

動作の段階では、ユーザが物体をロボットの手に載せると、ロボット 50 は学習の段階の動作としてほとんど同じ動作を繰返す。しかし、学習の段階とは異なり、物体の画像を収集した後、ロボット 50 は画像の画素 X_G の一定の格子の上でのガボールジェットをサンプリングし、ガボールジェットを計算し、それを特徴ベクタに加えることによって物体のための特徴ベクタを構築する。その特徴ベクタは分類器 86 に与えられ、その後、分類器 86 はその特徴ベクタに応じて物体の識別子を出力する。

【 0 0 9 3 】

[実験結果]

本実施の形態では、トレーニングと認識とのために画像を抽出することにおいて、「物体の発見」で述べられたベイズの技術と、「アプローチとシステムの構成の概要」の節で概略を述べた物体の操作とを組合せたものがどれだけ有効であるかを検査した。ロボットがその環境にどのように相互作用するかについて合理的な仮定であると信ずるところに基づいて、物体についての事前知識なしに物体の画像を収集することができた。シーン中の他の物体からある物体を識別するための手順は、ベイズのアプローチによってなされた仮定を満たす限りは、信頼できることが立証された。

【 0 0 9 4 】

提案されたアプローチが本当に物体の表現を学習するために使用され得ることを立証するために、画像の他の部分から物体を識別するために既知の色のテクスチャを使用したときに達成された分類の結果 (非特許文献 1) とそれとを比較した。SVM をトレーニングするために、14 個の異なる物体の 104 個の画像を収集した。それらの 4 つの外観の画像を提案されたアプローチを使用して抽出し、残りのものの画像はセグメンテーションのために色のテクスチャのモデルを適用して収集した。

【 0 0 9 5 】

14 個の物体のライブラリについて、回転及びスケールについて完全に不変である分類

10

20

30

40

50

器をトレーニングするために、次元が16080の1456の特徴ベクタを使った。非特許文献11による非線形マルチクラスSVMの実現例を使用し、これはユーザ定義のカーネルを許容するものである。その結果、「非線形マルチクラスサポートベクタマシン」の節で述べている特別に設計されたカーネルを使用することができた。

【0096】

【表1】

テーブル1 分類結果

	正しい	間違い	認識率
完全なライブラリ	7307	421	94.60%
色によって検出された物体	4897	303	94.20%
操作によって検出された物体	2410	118	95.30%

10

試験のために、ライブラリから物体の他の7728の外観の画像を収集した。テーブル1の結果は、この実施の形態によって収集された画像が、色のテクスチャの事前モデルを使用して収集した視界とまさしく同等に使用に適していることを立証している。この実施の形態の認識結果の方がわずかにより良好なほどであったが、これは視界を抽出するために色のテクスチャのセグメンテーションを用いた物体の分類率が比較的悪かったために引き起こされたものであった。この物体を除外したとき、認識率はほとんど同一となった。

20

【0097】

【結論】

この実施の形態によれば、物体についての事前知識を何ら持たず、画像に人手によって変更を加えることがないような、ヒューマノイドロボットによる認識のための完全な物体の表現を学習する手順が提供される。知られている限り、この実施の形態のロボット50は、もしそれが物体をつかむことができれば、その画像を全く自動的に収集することができるのはじめてのヒューマノイドロボットである。その実験は構築されたモデルが3次元の中で大きさと回転とに対して全く不変であること、及びこの実施の形態が、実施の形態のシステム上で場面の残りの部分からの画像を認めるために物体の色のテクスチャについての事前知識を使用したより以前のシステムに匹敵する認識率を達成していることを示している。

30

【0098】

特に照明条件がトレーニングと認識の段階とで異なるとき、ガボールジェットによって表現された画像の分類のためにこの実施の形態で開発された新しいカーネルはヒューマノイドロボット又はコントローラ62に、より確実に画像の分類を可能にする。画像を収集するためのこの実施の形態は全く一般的であり、開発された分類の技術に制限されないということにも注目されたい。このアプローチは非特許文献5で提案されているSIFTキー+ハフ変換のように、他に普及しているアプローチに適用可能であろう。

【0099】

検討すべき問題の1つは向きによってトレーニング用の画像を編成するための自己受容的な情報の使用である。動的な情報が認識のために有用であり得ることが示されている。物体を制御するロボットは視覚ベースの動的な物体の認識のための必要な入力を提供することができる。さらに、そのような情報は認識の後で物体の向きを推定するために使用することができる。

40

【0100】

上述の実施の形態は単なる例示であって制限的なものと解してはならない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内でのすべての変更を含む。

【図面の簡単な説明】

50

【 0 1 0 1 】

【図 1】操作の間に物体を観測しているこの発明の 1 つの実施の形態に関連したヒューマノイドロボットを示す図である。

【図 2】辺縁視と中心視とのカメラからの同時の視界を示す図である。

【図 3】ロボット 5 0 の機能のブロック図である。

【図 4】物体を操作するための理想的な位置を探すカメラ 5 2 からの視界を例示する図である。

【図 5】学習の段階の間の物体の動きの軌道を例示する図である。

【図 6】物体の外観の抽出の例を示す図である。

【図 7】コントローラ 6 2 を実現するソフトウェアの制御の流れを示すフロー図である。 10

【図 8】図 7 のステップ 1 8 0 で動作の詳細なフロー図である。

【図 9】図 7 のステップ 1 8 4 の詳細なフロー図である。

【符号の説明】

【 0 1 0 2 】

3 0 , 5 0 ヒューマノイドロボット

4 0 , 5 2 カメラ

4 2 ロボットの手

4 4 , 1 0 2 物体

4 6 ロボットの腕

6 0 フレームグラバ 20

6 2 コントローラ

6 4 アクチュエータドライバ

6 6 アクチュエータ

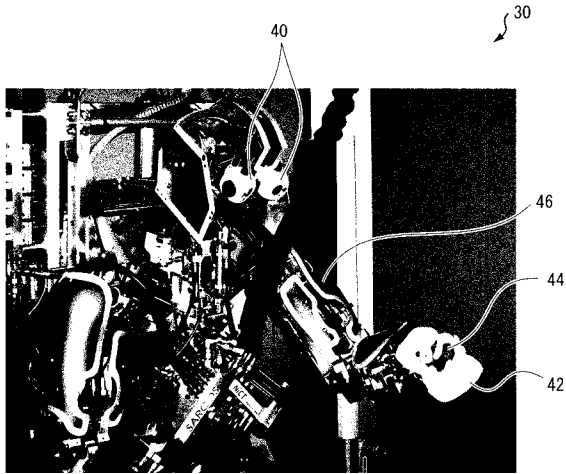
8 0 オブジェクト操作プロセス

8 2 アクチュエータパラメータ計算プロセス

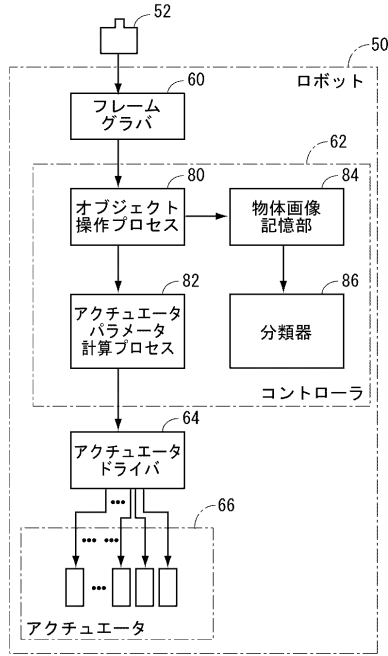
8 4 物体画像記憶部

8 6 分類器

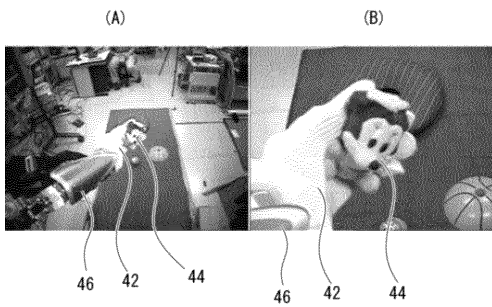
【図1】



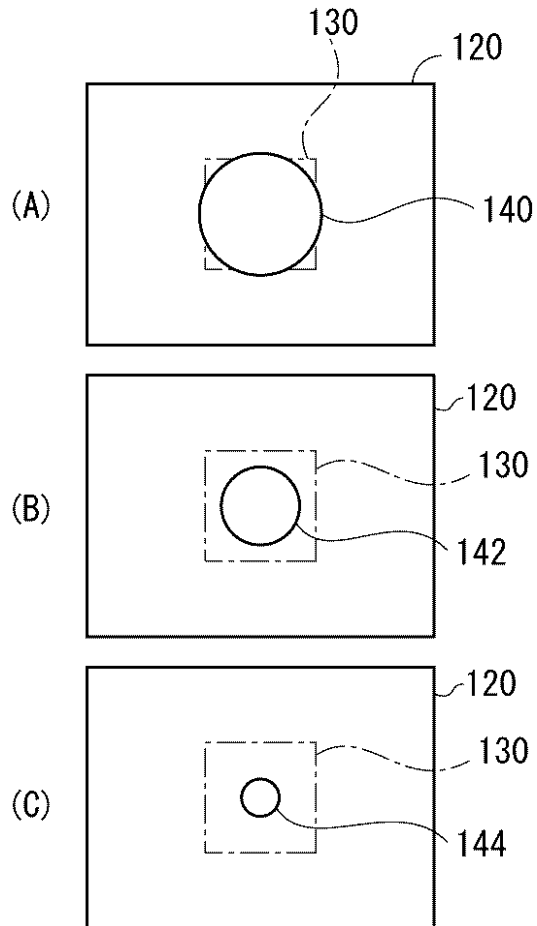
【図3】



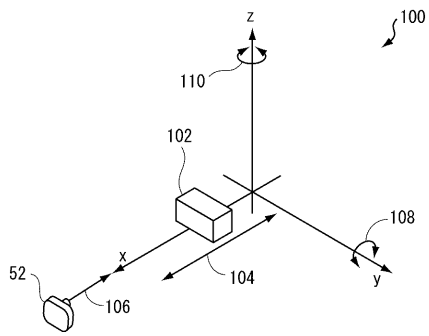
【図2】



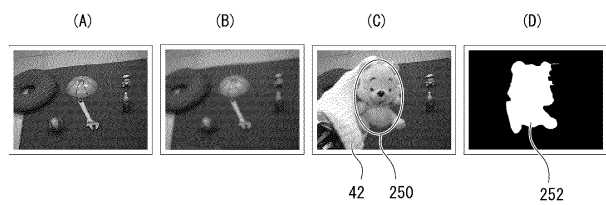
【図4】



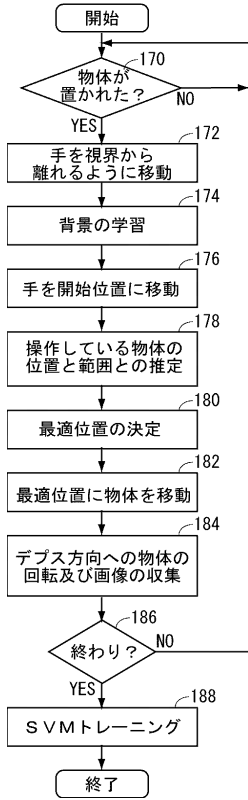
【図5】



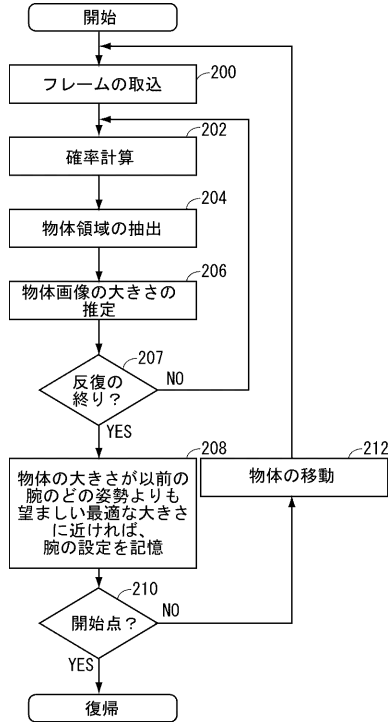
【図6】



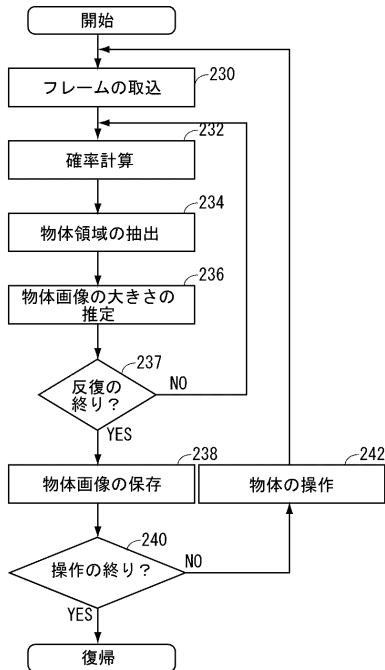
【図7】



【図8】



【図9】



フロントページの続き

- (72)発明者 カイ・ウェルケ
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 ジョシュア・ヘイル
京都府相楽郡精華町光台二丁目2番地2 独立行政法人科学技術振興機構内

審査官 松永 稔

- (56)参考文献 特開2003-346152(JP,A)
特開昭62-286189(JP,A)
特開平11-326235(JP,A)
特開2006-350575(JP,A)
特開2002-208013(JP,A)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|------|------|
| G06T | 7/00 |
| G06T | 1/00 |