

空間型 AR のための距離情報を用いた 基本形状の物体識別と位置姿勢推定

八馬遼^{†1} 斎藤英雄^{†2}

概要：本稿では、空間型 AR のための基本形状の物体識別と位置姿勢推定を行う手法を提案する。空間型 AR を行うためには投影対象物体を識別し位置姿勢推定を行うことが重要となる。しかし、物体の3次元形状に特徴が少ないため基本形状物体を識別するのは困難となる。そこで我々は法線ベクトルマップの勾配に着目し平面領域を抽出し、各投影対象物体の平面情報を既知とすることで各物体を識別する。我々の手法を用いることで、各対象物体が様々におかれた状況でも各物体を識別し位置姿勢推定を行うことができるようになった。さらに、各物体にプロジェクタを用いて画像を投影した。

キーワード：空間型 AR, 位置姿勢推定, 物体識別, RGB-D カメラ

Recognition and Pose Estimation of Primitive Shapes From Depth Images for Spatial Augmented Reality

RYO HACHIUMA^{†1} HIDEO SAITO^{†2}

1. はじめに

近年、実世界の物体に CG をプロジェクタで投影する空間型拡張現実感(Augmented Reality :AR)の研究が盛んにおこなわれている。AR では体験者がビデオスルー型ヘッドマウントディスプレイをつけることで実世界と仮想空間を同時に見ることができると、臨場感のある体験をすることが可能となる。しかし、ひとりひとりに対してデバイスを用意する必要があり不特定多数の人が同時に体験することは困難となる。そこで、空間型 AR を用いて実世界の物体にプロジェクタによって CG を投影することで不特定多数が同時に観察することを可能とした。空間型 AR は投影させる画像を変化させることで物体の見え方を変えることができるため、製品の試作デザインを容易に行うことができる[1]。

空間型 AR を行うためには、対象物体を識別し物体の位置姿勢を推定しなくてはならない。距離情報から物体の識別と位置姿勢推定を行うのは容易ではないが、10年以上研究されてきた。物体識別を行うために、3次元特徴量を記述する手法がある。例えば、Rusu らは Point Feature Histograms(PFH)特徴量を定義した[2]。PFH 特徴量は2つの法線ベクトルの関係を3次元点群に属するすべての点について計算した多次元特徴量であり、Tombari らは Signatures of Histograms of Orientations(SHOT)特徴量を定義した[3]。

複数の投影対象物体に対して空間型 AR を行うためには、撮影シーンから各物体の位置姿勢を推定しなくてはなら

ない。平面に物体が置かれたシーンで各物体の位置姿勢推定を行う Aldoma らの手法[4]があげられる。対象物体の3次元モデルを保持しそれぞれのモデルが持つ Clustered ViewPoint Feature Histogram(CVFH)特徴量とオンラインで検出された対象物体の3次元点群が持つ CVFH 特徴量を用いてマッチングし大まかな位置合わせを行った後、Iterative Closest Point(ICP)アルゴリズムによって高精度な位置合わせを行う。しかし、ICP アルゴリズムでは最近傍点を対応点として変換を行うため、多面体のように形状に特徴が少ない物体では誤対応が多くなり位置合わせをすることができない。また、物体の形状に幾何学的特徴が少ない物体の位置姿勢を求める手法として佐野らの手法[5]があげられる。佐野らの手法では RGB-D カメラから得られた Depth 情報を用いて立方体を追跡し、動的な空間型 AR を実現した。しかし、投影対象物体は立方体に限定されており、また3個程度の物体がお互いに接触しない状況においてのみ対応していた。よって、立方体に加えて角錐、直方体、円筒といった形状の違う物体が存在し、お互いに接触したり積み上げられたりしている状況には対応していなかった。

そこで、本稿ではこのように異なる形状の投影対象物体が複数存在し、かつそれらを積み上げられたり並べられたりしてもその物体群に対して空間型 AR 表示できるための手法を提案する。

2. 提案手法

空間型 AR ではプロジェクタにより物体にテクスチャを

^{†1} 慶應義塾大学
Keio University

投影し、テクスチャを変化させることがあるため投影対象物体の色情報を用いて位置姿勢推定を行っていくことは困難となる。そこで本手法では、RGB-D カメラから得られる Depth 情報と各投影対象物体（立方体，直方体，四角錐，円筒）を構成する面の法線情報と辺の大きさを既知とすることで，テクスチャのない投影対象物体の位置姿勢を推定する。

本手法は3つの流れからなっている。1つ目は RGB-D カメラから得られる3次元点群からの平面抽出，2つ目は平面ごとに分割された3次元点群からの物体識別，3つ目はそれぞれの物体での位置姿勢推定である。本手法の流れを図1に示す。

2.1 平面抽出

平面抽出の代表的な方法は得られた3次元点群から最小二乗法や RANSAC アルゴリズムを用いてすべての点群から一番誤差が少なるような点のまとまりを平面とする方法であるが，すべての点群から平面を抽出すると物体のエッジにおいてエッジ付近の点が間違っただけの平面に含まれてしまうことが多い。そこで，提案手法では，A. Ückermann らの手法[6]の着想を用いて局所的に法線ベクトルの勾配を計算することで画像を生成し，その画像に対して領域分割を行うことで平面を抽出する。

2.2 ノイズ除去

Kinect v2 から得られる Depth 値のノイズを除去するために2つのフィルタを適用する。まずは生の Depth 値に対して3×3のメディアンフィルタを適用する。次に時系列方向に個々の画素で過去10フレームにおいて平均化フィルタを適用する。そして得られた Depth 値を3次元座標に変換する。

2.3 法線ベクトル勾配画像の作成

得られた法線ベクトルの勾配に着目するため，法線ベクトルの内積を算出する。注目点の8近傍の点を取り出し注目点と各点で法線ベクトルの内積を取る。そして8つの内積を平均化し，その内積に対してしきい値を設けることで2値化された法線ベクトルの勾配画像を生成する（図1中央）。この画像では，白い画素付近では法線ベクトルが同じ方向を向いており，黒い画素付近では法線ベクトルが異なった方向を向いていることを示している。

2.4 領域分割

本研究では，領域分割の手法として領域拡張法を用いる。領域拡張法では，ある画素に着目した時にその8近傍にしきい値以上の画素が存在すれば同じ領域とし，ラベリングを行う。そしてラベリングを行っていない画素について上記を繰り返すことで画像を平面による領域ごとに分割する。（図1右）今回は，しきい値として勾配が8.79(°)以上のものに対してラベリングを行った。



図1 平面抽出の流れ（左：法線ベクトルマップ，中央：法線ベクトル勾配画像，右：平面領域抽出画像）

2.5 平面の方程式

RANSAC アルゴリズムによる平面の方程式導出について説明する。図3にフローチャートを示す。RANSAC アルゴリズムとは，データ中の外れ値を除いてパラメータを推定する手法のことである。平面の方程式を求めるには最低でも3点が必要となる。

まずは，ある平面を構成するすべての点から無作為に3点を選び，その3点を用いて平面の方程式のパラメータを求める。そして，その平面の方程式と平面を構成するすべての点においてユークリッド距離を算出し，距離がしきい値以内の点の個数を求める。その個数をパラメータの投票数とする。以上を試行回数だけ繰り返し一番投票数の多かったパラメータを仮決定パラメータとする。仮決定パラメータを用いて再度点とのユークリッド距離を計算し，しきい値以内の点をインライアとする。そのインライアを用いて最小二乗法により最終的なパラメータを決定する。今回は距離のしきい値として0.005(5mm)，試行回数は600回を用いた。

2.6 物体識別

次に平面ごとに分割された3次元点群からどこにどの物体があるかを識別する。識別をするために得られた平面に対して3つの手法で平面がどの物体を構成する平面かということを求める。そのためにまずは，領域分割画像から一番領域を構成する要素数が多いものを床平面としておく。1つ目は平面の分散を用いることで円筒を識別し，2つ目は平面間の法線ベクトルの角度を用いて四角錐と立方体を識別し，最後に平面群のユークリッド距離を用いて立方体と直方体を識別する。

2.7 平面と非平面の識別

まず行うのは分散を用いた識別である。各平面 S を構成する点の法線の分散 V_S は各点の法線 \mathbf{n}_i ，法線ベクトルの平均 $\bar{\mathbf{n}}$ と各平面を構成する点の数 N_S を用いて次のように表せる。

$$V_S = \frac{1}{N_S} \sum_{\mathbf{n}_i \in S} (\bar{\mathbf{n}} - \mathbf{n}_i)^2$$

円筒などを構成する曲面は面を構成する点の法線がばらつくため分散も大きくなる。よって，その分散がしきい値 ε ($\varepsilon = 0.1$) 以上のものは円筒を構成する曲面であるとする。

2.8 平面間の角度による識別

次は、残りの多面体を分類する。多面体を識別するには近傍にある平面間の角度と平面と床平面との角度を用いる。

本研究で用いる投影対象物体は多面体であり1辺の長さが20cm以内であるため、各平面の重心を算出し2平面の重心間距離が20cm以内の2平面の組み合わせにおいてのみ平面の交わる角度を計算する。2平面 S_i, S_j の法線ベクトル $\mathbf{n}_i, \mathbf{n}_j$ が成す角度を θ とし、 θ によってその面の組がどの多面体かを判断する。 $90 - \mu < \theta < 90 + \mu$ のとき2平面 S_i, S_j は立方体であるとし、 $68 - \mu < \theta < 68 + \mu$ のとき四角錐であるとする。 $(\mu = 5.0)$

2.9 ユークリッド距離による識別

最後に、直方体と立方体を識別する。この識別は2.2.2において立方体であると仮定された面 S_i が4面以上である場合においてのみ行う。分類は次の評価式に基づいて行う。

$$d = \arg \max \left(\frac{1}{n(\mathbf{E})} \sum_{S_i \in \mathbf{E}} \mathbf{p}_i - \frac{1}{n(\mathbf{V} \cap \bar{\mathbf{E}})} \sum_{S_i \in \mathbf{V} \cap \bar{\mathbf{E}}} \mathbf{p}_i \right)^2$$

$\mathbf{V} = \{S_i | i = 1, 2, \dots\}$, \mathbf{E} を \mathbf{V} の中から無作為に r 個抽出された面の集合、 \mathbf{p}_i を面 S_i の重心とする。 $n(\mathbf{V}) = 6$ のとき $r = 3$, $n(\mathbf{V}) = 4, 5$ のとき $r = 2$ とする。

まず面群 \mathbf{V} を無作為に選んだ面群 \mathbf{E} と \mathbf{E} を除く面群に分割する。それぞれの面群を構成する面の重心から面群の重心を求め、面群同士の重心間のユークリッド距離 d を算出する。これを $n(\mathbf{V})C_r$ 回繰り返す、 d が最大となるような面の選び方で面群 \mathbf{V} を分割する。本研究で用いる立方体は一辺が20cm, 直方体は長辺が20cm, 短辺が10cmなので次のように e_1, e_2 を定義する。

$$e_1 = \frac{1}{n(\mathbf{E})C_2} \sum_{\substack{S_i \in \mathbf{E}, S_j \in \mathbf{E} \\ i \neq j}} (\mathbf{p}_i - \mathbf{p}_j)^2$$

$$e_2 = \frac{1}{n(\mathbf{V} \cap \bar{\mathbf{E}})C_2} \sum_{\substack{S_i \in \mathbf{V} \cap \bar{\mathbf{E}}, S_j \in \mathbf{V} \cap \bar{\mathbf{E}} \\ i \neq j}} (\mathbf{p}_i - \mathbf{p}_j)^2$$

e_1 は $S_i \in \mathbf{E}$ から無作為に抽出した2面の重心間距離の平均、 e_2 は $S_i \in \mathbf{V} \cap \bar{\mathbf{E}}$ から抽出した重心間距離の平均である。 $e_1 > e_2$ ならば面群 \mathbf{E} が立方体となる。

2.10 各物体の位置姿勢推定

各物体の位置姿勢を推定する。各物体が予め定義された座標系(モデル座標系)からどれだけ回転し並進したかを求め変換行列を導出する。多面体における回転行列と並進ベクトルは一意に求まるのに対し、曲面を含む立体では並進ベクトルは一意に求まるが回転行列は一意には求まらない。そこで、それぞれで位置姿勢推定を行う。

2.11 多面体の位置姿勢推定

まずは、多面体の回転行列を求める。本手法では定義さ

れたモデルと検出物体の点と点の対応関係を見ていくのではなく、その多面体を構成する面の法線に着目し回転行列を算出する。回転行列は 3×3 の行列であり導出には計3つのベクトルが必要となる。よって、多面体を構成する2面 S_1, S_2 の法線ベクトル $\mathbf{n}_{S_1}, \mathbf{n}_{S_2}$ と床平面 S_f の法線ベクトル \mathbf{n}_{S_f} を用いて回転行列を算出する。 \mathbf{n}_{S_i} は平面 S_i の法線ベクトルである。モデル座標における法線ベクトルを $(\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_f)$ とすると回転行列 \mathbf{R} を用いて次のような式になる。

$$(\mathbf{n}_{S_1}, \mathbf{n}_{S_2}, \mathbf{n}_{S_f}) = \mathbf{R} \cdot (\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_f)$$

上式を解くことで回転行列を求める。

次に、多面体の並進ベクトルを求める。並進ベクトルはモデル座標における原点がどこに移動したかを求める。原点を各多面体の2平面と床平面の計3平面の交点とする。しかし、図2のようにある物体の上に別の物体が乗っている場合には、3平面の交点を用いると図4のように赤い点が原点と認識されるため、正しく並進ベクトルを求められない。

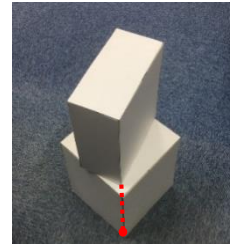


図2 原点の誤認識

よって、各物体を構成する面の重心から各物体の重心を求め、重心と床平面との距離から物体が床平面に接しているのか別の物体の上に乗っているのかを判定し、交点を求める際に用いる床平面を床平面の法線方向に平行移動させ再計算を行う。

2.12 円筒の位置姿勢推定

まずは、回転行列を求める。円筒の回転は円筒の軸方向に一意には定まらないため、2方向の回転成分を求める。回転行列を求めるために、多面体の時と同じく3つのベクトルを求める。一つ目として、円筒の軸方向と床平面の法線ベクトルが一致しているとし床平面の法線ベクトル \mathbf{n}_{S_f} , 2つ目としてその法線ベクトルと垂直に交わる $\mathbf{m}_{S_f} (\mathbf{n}_{S_f} \perp \mathbf{m}_{S_f})$, そして3つ目としてこの2つのベクトルの外積により求まる $\mathbf{l}_{S_f} (\mathbf{n}_{S_f} \times \mathbf{m}_{S_f})$. この3つのベクトルを用いて多面体の時と同じく回転行列を求める。

次に、並進ベクトルを求める。多面体の時と同じくある原点がどこに移動したかを決定する。原点を、円筒の軸と円筒の底面の交わる点とする。円筒の側面のある1点に着目した時に、その点をその法線ベクトルと反対方向に円筒の半径の長さだけ移動させると円筒の軸上に点がくる。以上を円筒の検出された側面の全点について行い、それらを平均し円筒の底面に移動させることで並進ベクトルを求める。

3. 実験

本稿では、実験を3つ行った。まず1つ目として、複数の投影対象物体が積み上げられたり重なったりして置かれていた場合においても各物体の位置姿勢推定が行えるかを実験した。2つ目として、各物体の位置姿勢推定を行った際に真値と比較し回転行列と並進ベクトルの誤差を調べた最後に、実際にプロジェクタを用いて物体に画像を投影することで物体識別が行えているかを確認した。実験環境は以下の通りである。

CPU: Intel Core i7-4790 3.60GHz,

RAM: 8.00GB,

RGB-D カメラ : Microsoft Kinect v2.0.

3.1 位置姿勢推定結果

提案手法を用いて、投影対象物体を多様に配置し各物体の位置姿勢推定が行えるかを実験した。また、位置姿勢推定結果を画像で確かめるためにモデル座標系で定義された基底ベクトル $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, $\mathbf{e}_3 = (0, 0, 1)$ を提案手法により求めた変換行列を用いて表示した。その結果を図3に示す。(\mathbf{e}_1 :青, \mathbf{e}_2 :赤, \mathbf{e}_3 :緑)

図3により、異なる形状の物体を床に置いたり物体を積み上げたりしても各物体の位置をそれぞれで推定できたことがわかる。提案手法では Depth 情報のみを用いて位置姿勢推定を行っているために、空間型 AR における位置姿勢推定を実現できた。また、提案手法の平均処理時間を処理内容ごとに表1に示す。

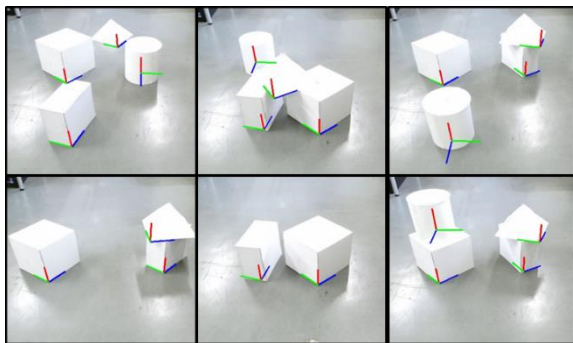


図3 位置姿勢推定結果

表1 提案手法の平均処理時間

処理内容	平均処理時間(s)
平面抽出	25.084
物体識別	0.024
位置推定	0.016
合計	25.124

最も時間がかかった処理は平面抽出の中でも領域分割であり処理時間は 25.084 (s) となった。よって、領域分割を高速化させることで処理時間の短縮が可能となる。

3.2 位置姿勢推定精度評価実験

位置姿勢推定の精度を評価するために回転行列と並進ベク

トルを真値と比較した。二つの回転行列の誤差を比較するために[7]にて定義されているリーマン距離を用いて比較を行った。また、ユークリッド距離を用いることで並進ベクトルの誤差とした。図4に各物体における位置姿勢推定の平均誤差をまとめる。

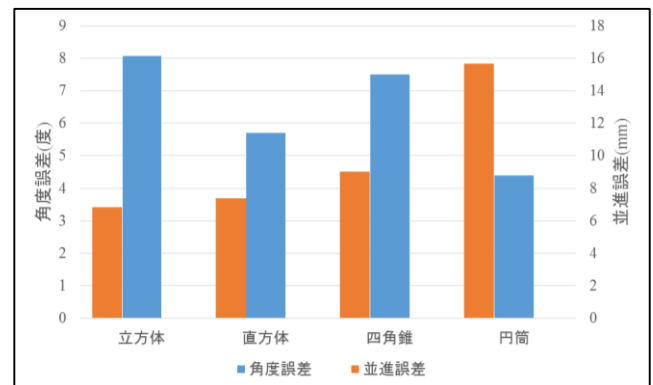


図4 位置姿勢推定誤差平均

上図より、並進誤差は円筒を除いて 1cm 以内、回転誤差はすべての対象物体で 10 度以内になったことがわかる。円筒の並進誤差だけ大きくなってしまった理由としては、多面体の並進ベクトルは面の交点を求めて算出したのに対し円筒の並進ベクトルは曲面の各点を用いて算出してしまったからであると考えられる。これにより円筒の並進誤差はおおよそ 1.4cm となってしまった。

3.3 物体識別結果

実験 3.1, 3.2 で位置姿勢推定結果の定性的評価、定量的評価と行ったが 3.3 では実際にプロジェクタを用いて対象物体の各面に画像を投影することで各物体を識別できたことを示す。その結果を図5にまとめる。

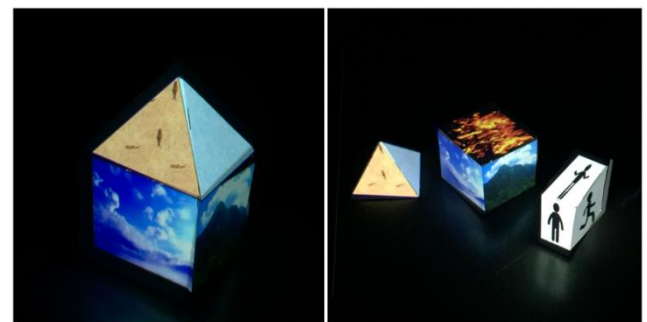


図5 物体識別結果

図5の左画像では、四角錐が立方体の上に乗っているシーンである。このようなシーンに対しても我々の手法を用いることで各物体を識別し位置姿勢推定を行えたことが図8によりわかる。

4. 考察

提案手法を用いると複数物体を上 (図5における y 軸方向) つみあげても各物体を識別し位置推定を行うことが

できた。しかし、例えば図6のように立方体と直方体が横方向に接しており多面体の側面2面が見えない場合は、位置姿勢を推定できないことが失敗例として挙げられる。

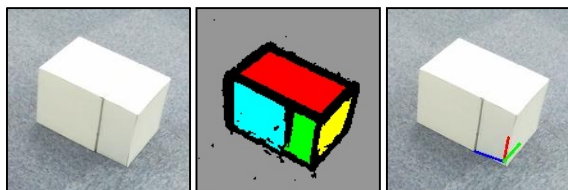


図6 位置姿勢推定の失敗例

(左:配置例, 中央:領域分割画像, 右:位置推定結果)

5. まとめ

本研究では RGB-D カメラの Depth 情報から得られる 3 次元点群を複数の平面に分割し、複数の平面を構成する点群の分散や平面同士の交わる角度によって複数の投影対象物体に分け、それぞれの投影対象物体において位置推定を行う手法を提案した。これを用いて複数の投影対象物体を識別し、空間型 AR のための位置推定を行うことができた。

今後の課題として、より精度を上げるために点群に対するノイズ除去を行いたい。また、投影対象物体が横方向に連結している場合でも識別し位置推定を行いたい。

謝辞

本研究の一部は、科学研究費 基盤研究(S) 24220004 の補助により行われた。

参考文献

- [1] C. Tonn, F. Petzold, O. Bimber, A. Grundhofer, and D. Donath. Spatial augmented reality for architecture—designing and planning with and within existing buildings. *International Journal of Architectural Computing*, 6(1):41–58, 2008.
- [2] Radu B. Rusu, et al. "Aligning point cloud views using persistent feature histograms." *Intelligent Robots and Systems, (IROS 2008)*, pp.3384-3391, 2008.
- [3] Federico Tombari, Samuele Salti, and Luigi Di Stefano. "Unique signatures of histograms for local surface description." *Computer Vision—ECCV2010*. Springer Berlin Heidelberg, pp.356-369, 2010.
- [4] Aldoma, A., Blodow, N., Gossow, D., Gedikli, S., Rusu, R.B., Vincze, M., Bradski, G.: CAD-Model Recognition and 6DOF Pose Estimation Using 3D Cues. In: 3DRR Workshop, ICCV (2011)
- [5] Sano, M. ; Matsumoto, K. ; Thomas, B.H. ; Saito, H., Rubix: Dynamic Spatial Augmented Reality by Extraction of Plane Regions with a RGB-D Camera. In: *Mixed and Augmented Reality (ISMAR)*, 2015 IEEE International Symposium
- [6] A. Ückermann, R. Haschke, H. Ritter, Real-Time 3D Segmentation of Cluttered Scenes for Robot Grasping, *Proc. Humanoids*, (2012)

- [7] M. Moakher, "Means and averaging in the group of rotations," in *SIAM J. Matrix Anal. Applicat.*, vol. 24, 2002, pp. 1–16.