

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6431404号
(P6431404)

(45) 発行日 平成30年11月28日(2018.11.28)

(24) 登録日 平成30年11月9日(2018.11.9)

(51) Int.Cl.		F I			
G06T	7/70	(2017.01)	G06T	7/70	
G06T	1/00	(2006.01)	G06T	1/00	3 1 5
G06T	7/00	(2017.01)	G06T	7/00	3 5 0 B

請求項の数 9 (全 17 頁)

(21) 出願番号	特願2015-33179 (P2015-33179)	(73) 特許権者	502324066
(22) 出願日	平成27年2月23日(2015.2.23)		株式会社デンソーアイティラボラトリ
(65) 公開番号	特開2016-157188 (P2016-157188A)		東京都渋谷区渋谷2-15-1 渋谷クロ
(43) 公開日	平成28年9月1日(2016.9.1)		スタワー28F
審査請求日	平成29年6月6日(2017.6.6)	(74) 代理人	100113549
			弁理士 鈴木 守
		(74) 代理人	230113882
			弁護士 本橋 たえ子
		(74) 代理人	100115808
			弁理士 加藤 真司
		(72) 発明者	関川 雄介
			東京都渋谷区渋谷二丁目15番1号 渋谷
			クロスタワー28F 株式会社デンソーア
			イティラボラトリ内

最終頁に続く

(54) 【発明の名称】 姿勢推定モデル生成装置及び姿勢推定装置

(57) 【特許請求の範囲】

【請求項1】

平行等位に設置された複数の撮像デバイスによって撮像されたRGB画像組を取得するRGB画像取得部と、

前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記RGB画像組に対応するデプス画像を生成するデプス画像生成部と、

前記RGB画像取得部にて取得されたRGB画像組における、同一の注目ピクセルの周辺の輝度分布及び/または、前記デプス画像生成部にて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の3次元座標との関係を学習することで、前記対象物の複数の前記3次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデル生成部と、

を備えた姿勢推定モデル生成装置。

【請求項2】

前記姿勢推定モデル生成部は、ランダムフォレストに基づいて、前記注目ピクセルから、前記RGB画像組を構成する画像ごとに異なる量でオフセットさせた周辺ピクセルにおける輝度値の差、または、前記デプス画像において、前記注目ピクセルから異なる量でオフセットさせた周辺ピクセルにおけるデプス値の差に応じて、前記注目ピクセルが属する部分領域の集合のノード分割を繰り返し行うことで、1以上の決定木を生成し、当該各決定木のリーフノードに、正解データとして、前記3次元座標を対応付けることで、姿勢推定モデルを生成する、請求項1に記載の姿勢推定モデル生成装置。

【請求項3】

前記姿勢推定モデル生成部は、前記デプス画像内のあるピクセルにおけるデプス値を d (p_i)、カメラ i によって撮像された画像のあるピクセルにおける RGB 輝度を I (p_i, γ) オフセット量を ω_1 、 ω_2 として、

【数9】

$$f^{da-d}(\theta, p_i) = d\left(p_i + \frac{\omega_1}{d_i}\right) - d\left(p_i + \frac{\omega_2}{d_i}\right)$$

【数10】

$$f^{da-rgb}(\theta, p_i, \beta) = I^\beta\left(p_i + \frac{\omega_1}{d_i}, \gamma_1\right) - I^\beta\left(p_i + \frac{\omega_2}{d_i}, \gamma_2\right),$$

$$\gamma \in \{R, G, B\}, \beta \in \{\text{camera1, camera2, ... camera n}\}$$

d_i : 注目ピクセルにおけるデプス値

ω_1, ω_2 : RGB 輝度

d_i : カメラ ID

で表される $f^{da-d}(\theta, p_i)$ または $f^{da-rgb}(\theta, p_i, \beta)$ と、分割閾値 θ の比較により、情報利得が最も大きくなるように、前記ノード分割を行う請求項2に記載の姿勢推定モデル生成装置。

【請求項4】

前記姿勢推定モデル生成部は、各前記決定木のリーフノードにおいて、前記部分領域を識別する ID と、前記正解データとを対応付け、

前記部分領域を識別する ID は、前記デプス画像生成部にて生成されたデプス画像における前記部分領域に付与された ID を用いる、請求項2に記載の姿勢推定モデル生成装置。

【請求項5】

前記姿勢推定モデル生成部は、各前記決定木のリーフノードにおいて、前記部分領域を識別する ID と、前記正解データとを対応付け、

前記部分領域を識別する ID は、各前記 RGB 画像における前記部分領域に付与された ID と、前記デプス画像生成部にて生成されたデプス画像における前記部分領域に付与された ID に対して、所定の重みづけ係数が設定された所定の算出式に基づいて決定される、請求項2に記載の姿勢推定モデル生成装置。

【請求項6】

平行等位に設置された複数の撮像デバイスによって撮像された RGB 画像組を取得する RGB 画像取得部と、

前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記 RGB 画像組に対応するデプス画像を生成するデプス画像生成部と、

前記 RGB 画像取得部にて取得された RGB 画像組における、同一の注目ピクセルの周辺の輝度分布及びノまたは、前記デプス画像生成部にて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の3次元座標との関係を学習することで、前記対象物の複数の前記3次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデル生成部と、

所定の推定モデルに基づいて、前記 RGB 画像組から、前記仮想的なデプス撮像装置によってデプス画像を撮像した場合におけるデプス値を推定するデプス推定部と、

前記姿勢推定モデル生成部にて生成された姿勢推定モデルを用いて、前記 RGB 画像組及び推定された前記デプス値に基づいて、前記対象物の姿勢推定を行う姿勢推定部とを備えた姿勢推定装置。

【請求項7】

前記デプス推定部は、前記対象物が、前記姿勢推定モデルの生成において学習対象とな

っていない場合には、推定デプス値をゼロとして出力する請求項 6 に記載の姿勢推定装置。

【請求項 8】

平行等位に設置された複数の撮像デバイスによって撮像された R G B 画像組を取得するステップと、

前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記 R G B 画像組に対応するデプス画像を生成するステップと、

前記 R G B 画像組を取得するステップにて取得された R G B 画像組における、同一の注目ピクセルの周辺の輝度分布及び/または、前記デプス画像を生成するステップにて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の 3 次元座標との関係を学習することで、前記対象物の複数の前記 3 次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデルを生成するステップと

を備えた姿勢推定モデル生成方法。

【請求項 9】

姿勢推定モデルを生成するためのプログラムであって、コンピュータに、

平行等位に設置された複数の撮像デバイスによって撮像された R G B 画像組を取得するステップと、

前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記 R G B 画像組に対応するデプス画像を生成するデプス画像を生成するステップと、

前記 R G B 画像組を取得するステップにて取得された R G B 画像組における、同一の注目ピクセルの周辺の輝度分布及び/または、前記デプス画像を生成するステップにて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の 3 次元座標との関係を学習することで、前記対象物の複数の前記 3 次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデルを生成するステップと

を実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、撮像した R G B 画像データから被写体の姿勢を推定するための、姿勢推定モデル生成装置等に関するものである。

【背景技術】

【0002】

従来、ばら積みされた部品を、ロボットアームでピックアップする、ピンピックアップ等の分野において、ワーク（対象物）の姿勢を推定する手法が種々提案されている。例えば、非特許文献 1 に記載の技術は、R G B 画像と、デプス画像を同時に取得することができる R G B - D カメラを用いてワークを撮像する。そして、対象ワーク画像におけるピクセルごとに、パーツ位置と対応付けられた 3 次元座標を推定し、この座標値に基づいてワークの姿勢を推定するものである。

【先行技術文献】

【非特許文献】

【0003】

【非特許文献 1】Eric Brachmann 他: Learning 6 D Object Pose Estimation Using 3 D Object Coordinates, September, 2014

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、非特許文献 1 に記載の技術は、R G B - D カメラの使用を前提とした技術であって、一般的なステレオ R G B カメラにそのまま適用することはできない。また、R G B

10

20

30

40

50

- Dカメラは、カメラ付近の赤外線照明を被写体に照射した場合に、被写体からの反射によって得られる輝度データに基づいてデプスを算出する。したがって、遠くの被写体や、屋外における被写体のデプスの推定に誤差が生じやすく、姿勢推定を精度よく行うことは困難である。

【0005】

本発明は、上記の問題に鑑みてなされたものであり、RGBカメラを用いて、パッシブな構成で、ワークの姿勢推定を精度よく行うことができる、姿勢推定モデル生成装置及び姿勢推定装置を提供することを目的とする。

【課題を解決するための手段】

【0006】

本発明の姿勢推定モデル生成装置は、平行等位に設置された複数の撮像デバイスによって撮像されたRGB画像組を取得するRGB画像取得部と、前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記RGB画像組に対応するデプス画像を生成するデプス画像生成部と、前記RGB画像取得部にて取得されたRGB画像組における、同一の注目ピクセルの周辺の輝度分布及び/または、前記デプス画像生成部にて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の3次元座標との関係を学習することで、前記対象物の複数の前記3次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデル生成部とを備えた構成を有している。

【0007】

ワークの各パーツは、ワークを撮像した画像において、輝度分布とデプス分布に特徴が現れる。したがって、画像におけるワークのパーツ位置を3次元座標で表現した場合、輝度分布とデプス分布の関係と、この3次元座標との関係を学習することで、撮像データからワークのパーツ位置の座標、ひいては、ワークの姿勢を推定することが可能となる。この構成によれば、ステレオカメラ等、平行等位に設置された複数の撮像デバイスによって撮像された左右の画像組における注目ピクセル周辺のRGB輝度分布及び/または仮想的なデプス画像におけるデプス値分布から、ワークの姿勢を推定することが可能なモデルを生成することができる。したがって、パッシブな構成で、姿勢推定を行うことができるモデルを生成することができる。

【0008】

本発明の姿勢推定モデル生成装置において、前記姿勢推定モデル生成部は、ランダムフォレストに基づいて、前記注目ピクセルから、前記RGB画像組を構成する画像ごとに異なる量でオフセットさせた周辺ピクセルにおける輝度値の差、または、前記デプス画像において、前記注目ピクセルから異なる量でオフセットさせた周辺ピクセルにおけるデプス値の差に応じて、前記注目ピクセルが属する部分領域の集合のノード分割を繰り返し行うことで、1以上の決定木を生成し、当該各決定木のリーフノードに、正解データとして、前記3次元座標を対応付けることで、姿勢推定モデルを生成してよい。

【0009】

輝度分布及びデプス分布と、オブジェクトの3次元座標との関係を直接数式化することは困難である。したがって、この構成のように、ランダムフォレストに基づいて、注目ピクセル周辺におけるRGB画像間の輝度差及び/またはデプス値の差に基づいて、注目ピクセルが属する部分領域集合のノード分割を繰り返し行うことで、輝度分布及び/デプス値分布からオブジェクトの姿勢を推定するためのモデルを簡易に生成することができる。なお、部分領域とは、画像を所定の大きさの領域に分割した場合における、1つの単位領域を意味する。

【0010】

本発明の姿勢推定モデル生成装置において、前記姿勢推定モデル生成部は、前記デプス画像内のあるピクセルにおけるデプス値を $d(p_i)$ 、カメラによって撮像された画像のあるピクセルにおけるRGB輝度を $I(p_i, \quad)$ オフセット量を \quad_1 、 \quad_2 として、

10

20

30

40

【数 1】

$$f^{da-d}(\theta, p_i) = d\left(p_i + \frac{\omega_1}{d_i}\right) - d\left(p_i + \frac{\omega_2}{d_i}\right)$$

【数 2】

$$f^{da-agb}(\theta, p_i, \beta) = I^\beta\left(p_i + \frac{\omega_1}{d_i}, \gamma_1\right) - I^\beta\left(p_i + \frac{\omega_2}{d_i}, \gamma_2\right),$$

$$\gamma \in \{R, G, B\}, \beta \in (\text{camera1}, \text{camera2}, \dots, \text{camera } n)$$

10

d_i : 注目ピクセルにおけるデプス値

1, 2 : R G B 輝度

: カメラ ID

で表される $f^{da-d}(\theta, p_i)$ または $f^{da-agb}(\theta, p_i, \beta)$ と、分割閾値 θ との比較により、情報利得が最も大きくなるように、前記ノード分割を行ってよい。

【0011】

この構成のように、注目ピクセル周辺における R G B 画像間の輝度差または 2 ピクセル間のデプス差と分割閾値 θ との比較によって、ノード分割後の情報利得が最も大きくなるように、ノード分割を繰り返し行うことで、輝度分布及びデプス分布から姿勢を推定するためのモデルを簡易に生成することができる。

20

【0012】

本発明の姿勢推定モデル生成装置において、前記姿勢推定モデル生成部は、各前記決定木のリーフノードにおいて、前記部分領域を識別する ID と、前記正解データとを対応付け、前記部分領域を識別する ID は、前記デプス画像生成部にて生成されたデプス画像における前記部分領域に付与された ID を用いてよい。

【0013】

部分領域の集合をノード分割していくことで、決定木のリーフノードには、まず、各部分領域が対応付けられることになる。これに正解データとしての 3 次元座標を対応付けることで、姿勢を推定するためのモデルを生成することができる。この構成によれば、各リーフノードに部分領域を対応付ける際、デプス画像における部分領域の ID を用いればよく、R G B 画像における部分領域に ID を付与する必要はない。したがって、簡易な構成で、姿勢推定モデルを生成することができる。

30

【0014】

本発明の姿勢推定モデル生成装置において、前記姿勢推定モデル生成部は、各前記決定木のリーフノードにおいて、前記部分領域を識別する ID と、前記正解データとを対応付け、前記部分領域を識別する ID は、各前記 R G B 画像における前記部分領域に付与された ID と、前記デプス画像生成部にて生成されたデプス画像における前記部分領域に付与された ID に対して、所定の重みづけ係数が設定された所定の算出式に基づいて決定されてよい。

【0015】

R G B 画像と仮想デプス画像とは、相互に撮像角度が異なるため、同じ位置の部分領域に写るオブジェクトの領域が異なる。この構成によれば、各リーフノードに部分領域を対応付ける際、このような画像のずれも考慮して部分領域 ID が決定されるので、精度よく姿勢推定モデルを生成することができる。

40

【0016】

本発明の姿勢推定装置は、平行等位に設置された複数の撮像デバイスによって撮像された R G B 画像組を取得する R G B 画像取得部と、前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記 R G B 画像組に対応するデプス画像を生成するデプス画像生成部と、前記 R G B 画像取得部にて取得された R G B 画像組における、同一の注目ピクセルの周辺の輝度分布及びノまたは、前記デプス画像生成

50

部にて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の3次元座標との関係を学習することで、前記対象物の複数の前記3次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデル生成部と、所定の推定モデルに基づいて、前記RGB画像組から、前記仮想的なデプス撮像装置によってデプス画像を撮像した場合におけるデプス値を推定するデプス推定部と、前記姿勢推定モデル生成部にて生成された姿勢推定モデルを用いて、前記RGB画像組及び推定された前記デプス値に基づいて、前記対象物の姿勢推定を行う姿勢推定部とを備えた構成を有している。

【0017】

この構成によれば、ステレオカメラ等、平行等位に設置された複数の撮像デバイスによって撮像された左右の画像組における注目ピクセル周辺のRGB輝度分布及び/または仮想的なデプス画像におけるデプス値分布から、ワークの姿勢を推定するモデルを生成し、姿勢推定を行うことができる。したがって、パッシブな構成で、姿勢推定を行うことができる。

10

【0018】

本発明の姿勢推定装置において、前記デプス推定部は、前記対象物が、前記姿勢推定モデルの生成において学習対象となっていない場合には、推定デプス値をゼロとして出力してよい。

【0019】

この構成によれば、推定デプス値ゼロの領域は、オブジェクトが存在しない領域と扱うことができ、例えば、推定デプス値ゼロの領域をマスクすることにより、姿勢推定対象領域から除外することができる。したがって、かかる構成により、高速かつ精度よく姿勢推定を行うことができる。

20

【0020】

本発明の姿勢推定モデル生成方法は、平行等位に設置された複数の撮像デバイスによって撮像されたRGB画像組を取得するステップと、前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記RGB画像組に対応するデプス画像を生成するステップと、前記RGB画像組を取得するステップにて取得されたRGB画像組における、同一の注目ピクセルの周辺の輝度分布及び/または、前記デプス画像を生成するステップにて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の3次元座標との関係を学習することで、前記対象物の複数の前記3次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデルを生成するステップとを備える。

30

【0021】

本発明のプログラムは、姿勢推定モデルを生成するためのプログラムであって、コンピュータに、平行等位に設置された複数の撮像デバイスによって撮像されたRGB画像組を取得するステップと、前記複数の撮像デバイスの中央に位置する仮想的なデプス撮像装置によって仮想的に撮像される、前記RGB画像組に対応するデプス画像を生成するデプス画像を生成するステップと、前記RGB画像組を取得するステップにて取得されたRGB画像組における、同一の注目ピクセルの周辺の輝度分布及び/または、前記デプス画像を生成するステップにて生成されたデプス画像内の当該注目ピクセル周辺のデプス値分布と、対象物の3次元座標との関係を学習することで、前記対象物の複数の前記3次元座標に基づいて、当該対象物の姿勢を推定するための姿勢推定モデルを生成するステップとを実行させる。

40

【発明の効果】

【0022】

本発明によれば、RGBカメラを用いて、ワークの姿勢推定を精度よく行うためのモデルを生成することができる。

【図面の簡単な説明】

【0023】

【図1】本発明の実施の形態における姿勢推定装置の構成を示すブロック図

50

【図2】本発明の実施の形態におけるデプス推定モデル生成装置の構成を示すブロック図

【図3】本発明の実施の形態におけるデプス推定モデル生成装置の動作フロー図

【図4】(a)~(c) 教師画像の一例を示す図

【図5】ノード分割を模式的に説明するための図

【図6】本発明の実施の形態におけるデプス推定モデル生成装置によって生成された回帰木の一例を示す図

【図7】本発明の実施の形態におけるデプス推定処理のフロー図

【図8】本発明の実施の形態における姿勢推定モデル生成処理のフロー図

【図9】本発明の実施の形態における姿勢推定モデル生成処理によって生成された決定木の構造を説明するための図

10

【図10】本発明の実施の形態における姿勢推定モデル生成処理によって生成された決定木の一例を示す図

【図11】本発明の実施の形態における姿勢推定処理のフロー図

【発明を実施するための形態】

【0024】

以下、本発明の実施の形態のデプス推定装置について、図面を参照しながら説明する。

【0025】

図1は、本発明の実施の形態の姿勢推定装置の構成を示す図である。図1に示すように、本発明の実施の形態の姿勢推定装置1は、RGB画像取得部10と、デプス画像生成部11と、ピクセル情報取得部12と、姿勢推定モデル生成部13と、デプス推定部14と、姿勢推定部15と、記憶部16とを備える。記憶部16は、学習データ記憶部17と、デプス推定モデル記憶部18と、姿勢推定モデル記憶部19とを備える。

20

【0026】

(姿勢推定装置の構成)

RGB画像取得部10は、ステレオカメラによって、様々なワーク(姿勢推定対象物)を、様々な距離、角度から撮像された左右のRGB画像を取得し、学習データ記憶部17に出力する。デプス画像生成部11は、仮想中央カメラによって撮像された場合のデプス画像をコンピュータグラフィックによって生成し、左右のRGB画像と対応付けて学習データ記憶部17に出力する。ここで、仮想中央カメラとは、RGB画像取得部10にて取得されたRGB画像を撮像したステレオカメラの左右のカメラの中央に、仮想的に設置されたカメラを意味する。すなわち、デプス画像生成部11にて生成されるデプス画像は、ステレオカメラの左右カメラ及び仮想的な中央カメラによって、同時に撮影を行った場合において、当該仮想的中央カメラから取得されるデプス画像に対応する。したがって、デプス画像生成部11は、RGB画像取得部にて取得した左右のRGB画像のペアの分だけ、デプス画像を生成する。

30

【0027】

ピクセル情報取得部12は、デプス画像生成部11にて生成されたデプス画像における、各ピクセル情報を取得し、RGB画像の各ペアと対応付けて、学習データ記憶部17に出力する。本実施の形態において、ピクセル情報とは、各ピクセルに対応付けられた、オブジェクト識別データ(ID等)及び各オブジェクトにおける位置を示す3次元座標データまたはオブジェクト外部の背景領域であることを示すデータである。なお、本実施の形態において、ピクセル情報は、ユーザにより入力されたデータを取得する。

40

【0028】

姿勢推定モデル生成部13は、RGB画像と、ピクセル情報と、仮想的デプス画像を用いて、姿勢推定モデルを生成し、姿勢推定モデル記憶部19に出力する。本実施の形態においては、左右RGB画像の輝度分布及びデプス値分布と、オブジェクトにおける位置情報との関係を学習し、RGB画像から、対象ワークの姿勢を推定することができるモデルを生成する。後述のように、本実施の形態において、姿勢推定モデルは、ランダムフォレストに基づいて生成される、仮想中央カメラから見たワークの3次元座標を出力する複数の決定木として生成される。

50

【 0 0 2 9 】

デプス推定部 1 4 は、デプス推定モデル記憶部 1 8 に記憶される推定モデルを用いて、上記の仮想中央カメラによって画像を撮像した場合における、当該画像のあるピクセルにおけるデプス値を推定する。後述のように、本実施の形態においては、デプス値の推定は、RGB 画像取得部 1 0 にて取得された左右の RGB 画像ペアを用いて行われる。

【 0 0 3 0 】

姿勢推定部 1 5 は、姿勢推定モデル生成部 1 3 にて生成され、姿勢推定モデル記憶部 1 9 に記憶されるモデルと、デプス推定部 1 4 にて算出された推定デプス値を用いて、ステレオカメラから取得した左右の RGB 画像について、仮想中央カメラから見たワークの姿勢推定を行う。具体的には、ランダムに選択した 3 つのピクセルを姿勢推定モデルに当てはめて、3 つのピクセルの 3 次元座標を求め、それら 3 ピクセルの位置関係に基づいて、1 つの姿勢候補とする。同様に、3 ピクセルのランダムサンプリングを行って、複数の姿勢候補を求め、最適化処理に基づき、1 つの姿勢候補を選択し、姿勢推定結果として出力する。

10

【 0 0 3 1 】

なお、図 1 に示す姿勢推定装置 1 は、CPU、RAM、ROM、HDD等を備えたコンピュータにより実現される。CPUがROMに記憶されたプログラムを読み出して実行することにより、姿勢推定モデル生成部 1 3 等の機能が実現される。このような姿勢推定装置 1 を実現するためのプログラムも本発明の範囲に含まれる。

【 0 0 3 2 】

また、上述のように、RGB 画像取得部 1 0 と、デプス画像生成部 1 1 と、ピクセル情報取得部 1 2 と、姿勢推定モデル生成部 1 3 と、学習データ記憶部 1 7 との協働により、姿勢推定モデルが生成される。したがって、これらの構成により、姿勢推定モデル生成装置としての機能が実現される。

20

【 0 0 3 3 】

(デプス推定モデル生成装置の構成)

ここで、本実施の形態における姿勢推定装置 1 のデプス推定部 1 4 にて用いられる、デプス推定モデルを生成する、デプス推定モデル生成装置の構成について説明する。図 2 は、デプス推定モデル生成装置 2 の構成を示すブロック図である。図 2 に示すように、デプス推定モデル生成装置 2 は、RGB 画像取得部 2 0 と、デプス画像生成部 2 1 と、デプス推定モデル生成部 2 2 と、記憶部 2 3 とを備える。記憶部 2 3 は、学習データ記憶部 2 4 と、デプス推定モデル記憶部 2 5 とを備える。

30

【 0 0 3 4 】

RGB 画像取得部 2 0 は、ステレオカメラにて撮像された左右の RGB 画像を取得し、学習データ記憶部 2 4 へ出力する。デプス画像生成部 2 1 は、RGB 画像取得部 2 0 にて取得された左右 1 組の RGB 画像に対応する、仮想カメラのデプス画像をコンピュータグラフィックスによって生成し、学習データ記憶部 2 4 へ出力する。ここでも、仮想カメラとは、RGB 画像取得部 2 0 にて取得された RGB 画像を撮像したステレオカメラの左右のカメラの中央に、仮想的に設置されたカメラを意味する。

【 0 0 3 5 】

デプス推定モデル生成部 2 2 は、学習データ記憶部 2 4 に記憶される、RGB 画像及びデプス画像生成部 2 1 にて生成されたデプス画像を用いて、デプス推定モデルを生成する。被写体をステレオカメラによって、距離を様々に変えて撮像した場合、カメラからの距離(デプス)と左右カメラ画像の輝度変化との間には、確率的な関数関係が存在する。本実施の形態においては、デプスと左右画像とのこのような関係性に着目し、左右画像データからデプス値へ変換する確率的関係の逆関数を、デプス推定モデルとして学習によって取得する。

40

【 0 0 3 6 】

(デプス推定モデル生成フロー)

まずは、デプス推定モデル記憶部 1 8 に記憶されるデプス推定モデルを生成する、デプ

50

ス推定モデル生成装置2の動作フローについて説明する。図3は、本実施の形態におけるデプス推定モデル生成装置2の動作フロー図である。デプス推定モデル生成装置2は、まず、RGB画像を取得し、学習データ記憶部24に記憶する(ステップS1)。ステップS1にて取得されるRGB画像は、ステレオカメラによって、視線の方向及び/または距離を様々に変えて撮像された左右のペア画像である。続いて、デプス推定モデル生成装置2は、ステップS1にて取得したRGB画像の各左右画像のペアについて、仮想カメラのデプス画像を生成し、各画像ペアに対応付けて、学習データ記憶部16に記憶する(ステップS2)。図4(a)、(b)は、ステップS1にて取得されたRGB画像の一例、図4(c)は、ステップS2にて生成されたデプス画像の一例を示す図である。これらのRGB画像のペア及びそれに対応するデプス画像が、1組の教師データとなる。

10

【0037】

図3に戻り、デプス推定モデル生成装置10は、各教師データの組について、注目領域(ROI)のサンプリングを行う(ステップS3)。以下では、サンプリングされた注目領域の中心のピクセルを「注目ピクセル」と呼ぶ。デプス推定モデル生成装置2は、注目ピクセルについて、左画像について ω_1 、右画像について ω_2 だけオフセットし(ステップS4)、オフセット後の左右画像のRGB輝度の差に基づいて、デプス値集合のノード分割を行う(ステップS5)。具体的には、各ノードにおいて、次式で表されるオフセット後の左右画像のRGB輝度の差 $f(\omega, \gamma, p_i)$

【数3】

$$f(\omega, \gamma, p_i) = I^L(p_i + \omega_1, \gamma_1) - I^R(p_i + \omega_2, \gamma_2), \gamma \in \{R, G, B\}$$

20

について、分割閾値を σ とし、分割後のデプス値の分散 $E(j)$

【数4】

$$E(j) = \sum_n (\sigma_n), n \in \{\leftarrow, \rightarrow\}$$

が最小となるように、上式のパラメータを決定して、ノード分割を行う。本実施の形態では、 i) RGBの組み合わせ(ω_1 及び ω_2 として、どのRGB輝度値を、単独でまたはいくつ組み合わせで用いるか)、 i) オフセット量 ω_1 、 ω_2 、 i) 分割閾値 σ が、学習により決定するパラメータとなる。なお、 i) RGBの組み合わせについては、例えば、右画像はR輝度値のみ、左画像はGとB輝度値のみ、等、左画像と右画像とで、異なる画素の輝度値を用いてもよい。

30

【0038】

図5は、ステップS4及びS5における、ノード分割を模式的に説明するための図である。今、あるノードJにおいて、図4(c)のデプス画像の注目ピクセル p_i におけるデプス値 d_i を2つの子ノード(ノードK、ノードL)のうちのいずれかに分類する例を考える。図5のRGB画像の左画像において、オフセット後のピクセル($p_i + \omega_1$)のR/G/Bの少なくとも1つの輝度値(2つ以上の輝度値を用いる場合は、その合計値)が、上式における $I^L(p_i + \omega_1, \gamma_1)$ であり、また、RGB画像の右画像において、オフセット後のピクセル($p_i + \omega_2$)のR/G/Bの少なくとも1つの輝度値(2つ以上の輝度値を用いる場合は、その合計値)が、上式における $I^R(p_i + \omega_2, \gamma_2)$ である。本実施の形態においては、これらの2つの輝度値の差が入力値となる。この入力値は、図5に模式的に示すように、同じ注目領域内の注目ピクセル p_i 周辺の2つのピクセル間の輝度差を意味する。

40

【0039】

p_i におけるデプス値 d_i が、ノードJからノードKに分類されたとすると、ノードKでは、注目ピクセル p_i からのオフセット量を変えて、 p_i 周辺の2つのピクセル間の輝度差に基づいて、デプス値 d_i を2つの子ノードのいずれに分類するかが決定される。このように、1回のノード分割でデプス値と対応付けられるのは、左右画像の2つのピクセル間の輝度差である。しかし、図5から理解されるように、ステップS4及びS5の処理を繰

50

り返してノードを順次分割していくことで、左右画像の同じ注目領域内の輝度分布と、中心ピクセルである注目ピクセルにおける仮想カメラから見たデプスとの関係を学習していくことになる。本願の発明者は、推定対象となる被写体にテクスチャが少ない場合であっても、左右画像の同じ注目領域内の画像の輝度分布は、対象物までの距離に応じて変化することを見出した。したがって、本実施の形態のように、ある注目ピクセルについて、オフセット量を変化させながら、繰り返し左右画像の輝度差とデプス値との対応付けを行う学習によって、推定対象にテクスチャが少ない場合でも、精度よくデプス値を推定することができるモデルを生成することができる。

【0040】

再び図3に戻り、デプス値推定モデル生成装置10は、ステップS6にて、所定の収束条件を満足したと判断されるまで、ステップS4及びS5の処理を繰り返す。なお、所定の収束条件としては、例えば、ノードが所定の深さに達することや、ノード分割後のデプス値の分散値の減少幅が所定の閾値以下になること、等を設定することができる。

10

【0041】

所定の収束条件を満たすと判断された場合には（ステップS6にてYes）、1つの回帰木の生成を終了し、所定数の回帰木が生成されるまで（ステップS7にてYes）、注目領域のサンプリング（ステップS3）に戻り、回帰木の生成を繰り返す。

【0042】

図6は、以上の処理によって生成された1つの回帰木の一例を示す図である。図6に示すように、生成された回帰木の各リーフには、デプス値のヒストグラムが対応付けられることになる。

20

【0043】

（デプス値推定フロー）

次に、本発明の実施の形態における姿勢推定装置1において、姿勢推定モデルを用いた姿勢推定を行う際に用いられる、デプス値の推定フローについて説明する。本実施の形態において、デプス値の推定は、上述の処理によって生成された、デプス推定モデルを用いて行われる。

【0044】

図7は、本実施の形態の姿勢推定装置1におけるデプス値推定処理の流れを示す図である。姿勢推定装置1は、まず、推定対象となるステレオ画像の入力を受け付ける（ステップS11）。そして、入力された左右画像の各ピクセルについて、複数の回帰木である推定モデルを適用し、（ステップS12）、デプス確率値の合計を算出する（ステップS13）。姿勢推定装置1は、さらに、デプス方向及び空間方向に平滑化フィルタ処理を行い、デプス出力値とする（ステップS14）。

30

【0045】

なお、ステップS14におけるフィルタ処理には、種々のアルゴリズムをベースとする線形フィルタまたは非線形フィルタを用いてよい。その際、例えば、空間方向にはバイラテラルフィルタ等のように、エッジを保存するような非線形フィルタを適用することもできる。

【0046】

また、本実施の形態では、デプス推定モデルの生成において学習したオブジェクト以外の物が存在する領域については、デプス値をゼロとして出力する。

40

【0047】

（姿勢推定モデル生成フロー）

次に、本発明の実施の形態における、姿勢推定モデル生成フローについて説明する。本実施の形態において、姿勢推定モデルの生成は、上述の処理にて推定されたデプス値を用いて行われる。

【0048】

図8は、本実施の形態における、姿勢推定モデルの生成処理の流れを示す図である。まず、RGB画像取得部10はRGB画像を取得し（ステップS201）、デプス画像生成

50

部 1 1 はデプス画像を生成する (ステップ S 2 0 2)。また、ピクセル情報取得部 1 2 は、デプス画像の各ピクセルについて、ピクセル情報を取得する (ステップ S 2 0 3)。次に、姿勢推定モデル生成部 1 3 は、デプス画像において、対象ワークが映っている領域を所定の大きさの部分領域に分割し、各部分領域にラベルを付与する (ステップ S 2 0 4)。ここで、ラベルとは、各部分領域の ID を意味する。

【 0 0 4 9 】

次に、姿勢推定モデル生成部 1 3 は、注目領域のサンプリングを行う (ステップ S 2 0 5)。以下では、この注目領域の中心ピクセルを「注目ピクセル」と呼ぶ。姿勢推定モデル生成部 1 3 は、注目ピクセルについて、オフセット後の左右画像の RGB 輝度の差 (注目ピクセル周辺の輝度差特徴) を算出し、また、デプス画像における注目ピクセルを異なる 2 ピクセルにオフセットした後の、デプス差 (注目ピクセル周辺のデプス差特徴) を算出する (ステップ S 2 0 6)。なお、輝度差特徴及びデプス差特徴を算出する際のオフセット量は、デプス画像の注目ピクセルにおけるデプス値で規格化した値を用いる。同じピクセル分だけオフセットさせても、カメラからオブジェクトまでの距離 (デプス) に応じて、オフセット後のピクセルが対応するオブジェクトの位置は異なる。オフセット量をデプス値で規格化することで、モデルを用いて姿勢推定を行う際に、カメラからオブジェクトまでの距離を学習時と同一にする必要はない。注目ピクセル周辺の輝度差特徴 f^{da-agb} (θ, p_i) は、具体的には、

【数 5】

$$f^{da-agb}(\theta, p_i) = I^L\left(p_i + \frac{\omega_1}{d_i}, \gamma_1\right) - I^R\left(p_i + \frac{\omega_2}{d_i}, \gamma_2\right), \gamma \in \{R, G, B\}$$

であり、注目ピクセル周辺のデプス差特徴は、

【数 6】

$$f^{da-d}(\theta, p_i) = d\left(p_i + \frac{\omega_1}{d_i}\right) - d\left(p_i + \frac{\omega_2}{d_i}\right)$$

である。

【 0 0 5 0 】

次に、ステップ S 2 0 6 にて算出された注目ピクセル周辺の輝度差特徴及びデプス差特徴と、分割閾値 τ とを比較し、ノード分割を行う (ステップ S 2 0 7)。具体的には、輝度差特徴とデプス差特徴のうち、分割閾値 τ の減算を行い、分割による情報利得がより大きくなる方を選択する。その上で、分割後の情報利得が最大となるように、上式 (選択した、輝度差特徴またはデプス差特徴のいずれか) のパラメータを決定して、ノード分割を行う (ステップ S 2 0 7)。オブジェクトの各パーツは、オブジェクトを撮像した画像において、輝度分布とデプス分布に特徴が現れるが、オブジェクトのパーツによって、輝度分布よりもデプス分布に、より大きく特徴が現れる場合と、デプス分布よりも輝度分布に、より大きく特徴が現れる場合がある。本実施の形態は、この点に着目し、入力値をより少ない回数で (つまり、高速に) 分類できるようにノード分割を行うものである。本実施の形態では、 i) RGB の組み合わせ (ω_1 及び ω_2 として、どの RGB 輝度値を、単独でまたはいくつ組み合わせで用いるか)、 i) オフセット量 ω_1 、 ω_2 、 i) 分割閾値 τ が、学習により決定するパラメータとなる。

【 0 0 5 1 】

ステップ S 2 0 8 にて、所定の収束条件を満足したと判断されるまで、ステップ S 2 0 6 及び S 2 0 7 の処理を繰り返す。なお、所定の収束条件としては、例えば、ノードが所定の深さに達することや、ノード分割前後の情報利得の利得差が所定の閾値以下になること、等を設定することができる。

【 0 0 5 2 】

所定の収束条件を満たすと判断された場合には (ステップ S 2 0 8 にて Yes)、1 つの回帰木の生成を終了し、所定数の回帰木が生成されるまで (ステップ S 2 0 9 にて Yes

10

20

30

40

50

s)、注目領域のサンプリング(ステップS205)に戻り、回帰木の生成を繰り返す。

【0053】

ステップS209までの処理で、図9に模式的に示すような、決定木の構造が生成される。この段階で、各リーフノードに対応付けられているのは、注目ピクセルが属する部分領域のラベルである。本実施の形態では、3次元座標を出力するためのモデルを生成することから、ステップS209に続いて、各リーフノードに、正解データとしての3次元座標の対応付けを行う。

【0054】

具体的には、まず、姿勢推定モデル生成部13は、ピクセルのランダムサンプリングを行う(ステップS210)。次に、姿勢推定モデル生成部13は、学習データ記憶部17
10
に記憶されるピクセル情報のうち、ステップS210で選択したピクセルに対応するピクセル情報を取得し、各リーフノードにピクセル情報を対応付ける(ステップS211)。具体的には、ステップS210にて選択したピクセルについて、ステップS209までの処理で生成された木構造をたどり、到達したリーフノードに、取得したピクセル情報を対応付ける。全てのリーフノードにピクセル情報を対応付けるまで(ステップS212にてYes)、ステップS210及びS211の処理を繰り返し、姿勢推定モデルを生成する。

【0055】

図10は、生成された1つの決定木の一例を示す図である。上述のように、ピクセル情報とは、各ピクセルが属するオブジェクトのID及び当該オブジェクトにおける3次元座
20
標データまたは背景に属することを示すデータである。したがって、1つのリーフノードに複数の3次元座標が割り当てられ得る。本実施の形態では、ミーンシフト等のクラスタリング手法を用いて、代表座標値を割り当てる。図10において、 $p(c|l^j)$ は、 j 番目の決定木におけるリーフノード l に分類されるピクセルが、オブジェクト c に属する確率、 $p(bg|l^j)$ は、同じリーフノードに分類されるピクセルが、背景に属する確率を示す。また、 $y_c(l^j)$ は、同じリーフノードに分類されるピクセルの、オブジェクト c における推定3次元座標値である。

【0056】

(姿勢推定フロー)

図11は、本実施の形態における、姿勢推定処理の流れを示すフロー図である。まず、
30
姿勢推定装置1は、推定対象となるワークを撮像したステレオRGB画像の入力を受け付ける(ステップS31)。次に、デプス推定部14は、デプス推定モデルを用いて、ステップS31にて取得したRGB画像に基づき、デプス値の推定を行う(ステップS32)。続いて、姿勢推定部15は、ステップS31にて取得したRGB画像及びステップS32にて算出された推定デプス値に基づき、姿勢推定モデルを用いて、オブジェクトの識別と、オブジェクトにおける3次元座標値を求める(ステップS33)。そして、姿勢推定部15は、最適化処理を行い、ワーク姿勢を決定する(ステップS34)。

【0057】

最適化処理は、オブジェクトの3Dモデルを生成したうえで、3次元座標と対応する3
40
Dモデルの3点をランダムサンプリングし、この3点からワーク姿勢候補を推定する。そして、 $i)$ レンダリング画像とのデプス差、 $ii)$ レンダリング画像との3次元座標の差、 $iii)$ オブジェクト存在確率、の3つの値の和をエネルギー(コスト)として、これを最小にする姿勢を求める。

【0058】

ところで、上述のように、本実施の形態において、デプス推定部14は、デプス推定モデルの生成の際に学習したオブジェクト以外のワークについては、デプス値ゼロを出力する。デプス値がゼロであるということは、そこにオブジェクトが存在しないことと同じとみてよい。このことを利用して、本実施の形態では、 $iii)$ クラス存在確率について、デプス値ゼロの領域は、クラス存在確率をゼロとする。また、上記の最適化処理において、3点をランダムサンプリングする際に、デプス値ゼロの領域からランダムサンプリング
50

が行われないようにする。これにより、不要な姿勢候補を予め除外することができ、また、エネルギー算出の負荷を軽減させることができる。すなわち、上述のデプス推定モデル生成装置にて生成されたモデルを用いて算出されたデプス推定値を姿勢推定に用いることで、高速に、精度よく、姿勢推定を行うことができる。

【0059】

以上、説明したように、上記の実施の形態の姿勢推定モデル生成装置は、ランダムフォレストに基づいて、注目ピクセル周辺におけるRGB画像間の輝度差及び/またはデプス画像における2ピクセル間のデプス差に基づいて、ラベル集合のノード分割を繰り返し行い、決定木の構造を生成する。そして、各リーフノードに、正解データとしての3次元座標とオブジェクトID（または、オブジェクトの存在しない背景領域）を対応付けることで、3次元座標を推定するためのモデルを生成し、推定された3次元座標に基づいてワークの姿勢推定を行う。したがって、RGBステレオカメラを用いて、パッシブな構成でワークの姿勢推定を行うことができるモデルを生成することができる。

10

【0060】

なお、上記の実施の形態では、デプス画像を生成し、該デプス画像に基づいて算出されたデプス値と、RGB画像の輝度データとに基づいて、姿勢推定モデルを生成する場合について説明したが、デプス画像の生成は行わずに、代わりに、デプス推定部にて、デプス推定モデルを用いて算出される推定デプス値を用いてもよい。

【0061】

また、上記の実施の形態では、生成されたデプス画像における部分領域のラベルIDを用いてノード分割を行う場合について説明したが、代わりに、RGB左右画像を同様に部分領域に分割して、部分領域のラベルIDを用いてもよい。また、デプス画像における部分領域のラベルIDに加えて、左右画像における部分領域のIDを用いてもよく、この場合、それぞれのIDに所定の重みづけをした上でラベルIDを決定してもよい。

20

【0062】

さらに、上記の実施の形態では、ステレオカメラによって撮影された2つの左右画像を輝度分布を求める教師データとして利用する場合について説明したが、本発明の範囲はこれに限られない。例えば、2つ以上の撮像デバイスによって撮像された、2つ以上の左右画像をセットとしてもよい。この場合、分割閾値と比較される輝度差特徴は、そのうちの1つまたは2つのカメラによって撮像された1つまたは2つの画像間の輝度値の差

30

【数7】

$$f^{da-agb}(\theta, p_i, \beta) = I^\beta \left(p_i + \frac{\omega_1}{d_i}, \gamma_1 \right) - I^\beta \left(p_i + \frac{\omega_2}{d_i}, \gamma_2 \right),$$

$$\gamma \in \{R, G, B\}, \beta \in (\text{camera1}, \text{camera2}, \dots, \text{camera } n)$$

としてもよい。また、分割閾値と比較する輝度差特徴の算出式は、上式に限られない。例えば、

【数8】

$$f^{da-agb}(\theta, p_i, \beta) = I^\beta \left(p_i + \frac{\omega_1}{d_i}, \gamma_1 \right) - I^\beta \left(p_i + \frac{\omega_2}{d_i}, \gamma_2 \right) + I^\beta \left(p_i + \frac{\omega_3}{d_i}, \gamma_3 \right),$$

40

$$\gamma \in \{R, G, B\}, \beta \in (\text{camera1}, \text{camera2}, \dots, \text{camera } n)$$

等、2つ以上の画像を用いて、各画像のオフセット後の輝度値を用いた所定の演算結果を分割閾値と比較してもよい。

【0063】

また、上記の実施の形態では、ステレオカメラの中央に仮想カメラを設置した場合のデプス画像を生成する場合について説明したが、仮想カメラの設置位置は、中央でなくてもよいし、仮想カメラは1つに限られない。例えば、RGB撮像デバイスが、等位平行に4つ設置されている場合、左端の撮像デバイスおよび右端の撮像デバイスによって取得したRGB画像に対し、左から2番目、3番目の撮像デバイスから見たデプス画像を生成して

50

もよい。この場合、輝度差特徴の変形例として説明したように、任意の仮想カメラを組み合わせてデプス差特徴を算出してもよく、また、2つ以上のデプス画像のデプス値を用いた所定の演算結果をデプス差特徴としてもよい。

【産業上の利用可能性】

【0064】

本発明は、RGBカメラを用いて、ワークの姿勢推定を精度よく行うことができる姿勢推定を行うことができるモデルを生成することができるという効果を有し、姿勢推定モデル生成装置等として有用である。

【符号の説明】

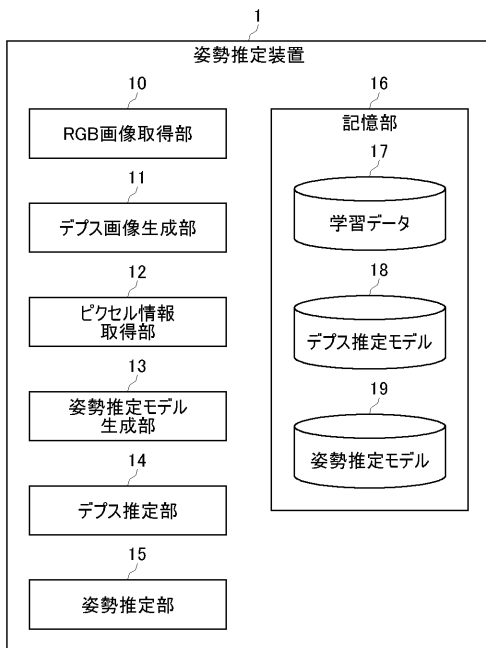
【0065】

- 1 姿勢推定装置
- 10 RGB画像取得部
- 11 デプス画像生成部
- 12 ピクセル情報取得部
- 13 姿勢推定モデル生成部
- 14 デプス推定部
- 15 姿勢推定部
- 16 記憶部
- 17 学習データ記憶部
- 18 デプス推定モデル記憶部
- 19 姿勢推定モデル記憶部

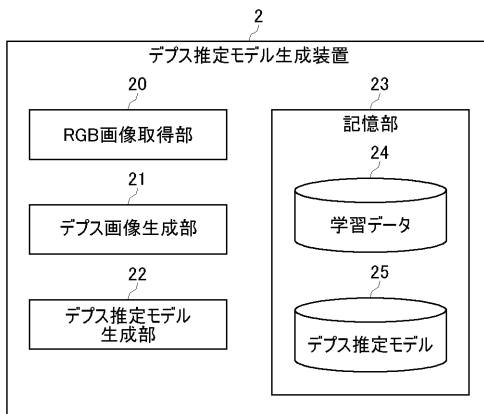
10

20

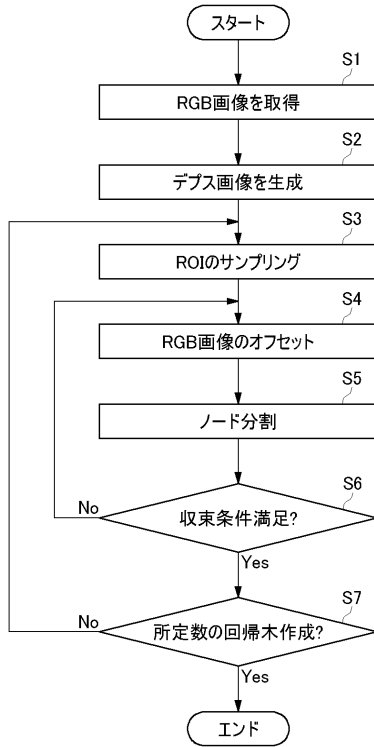
【図1】



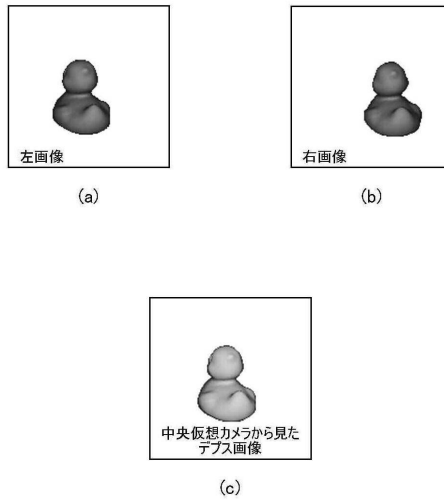
【図2】



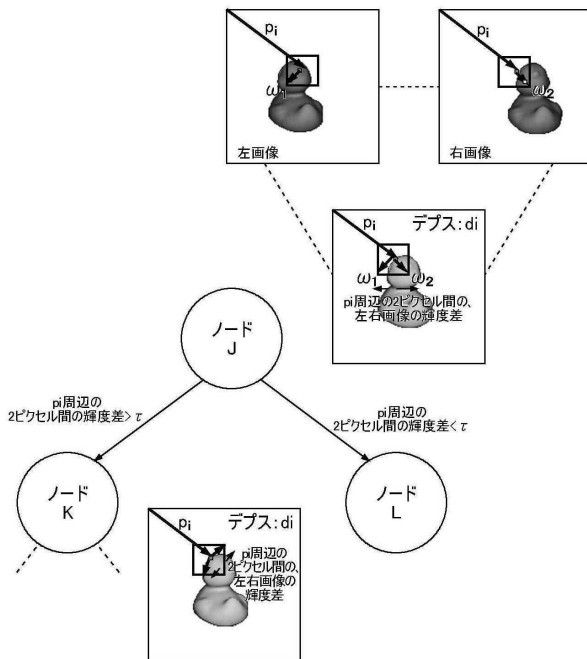
【図3】



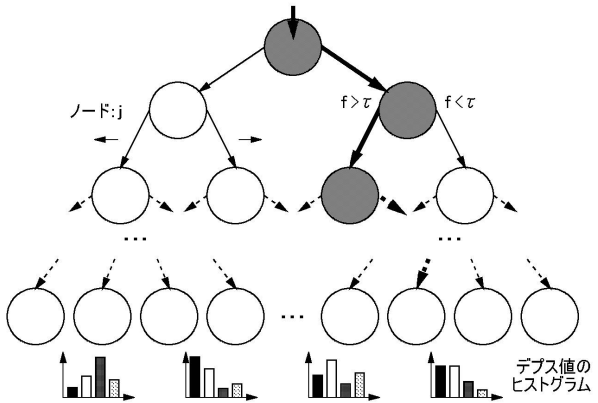
【図4】



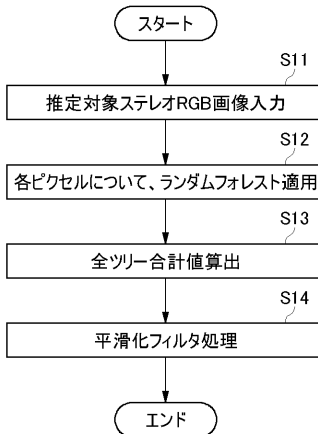
【図5】



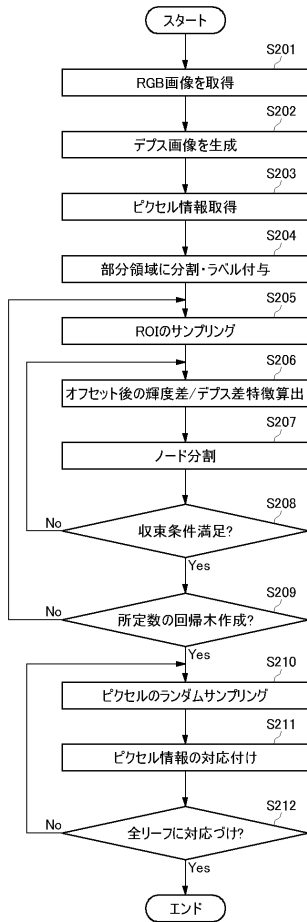
【図6】



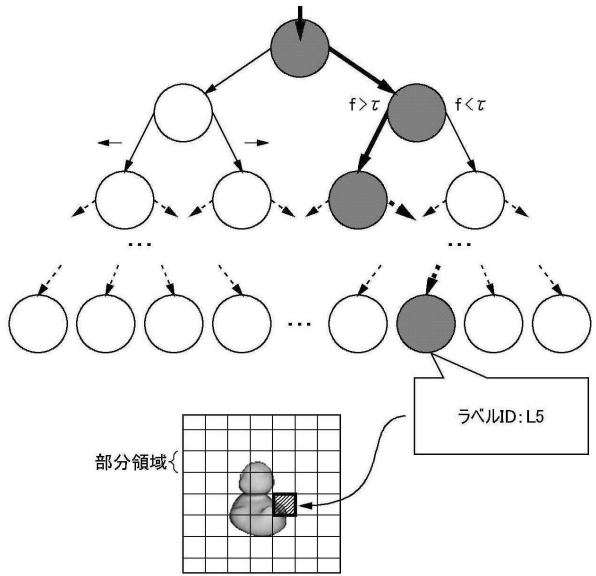
【図7】



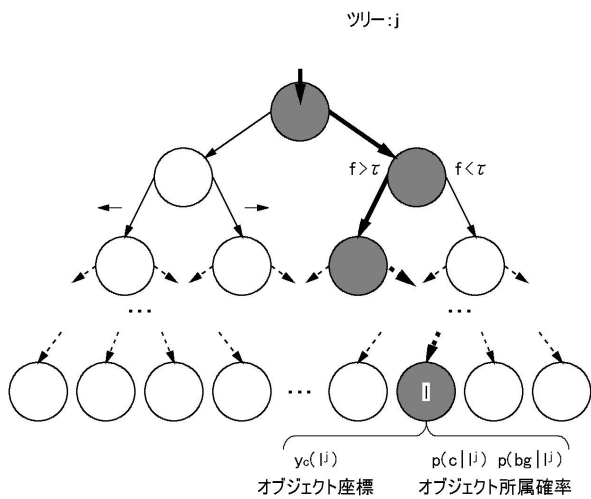
【図8】



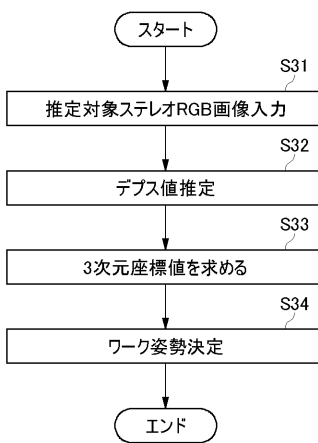
【図9】



【図10】



【図11】



フロントページの続き

(72)発明者 鈴木 幸一郎

東京都渋谷区渋谷二丁目15番1号 渋谷クロスタワー28F 株式会社デンソーアイティ-ラボ
ラトリ内

審査官 村松 貴士

(56)参考文献 特開2014-157509(JP,A)

特開2001-266128(JP,A)

特開2013-125529(JP,A)

特表2007-518970(JP,A)

Jamie Shotton et al., "Scene Coordinate Regression Forests for Camera Relocalization
in RGB-D Images", 2013 IEEE Conference on Computer Vision and Pattern Recognition, 米
国, IEEE, 2013年, pp.2930-2937

(58)調査した分野(Int.Cl., DB名)

G06T 1/00 - 7/90