

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7121017号
(P7121017)

(45)発行日 令和4年8月17日(2022.8.17)

(24)登録日 令和4年8月8日(2022.8.8)

(51)国際特許分類		F I		
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00	3 5 0 C
G 0 6 T	7/70 (2017.01)	G 0 6 T	7/70	Z
G 0 6 N	3/04 (2006.01)	G 0 6 N	3/04	1 4 5
		G 0 6 N	3/04	1 5 4

請求項の数 29 (全15頁)

(21)出願番号	特願2019-539805(P2019-539805)	(73)特許権者	507226592
(86)(22)出願日	平成30年1月17日(2018.1.17)		オックスフォード ユニヴァーシティ イ
(65)公表番号	特表2020-505695(P2020-505695		ノヴェーション リミテッド
	A)		英国 オーエックス2 0ジェイビー オ
(43)公表日	令和2年2月20日(2020.2.20)		ックスフォード ウエスト ウェイ3 バ
(86)国際出願番号	PCT/GB2018/050132	(74)代理人	110001999
(87)国際公開番号	WO2018/134587		特許業務法人はなぶさ特許商標事務所
(87)国際公開日	平成30年7月26日(2018.7.26)	(72)発明者	クラーク, ロナルド
審査請求日	令和3年1月15日(2021.1.15)		南アフリカ, 2 1 9 6 ハウテン州, サ
(31)優先権主張番号	20170100023		ントン, ウィロワイルド, 8 アデル ロ
(32)優先日	平成29年1月23日(2017.1.23)	(72)発明者	ワン, セン
(33)優先権主張国・地域又は機関	ギリシャ(GR)		中華人民共和国, ベイジン 1 0 1 1 0
(31)優先権主張番号	1703006.5		1, トンジョウ ディストリクト, サウ
(32)優先日	平成29年2月24日(2017.2.24)		最終頁に続く
	最終頁に続く		

(54)【発明の名称】 モバイル装置の位置特定方法

(57)【特許請求の範囲】

【請求項1】

カメラと少なくとも1つの慣性センサとを含むモバイル装置の位置を特定するための、コンピュータで実行される方法であって、

前記カメラを使用して、一定期間にわたって一連の画像を取り込むステップと、

前記一連の画像からの、複数の連続した一対の画像について、各一対の画像のうち1番目の画像が取り込まれた時間と、各一対の画像のうち2番目の画像が取り込まれた時間との間の、前記装置の動きを示す特徴を、第1のニューラルネットワークを用いて抽出するステップと、

前記少なくとも1つの慣性センサを使用して、前記装置の動きを示すデータを取り込むステップと、

前記少なくとも1つの慣性センサからの前記装置の動きを示す前記データから、前記装置の動きを示す特徴を、第2のニューラルネットワークを用いて抽出するステップと、

一連の連続した画像について、前記第1のニューラルネットワーク及び前記第2のニューラルネットワークによって抽出された特徴から、前記装置の位置を示す特徴を、第3のニューラルネットワークを用いて抽出するステップと、

一連の連続した画像について、前記第3のニューラルネットワークによって抽出された特徴から、前記装置の位置を特定するステップと、を含むことを特徴とする方法。

【請求項2】

前記装置の位置に加えて、前記装置の向きを特定することを特徴とする請求項1記載の

10

20

方法。

【請求項 3】

前記一連の画像の画像は、単眼画像であることを特徴とする請求項 1 又は 2 記載の方法。

【請求項 4】

前記少なくとも 1 つの慣性センサは、加速度計及び／又はジャイロスコープを含むことを特徴とする請求項 1 から 3 のいずれか 1 項記載の方法。

【請求項 5】

前記第 1 のニューラルネットワークは、畳み込みニューラルネットワークであることを特徴とする請求項 1 から 4 のいずれか 1 項記載の方法。

【請求項 6】

前記第 2 のニューラルネットワークは、リカレントニューラルネットワークであることを特徴とする請求項 1 から 5 のいずれか 1 項記載の方法。

【請求項 7】

前記第 2 のニューラルネットワークは、長・短期記憶ニューラルネットワークであることを特徴とする請求項 6 記載の方法。

【請求項 8】

前記第 3 のニューラルネットワークは、リカレントニューラルネットワークであることを特徴とする請求項 1 から 7 のいずれか 1 項記載の方法。

【請求項 9】

前記第 3 のニューラルネットワークは、長・短期記憶ニューラルネットワークであることを特徴とする請求項 8 記載の方法。

【請求項 10】

更に、前記第 2 のニューラルネットワークによって抽出された特徴から、連続して取り込まれた各一対の画像について、前記装置の相対的な位置及び向きを抽出するステップを含むことを特徴とする請求項 1 から 9 のいずれか 1 項記載の方法。

【請求項 11】

前記装置の位置を特定するステップは、前記第 3 のニューラルネットワークによって抽出された特徴から抽出される、前記相対的な位置及び向きを構成することを含むことを特徴とする請求項 10 記載の方法。

【請求項 12】

更に、連続して取り込まれた各一対の画像について、前記相対的な位置及び向きの情報に対応する不確実性の情報を抽出するステップを含むことを特徴とする請求項 10 又は 11 記載の方法。

【請求項 13】

前記一連の画像の各画像は、その対応する位置の情報に関連付けられており、更に、前記対応する位置の情報を使用して、前記第 1 のニューラルネットワーク、前記第 2 のニューラルネットワーク、及び前記第 3 のニューラルネットワークを訓練するステップを含むことを特徴とする請求項 1 から 12 のいずれか 1 項記載の方法。

【請求項 14】

更に、前記第 1 のニューラルネットワークによって抽出される特徴と、前記第 2 のニューラルネットワークによって抽出される特徴との間の相関が最大になるように、前記第 1 のニューラルネットワーク、前記第 2 のニューラルネットワーク、及び前記第 3 のニューラルネットワークを訓練するステップを含むことを特徴とする請求項 1 から 13 のいずれか 1 項記載の方法。

【請求項 15】

前記装置は、自律型ロボットであることを特徴とする請求項 1 から 14 のいずれか 1 項記載の方法。

【請求項 16】

モバイル装置であって、メモリ、プロセッサ、カメラ、及び少なくとも 1 つの慣性センサを含み、

10

20

30

40

50

前記カメラを使用して、一定期間にわたって一連の画像を取り込み、

前記一連の画像からの、複数の連続した一対の画像から、各一対の画像のうち1番目の画像が取り込まれた時間と、各一対の画像のうち2番目の画像が取り込まれた時間との間の、当該装置の動きを示す特徴を、前記プロセッサにより提供される第1のニューラルネットワークを用いて抽出し、

前記少なくとも1つの慣性センサを使用して、当該装置の動きを示すデータを取り込み、

前記少なくとも1つの慣性センサからの当該装置の動きを示す前記データから、当該装置の動きを示す特徴を、第2のニューラルネットワークを用いて抽出し、

一連の連続した画像について、前記第1のニューラルネットワーク及び前記第2のニューラルネットワークによって抽出された特徴から、当該装置の位置を示す特徴を、前記プロセッサにより提供される第3のニューラルネットワークを用いて抽出し、

一連の連続した画像について、前記第3のニューラルネットワークによって抽出された特徴から、当該装置の位置を特定するように構成されていることを特徴とする装置。

【請求項17】

当該装置の位置に加えて、当該装置の向きを特定するように構成されていることを特徴とする請求項16記載の装置。

【請求項18】

前記一連の画像の画像は、単眼画像であることを特徴とする請求項16又は17記載の装置。

【請求項19】

前記少なくとも1つの慣性センサは、加速度計及び/又はジャイロスコープを含むことを特徴とする請求項16から18のいずれか1項記載の装置。

【請求項20】

前記第1のニューラルネットワークは、畳み込みニューラルネットワークであることを特徴とする請求項16から19のいずれか1項記載の装置。

【請求項21】

前記第2のニューラルネットワークは、リカレントニューラルネットワークであることを特徴とする請求項16から20のいずれか1項記載の装置。

【請求項22】

前記第2のニューラルネットワークは、長・短期記憶ニューラルネットワークであることを特徴とする請求項21記載の装置。

【請求項23】

前記第3のニューラルネットワークは、リカレントニューラルネットワークであることを特徴とする請求項16から22のいずれか1項記載の装置。

【請求項24】

前記第3のニューラルネットワークは、長・短期記憶ニューラルネットワークであることを特徴とする請求項23記載の装置。

【請求項25】

更に、前記第2のニューラルネットワークによって抽出された特徴から、連続して取り込まれた各一対の画像について、当該装置の相対的な位置及び向きの情報を抽出するように構成されていることを特徴とする請求項16から24のいずれか1項記載の装置。

【請求項26】

前記第2のニューラルネットワークによって抽出された特徴から抽出される、前記相対的な位置及び向きの情報を構成することにより、当該装置の位置を特定するように構成されていることを特徴とする請求項25記載の装置。

【請求項27】

更に、連続して取り込まれた各一対の画像について、前記相対的な位置及び向きの情報に対応する不確実性の情報を抽出するように構成されていることを特徴とする請求項25又は26記載の装置。

【請求項28】

10

20

30

40

50

モバイル装置で実行されたときに、請求項 1 から 15 のいずれか 1 項記載の方法を実行するように構成されていることを特徴とするコンピュータプログラム。

【請求項 29】

モバイル装置で実行されたときに、請求項 16 から 27 のいずれか 1 項記載のモバイル装置を提供するように構成されていることを特徴とす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】

本発明は、モバイル装置の位置を特定することに関する。より詳細には、排他的ではないが、本発明は、モバイル装置のカメラによって取り込まれた画像と、モバイル装置の慣性センサによって取り込まれた動きの情報との組み合わせから、モバイル装置の位置を特定するために、ニューラルネットワークを使用することに関する。

10

【0002】

特に、本発明は、排他的ではないが、モバイル装置が自律型ロボットである場合に適用可能なものである。しかしながら、本発明は、携帯電話、スマートウォッチ、及びそれらの類のものといった、他の種類のモバイル及びウェアラブル装置にも適用可能である。

【0003】

本明細書で扱う「位置」は、モバイル装置が緯度及び経度によって規定される地球上の位置にある、といった絶対位置を指し、又、別の位置に対する相対位置を指すこともある（例えば、モバイル装置は最初の開始位置からの距離及び方向）。位置の特定には、例えば地球の磁場に関する絶対的な用語での、及び初期の向きに対する特定量の回転のような相対的な用語での、向きの特定も含まれることが多い。

20

【背景技術】

【0004】

GPS 信号が利用できない場所で、モバイル装置の位置を特定できることが望ましい。これは、正確な誘導を可能にする自律型ロボットの場合に特に当てはまる。既知のアプローチは、位置を特定するためにカメラからの画像を使用することである。しかしながら、そのようなシステムでは、位置を確実に特定する場合、カメラに非常に正確な校正が必要になる傾向がある。従来のビジュアル・オドメトリ技術には、スプース法（特徴の検出及び照合と動きの推定及び最適化とを含む、幾つかのステップを含む）、及び直接法（動きの推定及び最適化のステップを含む）が含まれる。そのような技術は、正確なカメラ校正を必要とする傾向があり、テクスチャの少ない環境（すなわち特徴が少ない環境）や、画像を取り込むカメラが急速に回転する場合に、失敗することがよくある。更に、このようなシステムは、通常、移動経路の形状を判定することはできるが、大抵はスケール、つまり実際の移動距離を推定することはできない。

30

【0005】

或いは、位置を特定するために、ニューラルネットワークを使用してカメラからの画像を処理することが知られている。モハンティ等作、DeepVO：単眼ビジュアル・オドメトリのための深層学習方法、arXiv 識別子：1611.06069、2016年11月18日公開の論文は、そのようなシステムを開示している。しかしながら、ニューラルネットワークを使用する既知のシステムには様々な問題がある。それらは、使用される特定の環境向けのトレーニングを必要とする傾向があるため、最初に適切なトレーニングが行われないと、新しい環境で使用することができない。

40

【0006】

カメラの姿勢と慣性データとの関係を学習するために、ニューラルネットワークを使用することが知られている。融合することを学ぶ：視覚慣性カメラの姿勢推定のための深層学習方法、IEEE ISMAR、2016年12月15日公開の論文は、そのようなシステムを開示している。しかしながら、ニューラルネットワークを使用した既知のアプローチには様々な問題がある。不都合なことに、それらは、正確な姿勢推定を得るために、従来の方法を使用してカメラデータを処理する必要があり、動作するためにマーカを必

50

要とする。更に、それらは、生データを入力として使用して、エンドツーエンドのトレーニングを行うことができない。

【 0 0 0 7 】

画像データと動きのデータとを組み合わせ、位置を特定するシステムが知られている。但し、そのようなシステムの問題は、カメラが画像を取り込むのと同じ間隔で動きのデータを提供するように慣性センサが設計されていないため、画像データと動きのデータとを時間同期させる必要があり、そのようなシステムのセットアップが難しくなることである。更に、カメラや慣性センサの較正にミスがあると、重大なエラーにつながる虞がある。

【 0 0 0 8 】

本発明は、上述した問題を軽減しようとするものである。又、代替的及び/又は付加的に、本発明は、モバイル装置の位置を特定する改善された方法を提供しようとするものである。

10

【 発明の概要 】

【 0 0 0 9 】

本発明の第1の態様によれば、カメラと少なくとも1つの慣性センサを含むモバイル装置の位置を特定するための、コンピュータで実行される方法であって、カメラを使用して、一定期間にわたって一連の画像を取り込むステップと、一連の画像からの、複数の連続した一対の画像について、各一対の画像のうち1番目の画像が取り込まれた時間と、各一対の画像のうち2番目の画像が取り込まれた時間との間の、装置の動きを示す特徴を、第1のニューラルネットワークを用いて抽出するステップと、少なくとも1つの慣性センサを使用して、装置の動きを示すデータを取り込むステップと、少なくとも1つの慣性センサからの装置の動きを示すデータから、装置の動きを示す特徴を、第2のニューラルネットワークを用いて抽出するステップと、一連の連続した画像について、第1のニューラルネットワーク及び第2のニューラルネットワークによって抽出された特徴から、装置の位置を示す特徴を、第3のニューラルネットワークを用いて抽出するステップと、一連の連続した画像について、第3のニューラルネットワークによって抽出された特徴から、装置の位置を特定するステップと、含む方法が提供される。

20

【 0 0 1 0 】

第1及び第2のニューラルネットワークを利用して、画像と動きのデータとの双方から特徴を抽出し、その上で、第3のニューラルネットワークを利用して、それらの特徴から位置を抽出することで、はるかに堅牢で信頼性の高い位置特定が可能になることが判明した。特に、とりわけ互いに関して、カメラと慣性センサとを別々に慎重に較正することは、それら2つの間の時間同期と同様に、必要がない。更に、システム全体をトレーニングすることで、全てのニューラルネットワークを同時にトレーニングすることができる。これの意味するところは、特に、実際には使用に最適なタイプの特徴ではない虞がある事前選択された属性で、動きの特徴を抽出するようにトレーニングするのではなく、システム全体の動作に最適な動きの特徴を抽出するように、第1のニューラルネットワーク及び第2のニューラルネットワークをトレーニングできることである。更に、第1のニューラルネットワーク及び第2のニューラルネットワークは、互いに組み合わせられたときに最適な特徴を抽出するようにトレーニングされ、換言すれば、第1のニューラルネットワーク及び第2のニューラルネットワークからの望ましい特徴の如何なる同期も、トレーニングの一部として自動的に処理されることになる。更に、第3のニューラルネットワークは、第1のニューラルネットワーク及び第2のニューラルネットワークから特徴を取得するため、発生する虞がある較正や同期の欠如を最適に処理するように、同時にトレーニングされることとなる。(第3のニューラルネットワークがあらゆる同期の欠如を処理するように訓練される範囲は、トレーニングから自動的に発生し、一般的に、同期がより良い位置特定につながる場合は、第1のニューラルネットワークと第2のニューラルネットワークとが同期するようにトレーニングされるが、第3のニューラルネットワークによって同期の欠如を効果的に処理できる場合は、そうするようにトレーニングされる。)

30

40

【 0 0 1 1 】

50

好ましくは、装置の位置に加えて装置の向きが特定される。従って、装置の「ポーズ」が特定される。

好ましくは、一連の画像の画像が単眼画像である。

【0012】

少なくとも1つの慣性センサは、加速度計及び/又はジャイロ스코プを含んでいる。それに加えて又は代替的に、磁力計、或いは他の適切な慣性センサが含まれる。

【0013】

好適には、第1のニューラルネットワークが畳み込みニューラルネットワークである。この種のニューラルネットワークは、画像データ等の多数のパラメータを有するデータを処理するのに特に適している。

【0014】

好適には、第2のニューラルネットワークがリカレントニューラルネットワークである。この場合、好ましくは、第2のニューラルネットワークが長・短期記憶ニューラルネットワークである。

【0015】

好適には、第3のニューラルネットワークがリカレントニューラルネットワークである。この場合、好ましくは、第3のニューラルネットワークが長・短期記憶ニューラルネットワークである。リカレントニューラルネットワーク、特に長・短期記憶タイプのもものは、時間依存型データの処理に特に適している。

【0016】

好ましくは、本方法は、更に、第2のニューラルネットワークによって抽出された特徴から、連続して取り込まれた各一对の画像について、装置の相対的な位置及び向きの情報を抽出するステップを含む。この場合、好ましくは、装置の位置を特定するステップが、第2のニューラルネットワークによって抽出された特徴から抽出される、相対的な位置及び向きの情報を構成することを含む。換言すれば、装置の位置は、連続した動きの推定からオドメトリ的に (odometrically) 特定される。

【0017】

好適には、本方法は、更に、連続して取り込まれた各一对の画像について、相対的な位置及び向きの情報に対応する不確実性の情報を抽出するステップを含む。不確実性の情報は、ポーズの情報と共に、同時ローカリゼーション及びマッピング (SLAM) アルゴリズムへの入力として使用することができる。

【0018】

一連の画像の各画像は、その対応する位置の情報に関連付けられていてもよく、本方法は、更に、対応する位置の情報を使用して、第1及び第2のニューラルネットワークを訓練するステップを含んでいてもよい。好ましくは、各画像は、その向きの情報にも関連付けられている。

【0019】

好適には、本方法は、更に、第1のニューラルネットワークによって抽出される特徴と、第2のニューラルネットワークによって抽出される特徴との間の相関が最大になるように、第1、第2、及び第3のニューラルネットワークをトレーニングするステップを含む。第1のニューラルネットワーク及び第2のニューラルネットワークの特徴の双方は、モバイル装置の動きを示しており、一般的に、それらの特徴間の相関は予想されるべきものであるため、そのような相関を最大化するトレーニングは有利である。

【0020】

装置は、自律型ロボットであってもよい。或いは、装置は、携帯電話、ウェアラブル装置、又は他の適切なモバイル装置であってもよい。

【0021】

本発明の第2の態様によれば、モバイル装置であって、メモリ、プロセッサ、カメラ、及び少なくとも1つの慣性センサを含み、カメラを使用して、一定期間にわたって一連の画像を取り込み、一連の画像からの、複数の連続した一对の画像から、各一对の画像のう

10

20

30

40

50

ち1番目の画像が取り込まれた時間と、各一对の画像のうち2番目の画像が取り込まれた時間との間の、装置の動きを示す特徴を、プロセッサにより提供される第1のニューラルネットワークを用いて抽出し、少なくとも1つの慣性センサを使用して、装置の動きを示すデータを取り込み、少なくとも1つの慣性センサからの装置の動きを示すデータから、装置の動きを示す特徴を、第2のニューラルネットワークを用いて抽出し、一連の連続した画像について、第1のニューラルネットワーク及び第2のニューラルネットワークによって抽出された特徴から、装置の位置を示す特徴を、プロセッサにより提供される第3のニューラルネットワークを用いて抽出し、一連の連続した画像について、第2のニューラルネットワークによって抽出された特徴から、装置の位置を特定するように構成されている装置が提供される。

10

【0022】

好ましくは、本装置は、装置の位置に加えて装置の向きを特定するように構成されている。

好ましくは、一連の画像の画像が単眼画像である。

【0023】

少なくとも1つの慣性センサは、加速度計及び/又はジャイロスコープを含んでいてもよい。

好適には、第1のニューラルネットワークが畳み込みニューラルネットワークである。

【0024】

好適には、第2のニューラルネットワークがリカレントニューラルネットワークである。この場合、好ましくは、第2のニューラルネットワークが長・短期記憶ニューラルネットワークである。

20

【0025】

好適には、第3のニューラルネットワークがリカレントニューラルネットワークである。この場合、好ましくは、第3のニューラルネットワークが長・短期記憶ニューラルネットワークである。

【0026】

好ましくは、本装置は、更に、第2のニューラルネットワークによって抽出された特徴から、連続して取り込まれた各一对の画像について、装置の相対的な位置及び向きの情報を抽出するように構成されている。この場合、好ましくは、本装置は、第2のニューラルネットワークによって抽出された特徴から抽出される、相対的な位置及び向きの情報を構成することによって、装置の位置を特定するように構成されている。

30

【0027】

好適には、本装置は、更に、連続して取り込まれた各一对の画像について、相対的な位置及び向きの情報に対応する不確実性の情報を抽出するように構成されている。

【0028】

本発明の第3の態様によれば、モバイル装置で実行されたときに、上述した何れかの方法を実行するように構成された、コンピュータプログラム製品が提供される。

【0029】

本発明の第4の態様によれば、モバイル装置で実行されたときに、上述した何れかのモバイル装置を提供するように構成された、コンピュータプログラム製品が提供される。

40

【0030】

当然のことながら、本発明の一態様に関連して説明された特徴が、本発明の他の態様に組み込まれてもよいことは、理解されるであろう。例えば、本発明の方法は、本発明のモバイル装置を参照して説明された何れの特徴をも組み込むことができ、その逆もまた同様である。

【0031】

ここで、本発明の実施形態について、添付の概略的な図面を参照しながら、単なる例として説明する。

【図面の簡単な説明】

50

【 0 0 3 2 】

【 図 1 】 本発明の実施形態に係る自律型ロボットの概略図である。

【 図 2 】 図 1 の自律型ロボットの、その位置を推定するための動作を示すフローチャートである。

【 図 3 】 図 2 の手順を表す概略図である。

【 図 4 】 図 2 及び図 3 の手順で使用される LSTM の概略図である。

【 発明を実施するための形態 】

【 0 0 3 3 】

本発明の実施形態に係る自律型ロボットの概略図を、図 1 に示している。自律型ロボット 1 は、プロセッサ 2 を含んでいる。異なる実施形態では、プロセッサ 2 がシングルプロセッサシステム、デュアルプロセッサシステム、或いはその他の適切なプロセッサシステムであってもよいことは、理解されるであろう。プロセッサ 2 は、カメラ 3、慣性センサ 5、及びメモリ 4 と通信を行い、メモリ 4 は、(とりわけ)カメラ 3 によって取り込まれた画像と、慣性センサ 5 によって取り込まれた動きのデータとを記憶する。

10

【 0 0 3 4 】

慣性センサ 5 は、ジャイロスコープ、加速度計、磁力計、又は他の適切な慣性センサであってもよく、或いは、その上でそのようなデバイスの組み合わせであってもよい。

【 0 0 3 5 】

ここで、自律型ロボット 1 がその位置を特定するための動作について、図 2 のフローチャートを参照して説明する。その手順を図 3 にも概略的に示している。各時間ステップにおいて、自律型ロボット 1 は、現在利用可能な情報に基づいてその位置を特定する。図 3 は、3 つの連続した時間ステップ t 、 $t + 1$ 、及び $t + 2$ を示しており、各時間ステップについての自律型ロボット 1 の「ポーズ (pose)」である、 $P o s e t$ 、 $P o s e t + 1$ 、及び $P o s e t + 2$ の夫々が特定され、ポーズは、位置と進路情報 (すなわち自律型ロボットが向いている向き) との組み合わせである。

20

【 0 0 3 6 】

まず、カメラ 3 によって連続的に取り込まれた一対の画像を取得する (ステップ 2 1、図 3 の部分 3 1)。次に、画像のサンプルセットから平均 RGB チャンネル値を減算することにより、各画像を前処理する (ステップ 2 2、図 3 の部分 3 2)。画像のセットは、例えば後に詳細に説明するように、自律型ロボット 1 が訓練される際の画像であってもよい。更に、画像を 6 4 の倍数になるようにサイズ変更する。但し、このような前処理はオプションであって、他の実施形態では存在しない。図 3 から分かるように、時間ステップ t の場合、最初の一対の画像は前処理された画像 $R G B t$ 及び $R G B t + 1$ になるが、時間ステップ $t + 1$ の場合、前処理された画像は $R G B t + 1$ 及び $R G B t + 2$ であり、他の時間ステップについても同じように続いていく。

30

【 0 0 3 7 】

プロセッサ 2 及びメモリ 4 によって実装される畳み込みニューラルネットワーク (CNN) は、前処理された一対の画像を取得し、それを使用して特徴を抽出する (ステップ 2 3、図 3 の部分 3 3)。CNN は、以下の説明のように実行されたそのトレーニングに従って、特徴を抽出する。

40

【 0 0 3 8 】

CNN は、ネットワーク構造に畳み込み層を組み込むニューラルネットワークの一種であるため、他の種類のニューラルネットワークで使用される全結合層とは対照的に、データの空間的規則性を活用することができる。つまり、CNN に必要なパラメータの数が大幅に削減され、高次元の入力 (生の画像データ等) を取り扱うことができるようになる。CNN では、各畳み込み層で複数の畳み込み演算が適用され、前の層の出力マップから多くの特徴が抽出される。例えば [3 8] で説明されているように、マップの畳み込みに使用されるフィルターカーネルは、トレーニング中に学習される。

【 0 0 3 9 】

CNN は、前処理された連続した一対の画像を積み重ねることによって生成されたテン

50

ソルを、入力として受け取る。CNNは、9つの畳み込み層で構成され、最後を除く夫々に正規化線形ユニット（ReLU）非線形活性化（non-linearity activation）が続くことで、合計で17層になる。それらの層は次のように構成される。

【表1】

層	受容野サイズ	パディング	ストライド	チャンネル数
Conv1	7x7	3	2	64
Conv2	5x5	2	2	128
Conv3	5x5	2	2	256
Conv3_1	3x3	1	1	256
Conv4	3x3	1	2	512
Conv4_1	3x3	1	1	512
Conv5	3x3	1	2	512
Conv5_1	3x3	1	1	512
Conv6	3x3	1	2	1024

10

20

【0040】

ネットワーク内の受容野のサイズは、7x7から5x5、そして3x3へと徐々に小さくなり、小さな興味深い特徴を捉えている。受容野の構成に適応するため、或いは、畳み込み後にテンソルの空間次元を保持するために、ゼロパディングが導入される。チャンネル数、すなわち、特徴検出用のフィルター数は、様々な特徴を学習するために増加する。

【0041】

本実施形態ではCNNが5500万の訓練可能な重みを有するが、他の実施形態では異なる数の重みを使用できることは、理解されるであろう。

30

そして、最終層（つまりConv6）からの特徴が、CNNの出力になる。

【0042】

同時に、第1のリカレントニューラルネットワーク（RNN）が慣性センサ5から動きデータを取得し、それを使用して特徴を抽出する（ステップ24、図3の部分34）。CNNと同様に、第1のRNNは、以下で説明するように実行されるそのトレーニングに従って、特徴を抽出する。

【0043】

RNNは、ニューラルネットワークの一種で、層は入力に対して作用するが、隠れ層及び/又は出力の遅延バージョンに対しても作用する。この方法において、RNNは、過去の入力及び対応する特定を追跡するための「メモリ」として使用できる、内部状態を有している。

40

【0044】

本実施形態では、図4に示すような長・短期記憶（LSTM）アーキテクチャを備えた複数のRNNが使用され（その様々なバリエーションが存在する）、図4において、丸中黒符号（の中にがある符号）は、要素単位の積を示し、丸囲み+符号（の中に+がある符号）は、2つのベクトルの加算を示している。メモリセルの内容は、 c_t に保存される。入力ゲート i_t は、現在の時間ステップで入力メモリの内容に入る方法を制御する。忘却ゲート f_t は、必要に応じてメモリセルをクリアする制御信号0~1を生成することにより、メモリセルを空にするタイミングを決定する。最後に、出力ゲート o_t は、メモリセルの内容を現在の時間ステップで使用するか否かを決定する。RNNの動作は、

50

以下の式で説明される。

【 0 0 4 5 】

【数 1】

$$i_t = \sigma(\mathbf{W}_{xi}x_t + \mathbf{W}_{hi}h_{t-1} + \mathbf{W}_{ci}c_{t-1} + b_i)$$

【数 2】

$$f_t = \sigma(\mathbf{W}_{xf}x_t + \mathbf{W}_{hf}h_{t-1} + \mathbf{W}_{cf}c_{t-1} + b_f)$$

10

【数 3】

$$z_t = \tanh(\mathbf{W}_{xc}x_t + \mathbf{W}_{hc}h_{t-1} + b_c)$$

【数 4】

$$c_t = f_t \odot c_{t-1} + i_t \odot z_t$$

【数 5】

$$o_t = \text{act}(\mathbf{W}_{xo}x_t + \mathbf{W}_{ho}h_{t-1} + \mathbf{W}_{co}c_t + b_o)$$

20

【数 6】

$$h_t = o_t \odot \tanh(c_t)$$

【 0 0 4 6 】

パラメータ $W_{i,j}$ 及び b_i は、RNNの動作を完全にパラメータ化し、トレーニング中に学習される。再帰的な隠れ層により、ネットワークは、入力データの時間的規則性を利用してパフォーマンスを向上させることができる。

30

【 0 0 4 7 】

本実施形態において、LSTMは、2000ユニットのセルを備えた2つの層を有するが、他の実施形態では、異なる数の層及びユニットを使用できることは、理解されるであろう。

【 0 0 4 8 】

次に、CNNと第1のRNNとによって抽出された特徴が第2のRNNに渡され、第2のRNNは、それらを使用して動きの特徴を抽出する(ステップ25、図3の部分35)。本実施形態において、第2のRNNは、前の時間ステップから引き継がれた隠れ状態を、同時に入力としても取得し、前の時間ステップについて特定されたポーズが、RNNへ直接供給される。これは図3で確認することができ、この図では、次の時間ステップのために、時間ステップのポーズがLSTMボックスに送られる。これを行う理由は、位置推定の場合、出力が基本的に各時間ステップでの連続する変位の積分であるためである。従って、前の時間ステップで特定されたポーズは、特に重要である。

40

【 0 0 4 9 】

次に、(高次元の)RNNによって抽出された動きの特徴は、全結合層に渡され(ステップ26)、それは低次元(少なくともポーズについて6、不確実性について6、更にポーズ及び不確実性の推定にガウス混合を使用する場合は各々についてより多いかもしれない)の特徴を出力する。

【 0 0 5 0 】

次に、全結合層からの低次元の特徴は、SE(3)層に渡される(ステップ27、図3

50

の部分 3 4 の $SE(3)$ ボックス)。 $SE(3)$ は、各時間ステップで自律型ロボット 1 の位置（実際には、時間ステップ t についての Pos_e_t 等のポーズ）を特定するために、各時間ステップで連続した動きの特徴を構成する。

【 0 0 5 1 】

$SE(3)$ は、その要素が変換行列である特別なユークリッド群であって、特別な直交群 $SO(3)$ からの回転と並進ベクトルとで構成される。

【数 7】

$$T = \left\{ \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \mid R \in SO(3), T \in \mathcal{R}^3 \right\}$$

10

【 0 0 5 2 】

$SO(3)$ コンポーネントが直交行列である必要があるため、 $SE(3)$ に属する変換推定値の生成は簡単ではない。しかしながら、 $SE(3)$ のリー代数 $se(3)$ は、直交性の制約を受けないコンポーネントによって記述することができる。

【数 8】

$$\frac{\xi}{dt} = \left\{ \begin{pmatrix} [\omega]_{\times} & v \\ 0 & 0 \end{pmatrix} \mid v \in so(3), v \in \mathcal{R}^3 \right\}$$

20

【 0 0 5 3 】

そして、 $se(3)$ と $SE(3)$ との間の変換は、指数マップを使用して実行できる。

【数 9】

$$\exp: se(3) \rightarrow SE(3)$$

別の実施形態では、行列表現の代わりに回転の四元数表現が使用される。特に、成分はベクトルに変換される。

【数 10】

$$w_t = [0, \omega_x, \omega_y, \omega_z]$$

30

【数 11】

$$\dot{q}_t = \frac{1}{2} w_t \times q_t$$

【数 12】

$$q_t = q_t + dt \dot{q}_t$$

40

【 0 0 5 4 】

これらの量の勾配の計算は、単純な線形代数演算のみを使用して実行することができる。更に、指数マップを計算するために必要な、高価な固有値の代償 (expensive eigenvalue decomposition) が回避される。

従って、このようにして自律型ロボット 1 は、カメラ 3 からの画像を使用して、その位置、特にそのポーズを推定する。

【 0 0 5 5 】

連続したセンサ測定値からの位置の推定 (すなわちオドメトリ) は、必然的にドリフトの影響を受ける。結果的に、それをループ閉じ込み、マップマッチング、又はポーズグラ

50

フ最適化手法と組み合わせて使用して、同時ローカリゼーション及びマッピング（SLAM）システムを作成するのが一般的である。オドメトリ測定をそのようなシステムに統合する重要な側面は、不確実性の推定値の可用性である。

【0056】

そのような推定値を提供するために、全結合層の出力が使用される（SE（3）層の前）。全結合層によって生成された推定値は、訓練データからのグラウンドトゥルス・ポーズ情報と比較され、ポーズ（位置及び向き）の誤差分布をもたらす。そして、不確実性を表すガウス分布の混合の予測をトレーニングするために、最尤法が利用される。

【0057】

動作させるためには、当然のことながら、ニューラルネットワークを訓練する必要があり、これは、テストデータと最小化される目的関数（cost function）とを提供することによって行われる。現在説明されている自律型ロボット1のCNNと第1及び第2のRNNとのトレーニングは、実際、その全てが同時に訓練される。

10

【0058】

上述したように、本実施形態のシステムは、ポーズと不確実性との双方を推定する。テストデータは、一連の画像であり、それらの画像の「グラウンドトゥルス」ポーズ（つまり正しいポーズ）が含まれる。トレーニングのための目的関数は、2つの部分で構成され、最初の部分がポーズの推定に関連し、2番目の部分が不確実性の推定に関連している。ポーズの推定については、推定されたポーズとグラウンドトゥルス・ポーズとの差を最小化するように、目的関数の最初の部分でシステムを訓練する。不確実性の推定については、ニューラルネットワークの出力をポーズラベルと比較することにより、目的関数の2番目の部分でシステムを訓練する。その後、時間を通して誤差逆伝播によるトレーニングを行い、CNNと第1及び第2のRNNとの重みを調整して、目的関数の結果を最適に最小化する。

20

【0059】

このように、CNN及び第1のRNNの双方は、第2のRNNへの入力に最も適した特徴を提供するように訓練され、同時に第2のRNNは、それらの特徴（及び以前の特定）から自律型ロボット1のポーズ（及びその不確実性）を最も正確に特定するように訓練されることが分かる。特にCNN及び第1のRNNは、特定の種類の特徴又は特定の性質を備えた特徴を最適に提供するには訓練されず、むしろ、システム全体の運用に最適な特徴を提供するように単純に訓練される。しかしながら、特定の実施形態では、初期の訓練プロセスを高速化するために、CNNが最初に単独で訓練され（或いはそのような訓練の効果をもつ重みが与えられ）、連続する画像間の動きを示す特徴を提供する。これにより、CNNの初期状態が提供され、システムが全体として訓練されるにつれて、更に最適な訓練が行われる。

30

【0060】

特に有利な実施形態では、ニューラルネットワークの教師なし訓練が、相関（correlation）を使用して実行される。特に、カメラ3からの画像と慣性センサ5からの動きのデータとの双方は、自律型ロボット1の同じ動きを示す必要があるため、CNNによって画像から抽出される特徴と、第1のRNNによって動きのデータから抽出される特徴との間の相関を最大化するように、訓練を行うことができる。換言すれば、画像が大きな動きを示唆している場合は動きのデータもそうあるべきであり（及びその逆も同様）、反対に画像が小さい動きを示唆している場合は動きのデータもそうあるべきである（及びその逆も同様）ため、一般的に、CNNと第1のRNNとからの特徴間には相関があるはずであり、従って、相関を最大化する訓練が有利である。

40

【0061】

特定の実施形態を参照しながら、本発明について説明及び図示してきたが、当業者は、本明細書に具体的に示されていない多くの異なる変形例に、本発明が適していることを理解されるであろう。

【0062】

50

前述の説明において、既知の、明白な又は予見可能な同等物を有する完全形 (integer) や要素が言及されている場合、そのような同等物は、個別に記載されたものとして本明細書に組み込まれる。本発明の真の範囲の判定には特許請求の範囲を参照すべきであり、そのような同等物を包含するように解釈されるべきである。又、読者は、好ましいもの、有利なもの、便利なもの、又はそれらの類のものとして説明されている、本発明の完全形や特徴が任意のものであり、独立請求項の範囲を限定するものではないことは、理解されるであろう。更に、そのような任意の完全形や特徴は、本発明の幾つかの実施形態では可能な利点があるが、他の実施形態では望ましくないことがあり、従って存在しない場合があることを理解されたい。

【符号の説明】

【 0 0 6 3 】

1 : 自律型ロボット、 2 : プロセッサ、 3 : カメラ、 4 : メモリ、 5 : 慣性センサ

10

20

30

40

50

【図面】

【図 1】

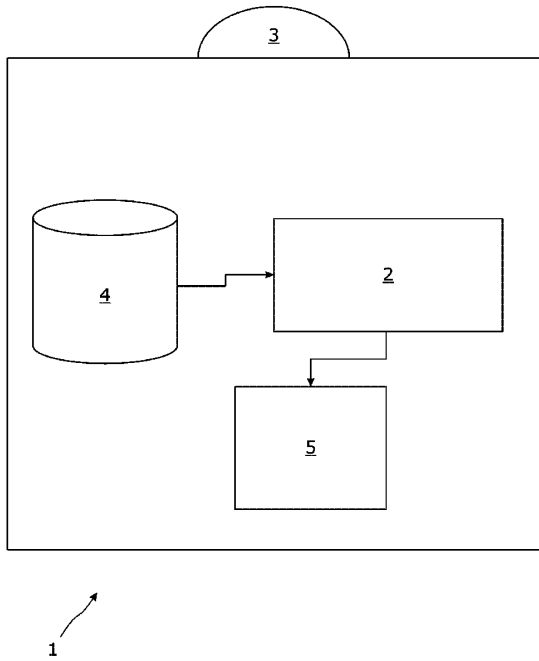
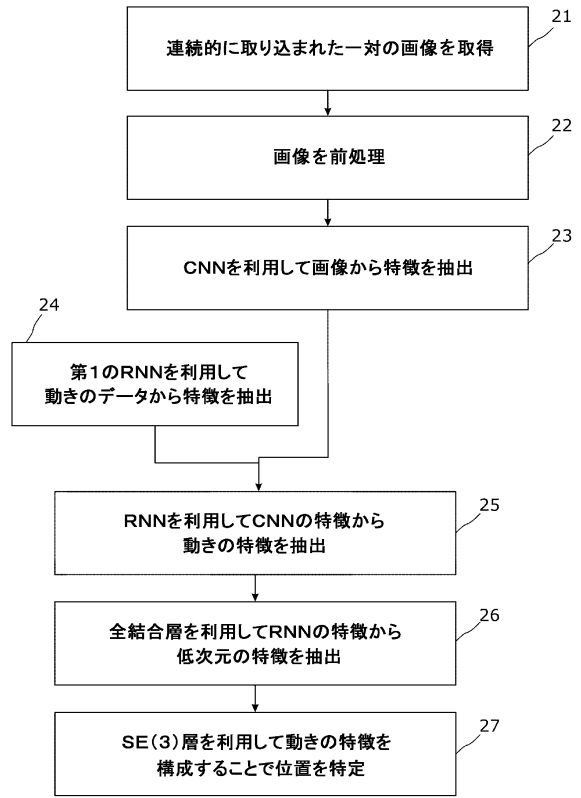


Fig. 1

【図 2】



10

20

【図 3】

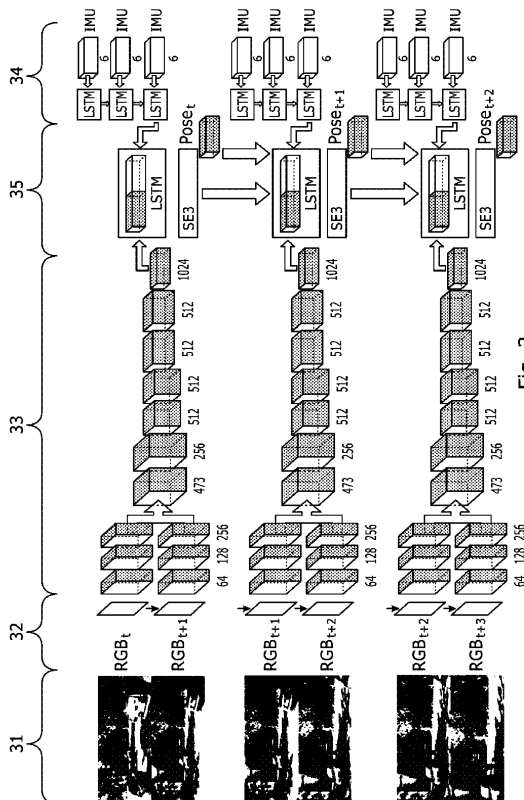
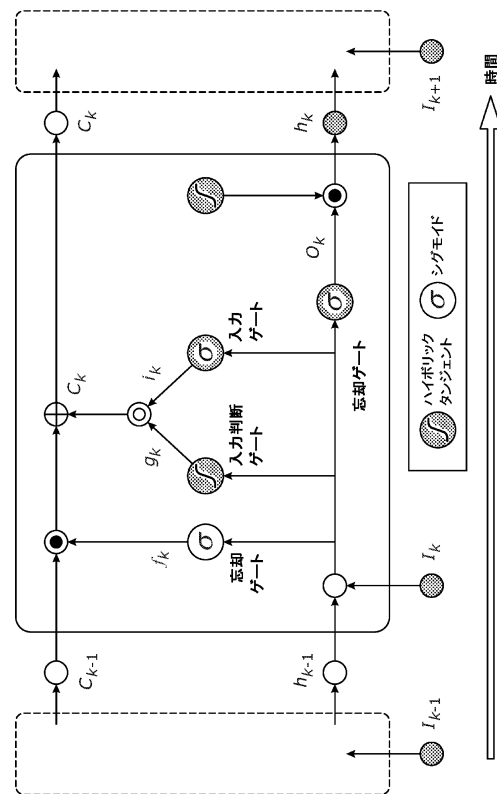


Fig. 3

【図 4】



30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

英国(GB)

ザン ディストリクト オブ シンホワリアン コミュニティ, ブルジ 20, ルーム 262

(72)発明者 トリゴニ, ニキ

イギリス, オックスフォード オーエックス3 7エスエー, 54 フランクリン アールディー

審査官 笠田 和宏

(56)参考文献 特開平10 - 170299 (JP, A)

特開平11 - 110542 (JP, A)

(58)調査した分野 (Int.Cl., DB名)

G06T 7/00

G06T 7/70

G06N 3/04