

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5352738号
(P5352738)

(45) 発行日 平成25年11月27日(2013.11.27)

(24) 登録日 平成25年8月30日(2013.8.30)

(51) Int.Cl. F I
G O 6 T 7/00 (2006.01) G O 6 T 7/00 3 5 0 B

請求項の数 20 (全 14 頁)

(21) 出願番号	特願2012-518603 (P2012-518603)	(73) 特許権者	000005326
(86) (22) 出願日	平成22年6月30日 (2010.6.30)		本田技研工業株式会社
(65) 公表番号	特表2012-532382 (P2012-532382A)		東京都港区南青山二丁目1番1号
(43) 公表日	平成24年12月13日 (2012.12.13)	(74) 代理人	100064414
(86) 国際出願番号	PCT/US2010/040678		弁理士 磯野 道造
(87) 国際公開番号	W02011/002938	(74) 代理人	100111545
(87) 国際公開日	平成23年1月6日 (2011.1.6)		弁理士 多田 悦夫
審査請求日	平成25年6月28日 (2013.6.28)	(72) 発明者	ハイゼレ, ベルント
(31) 優先権主張番号	61/222, 245		アメリカ合衆国、オハイオ州 43067
(32) 優先日	平成21年7月1日 (2009.7.1)		-9705、レイモンド、ステート ルー
(33) 優先権主張国	米国 (US)		ト 739 21001、ホンダ パテン
早期審査対象出願			ツ アンド テクノロジーズ ノース ア
			メリカ エルエルシー気付

最終頁に続く

(54) 【発明の名称】 3次元モデルを使用した物体認識

(57) 【特許請求の範囲】

【請求項 1】

コンピュータにより実行する、ビュー・ベースの物体認識分類器をトレーニングするための方法であって、前記方法は、

三次元モデルと、複数のレンダリングパラメータの値からなるセットと、に基づいて生成される複数の画像からなる、画像の初期トレーニングセットを生成するステップであって、前記複数のレンダリングパラメータが、視点の方位角位置、視点の高度位置、光学軸回りの視点の回転、の少なくとも一つを含むステップと、

前記初期トレーニングセットを使用して、分類器をトレーニングするステップと、

前記分類器の精度を判定するステップと、

前記分類器の出力の一つ以上のローカルミニマムのセットを判定するステップと、

前記ローカルミニマムのセットのそれぞれのローカルミニマムに関連付けられた複数のレンダリングパラメータの値からなるセットを決定するステップであって、前記ローカルミニマムに関連付けられた複数のレンダリングパラメータが、視点の方位角位置、視点の高度位置、光学軸回りの視点の回転、の少なくとも一つを含むステップと、

前記ローカルミニマムのセットのそれぞれのローカルミニマムについて、前記三次元モデルと前記決定された複数のレンダリングパラメータの値からなるセットに基づいて、追加の画像を生成するステップと、

前記初期トレーニングセットと前記追加で生成された画像を使用して、前記分類器をトレーニングするステップと、を含むことを特徴とする方法。

10

20

【請求項 2】

前記分類器は、統計的分類器を含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記分類器は、サポートベクターマシンの含むことを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記分類器は、画像の光譜調度のヒストグラムの特徴ベクトルに基づいて、前記画像を分類することを特徴とする請求項 1 に記載の方法。

【請求項 5】

前記レンダリングパラメータが、点光源の方位角位置、点光源の高度位置、周辺光と点光源間の光強度比を含むグループのうちの一つの要素を更に含むことを特徴とする請求項 1 に記載の方法。

10

【請求項 6】

前記初期トレーニングセット内の画像のためのレンダリングパラメータの値は、ランダム又は疑似ランダムに決定されることを特徴とする請求項 1 に記載の方法。

【請求項 7】

前記初期トレーニングセット内の画像は、前記画像が肯定例か否定例かを示すラベルに関連付けられていることを特徴とする請求項 1 に記載の方法。

【請求項 8】

前記分類器の出力の前記一つ以上のローカルミニマムのセットを決定するステップは、最適化アルゴリズムを実行するステップを含むことを特徴とする請求項 1 に記載の方法。

20

【請求項 9】

プロセッサによって実行された時、プロセッサにビュー・ベースの物体認識分類器をトレーニングするための方法を実行させる命令がコード化された、機械が読み取り可能な記憶媒体であって、前記方法は、

三次元モデルと、複数のレンダリングパラメータの値からなるセットと、に基づいて生成される複数の画像からなる、画像の初期トレーニングセットを生成するステップであって、前記複数のレンダリングパラメータが、視点の方位角位置、視点の高度位置、光学軸回りの視点の回転、の少なくとも一つを含むステップと、

前記初期トレーニングセットを使用して、分類器をトレーニングするステップと、

30

前記分類器の精度を判定するステップと、

前記分類器の出力の一つ以上のローカルミニマムのセットを判定するステップと、

前記ローカルミニマムのセットのそれぞれのローカルミニマムに関連付けられた複数のレンダリングパラメータの値からなるセットを決定するステップであって、前記ローカルミニマムに関連付けられた複数のレンダリングパラメータが、視点の方位角位置、視点の高度位置、光学軸回りの視点の回転、の少なくとも一つを含むステップと、

前記ローカルミニマムのセットのそれぞれのローカルミニマムについて、前記三次元モデルと前記決定された複数のレンダリングパラメータの値からなるセットに基づいて、追加の画像を生成するステップと、

前記初期トレーニングセットと前記追加で生成された画像を使用して、前記分類器をトレーニングするステップと、を含むことを特徴とする機械が読み取り可能な記憶媒体。

40

【請求項 10】

前記分類器は、統計的分類器を含むことを特徴とする請求項 9 に記載の記憶媒体。

【請求項 11】

前記分類器は、サポートベクターマシンの含むことを特徴とする請求項 9 に記載の記憶媒体。

【請求項 12】

前記分類器は、画像の光譜調度のヒストグラムの特徴ベクトルに基づいて、前記画像を分類することを特徴とする請求項 9 に記載の記憶媒体。

【請求項 13】

50

レンダリングパラメータが、点光源の方位角位置、点光源の高度位置、周辺光と点光源間の光強度比を含むグループのうちの一つの要素を更に含むことを特徴とする請求項 9 に記載の記憶媒体。

【請求項 14】

前記初期トレーニングセット内の画像のためのレンダリングパラメータの値は、ランダム又は疑似ランダムに決定されることを特徴とする請求項 9 に記載の記憶媒体。

【請求項 15】

前記初期トレーニングセット内の画像は、前記画像が肯定例か否定例かを示すラベルと関連付けられていることを特徴とする請求項 9 に記載の記憶媒体。

【請求項 16】

前記分類器の出力の前記一つ以上のローカルミニマムのセットを決定するステップは、最適化アルゴリズムを実行するステップを含むことを特徴とする請求項 9 に記載の記憶媒体。

【請求項 17】

三次元モデルと、複数のレンダリングパラメータの値からなるセットと、に基づいて生成される複数の画像からなる、画像の初期トレーニングセットを生成するステップであって、前記複数のレンダリングパラメータが、視点の方位角位置、視点の高度位置、光学軸回りの視点の回転、の少なくとも一つを含むステップと、

前記初期トレーニングセットを使用して、分類器をトレーニングするステップと、

前記分類器の精度を判定するステップと、

前記分類器の出力の一つ以上のローカルミニマムのセットを判定するステップと、

前記ローカルミニマムのセットのそれぞれのローカルミニマムに関連付けられた複数のレンダリングパラメータの値からなるセットを決定するステップであって、前記ローカルミニマムに関連付けられた複数のレンダリングパラメータが、視点の方位角位置、視点の高度位置、光学軸回りの視点の回転、の少なくとも一つを含むステップと、

前記ローカルミニマムのセットのそれぞれのローカルミニマムについて、前記三次元モデルと前記決定された複数のレンダリングパラメータの値からなるセットに基づいて、追加の画像を生成するステップと、

前記初期トレーニングセットと前記追加で生成された画像を使用して、前記分類器をトレーニングするステップと、を含む方法を実行する、機械が読み取り可能な命令がコード化された、機械が読み取り可能な記憶媒体と、

前記機械が読み取り可能な記憶媒体にコード化された、機械が読み取り可能な命令を実行するようになされたプロセッサとを含むことを特徴とする、ビュー・ベースの物体認識分類器をトレーニングするためのシステム。

【請求項 18】

レンダリングパラメータが、点光源の方位角位置、点光源の高度位置、周辺光と点光源間の光強度比を含むグループのうちの一つの要素を更に含むことを特徴とする請求項 17 に記載のシステム。

【請求項 19】

前記初期トレーニングセット内の画像のためのレンダリングパラメータの値は、ランダム又は疑似ランダムに決定されることを特徴とする請求項 17 に記載のシステム。

【請求項 20】

前記初期トレーニングセット内の画像は、前記画像が肯定例か否定例かを示すラベルに関連付けられていることを特徴とする請求項 17 に記載のシステム。

【発明の詳細な説明】

【関連出願】

【0001】

本願は、「3Dモデルを使用した物体認識」という発明の名称の米国特許仮出願番号 61/222,245 (出願日 2009 年 7 月 1 日) の優先権を主張するものであり、その記載の全てを本明細書に引用する。

10

20

30

40

50

【技術分野】

【0002】

本発明は、物体認識を含むマシンビジョンに関するものであり、より詳しくは、ビュー・ベースの物体認識を実行できるように分類器をトレーニングする技術に関するものである。

【背景技術】

【0003】

ここ数年、物体認識の分野では、各クラスごと数個の例から、何百もの物体のクラスを認識するように学習するシステムを構築するという課題に取り組んでいる。これらのシステムを評価するために使われる標準的なデータのセットは（物体検出における初期の取り組みで使われた、細かく分類された何千もの物体の画像のセットに対して）、平均で各クラスごとに二百個未満の画像を有するだけである。そのような数の少ないデータのセットに基づいて評価を行うことは、本質的に問題がある。そのテストの結果は、一般化できないし、誤った解釈を招く可能性もあるからである。

【0004】

手入力で注釈付けられた自然画像を記憶した、データ量の多いデータベースを作る取組もされてきた。しかしながら、画像が幾分恣意的に選択されることと、グランドトゥースがないことから、姿勢、スケール、位置、照明に対する不変性などの、物体認識システムの特定の性質を体系的に分析するのが困難であった。これらの問題を解決した、形状を基にした物体認識用のデータベースが、ニューヨーク大学クーラント研究所で開発されたNORBである。（NORBにおいては）、物体の画像は、視点と照明を考慮して撮像される。物体の回転、背景、「ディストラクタ」などを変更した、画像のバリエーションを増やすために、その画像は合成的に修正される。

【発明の概要】

【発明が解決しようとする課題】

【0005】

画像生成制御の考え方をもう一歩進めると、現実的な三次元（3D）コンピュータグラフィックモデルからレンダリングされた完全合成画像に至る。いくつかのビュー・ベースの物体認識システムは、合成画像によってトレーニングされ、評価されてきた。少なくとも1つの顔認識システムと1つの物体認識システムは、3Dモデルの画像でトレーニングされ、実際の画像でテストされてきた。3Dモデルはまた、物体認識への生成的アプローチにおいても使用されてきた。その生成的アプローチにおいては、合成画像が所与の写真画像に最も合うように、レンダリングパラメータの値が最適化される。（しかし）、この合成解析のアプローチにおいては、ローカルミナに陥ることを避けるために、レンダリングパラメータの値の初期の推定値の精度が高くないため、多くの物体認識、検出タスクにとって不適當である。

【課題を解決するための手段】

【0006】

（本発明の）ビュー・ベースの物体認識システムは、物体の外見のモデルを使用して、所与の画像にその物体があるかどうかを判定する。ビュー・ベースの物体認識システムは、トレーニングデータのセットを使用してシステムをトレーニングするとき、そのモデルを生成（学習）する。トレーニングデータのセットは、肯定例の画像（つまり、ターゲットとなる物体が存在している画像例）と否定例の画像（つまり、ターゲットとなる物体が存在していない画像例）を含んでいる。それぞれのトレーニング画像は、その画像が肯定例の画像か否定例の画像かについて、正しくラベル付けされている。

【0007】

他のあらゆる画像と同じように、トレーニング画像は、自然画像か合成画像のどちらかである。自然画像は現実世界の物体の外見を忠実に表わすものであり、例えば、カメラや他の画像センサを使用してその物体の写真を撮ることにより生成される。合成画像は、自然画像以外のあらゆる画像である。例えば、合成画像は、コンピュータ支援設計（CAD）

10

20

30

40

50

モデルをレンダリングすることにより生成される。合成画像を使用して、ビュー・ベースの物体認識システムをトレーニングすることは（自然画像を使用してトレーニングすることに対して）、いくつかの利点がある。

【 0 0 0 8 】

本明細書では、反復トレーニングプロセスの中で、合成トレーニング画像の生成が制御される、「能動学習」という方法を提示する（「能動学習」という用語は、本明細書では、自身のトレーニングデータを能動的に生成する学習方法という意味で使用されている）。能動学習の第一義的な考え方は、低次元レンダリング空間において、重要で有益な物体の画像を見つけ、その物体の画像をトレーニングセットに反復的に加えることである。それぞれの反復処理においては、（所与の画像のデータベースからある特定のサンプル画像を加えることにより）分類器を「ブートストラップ」する代わりに、新しい画像が生成される。（画像数の）少ないトレーニングセットが、レンダリングパラメータの低次元空間における、分類器の出力のローカルミニマを探ることにより、反復的に生成される。

10

【 0 0 0 9 】

一実施形態においては、能動学習方法は、ビュー・ベースの物体認識のための小型分類器をトレーニングする。初期トレーニングセットが生成され、そのトレーニングセットを使用して、分類器をトレーニングする。低次元レンダリング空間における分類器の出力のローカルミニマが発見され、画像はそのローカルミニマでレンダリングされる。新しくレンダリングされた画像は、トレーニングセットに付け加えられる。修正されたトレーニングセットを使用して、分類器が再度トレーニングされるように、この処理が繰り返される。

20

【 0 0 1 0 】

本明細書で記載された特徴及び利点は、包括的なものではない。特に、当業者は、図面、明細書、請求の範囲を参照することにより、多くの付加的な特徴と利点を理解するだろう。

【図面の簡単な説明】

【 0 0 1 1 】

【図 1】図 1 は、本発明の一実施形態による、ビュー・ベースの物体認識のための小型分類器をトレーニングする能動学習方法のフローチャートを示す図である。

【図 2】図 2 は、本発明の一実施形態による 5 つのコンピュータグラフィックのモデルを示す図である。

30

【図 3】図 3 は、図 2 に示された本発明の一実施形態による 5 つのモデルのプリントアウトの写真を示す図である。

【図 4】図 4 は、図 1 に示した本発明の一実施形態による能動学習方法を実行するためのシステムのブロック図である。

【図 5】図 5 は、本発明の一実施形態による、スケーリングとスムージングが行われた後のレンダリングされた画像の例を示す図である。

【図 6】図 6 は、本発明の一実施形態による、初期に選択された画像と近傍のローカルミニマにおける画像を示す図である。

【 0 0 1 2 】

40

図面は、説明を目的として、本発明の様々な実施形態を示している。当業者は、以降の説明から、本明細書に記載された本発明の精神から逸脱することなく、本明細書に記載された構造及び方法に代わる実施形態を用いることができることを容易に理解するであろう。

【発明を実施するための形態】

【 0 0 1 3 】

以下に、添付の図面を参照して様々な実施形態を説明する。図面中、同じ参照番号を付したものはまったく同一もしくは機能的に類似した構成要素を表している。また、各参照番号の最左端の数字は、その参照番号が初めて使われた図面に対応している。

【 0 0 1 4 】

50

明細書中の「一実施形態」もしくは「実施形態」という言葉は、実施形態と関連して説明されるある特定の機能、構造、特徴が本発明の実施形態の少なくとも一つに含まれる、という意味であり、常に同一の実施形態を表すものではない。

【 0 0 1 5 】

以下の詳細な説明には、コンピュータ・メモリ内のデータビット操作のアルゴリズム及び記号表現によって表された部分がある。このようなアルゴリズム記述及び表現は、データ処理技術分野において、当業者が研究内容を他の当業者に最も効率的に伝えるために用いる手段である。あるアルゴリズムがあり、それが概して期待した結果に至る筋の通ったステップの系列だと理解されたとする。それらのステップは、物理量の物理的处理を要するステップである。必ずという訳ではないが、これらの物理量は記憶、転送、結合、比較、その他処理が可能な電氣的または磁氣的信号の形をとるのが普通である。これらの信号をビット、値、要素、記号、文字、術語、番号などで表わすのが、主に慣用上の理由から便利な場合があることが分かっている。さらに、一般性を損なうわけではないが、物理量もしくは物理量を表したものの、物理的处理や変換を要するステップのある特定の配列を、モジュールもしくはコードデバイスなどと表すのが便利な場合があることが分かっている。

10

【 0 0 1 6 】

しかしながら、このような術語や同様の用語はすべて適切な物理量に関連付けられるべきであり、また、それら物理量に付けた便宜上のラベルに過ぎないということに留意すべきである。以下の説明より明らかなように、特に断わらない限り、“処理”、“演算”、“計算”、“判定”、“表示”などの術語によって論じられることは、コンピュータシステムのレジスタ及びメモリの内部の物理的（電子的）な量として表現されたデータを処理して、コンピュータシステムのメモリやレジスタ、その他同様の情報記憶装置、情報伝送装置又は表示装置の内部の同様に物理量として表現された他のデータへ変換する、コンピュータシステムや（特定の計算機などの）同様の電子的演算装置の作用及びプロセスを意味する。

20

【 0 0 1 7 】

本発明のある側面は、アルゴリズムの形で記述されたプロセス・ステップ及び命令を含んでいる。本発明におけるそれらのプロセス・ステップ及び命令は、ソフトウェア、ファームウェア、もしくはハードウェアのいずれでも実現することが可能であり、ソフトウェアによって実現する場合、ダウンロードして様々なオペレーティングシステム上の各種プラットフォームに駐在させて動作させることも出来る。本発明は、コンピュータシステムで実行可能なコンピュータプログラム内にあってもよい。

30

【 0 0 1 8 】

本発明は、本明細書に述べた処理を実行するための装置にも関係するものである。このような装置は、所要目的（たとえば、特定のコンピュータ）のために専用に作られてもよいし、汎用コンピュータを内蔵のコンピュータ・プログラムによって選択駆動もしくは再構成したものでもよい。そのようなコンピュータ・プログラムは、コンピュータが読み取り可能な記憶媒体、限定するわけではないが例えば、フロッピー（登録商標）ディスク、光ディスク、CD-ROM、光磁気ディスクなどの任意の種類のディスク、リードオンリーメモリ（ROM）やランダムアクセスメモリ（RAM）、EPROM、EEPROM、磁気カード又は光カードなど、コンピュータのシステムバスに接続された電子的命令の記憶に適した任意種類の媒体に格納することができる。メモリは、情報、データまたはプログラムを記憶可能な、上述した機器もしくはその他の機器でありうる。さらに、明細書中でコンピュータと呼ばれるものは、1つのプロセッサのみを含むものでもよいし、演算能力の向上のために設計されたマルチプロセッサ構成を実装したものでもよい。

40

【 0 0 1 9 】

本明細書で提示したアルゴリズム及び表示は、本質的に、いかなる特定のコンピュータ、その他の装置とも関わりがない。様々な汎用マシンを、本明細書に述べる内容に従ったプログラムで使用し得るが、所要の手順のステップの実行のために、より特化した装置を

50

作るのが好都合であるかもしれない。そのような多様なシステムに必要とされる構造は以下の説明から明らかになる。さらに、どのような特定のプログラミング言語とも関連付けることなく本発明を説明する。本明細書に述べる本発明の内容を実現するために様々なプログラミング言語を使用し得ることを理解されよう。本明細書における、特定のプログラミング言語に関する言及は、本発明の実施可能性及び最良の実施形態を開示するためである。

【0020】

さらに、本明細書における用語は原則的に可読性及び説明の便宜上選択されたものであり、発明の主題を特定、制限するために選択されたものではない。したがって、本発明の開示は説明のためのものであり、請求の範囲に記載された発明の範囲を制限するものではない。

10

【0021】

本明細書では、本発明の特定の実施形態と適用例を説明するが、本発明は、明細書に記載される構成や構成要素と正確に一致するものに限定されるわけではなく、添付の請求の範囲に定義される、本発明の精神と範囲から逸脱することなく、本発明の方法及び装置の配置、動作および詳細部について様々な修正や変更を行うことができることを理解されたい。

【0022】

ビュー・ベースの物体認識システムは、所与の画像の中に物体があるかどうかを判定するために、物体の外観のモデルを使用する。そのモデルは、画像中のその物体の、様々な条件下（例えば、様々な姿勢や照明）における外観を表すものである。ビュー・ベースの物体認識システムは、トレーニングデータセットを使用してトレーニングされるときに、そのモデルを生成（学習）する。そのトレーニングデータのセットは、肯定例である画像（つまり、対象物体がある）と否定例である（つまり、対象物体がない）画像とを含んでいる。それぞれのトレーニング画像は、自身が肯定例か否定例かということについて、正しくラベル付けされている。

20

【0023】

他の画像と同様に、トレーニング画像は、自然画像であってもよいし合成画像であってもよい。自然画像は、現実世界の物体の外観を忠実に表現したものであり、例えば、カメラや画像センサを使用して物体の写真を撮ることにより生成されるものである。合成画像は、自然画像以外のあらゆる画像である。合成画像の一つのタイプは、自然画像に基づいて生成される。例えば、何らかの方法で自然画像を歪めたり、変形させたりして、合成画像を生成することが出来る。他のタイプの合成画像は、自然画像とは無関係に生成されるものである。例えば、コンピュータ支援設計（CAD）モデルをレンダリングして、合成画像を生成することができる。これらの2つのタイプの合成画像を区別するために、第1のタイプの合成画像は、（自然画像に基づいて生成されているため）、部分合成画像と称し、第2のタイプの合成画像は（自然画像とは無関係に生成されているため）、完全合成画像と称する。

30

【0024】

合成画像を使用してビュー・ベースの物体認識システムをトレーニングすることは（自然画像を使用することに比べて）、いくつかの利点がある。まず始めに、多数のトレーニング画像を自動で生成することが可能であるということである。次に、内部または外部カメラのパラメータ、照明、景色の構成、景色の変化などを含む、画像生成パラメータを完全に制御できるということが挙げられる。最後に、それぞれの物体の位置、スケール、方向に関して、グラウンドトゥースが既知であるということである。映像のシーケンスにおいては、フレームレート、カメラの動き、物体の動きが既知である。

40

【0025】

ビュー・ベースの物体認識の課題を解決するために、統計的分類器などの、教師付き機械学習方法が使用されてきた。完全合成トレーニング画像（例えば、三次元（3D）モデルに基づいて生成された画像）を使用して、ビュー・ベースの分類器をトレーニングする

50

場合を考えてみる。そのトレーニングは、2つの課題に遭遇する可能性がある。まず第一に、トレーニングセットが大きくなってしまい、それにより、学習アルゴリズムが中断されてしまう可能性があることが挙げられる。第二に、解が少なくないため、ランタイムに分類器が遅くなるということが挙げられる。現代の機械学習アルゴリズムの最も大きなグループの一つは、カーネルベースの機械学習アルゴリズムのグループであり、カーネルベースの機械学習アルゴリズムは、サポートベクターマシーンなどを含んでいる。 $N \times N$ のサイズの二次行列 (N は、カーネルベースの機械学習アルゴリズムを「中断」する可能性のあるサンプルの数である) を算出することが出来る (O. Chapelle 著、「Training a Support Vector Machine in the Primal」、Neural Computation、第19巻5号、1155-1178 ページ、2007 三月発行参照)。カーネルアルゴリズムにおける計算時間は、サンプルの数の多項式であるため、多大な数のサンプルは、トレーニングを遅くしてしまう。ランタイムにおける計算時間は、サポートベクターマシーン (及び他のカーネルベースのアルゴリズム) によってもまた影響を受ける。分類結果を算出する際に使用するトレーニングデータポイントの数は、トレーニングセットの数に対して線形的に増加する。基本的に、分類時におけるランタイムの計算時間は、トレーニングサンプルのサイズに対して線形的である。必要とされているのは、わずかなデータのセットから、小型分類器をトレーニング (例えば、構築または算出) することが可能な技術である。

【0026】

ここで、「能動学習」方法について説明する。「能動学習」においては、反復的なトレーニングプロセスにおいて、合成トレーニング画像の生成が制御される (機械学習の分野において、「能動学習」という言葉は隠されたラベルのついたデータからの学習という意味で使用されることに注意されたい。本明細書では、「能動学習」という用語は、それとは違う意味で使われており、自身のトレーニングデータを自発的に生成する学習方法という意味で使われている)。能動学習の第一義的な考え方は、低次元レンダリング空間における重要で有益な物体の画像を発見し、それらの物体の画像を反復的にトレーニングセットに加えることである。それぞれの反復処理においては (所与の画像のデータベースから、特定のサンプル画像を加えることによって)、分類器を「ブートストラップ」する代わりに、新しい画像が生成される。(画像の数が) 少ないトレーニングセットは、レンダリングパラメータの低次元空間における分類器の出力のローカルミニマを探すことにより反復的に生成される。

【0027】

図1は、本発明の一実施形態による、ビュー・ベースの物体認識のための小型分類器をトレーニングする能動学習方法のフローチャートを示している。ステップ110では、初期トレーニングセットが生成される。ステップ120では、トレーニングセットを使用して、分類器がトレーニングされる。ステップ130では、低次元レンダリング空間における、分類器の出力のローカルミニマが発見される。ステップ140では、画像がローカルミニマでレンダリングされる。ステップ150では、新しくレンダリングされた画像が、そのトレーニングセットに加えられる。ステップ160では、修正されたトレーニングセットを使用して分類器を再度トレーニングするように、ステップ120から上記の処理が繰り返される。

【0028】

分類タスクと図1に示した能動学習方法を実行するためのシステムの一例を使用して、ステップ110から160について以下に詳述する。この例においては、物体認識は5つのクラスの物体に対して実行される。それぞれの物体は、30,000の表面三角形を有する質感のない3Dモデルによって表されている。図2は、本発明の一実施形態による5つのコンピュータグラフィックモデルを示している。図2において、この例示されたモデルは、(左から右へ) 女神アフロディーテの胸部、横たわっている熊、立位の象、立位の馬、靴のソールを表している。この5つのモデルは、自然画像で物体認識システムをテストすることができるように、それぞれ3Dプリンタでプリントアウトされる。図3は、本発明の一実施形態による、5つのプリントアウトされたモデルの写真を示している。

【 0 0 2 9 】

この例において、（合成画像、自然画像ともに）画像を生成するために、（仮想であろうが、実際であろうが）カメラの視点をそのモデルの中心に向かわせながら、カメラの視点をそのモデルの周りの球面上で移動させる。そのモデルは、（そのモデルの球面上に位置させることができる）点光源と周辺光によって照らされる。この画像生成の設定により、低次元のレンダリング空間（具体的には、六次元レンダリング空間）が生成される。6つのレンダリングパラメータは、1）方位角における視点の位置、2）高度における視点の高さ位置、3）光学軸回りの視点の回転、4）方位角における点光源の位置、5）高度における点光源の高さ位置、6）周辺光と点光源の光強度の比である。

【 0 0 3 0 】

図4は、図1に示す能動学習方法を実行するための、本発明の一実施形態によるシステムのブロック図である。システム400は、ビュー・ベースの物体認識のための小型分類器をトレーニングすることが出来る。図示されたシステム400は、能動学習モジュール405、レンダラー410、分類器トレーナー415、分類器テスター420、記憶部425及びオペレーティングシステムツール430を備える。

【 0 0 3 1 】

一実施形態においては、能動学習モジュール405（およびそのコンポーネントモジュール）、レンダラー410、分類器トレーナー415、分類器テスター420、オペレーティングシステムツール430は、一つ以上のコンピュータが読み取り可能な記憶媒体に記憶され、一つ以上のプロセッサで実行される、一つ以上のコンピュータプログラムモジュールである。記憶部425（及びその内容）は、一つ以上のコンピュータが読み取り可能な記憶媒体に格納されている。加えて、能動学習モジュール405（及びそのコンポーネントモジュール）、レンダラー410、分類器トレーナー415、分類器テスター420、オペレーティングシステムツール430及び記憶部425は、少なくともデータを互いに送受可能な程度に、互いに通信可能に接続されている。一実施形態においては、オペレーティングシステムツール430は、パーソナルコンピュータ上のオペレーティングシステムの一つ以上のパーツとして実行され、能動学習モジュール405、レンダラー410、分類器トレーナー415、分類器テスター420は、同じパーソナルコンピュータで実行される。

【 0 0 3 2 】

図示された記憶部425は、分類器445、トレーニングセット450、一つ以上の3Dモデル460を記憶している。分類器445は、サポートベクターマシン（SVM）などの、統計的分類器または最近近傍分類器である。一実施形態においては、分類器445は、ガウスクERNELを有するサポートベクターマシンである。サポートベクターマシンパラメータの値は、初期実験により最適化され、その後固定したまま維持される（例えば、 $s = 2:0$ 、 $C = 10$ 、ただし、 s はカーネル幅を表し、 C は、エラーコストを表している）。サポートベクターマシンについては、V. Vapnik著、「Statistical Learning Theory」、Wiley-Interscience、1998年発刊に説明がある。サポートベクターマシンの実装には、例えば、LIBサポートベクターマシンライブラリなどがある（C-C. Chang et al著、「LIB Support Vector Machine: a Library for Support Vector Machines」、2001発刊参照）。

【 0 0 3 3 】

分類器445は、画像特徴値のベクトル「特徴ベクトル」を使用して、動作する。濃淡値、正規化濃淡値、もしくは光の諧調度のヒストグラムなどの任意の画像特徴を使用することが出来る。一実施形態においては、特徴ベクトルとして、方向ヒストグラムや、光諧調度のヒストグラムが使用される。それらのヒストグラムは、 23×23 画像の中の5つの固定位置で算出され、ユニット長さに正規化された640次元特徴ベクトルとなる。128次元ヒストグラムは、 $(9/9)$ 、 $(15/9)$ 、 $(12/12)$ 、 $(9/15)$ 、 $(15/15)$ の (x/y) 位置にて算出される。

【 0 0 3 4 】

分類器 445 は、分類器トレーナー 415 を使用して、トレーニングされ、分類器テスター 420 を使用して、テストされる。

【0035】

トレーニングセット 450 は、分類器トレーナー 415 を使用して分類器 445 をトレーニングするのに使う画像の集合である。トレーニングセット 450 における画像は、レンダラー 410 を使用して 3D モデル 460 をレンダリングすることによって生成された合成画像である。

【0036】

3D モデル 460 は、三次元物体のモデルであり、レンダラー 410 を使用してレンダリングして、合成画像を生成することが出来る。

10

【0037】

オペレーティングシステムツール 430 は、乱数生成器 455 を備えている。乱数生成器 455 は乱数（もしくは擬似乱数）を生成することが可能である。

【0038】

レンダラー 410 は、3D モデル 460 をレンダリングすることが出来る、Blender（3D モデリングおよびレンダリングのためのオープンソース・ソフトウェアパッケージ）などの、従来のソフトウェアアプリケーションである。レンダラー 410 は、レンダリングパラメータの値のセットに基づいて、モデル 460 をレンダリングする。上述した例においては、6 つのレンダリングパラメータ（方位角及び高度における視点の位置、光軸回りの視点の回転、方位角及び高度における点光源の位置、周辺光と点光源の間の光強度の比）がある。

20

【0039】

一実施形態においては、レンダラー 410 は、 100×100 ピクセルの解像度で 3D モデル 460 をレンダリングする。レンダラー 410 は、また、レンダリングされた（合成）画像をスケーリングおよびスムージングすることにより、 23×23 ピクセル濃淡（合成）画像を生成する。図 5 は、本発明の一実施形態による、スケーリングとスムージングが施された後のレンダリングされた（合成）画像の例を示している。

【0040】

分類器トレーナー 415 は、トレーニングセット 450（例えば、トレーニング画像のセット）が与えられれば、分類器 445 をトレーニングすることが出来る、従来からあるソフトウェアアプリケーションである。

30

【0041】

分類器テスター 420 は、分類器 445 をテストして、分類器の分類精度を判定することが出来る、従来のソフトウェアアプリケーションである。一実施形態において、分類器 445 は、各クラスごとに、レンダリングパラメータの低次元空間からランダムに描かれた $40,000$ の画像を使用してテストされる。分類器の出力は、サポートベクターマシンの実数値の出力を使用することにより算出される（ネガティブクラス（ラベル「-1」）のサンプルに関しては、出力値は -1 を乗算される）。（そして）、各クラスにおける最も難しい画像（例えば、最低精度を有する 100 個の画像）が判定される。

【0042】

40

能動学習モジュール 405 は、コントロールモジュール 435 及びローカルミニマファインダ 440 を備えている。コントロールモジュール 435 は、能動学習モジュール 405 の動作をコントロールして、能動学習モジュール 405 にビュー・ベースの物体認識のための小型分類器 445 をトレーニングさせる。図 1 を参照して、コントロールモジュール 435 について詳述する。

【0043】

ローカルミニマファインダ 440 は、低次元レンダリング空間における分類器の出力のローカルミニマ（及びそれらに対応するレンダリングパラメータの値）を発見する。1 つのセットの画像（例、各クラスから最も難しい 100 個の画像）が、最適化アルゴリズム（例、ネルダメッドシンプレックスアルゴリズム）の初期点として使われる。6 つのレン

50

ダリングパラメータの値は、全ての画像に関して既知であることに注意されたい。レンダリング空間における分類器の出力のローカルミニマ（及びそれらに対応するレンダリングパラメータの値）を発見するために、アルゴリズムが幾度も（例えば、10回）反復して実行される。分類器の出力は、サポートベクターマシンの実数値の出力（ネガティブクラス（ラベル「-1」）のサンプルについては、出力は-1が乗算される）を使用して算出される。図6は、本発明の一実施形態による、初期に選択された（つまり、初期点の）画像と近傍ローカルミニマにおける画像を示している。図6においては、初期画像が上側に、近傍ローカルミニマにおける画像が下側になるように、各ペアの例を垂直に配置している。

【0044】

一般的に、能動学習は、対象の画像をランダムに選択してトレーニングする場合に比べて、大幅に少ないトレーニングセットとサポートベクトルしか使わないにも関わらずに、エラー率は、対象の画像をランダムに選択してトレーニングする場合と同じになる（例、同一エラー率）。言い換えると、能動学習は、「小型」分類器をトレーニングすると言える。

【0045】

図1に戻り、ステップ110では初期トレーニングセット450が生成される。初期トレーニングセット450は、一つ以上の合成画像を有している。レンダラー410は、3Dモデル460とレンダリングパラメータの値のセットに基づいて、それぞれの合成画像を生成する。一実施形態において、初期トレーニングセット450は、レンダリングパラメータの低次元空間からランダムに選ばれたサンプルによって構成されている。

【0046】

一実施形態においては、コントロールモジュール435がステップ110を実行する。例えば、コントロールモジュール435は、乱数生成器455を使用して、レンダリングパラメータのそれぞれについて、ランダムに選ばれた値を得る（上述した例においては、6つのレンダリングパラメータがあるため、ランダムに選ばれたサンプルの一つ一つは、これらの6つのパラメータについて、それぞれ一つづつランダムに選択された値を有する）。コントロールモジュール435は、次に、レンダリングパラメータの値に基づいて、レンダラー410を使用して、3Dモデル460の合成画像を生成する。そして、その合成画像が、初期トレーニングセット450に加えられる。一実施形態において、初期トレーニングセット450は、各クラスについて200個のサンプル（例えば、各3Dモデル460について、200個の合成画像）を有している。

【0047】

ステップ120において、トレーニングセット450を使用して、分類器445をトレーニングする。一実施形態においては、コントロールモジュール435は、ステップ120を実行する。例えば、コントロールモジュール435は、分類器トレーナー415を使って、トレーニングセット450を使用して分類器445をトレーニングする。ステップ120が最初に実行される時、トレーニングセット450は初期トレーニングセットである。ステップ120が二回目以降に実行される時は、ステップ150と合わせて以下に説明するように、トレーニングセット450は修正されているはずである。

【0048】

ステップ130では、低次元レンダリング空間における分類器の出力のローカルミニマが発見される。一実施形態においては、コントロールモジュール435は、ステップ130を実行する。例えば、コントロールモジュール435は、まず、分類器テスター420を使用して、分類器445をテストし、分類器の分類精度を決定する。分類器テスター420は、各クラスから最も難しい画像（例えば、最低精度を有する100個の画像）を決定する。コントロールモジュール435は、その次に、ローカルミニマファインダ440を使用して、その最も難しい画像を初期点として、分類器の出力のローカルミニマ（及びそれらに対応するレンダリングパラメータの値）を見つける。

【0049】

ステップ 140 において、画像が、ローカルミニマにおいて（それらに対応するレンダリングパラメータの値を使用して）描画される。一実施形態において、コントロールモジュール 435 は、ステップ 140 を実行する。例えば、コントロールモジュール 435 は、レンダラー 410 を使用して、ステップ 130 で発見されたローカルミニマで、物体の画像（合成画像）をレンダリングする。一実施形態においては、200 個の物体の画像がレンダリングされる。コントロールモジュール 435 はまた、レンダリングされた画像の方向ヒストグラムを算出する。方向ヒストグラムは、物体の特徴（つまり、分類の基礎）として使われる。

【0050】

ステップ 150 では、新しくレンダリングされた画像が、トレーニングセットに加えられる。一実施形態においては、コントロールモジュール 435 は、ステップ 150 を実行する。例えば、コントロールモジュール 435 は、（ステップ 140 にて生成された）新しくレンダリングされた画像を適切に記憶することによって、新しくレンダリングされた画像をトレーニングセット 450 に追加する。これにより、ステップ 140 でレンダリングされた画像と、ステップ 120 で分類器をトレーニングする際に使われた画像の両方を含む修正トレーニングセット 450 が生成される。

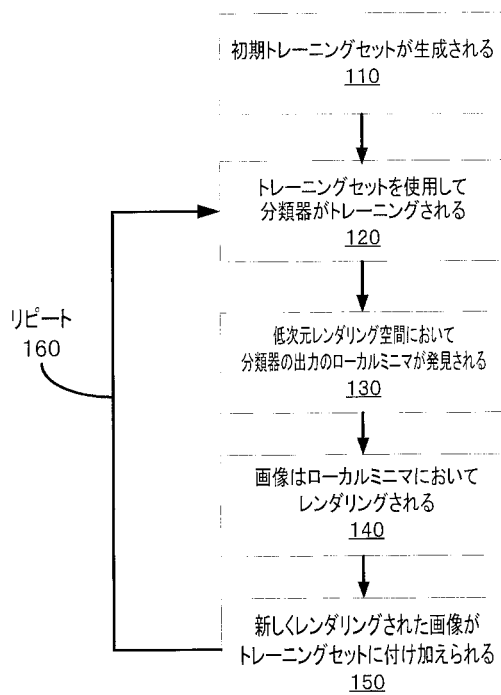
【0051】

ステップ 160 では、ステップ 120 から処理が繰り返され、修正されたトレーニングセット 450 を使用して分類器 445 がトレーニングされる。

【0052】

本発明は、ある実施形態を参照して、詳細に説明してきたが、当業者が理解するように、他の実施形態も可能である。例えば、他の実施形態が、「Object recognition with 3 Dmodels」 B. Heisele, G. Kim, A. Meyer 著、Proceedings of the 2009 British Machine Vision Conference (BMVC), London, England, September 7 - 10 に記載されており、その記載を本明細書に引用する。

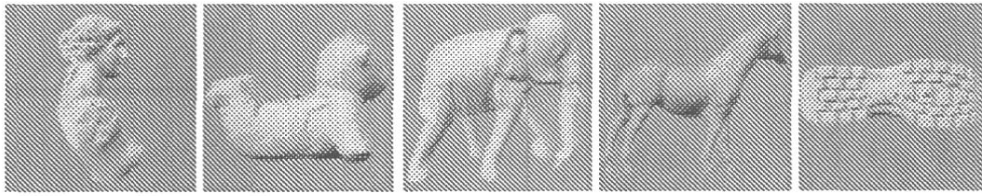
【図 1】



【図 4】



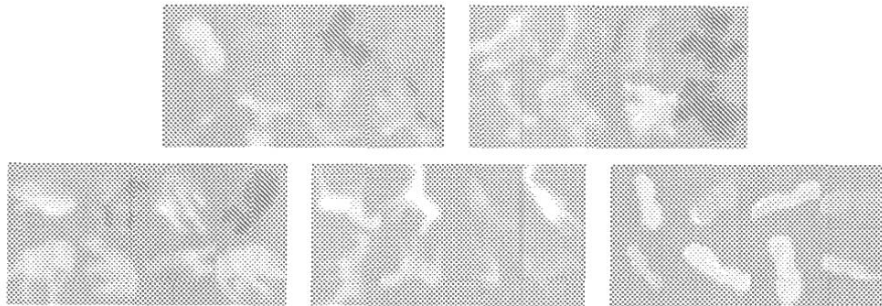
【図 2】



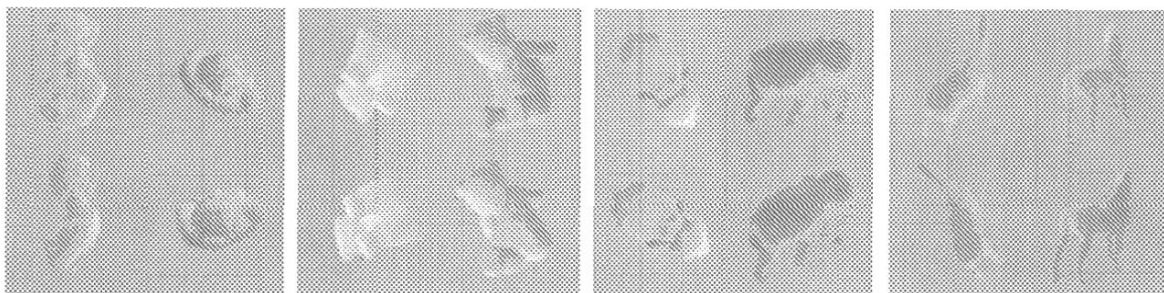
【図 3】



【図 5】



【図 6】



フロントページの続き

(72)発明者 キム, ガンヒ

アメリカ合衆国、オハイオ州 43067-9705、レイモンド、ステート ルート 739
21001、ホンダ パテント アンド テクノロジーズ ノース アメリカ エルエルシー気付

(72)発明者 メイアー, アンドリュー, ジェイ

アメリカ合衆国、オハイオ州 43067-9705、レイモンド、ステート ルート 739
21001、ホンダ パテント アンド テクノロジーズ ノース アメリカ エルエルシー気付

審査官 ▲広▼島 明芳

(56)参考文献 特開2009-93611(JP,A)

特表2010-522469(JP,A)

B.Weyrauch 外3名, Component-based Face Recognition with 3D Morphable Models, Computer
Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on, 米国, 2004
年

Heinrich H. Bulthoff 外2名, View-based dynamic object recognition based on human perc
eption, Pattern Recognition, 2002. Proceedings. 16th International Conference on, 米国,
2002年, p.768-776

(58)調査した分野(Int.Cl., DB名)

G06T 7/00