

点群の空間占有率とマルチスケールシェルを用いた SHORT 特徴量の提案とその3次元物体認識への応用に関する研究

情報科学研究科 情報科学専攻 H11408M 武井翔一

指導教員：橋本学

1. はじめに

ロボットの知能化には、それに搭載されるビジョンシステムが重要であり、3次元物体認識に関する技術が研究されている。特にこの技術では、認識対象の物体モデルと入力データそれぞれに対して、キーポイントと呼ばれる3次元点を検出し、認識のための手がかりとなる特徴量を記述するモデルベースの手法が多く提案されている。この手法では、キーポイントおよび特徴量の表現力を高めるためには複雑な計算を必要とするため、計算コストとの適切なバランスの実現が課題である。

従来のキーポイント検出法 [1] は、物体の局所領域内の点群から形状評価のための統計量を算出し、特定形状の点を検出する。また、特徴量記述の従来法 [2] は、球状の局所領域に存在する多数の点を使用して、法線方向分布などを特徴量として記述する。これらの手法は、統計量や法線ベクトルの算出が必要であることと、領域内の多数の点を使用するため、計算コストが高いという問題がある。

そこで本研究では、従来手法の認識性能を下げることなく、高速化を図ったキーポイント検出および特徴量記述の新手法、SHORT (Shell Histograms and Occupancy from Radial Transform) 法を提案する。

2. SHORT: Shell Histograms and Occupancy from Radial Transform

2.1 キーポイント検出および特徴量記述のための要件と基本アイデア

本研究におけるキーポイント検出と特徴量記述が具備すべき要件を表1にまとめる。

表1 キーポイント検出と特徴量記述が具備すべき要件

	要件
キーポイント検出	1. キーポイントが凸形状である 2. 形状の評価が高速である
特徴量記述	1. 特徴量が姿勢不変である 2. 特徴量のアイデンティティが高い 3. 計算コストが低い

物体認識のタスクにおいては、認識対象の物体の姿勢は不明であることから、検出すべきキーポイントは、さまざまな視点から見えやすい部分、すなわち凸形状の点であると考えられる。そして、注目点が凸形状か否かの判定のためには、その点の局所形状の評価が必要であり、この処理の高速化がキーポイント検出の高速化につながる。したがって、キーポイント検出では、1. 凸形状の点を検出することと、2. 形状を高速に評価する必要がある。

また、位置姿勢が不明な対象物体を認識するための特徴量の設計では、1. 姿勢不変であることと、2. アイデンティティが高いことが重要である。一般的に、これらを得るためには、複雑な計算を必要とするため、計算コストが高くなりやすい。したがって、特徴量記述では、姿勢不変性と高いアイデンティティの取得だけでなく、さらに計算

コストの低減が求められる。

本研究では、キーポイント検出と特徴量記述処理に関するこれらの要件を満たす手法を提案する。このためのアイデアは、局所領域内の点数の割合（占有率）から形状を評価することと、設定する局所領域を球の最外殻（シェル）領域に限定することである。特に特徴量記述では、占有率と、異なる半径のシェル（マルチスケールシェル）領域を用いることによって、姿勢不変であり、さらに高いアイデンティティを持つ特徴量を高速に算出する。

2.2 占有率を用いたキーポイント検出

占有率とシェル領域を用いてキーポイントを高速に検出する。図1に占有率算出の模式図を示す。

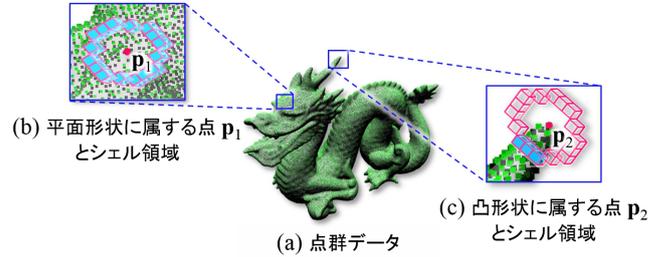


図1 占有率の算出

図1(a)は点群データであり、図1(b)(c)は(a)の局所領域を拡大した図である。ここでは、(b)の p_1 と(c)の p_2 は3次元点を意味し、それぞれ平面形状と凸形状に属する。また、 p_1 と p_2 の周辺のボクセル領域はシェル領域であり、表示の都合上、一部を示している。図1のように、形状によって点群の空間的な広がり異なるため、占有率も異なり、平面的な形状と比較すると凸形状は占有率が低い。この占有率 o_i は、3次元点群 $P = \{p_i \mid i = 1, 2, \dots, N\}$ の各点に対して、式(1)によって算出される。

$$o_i = \frac{1}{M} \sum_{s_j \in R_i} C(s_j) \quad (1)$$

s_j は o_i を算出するためのボクセル領域であり、 R_i は s_j の集合である。また、 M は領域の総数であり、 $C(s_j)$ は s_j に含まれる点数である。式(1)を用いて各点の占有率を算出し、占有率マップ $F = \{o_i \mid i = 1, 2, \dots, N_v\}$ を生成する。そして、本研究では凸形状をキーポイントとして優先的に検出するため、 F から o_i の極小値をキーポイント $P_k = \{p_k \mid k = 1, 2, \dots, K\}$ として検出する。このように、占有率の極小値を検出することは、物体全体から凸形状を検出することに相当するため、物体モデルと入力データ間の対応点が得られやすくなり、認識性能の向上に寄与する。

2.3 マルチスケールシェル領域を用いた特徴量記述

マルチスケールシェル領域内の点と、その占有率を用いて特徴量を高速に記述する。姿勢不変な特徴量を記述するために、まず主方向ベクトル d_k を式(2)、式(3)によって算出する。

$$\mathbf{g}_k = \frac{1}{L} \sum_{\mathbf{p}_j \in B_k} o_j (\mathbf{p}_j - \mathbf{p}_k) \quad (2)$$

$$\mathbf{d}_k = \frac{\mathbf{g}_k}{\|\mathbf{g}_k\|} \quad (3)$$

\mathbf{p}_j はシェル領域に存在する点を示し、 o_j は \mathbf{p}_j が持つ占有率である。また、 B_k は \mathbf{p}_j の集合であり、 L は領域内の点数である。式 (2) のように、占有率を重みとすることによって、占有率が低いキーポイントから占有率が高い点への勾配方向を考慮して主方向ベクトルを算出する。

次に、マルチスケールの領域内の点群とその占有率、主方向ベクトル \mathbf{d}_k を用いて特徴量を記述する。特徴量は、 \mathbf{d}_k と、キーポイントからマルチスケール領域内の点までの方向ベクトル \mathbf{x}_{r_j} の内積ヒストグラムによって構成される。図 2 に特徴量記述の模式図を示す。

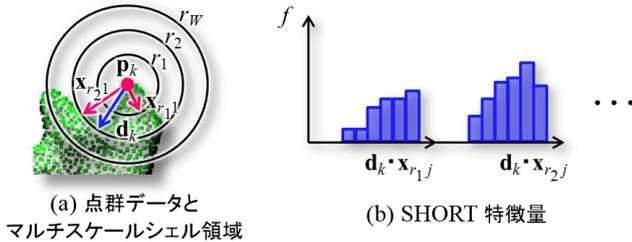


図 2 SHORT 特徴量の記述

図 2(a) は、物体の局所点群とマルチスケールのシェル領域の模式図を示している。また、図 2(b) は、SHORT 特徴量の模式図である。この特徴量記述では、まずスケールごとに内積ヒストグラム $\mathbf{f}_r = (f_{r1}, f_{r2}, \dots, f_{rh})$ を算出する。このとき、領域内の各点を持つ占有率を重みとして投票する。そして、各スケールの内積ヒストグラム \mathbf{f}_r を連結し、正規化することによって最終的な特徴量 $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_W)$ を記述する。 \mathbf{f} は異なるスケールのシェル領域から記述されるので、高いアイデンティティを持つことが期待される。また本研究では、物体モデルの特徴量の独自性が高くなるスケールのシェル領域のみを認識に用いることによって、照合の信頼性を向上しつつ、高速化を図る。

3. 実験と考察

文献 [2] で使用されたデータセット [3] から、次の 2 種類を用いて物体認識の性能を評価した。

- (1) Dataset A (シミュレーションデータ) :
Descriptor Matching Dataset 1 (Stanford)
- (2) Dataset B (実データ) :
Descriptor Matching Dataset 5 (Kinect)

Dataset A はシミュレーションデータであり、6 種類の物体モデルと 45 シーンの入力データで構成される。各入力データは、位置姿勢がランダムに変換された 3~5 種類の物体で構成され、ノイズが付加されている。また、Dataset B は実データであり、6 種類の物体モデルと 17 シーンの入力データで構成される。入力データは、Microsoft 社の Kinect [®] によって取得されたデータである。また、入力データには複数の物体が存在し、オクルージョンも発生している。

比較手法は、キーポイント検出の従来手法として高速手法の代表例である ISS キーポイント検出法 [1] と、SHOT 特徴量 [2] を用いた手法である。また、認識アルゴリズム

としては投票型のアプローチ [4] を提案手法と比較手法に適用した。なお、提案手法については、定数倍の大ききで設定したマルチスケールシェル領域を用いる手法を SHORT 法 (Without SMS) とし、物体モデルの特徴量の独自性が高くなるマルチスケールシェル領域を用いる手法を SHORT 法 (Using SMS) とする。各手法の処理時間 $T[\text{msec}]$ と認識成功率 $Pr[\%]$ の関係を表 2 に、提案手法による認識結果の例を図 3 に示す。

表 2 各手法の処理時間 $T[\text{msec}]$ と認識成功率 $Pr[\%]$

	データセット	$T[\text{msec}]$	$Pr[\%]$
ISS 法 [1] + SHOT 法 [2]	A	840.2	98.9
	B	4146.9	85.7
SHORT 法 (Without SMS)	A	220.8	99.4
	B	462.2	85.7
SHORT 法 (Using SMS)	A	211.7	100.0
	B	460.2	91.8

(CPU : Intel [®] CORE [™] i7-3770, メモリ : 16GB)

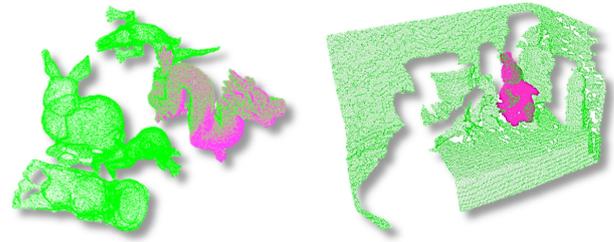


図 3 認識結果例

SHORT 法 (Using SMS) は、物体モデルの特徴量の独自性が高くなるスケールのシェル領域を用いるため、比較手法と比べて認識成功率が高く、さらに高速であった。Dataset B に対するキーポイント検出の処理時間としては、ISS 法は 299.0[msec]、提案手法は 20.0[msec] であり、特徴量記述の処理時間としては、SHOT 法は 1571.7[msec]、SHORT 法 (Using SMS) は 182.6[msec] であった。このように、実データに対しても提案手法は、比較手法よりも認識成功率が高く、処理速度は約 9 倍高速であることを確認した。

4. おわりに

本研究では、高速なキーポイント検出および特徴量記述の手法として SHORT 法を提案し、3 次元物体認識に応用した。実データを用いた認識性能の評価実験により、提案手法は、従来手法と比較して認識成功率が向上し、処理速度は約 9 倍と大幅に高速化できたことを確認した。

参考文献

- [1] Y. Zhong, Intrinsic Shape Signatures: A Shape Descriptor for 3D Object Recognition, ICCVW, pp.689-696, Kyoto, 2009/9.
- [2] S. Salti, et al., SHOT: Unique Signatures of Histograms for Surface and Texture Description, CVIU, Vol.125, pp.251-264, 2014/8.
- [3] Descriptor Matching Dataset, <http://vision.deis.unibo.it/research/78-cvlab/80-shot>
- [4] F. Tombari, et al., Object Recognition in 3D Scenes with Occlusions and Clutter by Hough Voting, PSIVT, pp.349-355, Nanyang Technological University, Singapore, 2010/11.