

(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2018-185655
(P2018-185655A)

(43) 公開日 平成30年11月22日(2018.11.22)

(51) Int. Cl.	F 1	テーマコード (参考)
G06T 7/00 (2017.01)	G06T 7/00 350B	5B057
G06T 7/70 (2017.01)	G06T 7/70 B	5L096
G06T 1/00 (2006.01)	G06T 1/00 315	

審査請求 未請求 請求項の数 14 O L (全 20 頁)

(21) 出願番号 特願2017-86992(P2017-86992)
(22) 出願日 平成29年4月26日(2017.4.26)

(71) 出願人 504133110
国立大学法人電気通信大学
東京都調布市調布ヶ丘一丁目5番地1
(74) 代理人 100107766
弁理士 伊東 忠重
(74) 代理人 100070150
弁理士 伊東 忠彦
(72) 発明者 橋本 直己
東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
(72) 発明者 小林 大祐
東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
Fターム(参考) 5B057 BA02 DA12 DB02 DC08 DC16 DC30 DC33 DC40

最終頁に続く

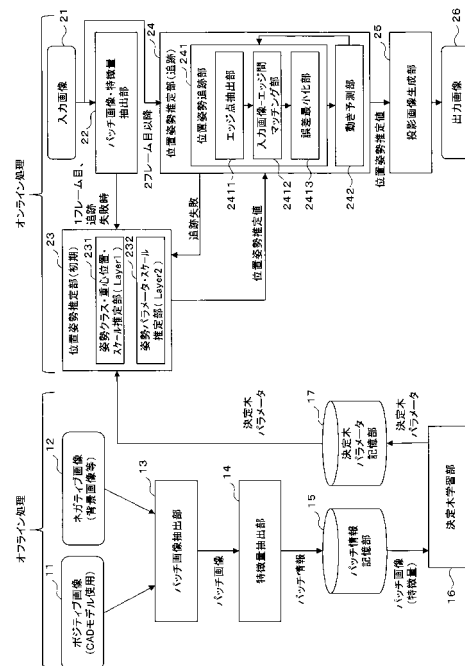
(54) 【発明の名称】 物体検出装置、物体検出方法、物体検出プログラムおよび物体検出システム

(57) 【要約】

【課題】照明変化や対象物体の位置姿勢の変動や自己遮蔽に対する頑健さと位置姿勢の推定の精度の向上とを同時に満たすことができる手法を提供する。

【解決手段】検出の対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、前記対象物体を撮影して得た入力画像のパッチ画像をいずれかの姿勢クラスに分類する第1の識別器と、前記対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、姿勢クラスが推定された入力画像のパッチ画像をいずれかの姿勢パラメータに分類する第2の識別器とを備える。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

検出の対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、前記対象物体を撮影して得た入力画像のパッチ画像をいずれかの姿勢クラスに分類する第 1 の識別器と、

前記対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、姿勢クラスが推定された入力画像のパッチ画像をいずれかの姿勢パラメータに分類する第 2 の識別器と

を備えたことを特徴とする物体検出装置。

【請求項 2】

前記入力画像を赤外線カメラから入力することを特徴とする請求項 1 に記載の物体検出装置。

【請求項 3】

前記特徴量として累積勾配方向特徴量または量子化勾配方向特徴量を用いることを特徴とする請求項 1 または 2 に記載の物体検出装置。

【請求項 4】

前記第 1 の識別器および前記第 2 の識別器は、前記入力画像の個々のパッチ画像による分類結果の総合的な投票結果に基づいて分類を行うことを特徴とする請求項 1 乃至 3 のいずれか一項に記載の物体検出装置。

【請求項 5】

前記第 1 の識別器および前記第 2 の識別器は、前記第 1 の識別器および前記第 2 の識別器を構成する決定木を Random Ferns 形式で構成することを特徴とする請求項 1 乃至 4 のいずれか一項に記載の物体検出装置。

【請求項 6】

前記第 2 の識別器により推定された姿勢パラメータを初期値として、前記入力画像により前記対象物体の位置姿勢を追跡する位置姿勢追跡部と、

前記対象物体の過去の位置姿勢の変化から所定の遅延後の前記対象物体の位置姿勢を予測する動き予測部と

を備えたことを特徴とする請求項 1 乃至 5 のいずれか一項に記載の物体検出装置。

【請求項 7】

前記位置姿勢追跡部は、前記初期値における前記対象物体の CAD モデル上のエッジ点と前記入力画像から抽出した前記対象物体のエッジ点との誤差を最小化するように位置姿勢に補正を加える

ことを特徴とする請求項 6 に記載の物体検出装置。

【請求項 8】

対象物体の姿勢パラメータの初期値を入力し、前記対象物体を撮影して得た入力画像により前記対象物体の位置姿勢を追跡する位置姿勢追跡部と、

前記対象物体の過去の位置姿勢の変化から所定の遅延後の前記対象物体の位置姿勢を予測する動き予測部と

を備えたことを特徴とする物体検出装置。

【請求項 9】

検出の対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、前記対象物体を撮影して得た入力画像のパッチ画像をいずれかの姿勢クラスに分類する第 1 の識別手順と、

前記対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、姿勢クラスが推定された入力画像のパッチ画像をいずれかの姿勢パラメータに分類する第 2 の識別手順と

をコンピュータが実行することを特徴とする物体検出方法。

【請求項 10】

対象物体の姿勢パラメータの初期値を入力し、前記対象物体を撮影して得た入力画像に

10

20

30

40

50

より前記対象物体の位置姿勢を追跡する位置姿勢追跡手順と、

前記対象物体の過去の位置姿勢の変化から所定の遅延後の前記対象物体の位置姿勢を予測する動き予測手順と

をコンピュータが実行することを特徴とする物体検出方法。

【請求項 1 1】

検出の対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、前記対象物体を撮影して得た入力画像のパッチ画像をいずれかの姿勢クラスに分類する第 1 の識別手順と、

前記対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、姿勢クラスが推定された入力画像のパッチ画像をいずれかの姿勢パラメータに分類する第 2 の識別手順と

をコンピュータに実行させることを特徴とする物体検出プログラム。

【請求項 1 2】

対象物体の姿勢パラメータの初期値を入力し、前記対象物体を撮影して得た入力画像により前記対象物体の位置姿勢を追跡する位置姿勢追跡手順と、

前記対象物体の過去の位置姿勢の変化から所定の遅延後の前記対象物体の位置姿勢を予測する動き予測手順と

をコンピュータに実行させることを特徴とする物体検出プログラム。

【請求項 1 3】

1 以上の情報処理装置を有する物体検出システムであって、

検出の対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、前記対象物体を撮影して得た入力画像のパッチ画像をいずれかの姿勢クラスに分類する第 1 の識別器と、

前記対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、姿勢クラスが推定された入力画像のパッチ画像をいずれかの姿勢パラメータに分類する第 2 の識別器と、

前記第 2 の識別器により推定された姿勢パラメータを初期値として、前記入力画像により前記対象物体の位置姿勢を追跡する位置姿勢追跡部と、

前記対象物体の過去の位置姿勢の変化から所定の遅延後の前記対象物体の位置姿勢を予測する動き予測部と

を備えたことを特徴とする物体検出システム。

【請求項 1 4】

対象物体の姿勢パラメータの初期値を入力し、前記対象物体を撮影して得た入力画像により前記対象物体の位置姿勢を追跡する位置姿勢追跡部と、

前記対象物体の過去の位置姿勢の変化から、前記対象物体が投影されるまでに発生する遅延後の位置姿勢を予測する第 1 動き予測部と、

前記対象物体の過去の位置姿勢の変化から、前記対象物体が次に撮影されるまでの遅延後の位置姿勢を予測する第 2 動き予測部と

を備えたことを特徴とする物体検出装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、物体検出装置、物体検出方法、物体検出プログラムおよび物体検出システムに関する。

【背景技術】

【0002】

ファクトリ・オートメーション、拡張現実感 (AR: Augmented Reality)、映像投影を用いた空間演出、プロジェクションマッピング等のエンターテインメント等において、対象物体の位置姿勢 (位置、方向) の検出が必要となる場面がある。例えば、ファクトリ・オートメーションにおいては、生産ラインを流れる部品・製品等の外観から部品・製品

10

20

30

40

50

等の特定や載置された位置・方向を検出し、その部品・製品等に対するその後の処理を決定する場合がある。また、拡張現実感、映像投影を用いた空間演出、プロジェクションマッピング等のエンターテインメント等では、映像を重ねる対象物体の位置姿勢の検出が必須となる。

【0003】

従来、画像中から対象物体の位置姿勢を検出（推定）する手法として、特徴点マッチングによる手法と、テンプレートマッチングによる手法と、投票ベースによる手法とが用いられていた。なお、ここでは対象物体の形状は変化しないものとする。

【0004】

特徴点マッチングによる手法は、予め登録しておいた対象物体の特徴点の3次元位置と入力画像から検出した特徴点との複数の対応関係から位置姿勢を推定するものである。この手法では、照明変化や部分的な特徴点の遮蔽（自己遮蔽）に頑健であるが、表面に模様等が存在しないか少ないテクスチャレス物体に対しては、有効な特徴点が抽出しづらく、安定した位置姿勢の推定が行えないという問題がある。プロジェクションマッピング等では、投影による映像効果を高めるため、単色（白色等）の対象物体が用いられることが多く、テクスチャレス物体への対応は重要である。

10

【0005】

テンプレートマッチングによる手法は、入力画像上を走査し、予め位置姿勢に対応させて登録しておいた2次元画像のテンプレートデータベースから類似度の高いテンプレートを選択することによって位置姿勢の推定を行うものである。この手法では、テクスチャレス物体に対しても有効であるが、ある位置姿勢における対象物体の全体の画像に基づいたテンプレートを用いるため、対象物体の微小な変動や自己遮蔽に対して頑健ではないという問題がある。

20

【0006】

この点につき、位置姿勢の変動に対応する手法として、微小な変動を考慮したテンプレートマッチングによる手法が提案されている（例えば、特許文献1、非特許文献1等を参照）。これは、3次元のCAD（Computer-Aided Design）データからテンプレート画像のデータベースを作成する際に、CADモデルを微小に変動させた際の輝度勾配方向を累積させることによって、3次元の姿勢の変動に頑健な特徴量を生成し、それを推定に用いるものである。この手法では、CADモデルを変動させる際に観測される輝度勾配の出現の頻度によって画素に重みを加えているが、CADモデルの重心から離れるほど変動量が増えるため、特徴量として選択されにくくなる。そのため、重心から離れた画像の特徴量が推定に反映されず、推定の精度を高められないという問題がある。また、この手法も、自己遮蔽に関しては考慮されていない。

30

【0007】

投票ベースによる手法は、画像を小さなサイズのパッチ画像に分け、対象物体かどうかのクラス確率とその物体中心へのオフセット量を複数の決定木で学習（機械学習）する。そして、識別時に決定木による識別結果を画像空間に投票することで、投票密度の高い点から物体中心を求め、位置姿勢を推定するものである。この手法は、対象物体の微小な変動や自己遮蔽に対しては頑健であるが、一元的な処理により推定を行うことと、処理能力の関係から学習に用いることのできるパッチ数に限界があることから、位置姿勢の推定の精度が低いという問題がある。

40

【先行技術文献】

【特許文献】

【0008】

【特許文献1】特開2015 007972号公報

【非特許文献】

【0009】

【非特許文献1】小西嘉典，半澤雄希，川出雅人，橋本学：“階層的統合モデルを用いた単眼カメラからの高速3次元物体位置・姿勢認識，Vision Engineering workshop (2015)

50

【発明の概要】

【発明が解決しようとする課題】

【0010】

上述したように、テクスチャレス物体に対しては、輝度勾配方向を累積させたテンプレートマッチングによる手法や、投票ベースによる手法が有利と考えられる。しかし、照明変化や対象物体の位置姿勢の変動や自己遮蔽に対する頑健さと位置姿勢の推定の精度の向上とを同時に満たすことができる手法は存在しなかった。

【0011】

本発明は上記の従来の問題点に鑑み提案されたものであり、その目的とするところは、照明変化や対象物体の位置姿勢の変動や自己遮蔽に対する頑健さと位置姿勢の推定の精度の向上とを同時に満たすことができる手法を提供することにある。

【課題を解決するための手段】

【0012】

上記の課題を解決するため、本発明にあっては、検出の対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、前記対象物体を撮影して得た入力画像のパッチ画像をいずれかの姿勢クラスに分類する第1の識別器と、前記対象物体の様々な姿勢の画像から抽出したパッチ画像の特徴量に基づいて学習され、姿勢クラスが推定された入力画像のパッチ画像をいずれかの姿勢パラメータに分類する第2の識別器とを備える。

【発明の効果】

【0013】

本発明にあっては、照明変化や対象物体の位置姿勢の変動や自己遮蔽に対する頑健さと位置姿勢の推定の精度の向上とを同時に満たすことができる。

【図面の簡単な説明】

【0014】

【図1】プロジェクションマッピングに適用した一実施形態のシステム構成例を示す図である。

【図2】実施形態の機能構成例を示す図である。

【図3】情報処理装置のハードウェア構成例を示す図である。

【図4】オフライン処理の例を示すフローチャートである。

【図5】3Dモデルからポジティブ画像を生成する例を示す図である。

【図6】パッチ画像の抽出の例を示す図である。

【図7】特徴量の例を示す図である。

【図8】記憶されるパッチ情報のデータ構造例を示す図である。

【図9】機械学習に用いられる決定木の例を示す図である。

【図10】オンライン処理の例を示すフローチャートである。

【図11】あるスケールに対応する投票空間への投票結果の例を示す図である。

【図12】エッジ点の例を示す図である。

【図13】対象物体への投影の例を示す図である。

【図14】第2実施形態の機能構成例を示す図である。

【図15】第2実施形態におけるオンライン処理の例を示すフローチャートである。

【図16】処理結果例を示す図である。

【図17】実験に用いた対象物体を示す図である。

【図18】平行移動の実験結果を示す図である。

【図19】回転運動の実験結果を示す図である。

【発明を実施するための形態】

【0015】

以下、本発明の好適な実施形態につき説明する。

【0016】

(第1実施形態)

【0017】

<構成>

図1はプロジェクションマッピングに適用した一実施形態のシステム構成例を示す図である。図1において、事前に行われるオフライン処理のためのPC(Personal Computer)等の情報処理装置1と、本番におけるオンライン処理のためのPC等の情報処理装置2とが設けられている。なお、情報処理装置1によるオフライン処理の結果は、決定木パラメータとして情報処理装置2に引き渡される。なお、情報処理装置1と情報処理装置2は同じ装置を用いてもよく、その場合は決定木パラメータの引き渡しは必要ない。

【0018】

オンライン処理においては、情報処理装置2のほかに、カメラ3とプロジェクタ4と赤外照明5とが設けられ、対象物体Oをカメラ3により撮影した入力画像が情報処理装置2に入力され、情報処理装置2からは出力画像(投影映像)がプロジェクタ4に出力される。なお、カメラ3とプロジェクタ4は、チェッカーボード等を用いたキャリブレーションが予め行われ、画素位置の対応付けがなされる。また、カメラ3は、プロジェクタ4により対象物体O上に投影される画像や外光による影響を受けないように、赤外線カメラが用いられる。更に、対象物体Oの動きへの追跡が容易となるように、カメラ3には高速度(フレームレートが高い)のものが用いられる。

【0019】

図2は実施形態の機能構成例を示す図である。図2において、オフライン処理を実行する情報処理装置1による機能構成として、パッチ画像抽出部13と特徴量抽出部14と決定木学習部16とを備えている。パッチ画像抽出部13は、CADモデルを使用して生成されたポジティブ画像11と、背景画像等のネガティブ画像12とを入力し、複数(多数)の小サイズのパッチ画像を抽出する機能を有している。特徴量抽出部14は、パッチ画像抽出部13により抽出されたパッチ画像から画像の特徴量を抽出し、学習時および識別(オンライン処理における初期の位置姿勢推定)時に用いる他の情報を付加したパッチ情報をパッチ情報記憶部15に格納する機能を有している。特徴量としては、ポジティブ画像11については主に累積勾配方向特徴量を用い、ネガティブ画像12については量子化勾配方向特徴量を用いている。なお、ポジティブ画像11について累積勾配方向特徴量を用いることで効率的な学習が可能になるが、量子化勾配方向特徴量を用いてもよい。累積勾配方向特徴量と量子化勾配方向特徴量の詳細については後述する。決定木学習部16は、パッチ情報記憶部15に格納されたパッチ情報に基づき、決定木のパラメータ(決定木パラメータ)を機械学習し、学習結果の決定木パラメータを決定木パラメータ記憶部17に格納する機能を有している。

【0020】

一方、オンライン処理を実行する情報処理装置2による機能構成として、パッチ画像・特徴量抽出部22と位置姿勢推定部(初期)23と位置姿勢推定部(追跡)24と投影画像生成部25とを備えている。位置姿勢推定部23は、姿勢クラス・重心位置・スケール推定部231と姿勢パラメータ・スケール推定部232とを備えている。位置姿勢推定部24は、位置姿勢追跡部241と動き予測部242とを備えている。位置姿勢追跡部241は、エッジ点抽出部2411と入力画像-エッジ間マッチング部2412と誤差最小化部2413とを備えている。

【0021】

パッチ画像・特徴量抽出部22は、カメラ3による撮影で取得された画像を複数のスケールにした入力画像21からパッチ画像を抽出し、その特徴量を抽出する機能を有している。特徴量としては、量子化勾配方向特徴量を用いている。複数のスケールの入力画像21とするのは、対象物体Oのカメラ3からの距離を推定するためである。

【0022】

位置姿勢推定部23は、入力画像21の1フレーム目または追跡失敗後の先頭フレームからパッチ画像・特徴量抽出部22により抽出されたパッチ画像の特徴量に基づき、オフ

10

20

30

40

50

ライン処理で学習された決定木パラメータに基づいて対象物体Oの初期の位置姿勢を推定する機能を有している。姿勢クラス・重心位置・スケール推定部231は、第1段階(Layer1)の推定として、対象物体Oの姿勢クラスと重心位置とスケールを推定する機能を有している。スケールは、パッチ画像の生成時の仮想カメラと対象物体Oの関係から距離に変換することが可能であり、カメラ3と対象物体Oの距離の表現方法の一つである。この姿勢クラス・重心位置・スケール推定部231は、入力画像21のパッチ画像を姿勢クラスに分類する第1の識別器として動作する。姿勢パラメータ・スケール推定部232は、第2段階(Layer2)の推定として、姿勢クラス・重心位置・スケール推定部231により推定された対象物体Oの姿勢クラスと重心位置とスケールに基づき、詳細な姿勢パラメータとスケール(第1段階よりも細分化したもの)を推定する機能を有している。第2段階で最終的に推定されたスケールから、カメラ3と対象物体Oの距離が求められる。この姿勢パラメータ・スケール推定部232は、姿勢クラス・重心位置・スケール推定部231により推定された姿勢クラス内で、入力画像21のパッチ画像を詳細な姿勢パラメータに分類する第2の識別器として動作する。

10

【0023】

位置姿勢推定部24は、初期値の例となる、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値に基づき、位置姿勢の誤差の補正およびその後の対象物体Oの追跡を行う機能を有している。追跡が失敗した場合、位置姿勢推定部24は位置姿勢推定部23に対して追跡失敗を通知する。位置姿勢追跡部241は、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値によるCADモデル上のエッジ点と入力画像21のエッジ点とを比較することにより、推定後に変化した対象物体Oの位置姿勢に補正する機能を有している。なお、この位置姿勢の補正は、位置姿勢推定部23による位置姿勢の推定の誤差を補正することにもなり、位置姿勢の精度向上に寄与する。

20

【0024】

エッジ点抽出部2411は、入力画像21から対象物体Oの輪郭を示すエッジ点を抽出するとともに、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値によるCADモデル上のエッジ点を抽出する機能を有している。入力画像-エッジ間マッチング部2412は、CADモデル上のエッジ点と入力画像21のエッジ点とを対応付ける機能を有している。誤差最小化部2413は、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値によるCADモデル上のエッジ点と入力画像21のエッジ点との誤差が最小化するように位置姿勢を補正する機能を有している。

30

【0025】

動き予測部242は、追跡中の対象物体Oの位置姿勢から、後続の投影画像の生成および対象物体Oへの投影に要する遅延時間後の対象物体Oの位置姿勢を予測する機能を有している。

【0026】

投影画像生成部25は、位置姿勢推定部24により推定された対象物体Oの位置姿勢に基づいて、その位置姿勢に整合させた投影画像を生成し、出力画像26として出力する機能を有している。

【0027】

なお、オンライン処理においては、初期の位置姿勢推定と、その後の追跡における位置姿勢推定とを同時に実施する場合について記載しているが、それぞれを単独で実施することもできる。例えば、追跡が必要ない場合または他の手法により追跡を行う場合は、初期の位置姿勢推定を単独で実施することができる。また、初期の位置姿勢推定を他の手法により行う場合は、追跡における位置姿勢推定を単独で実施することができる。

40

【0028】

図3は情報処理装置1、2のハードウェア構成例を示す図である。図3において、情報処理装置1、2は、バス107を介して相互に接続されたCPU(Central Processing Unit)101、ROM(Read Only Memory)102、RAM(Random Access Memory)103を備えている。なお、CPU101には、汎用的なCPUの他に、GPU(Graphic

50

Processing Unit) も含まれるものとする。また、情報処理装置 1、2 は、HDD (Hard Disk Drive) / SSD (Solid State Drive) 104、接続 I/F (Interface) 105、通信 I/F 106 を備えている。CPU 101 は、RAM 103 をワークエリアとして ROM 102 または HDD / SSD 104 等に格納されたプログラムを実行することで、情報処理装置 1、2 の動作を統括的に制御する。接続 I/F 105 は、情報処理装置 1、2 に接続される機器とのインタフェースである。通信 I/F 106 は、ネットワークを介して他の情報処理装置と通信を行うためのインタフェースである。

【0029】

図 2 で説明した情報処理装置 1、2 の機能は、CPU 101 において所定のプログラムが実行されることで実現される。プログラムは、記録媒体を経由して取得されるものでもよいし、ネットワークを経由して取得されるものでもよいし、ROM 組込でもよい。処理に際して参照・更新されるデータは、RAM 103 または HDD / SSD 104 に保持される。

10

【0030】

<動作>

図 4 はオフライン処理の例を示すフローチャートである。図 4 において、情報処理装置 1 では、検出対象となる対象物体 O の CAD モデルからポジティブ画像 11 を生成する (ステップ S11)。なお、他の情報処理装置においてポジティブ画像 11 を生成し、それを情報処理装置 1 で取得するようにしてもよい。

【0031】

図 5 は 3D モデルからポジティブ画像を生成する例を示す図である。図 5 において、対象物体 O の CAD による 3D モデルを中心とした所定半径の仮想球面上に仮想カメラ VC を置き、様々な位置からのポジティブ画像 11 を取得する。仮想カメラ VC の位置を v_x 、 v_y 、 v_z 、仮想カメラ VC の光軸回りの回転角を r_o とすると、姿勢パラメータは、

20

$$= \{v_x, v_y, v_z, r_o\}$$

と表すことができる。

【0032】

また、2 段階の機械学習における第 1 段階の機械学習に対応させるため、仮想カメラ VC を置く球面を例えば 8 つの領域 (クラス) に区分する。8 つの領域は、例えば、球面を北半球と南半球に分けた上で、それぞれを経度方向に 4 つに区分する。そして、8 つの領域内において、仮想カメラ VC の位置と回転を均等に变化させてポジティブ画像 11 を取得する。なお、ポジティブ画像 11 の特徴量に用いる累積勾配方向特徴量を得ることができるように、位置姿勢を微小に変化させたポジティブ画像 11 を併せて取得する。ただし、対象物体 O の重心を中心に位置姿勢を变化させた場合には重心から遠くなる点が特徴量に反映されにくくなるため、空間的に均等に配置されたサンプリング点を中心に位置姿勢を微小に変化させる。なお、照明の強度についても一様乱数で変化させる。

30

【0033】

図 4 に戻り、背景画像や、対象物体 O 以外の誤検出される可能性のある物体についてネガティブ画像 12 を取得する (ステップ S12)。ネガティブ画像 12 は、デジタルカメラ等により撮影したもの等を用いることができる。

40

【0034】

次いで、情報処理装置 1 のパッチ画像抽出部 13 は、ポジティブ画像 11 およびネガティブ画像 12 からそれぞれパッチ画像を抽出する (ステップ S13)。抽出したパッチ画像は、相対位置 (ポジティブ画像 11 にあっては対象物体 O の重心からのオフセット) と対応付けておく。図 6 はパッチ画像の抽出の例を示しており、対象物体 O からパッチ画像 P を抽出する様子を示している。パッチ画像 P は、重複を許容し、縦横に数ピクセルずつずらしながら、多数抽出する。

【0035】

図 4 に戻り、情報処理装置 1 の特徴量抽出部 14 は、パッチ画像抽出部 13 により抽出

50

されたパッチ画像から画像の特徴量を抽出し（ステップS14）、学習時および識別時に用いる他の情報を付加したパッチ情報をパッチ情報記憶部15に格納する（ステップS15）。

【0036】

図7は、パッチ画像Pをグリッド状に分割した各格子内における輝度勾配（矢印で示す）の例を示している。輝度勾配は画像にsobelフィルタを適用することで求めることができる。各格子内で輝度勾配の大きさが所定の閾値を超えるものの勾配方向を例えば8つの方向に量子化したものが量子化勾配方向特徴量である。また、ポジティブ画像11の生成時にサンプリング点を中心に位置姿勢を微小に変化させた複数のポジティブ画像11から抽出した近傍にある複数のパッチ画像における量子化勾配方向特徴量を累積し、出現頻度が所定の閾値を超えるものを抽出したものが累積勾配方向特徴量である。また、その際の出現頻度は累積勾配方向特徴量の重みとする。

10

【0037】

図8はパッチ情報記憶部15に記憶されるパッチ情報のデータ構造例を示す図である。ポジティブ画像11に対するパッチ情報は、「量子化勾配方向特徴量」「累積勾配方向特徴量」「累積勾配方向特徴量の重み」「パッチのクラスラベル」「パッチの姿勢方向ラベル」「オフセットベクトル」「姿勢パラメータ」「対象物体との距離」等を含んでいる。ネガティブ画像12に対するパッチ情報は、「量子化勾配方向特徴量」「パッチのクラスラベル」等を含んでいる。この場合の「パッチのクラスラベル」は、ポジティブ画像11の位置姿勢（図5において撮影を行う8つの領域に対応）のクラスラベル（例えば、1～8）とは異なるクラスラベル（例えば、0）が設定される。

20

【0038】

図4に戻り、情報処理装置1の決定木学習部16は、パッチ情報記憶部15に格納されたパッチ情報に基づいて2段階（2層）の機械学習を行い（ステップS16）、学習結果の決定木パラメータを決定木パラメータ記憶部17に格納する（ステップS17）。なお、この決定木パラメータが示す分岐条件に基づいて、決定木は、パッチ画像を分類し、分類結果を出力する。

【0039】

図9は機械学習に用いられる決定木の例を示す図であり、決定木は複数設けられ、各決定木はルートノードから2つに分岐して行き、以降のノードでも2つに分岐し、末端ノードに達する。各ノードには分岐関数が設定され、判断結果により左か右に分岐する。各ノードの分岐関数は、学習サンプルとなるパッチ画像と、比較対象としてランダムに選択されるパッチ画像の特徴量とから類似度を計算し、類似度を所定の閾値と比較して、閾値以上であるか否かの判断を行う。なお、一般にはRandom Forestsと呼ばれる、各ノードの分岐関数が異なるものが用いられるが、本実施形態では、演算処理の高速化のために、1つの決定木において、同じ階層のノードにおける分岐関数を同じにしたRandom Fernsと呼ばれる形式を用いている。

30

【0040】

第1段階（Layer1）の学習では、パッチ情報記憶部15に格納された多数のパッチ情報からランダムにサンプリングしたデータセットと、サンプル内からランダムに取り出したポジティブ画像のパッチ情報とに基づいて決定木で分岐する。第2段階（Layer1）の学習では、クラスラベル（例えば、1～8）毎に、各クラスに属するパッチ情報のデータセットと、同じクラス内からランダムに取り出したポジティブ画像のパッチ情報とに基づいて決定木で分岐する。そして、第1段階および第2段階のいずれにおいても、ポジティブ画像のパッチ情報と分岐関数の閾値とをランダムに変動させ、分岐結果のエントロピーが最小になるように各ノードのポジティブ画像のパッチ情報と閾値を決定する。

40

【0041】

第1段階（Layer1）の決定木は、並列的に複数（例えば、20）設けられ、各決定木の末端ノードにはクラスラベル（例えば、0、1～8）が割り当てられ、更に「クラス確率」と「オフセットベクトル」が保持される。「クラス確率」は、末端ノードに割り当

50

てられたクラスラベルに実際に分類された同クラスラベルのパッチ画像の比率である。例えば、クラスラベル「4」が割り当てられた末端のノードに10個のパッチ画像が分類され、そのうちクラスラベル「4」のパッチ画像が3個ある場合、クラス確率は0.3(=3÷10)となる。「オフセットベクトル」は、末端のノードに割り当てられたクラスラベルに実際に分類された同クラスラベルのパッチ画像のオフセットベクトルの平均である。各ノードにおける比較対象のパッチ情報と閾値と、末端のノードのクラスラベルとクラス確率とオフセットベクトルは、第1段階の決定木の決定木パラメータとして決定木パラメータ記憶部17に格納される。

【0042】

第2段階(Layer2)の決定木は、ポジティブ画像に対応するクラスラベル(例えば、1~8)のそれぞれに複数(例えば、20)設けられ、決定木の末端のノードには「姿勢パラメータ」が保持される。「姿勢パラメータ」は、末端のノードに分類されたパッチ画像の姿勢パラメータの平均である。各ノードにおける比較対象のパッチ情報と閾値と、末端のノードの姿勢パラメータは、第2段階の決定木の決定木パラメータとして決定木パラメータ記憶部17に格納される。

【0043】

図10はオンライン処理の例を示すフローチャートである。図10において、情報処理装置2のパッチ画像・特徴量抽出部22は、カメラ3による撮影で取得された画像を複数のスケールにした入力画像21からパッチ画像を抽出し、その特徴量を抽出する(ステップS201)。特徴量としては、量子化勾配方向特徴量を用いる。

【0044】

次いで、位置姿勢推定部(初期)23は、入力画像21の1フレーム目または追跡失敗後の先頭フレームからパッチ画像・特徴量抽出部22により抽出されたパッチ画像の特徴量に基づき、オフライン処理で学習された決定木パラメータに基づいて対象物体0の初期の位置姿勢を推定する(ステップS202)。

【0045】

すなわち、位置姿勢推定部23の姿勢クラス・重心位置・スケール推定部231は、第1段階(Layer1)の推定として、対象物体0の姿勢クラスと重心位置とスケールを推定する(ステップS203)。より具体的には、次のような処理を行う。まず、各スケールおよび姿勢方向クラスに対するx y空間の投票空間(投票平面)(より具体的には、スケール毎の投影平面(x y空間)が、スケール分だけ重なったような3次元空間)を作成しておく。入力画像21から抽出したパッチ画像を第1段階の決定木パラメータに基づく決定木に入力し、各ノードの分岐関数に基づいて分岐させる。末端のノードに辿りついた際に、格納されている姿勢方向のクラスおよびスケールに対応する投票空間に投票する。図11はあるスケールに対応する投票空間への投票結果の例を示す図であり、台風の目のように見える点が極大値(あるスケールでの重心)を示しており、x, y, scaleで構築される3次元空間の中なら、mean shift法を使って極大が求められる。全ての決定木の結果を投票した上で、極大が求められ、その位置、スケールおよび姿勢方向クラスが第1段階の推定の結果として出力される。なお、姿勢クラスには別に投票処理が用意され、末端に到達したパッチ数と、末端に保持されているクラス確率とが掛け合わされ、全末端ノード分を足し合わせた中から最大となるクラスが求められる。

【0046】

図10に戻り、位置姿勢推定部23の姿勢パラメータ・スケール推定部232は、第2段階(Layer2)の推定として、姿勢クラス・重心位置・スケール推定部231により推定された対象物体0の姿勢クラスと重心位置とスケールに基づき、詳細な姿勢パラメータとスケール(第1段階よりも細分化したもの)を推定する(ステップS204)。より具体的には、次のような処理を行う。まず、各スケール(第1段階よりも細分化したもの)および姿勢パラメータに対応するx y空間の投票空間(各スケール毎に投票平面を考え、これを積み重ねた3次元空間)を作成しておく。第1段階の推定で得られた姿勢方向クラスに対応する第2段階の決定木に対して、第1段階で検出した領域内(第1段階で検出した

10

20

30

40

50

重心を中心とした、対象物体が含まれると想定される領域内)のパッチ情報を入力して分岐させる。末端のノードに辿りついた際に、スケールに対応する投票空間(スケールと、それに対応する重心(x, y)で構成される3次元空間)に投票する。姿勢パラメータに対しては、投票空間に、決定木の末端に設定された姿勢パラメータに、到達したパッチ画像数を重みとして、平均を求めて、姿勢パラメータを加えていく。全ての決定木の結果を投票した上で、極大を求め、その位置、スケールおよび加重平均した姿勢パラメータが最終的な結果として出力される。順番的には、まずスケールと重心を全ての木の結果を総合して求め、それに対応する姿勢パラメータ(つまり回転)を求める。推定されたスケールからは、学習時にサンプルを撮影した距離を利用して、距離が算出される。

【0047】

次いで、位置姿勢推定部(追跡)24は、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値に基づき、位置姿勢の誤差の補正およびその後の対象物体Oの追跡を行う(ステップS205)。すなわち、位置姿勢推定部24の位置姿勢追跡部241のエッジ点抽出部2411は、入力画像21から対象物体Oの輪郭を示すエッジ点を抽出するとともに、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値によるCADモデル上のエッジ点を抽出する(ステップS206)。次いで、入力画像-エッジ間マッチング部2412は、CADモデル上のエッジ点と入力画像21のエッジ点とを対応付ける(ステップS207)。そして、誤差最小化部2413は、位置姿勢推定部23により推定された対象物体Oの位置姿勢推定値によるCADモデル上のエッジ点と入力画像21のエッジ点との誤差(位置誤差の総和)が最小化するように対象物体Oの位置姿勢を補正する(ステップS208)。図12はエッジ点の例を示しており、入力画像21から得られた輪郭をE1、CADモデルから得られた輪郭をE2で示している。CADモデルの位置姿勢を変化させて入力画像21から得られたエッジ点とできるだけ一致させることで、対象物体Oの位置姿勢を補正する。

【0048】

図10に戻り、エッジ点間の誤差が所定の閾値以下であって補正可である場合(ステップS209のYes)、過去の対象物体Oの動きの変化から所定の遅延後の対象物体Oの位置姿勢を予測して出力し(ステップS210)、位置姿勢の追跡(ステップS205)を繰り返す。カメラ3により撮影された入力画像21による対象物体Oの位置姿勢の推定は、撮影後の処理による遅延により既に実際の位置姿勢から遅延したものであり、更に、その後には更に処理の遅延が起きるため、それらの合計の遅延に相当する予測を行う。位置姿勢の予測は、例えば、直前までの対象物体Oの並行移動の速度および回転の角速度から予測する。また、誤差が所定の閾値より大きく補正不可である場合(ステップS209のNo)、パッチ画像および特徴量の抽出(ステップS201)および初期の位置姿勢の推定(ステップS202)から処理を繰り返す。

【0049】

一方、投影画像生成部25は、出力された位置姿勢に基づいて投影画像を生成して出力する(ステップS211)。図13は対象物体Oへの投影の例を示しており、テクスチャレス物体である対象物体Oに顔の画像を投影した状態を示している。対象物体Oの位置姿勢はリアルタイムに推定・予測され、その位置姿勢に応じた投影画像が生成されて投影されるため、対象物体Oを動かしても、自然な投影を行うことができる。

【0050】

<総括>

以上説明したように、本実施形態によれば、照明変化や対象物体の位置姿勢の変動や自己遮蔽に対する頑健さと位置姿勢の推定の精度の向上とを同時に満たすことができる。

【0051】

(第2実施形態)

第2実施形態は、例えば、第1実施形態と同様のシステム構成及びハードウェア構成によって実現できる。以下、第1実施形態と同様のシステム構成及びハードウェア構成によ

10

20

30

40

50

る例で説明し、重複する説明を省略する。また、同一の構成は、同一の符号を付し、説明を省略する。以下、第1実施形態と異なる点を中心に説明する。具体的には、第1実施形態と比較すると、以下のように、機能構成及びオンライン処理が異なる。

【0052】

<構成>

図14は第2実施形態の機能構成例を示す図である。図2と比較すると、図14に示す機能構成例は、動き予測部242が、第1動き予測部2421と、第2動き予測部2422となる点が異なる。

【0053】

第1動き予測部2421は、第1実施形態における動き予測部242と同一の構成である。

10

【0054】

第2動き予測部2422は、例えば、第1動き予測部2421と同様の方法によって動きを予測する。第1動き予測部2421が、後続における投影画像生成部25によって出力画像が生成され、投影されるまでの遅延時間分を予測するのに対して、第2動き予測部2422は、次の入力画像21が撮影されるまでの遅延時間分を予測する点が異なる。

【0055】

具体的には、まず、入力画像21がカメラによって撮影される時刻を「時刻a」とし、位置姿勢追跡部241によって追跡処理が終了する時刻を「時刻b」とする。

【0056】

この場合には、第1動き予測部2421は、時間「b - a」に加えて、出力画像26が投影されるまでの遅延分を補償する。一方で、第2動き予測部2422は、時間「b - a」に加えて、「時刻b」から次の入力画像21が取得されるまでの遅延分を補償する。

20

【0057】

なお、第1動き予測部2421及び第2動き予測部2422による処理は、例えば、並列の関係となる。

【0058】

<動作>

図15は第2実施形態におけるオンライン処理の例を示すフローチャートである。図10、すなわち、第1実施形態と比較すると、図15は、ステップS301で追加される点が異なる。

30

【0059】

第2動き予測部2422は、対象物体の過去の位置姿勢の変化から、対象物体が次に撮影されるまでの遅延後の位置姿勢を予測する(ステップS301)。補正可と判断されると(ステップS209でYES)、ステップS208のように位置姿勢が補正される。

【0060】

ステップS301では、第2動き予測部2422は、ステップS208によって補正された位置姿勢の時点から、次の入力画像が撮影される時点までの遅延後の位置姿勢を予測する。そして、ステップS206では、ステップS301で予測された位置姿勢におけるCADモデル上のエッジ点が抽出される。

40

【0061】

図16は処理結果例を示す図である。以下、図16(A)に示すような対象物体O2を例に説明する。まず、前フレームの入力画像による補正又は初期の位置姿勢の推定結果に基づいて、対象物体O2が、図16(A)に示すような位置姿勢であると推定されるとする。図16(A)に示す位置姿勢は、前フレームの入力画像等によって推定された位置姿勢であるため、現時点では、位置姿勢が変化している場合がある。

【0062】

そこで、第2動き予測部2422は、前フレームの入力画像等の時点から、次の入力画像が撮影されるまでの遅延分を予測する(ステップS301)。次に、ステップS206では、ステップS301で予測された3DモデルMDLからエッジ点を抽出する。

50

【 0 0 6 3 】

具体的には、ステップ S 2 0 6 では、図示するように、まず、3 D モデル M D L から、3 D モデルの輝度勾配 L U が検出される。次に、輝度勾配 L U から、エッジ点 E D が抽出される。このようにステップ S 3 0 1 で予測された 3 D モデル M D L からエッジ点を抽出されると、より実際の状態に近い状態から、エッジ点 E D が抽出できる。

【 0 0 6 4 】

そのため、図 1 6 (B) に示すように、入力画像と、エッジ点 E D との対応付けにおいて、マッチング精度が向上する。このようにして、図 1 6 (C) に示すように、誤差を最小化した位置姿勢に更新できる。

【 0 0 6 5 】

< 実験結果 >

図 1 7 は実験に用いた対象物体を示す図である。以下、奥行き方向を「Z 軸」とする。また、高さ方向を「Y 軸」とする。さらに、右手方向を「X 軸」とする。

【 0 0 6 6 】

また、対象物体は、Z 軸に 3 3 センチメートル、Y 軸に 1 8 センチメートル及び X 軸に 2 0 センチメートルの大きさである。

【 0 0 6 7 】

図示するような対象物体を X 軸方向へ 5 0 センチメートル程度往復運動させた「平行移動」の実験結果及び Z 軸方向において (R o l l 回転) 0 度乃至 9 0 度の回転をさせた「回転運動」の実験結果を以下に説明する。なお、値は、平均誤差である。

【 0 0 6 8 】

図 1 8 は平行移動の実験結果を示す図である。図 1 8 (A) 及び図 1 8 (B) に示す実験結果は、追跡精度である。追跡精度は、入力画像から位置姿勢を追跡した際の実際的位置との誤差、すなわち、投影される前段階における精度である。したがって、図は、値が小さいほど精度が良いことを示す。

【 0 0 6 9 】

また、図 1 8 は X 軸方向への並進運動であるため、X 軸方向に大きな位置ずれが生じやすい運動である。

【 0 0 7 0 】

図 1 8 (C) 及び図 1 8 (D) に示す実験結果は、投影精度である。投影精度は、プロジェクタによって投影された画像の位置と、対象物体との誤差である。

【 0 0 7 1 】

図では、「位置」は、X 軸、Y 軸又は Z 軸における位置を示し、「回転」は、X 軸 (T i l t 軸)、Y 軸 (P a n 軸) 又は Z 軸 (R o l l 軸) における角度を示す。

【 0 0 7 2 】

図示するように、追跡精度は、比較対象とする 3 D センサカメラ程度に近い精度を出すことができた。一方で、投影精度は、3 D センサカメラより精度良くできた。

【 0 0 7 3 】

なお、実験では、3 D センサカメラは、3 0 f p s (フレーム毎秒) であった。一方で、本実施形態は、1 3 3 f p s の更新速度であった。

【 0 0 7 4 】

図示するように、3 D センサカメラでは、投影における遅延に対して予測がかけにくいいため、3 D センサカメラは、投影精度が悪くなる。一方で、本実施形態は、投影における遅延があっても精度良く画像を投影できる。また、本実施形態では、上記のような更新速度によって、より精度良く予測ができる。さらに、本実施形態は、このような更新速度による予測によって、追跡の際と同じ程度の誤差とすることができる。

【 0 0 7 5 】

以上のように、大きな位置ずれが生じやすい X 軸方向への並進運動において、本実施形態は、X 軸方向における誤差を少なく抑えることができる。なお、本実施形態は、他の軸においては、誤差を計測誤差程度に抑えることができた。

10

20

30

40

50

【 0 0 7 6 】

図 1 9 は回転運動の実験結果を示す図である。図 1 8 (A) 及び図 1 8 (B) と同様に、図 1 9 (A) 及び図 1 9 (B) は、回転運動における追跡精度を示す。

【 0 0 7 7 】

また、図 1 8 (C) 及び図 1 8 (D) と同様に、図 1 9 (C) 及び図 1 9 (D) は、回転運動における投影精度を示す。

【 0 0 7 8 】

図示するように、Z 軸回転の誤差が生じやすい回転運動であっても、追跡精度は、Z 軸回転において、3 D センサカメラ程度に近い精度を出すことができた。また、投影精度は、3 D センサカメラより精度良くできた。

10

【 0 0 7 9 】

< 他の実施形態 >

なお、実施形態は、1 以上の情報処理装置を有する物体検出システムによって実現されてもよい。すなわち、物体検出システムは、図 1 に示すように、オフライン処理を行う情報処理装置と、オンライン処理を行う情報処理装置とを有する構成である。なお、情報処理装置は、1 台でもよく、複数台で構成されてもよい。

【 0 0 8 0 】

以上、本発明の好適な実施の形態により本発明を説明した。ここでは特定の具体例を示して本発明を説明したが、特許請求の範囲に定義された本発明の広範な趣旨および範囲から逸脱することなく、これら具体例に様々な修正および変更を加えることができることは明らかである。すなわち、具体例の詳細および添付の図面により本発明が限定されるものと解釈してはならない。

20

【 符号の説明 】

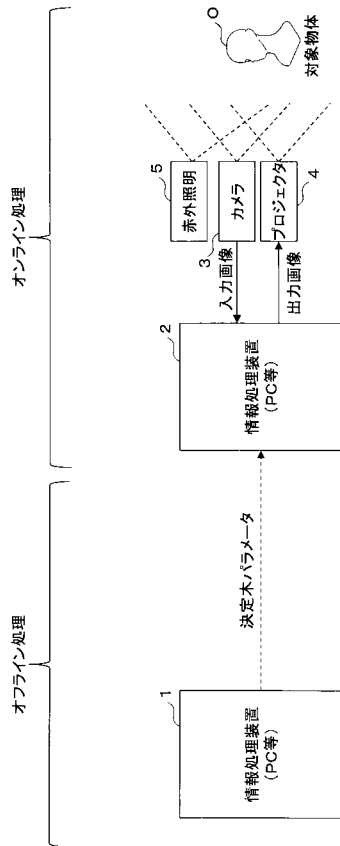
【 0 0 8 1 】

- | | | |
|---------|--------------------|----|
| 1 | 情報処理装置 | |
| 1 1 | ポジティブ画像 | |
| 1 2 | ネガティブ画像 | |
| 1 3 | パッチ画像抽出部 | |
| 1 4 | 特徴量抽出部 | |
| 1 5 | パッチ情報記憶部 | 30 |
| 1 6 | 決定木学習部 | |
| 1 7 | 決定木パラメータ記憶部 | |
| 2 | 情報処理装置 | |
| 2 1 | 入力画像 | |
| 2 2 | パッチ画像・特徴量抽出部 | |
| 2 3 | 位置姿勢推定部 | |
| 2 3 1 | 姿勢クラス・重心位置・スケール推定部 | |
| 2 3 2 | 姿勢パラメータ・スケール推定部 | |
| 2 4 | 位置姿勢推定部 | |
| 2 4 1 | 位置姿勢追跡部 | 40 |
| 2 4 1 1 | エッジ点抽出部 | |
| 2 4 1 2 | 入力画像-エッジ間マッチング部 | |
| 2 4 1 3 | 誤差最小化部 | |
| 2 4 2 | 動き予測部 | |
| 2 4 2 1 | 第 1 動き予測部 | |
| 2 4 2 2 | 第 2 動き予測部 | |
| 2 5 | 投影画像生成部 | |
| 2 6 | 出力画像 | |
| 3 | カメラ | |
| 4 | プロジェクタ | 50 |

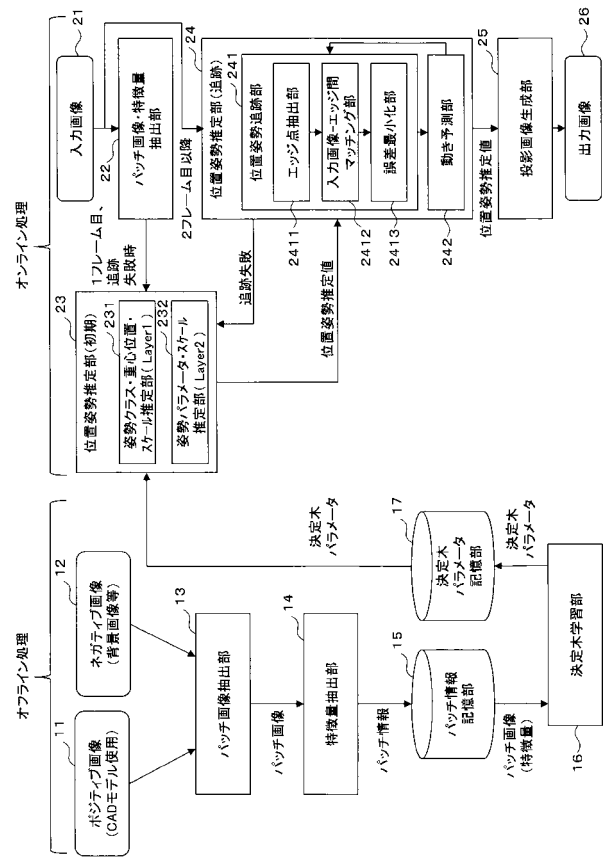
5
O、O 2

赤外照明
対象物体

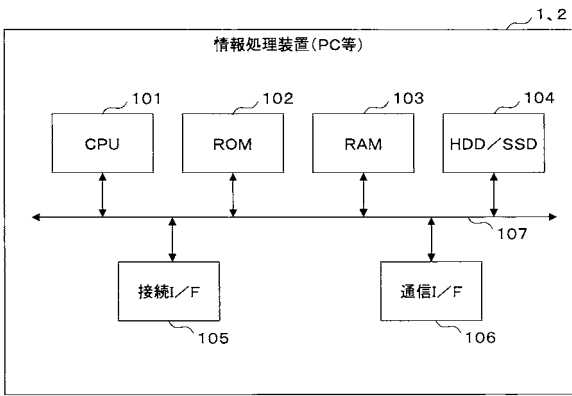
【 図 1 】



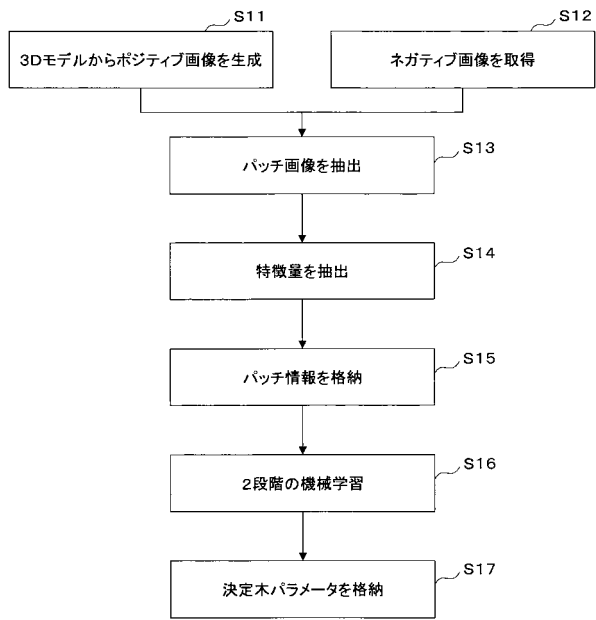
【 図 2 】



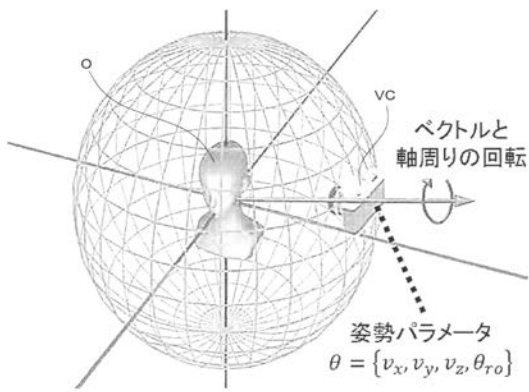
【 図 3 】



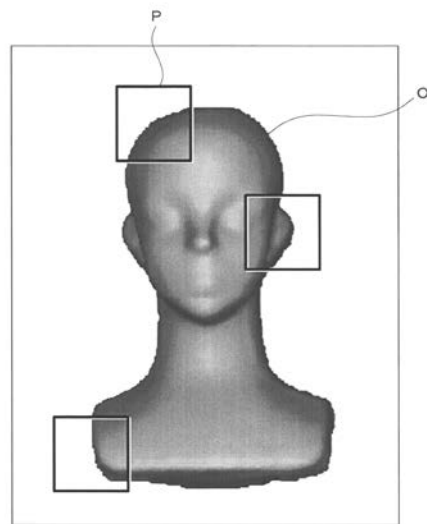
【 図 4 】



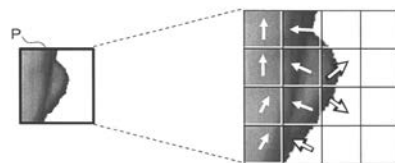
【 図 5 】



【 図 6 】



【 図 7 】



【 図 8 】

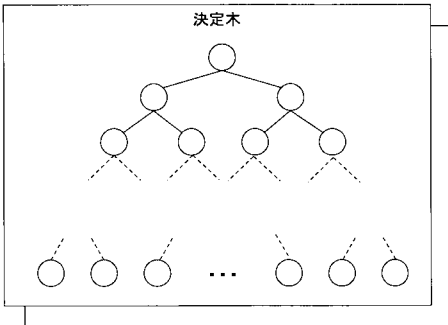
ポジティブ画像に対するパッチ情報

量子化 勾配方向 特徴量	累積 勾配方向 特徴量	累積勾配 方向特徴量 の重み	パッチの クラス ラベル	パッチの 姿勢方向 ラベル	オフセット ベクトル	姿勢 パラメータ	対象物体 との距離	...
--------------------	-------------------	----------------------	--------------------	---------------------	---------------	-------------	--------------	-----

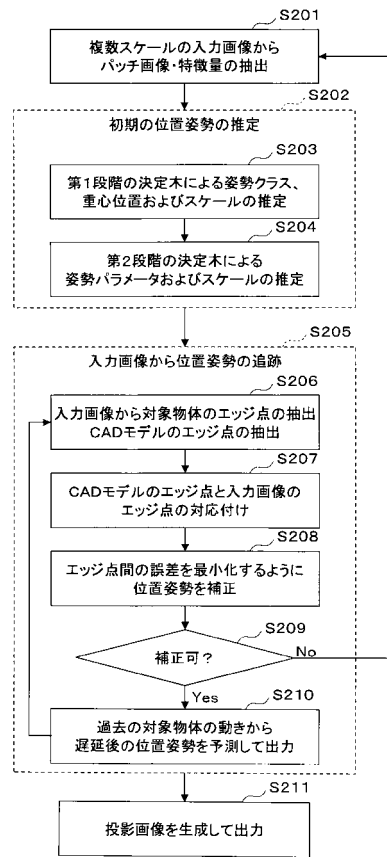
ネガティブ画像に対するパッチ情報

量子化勾配方向 特徴量	パッチの クラスラベル	...
----------------	----------------	-----

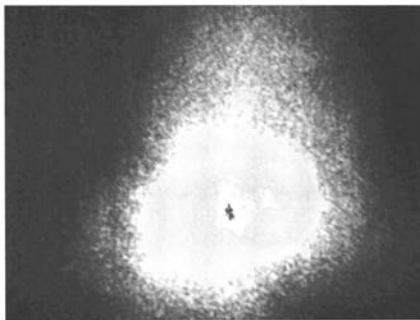
【 図 9 】



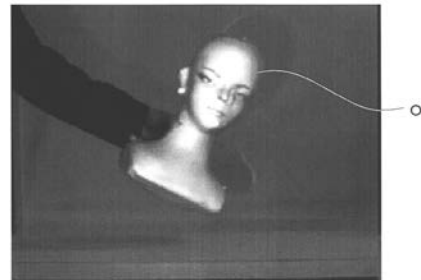
【 図 1 0 】



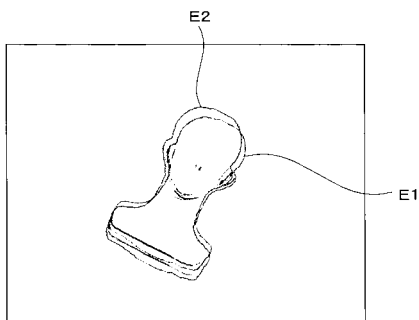
【 図 1 1 】



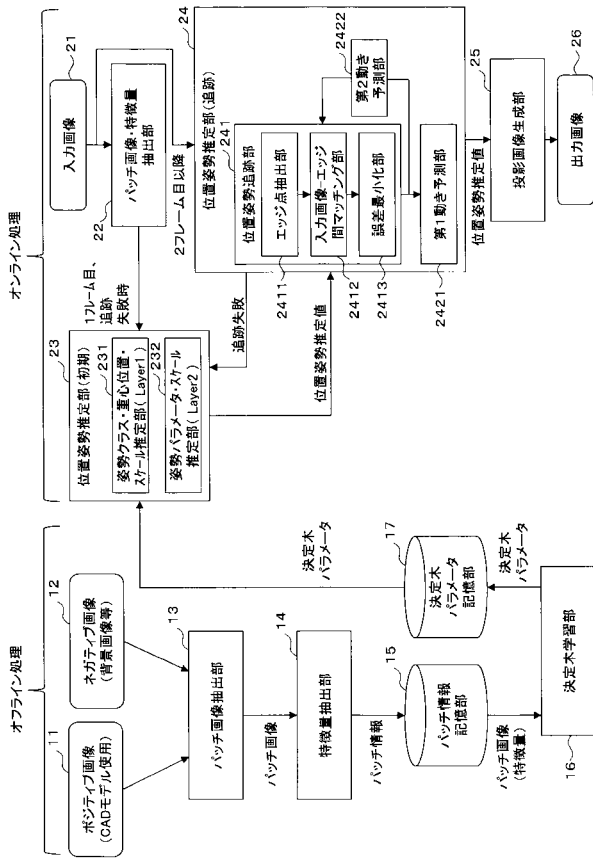
【 図 1 3 】



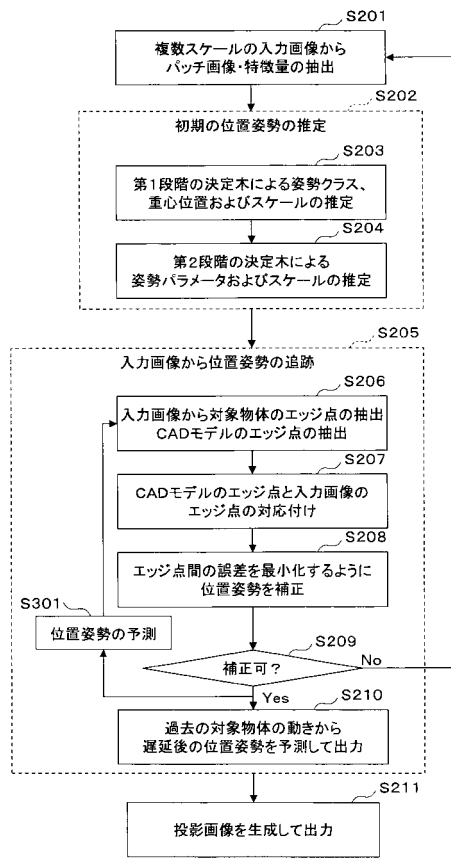
【 図 1 2 】



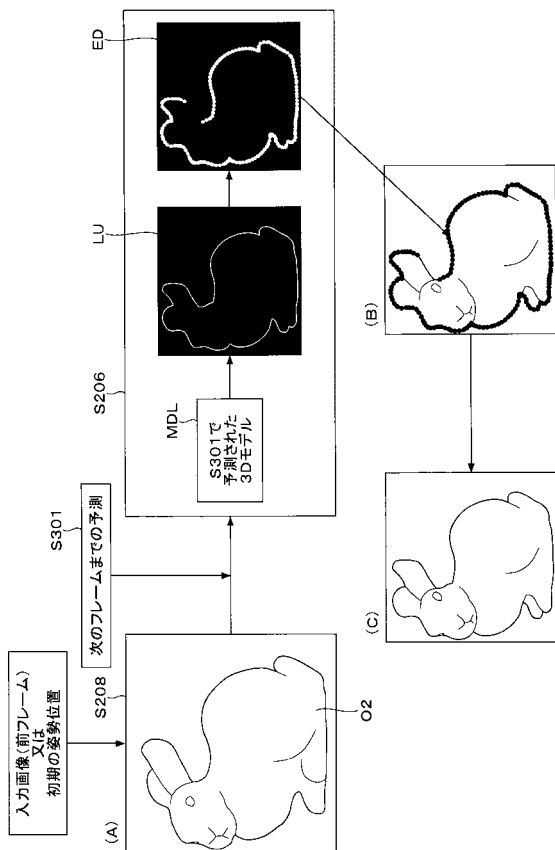
【図14】



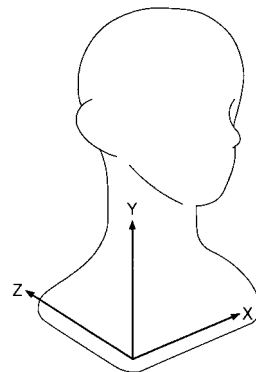
【図15】



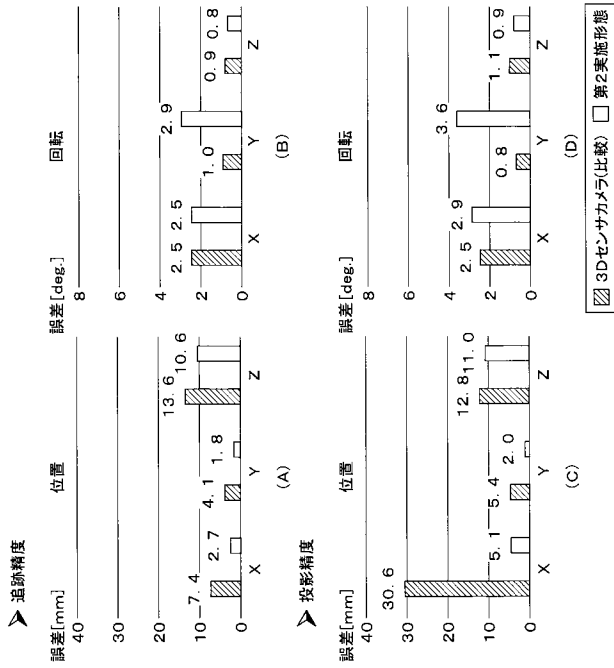
【図16】



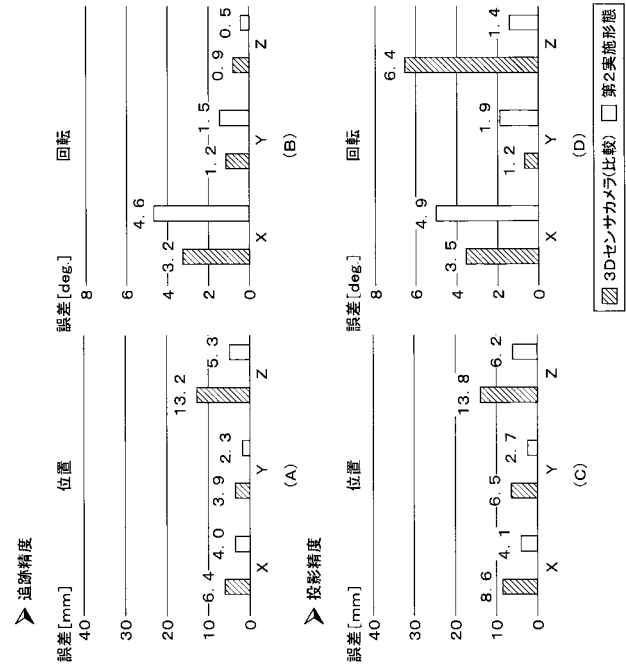
【図17】



【 図 18 】



【 図 19 】



フロントページの続き

Fターム(参考) 5L096 CA04 FA06 FA67 FA74 HA05 HA09 JA11 KA04