

幾何特徴量を用いた視点計画に基づく物体認識

○中里一幾 諸岡健一 辻徳生 倉爪亮(九州大学)

1. はじめに

生活支援を行うサービスロボットにとって、日用品を認識する機能は、重要且つ必要不可欠である。従来の物体認識の研究では、与えられた画像情報を使って、対象物体を特定している。一方、ロボットによる物体認識は、ロボットが対象物を直接扱うことができるため、ロボットの視線方向もしくは対象物の姿勢を変えながら、色々な方向から対象物を観測することが可能である。そこで、ロボットが物体認識に有効な情報を積極的に取得・認識する機能、すなわち能動的物体認識を備えることで、生活支援ロボットの実現性が高まることが期待される。

能動的物体認識を実現するためには、物体認識に有効な情報が得られる視点を自動的に計画する枠組みが必要である。物体形状を効率的に計測するための視点計画法は、これまで多くの研究が行われている [1, 2]。一方、物体認識のための視点計画に関する研究は少ない。その中で、従来の能動的物体認識法として、複数台の RGB カメラを物体の周囲に配置し、各カメラの視点位置・視線方向を決定するものがある [3, 4]。しかし、ロボットに取り付けるセンサを増やすと、高コスト化やセンサの配置によるロボットの稼働範囲の制限などが生じるため、センサ数は極力抑えるべきである。そこで、本研究では、ロボットには1台のカメラのみが備わっているとす。また、従来法では、既知の物体、すなわち、事前にデータが与えられている物体を認識することを目的としており、未知物体に対する認識精度は保証されていない。一方、生活支援ロボットが扱う日用品は、同一種類であっても様々な形状があり、あらゆる日用品のデータが事前に与えられているとは限らない。

そこで、本研究では、形状が未知の日用品に対し、ロボットが有する1台の距離センサを使って複数の方向から観測しながら、安定して認識する能動的物体認識法を提案する。提案手法は、Farshidiら [4] のカメラ配置法に基づく。この配置法では、対象物の曖昧さをエントロピーで定量化し、その曖昧さが減少するよう、物体の周囲に配置した2台の RGB カメラの視点位置・視線方向を決定する。しかし、上述のように、文献 [4] は既知物体を認識することを目的としているため、提案手法では、文献 [4] の手法を拡張し、距離センサから得られる形状幾何特徴量を用いて、認識に有効な視点位置を決定する。

2. 能動的物体認識法

提案する能動的物体認識の流れを図1に示す。

2.1 形状幾何特徴量

本研究では、形状幾何特徴量として、物体表面上の3次元点群から得られる Clustered Viewpoint Feature

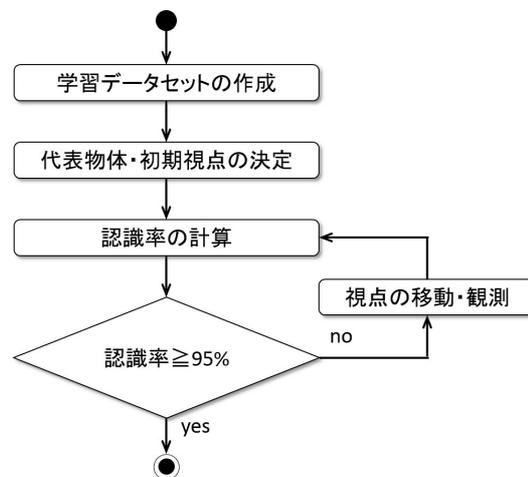


図1 認識終了までの流れ

Histogram(CV FH)[5]を用いる。距離センサで計測した物体表面の点群データが与えられると、CVFHの算出では、まず、各点ごとに、計測された表面の平均法線ベクトルに対する、その点での法線ベクトルの傾きを求める。そして、全点群データの傾きを数え上げたヒストグラムを作成し、このヒストグラムの度数を並べたベクトルが、CVFHである。CVFHの計算は、物体表面の曲率が高い箇所の点を予め取り除き、残った滑らかな面のみを用いる。これにより、センサノイズやオクルージョンに頑健な特徴記述を実現している。

2.2 学習データセットの作成

学習データセットの作成は、まず、1つのカテゴリに含まれる物体の3次元形状モデルを複数用意する。次に、各モデルに対して、図2に示すように、80個の視点から物体を計測したときの点群データを仮想的に生成し、各点群データごとにCVFHを求める。80個の視点は、以下の方法で決定する。まず、物体を覆う正20面体を設定し、正20面体の各三角パッチを4分割する処理を再帰的に2回繰り返す。その結果得られる正80面体の各パッチの重心を、視点として用いる。図2中、白点は視点を表し、黄線は視線方向を示す。本研究では、視線は、視点から物体の重心方向に向かうと仮定している。したがって、本手法の視点計画では、視点の位置のみを決定し、各学習データは、モデルIDと視点の組み合わせである。

モデルから得られるCVFHは、雑音がない理想的なデータであるが、実際のデータには雑音が含まれる。この問題を解決するために、Farshidiら [4] は、確率モデルを導入している。 k 番目の視点 $v_k(k=1,2,\dots,80)$ から、カテゴリ $c_i(i=1,2,\dots,N_c; N_c: \text{カテゴリ数})$ の日用品の d 次元特徴ベクトルCVFH g_{ik} が得られる確

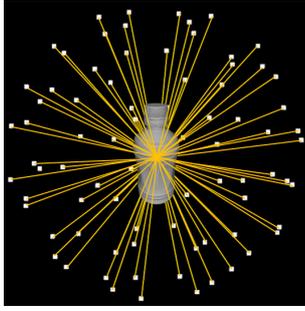
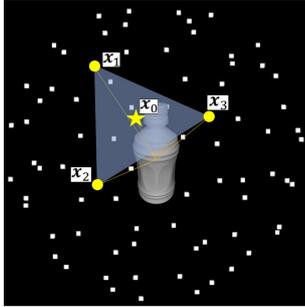


図2 物体の3次元形状モデルを観測する80視点



★:注目視点 ●:近傍視点

図3 分布を作成する4視点の例

率 \mathcal{N} を

$$\mathcal{N}(\mathbf{g}_{ik}, \boldsymbol{\mu}_{ik}, \sigma_{ik}^2) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma_{ik}} \exp\left\{-\frac{|\mathbf{g}_{ik} - \boldsymbol{\mu}_{ik}|^2}{2\sigma_{ik}^2}\right\} \quad (1)$$

とモデル化する。カテゴリ c_i に含まれるモデルを $M_{ij} (j=1, 2, \dots, N_m^{(i)})$, モデル数を $N_m^{(i)}$ とすると、図3に示すように、平均ベクトル $\boldsymbol{\mu}_{ik}$, 分散 σ_{ik}^2 は、注目視点 \mathbf{v}_k と、その3近傍の視点の計4視点のCVFHを使って求める:

$$\boldsymbol{\mu}_{ik} = \frac{1}{4N_m^{(i)}} \sum_{j=1}^{N_m^{(i)}} \sum_{\mathbf{v}_{k'} \in L_k} \mathbf{x}_{ijk'}; \quad (2)$$

$$\sigma_{ik}^2 = \frac{1}{4N_m^{(i)}} \sum_{j=1}^{N_m^{(i)}} \sum_{\mathbf{v}_{k'} \in L_k} |\mathbf{x}_{ijk'} - \boldsymbol{\mu}_{ik}|^2. \quad (3)$$

\mathbf{x}_{ijk} は、 k 番目の視点からモデル M_{ij} を観測した時、得られる特徴ベクトルである。また、 L_k は、 k 番目の視点と、その3近傍の視点の集合である。

3. CVFHを用いた物体認識

本節では、ある視点で観測された物体表面の点群データから、認識率を求める方法について述べる。

3.1 代表物体および初期視点の決定

初期視点から物体を観測し、CVFH $\tilde{\mathbf{g}}$ が得られた時、各カテゴリから、 $\tilde{\mathbf{g}}$ に類似した特徴量を持つ物体候補と、その特徴量が得られる視点を決定する。具体的には、カテゴリ c_i 内の全モデルのCVFHと $\tilde{\mathbf{g}}$ の距離を求め、その距離が小さい順に10個の学習データを選択

する。選択した10個の学習データと観測データの位置合わせを行い、位置合わせ誤差が最小となる学習データ、つまりモデルと視点の組み合わせを求める。選ばれたモデルと視点を、 c_i の代表モデル \tilde{M}_i , および初期視点 $\tilde{\mathbf{v}}_i^{(0)}$ とする。全カテゴリから、代表モデルと初期視点を決定し、これらの代表モデルと初期視点を使って視点計画を行う。

3.2 認識率の計算

観測を n 回行った時、それまでの視点 $\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(n-1)}$ と、その時に得られた特徴量 $\mathbf{g}^{(0)}, \dots, \mathbf{g}^{(n-1)}$ は既知である。この事前情報を用いて、 $(n+1)$ 回目の観測で視点 $\mathbf{v}^{(n)}$ から特徴量 $\mathbf{g}^{(n)}$ が得られた時、それがカテゴリ $c^{(n)}$ である確率は、ベイズ推定により事後確率密度関数 $p_{post}(c^{(n)})$ で表される:

$$p_{post}(c^{(n)}) = \frac{p(\mathbf{g}^{(n)} | c^{(n)}, \mathbf{v}^{(n)}) p_{pri}(c^{(n)})}{p(\mathbf{g}^{(n)} | \mathbf{v}^{(n)}, \mathbf{g}^{(n-1)}, \mathbf{v}^{(n-1)}, \dots, \mathbf{g}^{(0)}, \mathbf{v}^{(0)})} \quad (4)$$

$$P_{post}(c^{(n)}) = P(c^{(n)} | \mathbf{g}^{(n)}, \mathbf{v}^{(n)}, \mathbf{g}^{(n-1)}, \mathbf{v}^{(n-1)}, \dots, \mathbf{g}^{(0)}, \mathbf{v}^{(0)}) \quad (5)$$

$$P_{pri}(c^{(n)}) = P(c^{(n)} | \mathbf{g}^{(n-1)}, \mathbf{v}^{(n-1)}, \dots, \mathbf{g}^{(0)}, \mathbf{v}^{(0)}) \quad (6)$$

本手法では、式(5)の事後確率を、物体の認識率とみなす。ここで、式(4)の分母の確率は、

$$p(\mathbf{g}^n | \mathbf{v}^{(n)}, \mathbf{g}^{(n-1)}, \mathbf{v}^{(n-1)}, \dots, \mathbf{g}^{(0)}, \mathbf{v}^{(0)}) = \sum_{i=1}^{N_c} p(\mathbf{g}^{(n)} | c_i^{(n)}, \mathbf{v}^{(n)}) P_{pri}(c_i^{(n)}) \quad (7)$$

と表される。また、事前確率 $P_{pri}(c^{(n)})$ は、1ステップ前の事後確率と等しいので、

$$P_{pri}(c^{(n)}) = P_{post}(c^{(n-1)}) \quad (8)$$

で得られる。式(4)~(8)を用いると、カテゴリ c_i の日用品を視点 \mathbf{v}_k から観測した時、特徴ベクトル \mathbf{g}_{ik} が得られる尤度 $p(\mathbf{g}_k | s, \mathbf{v}_k)$ は、次の式で表される。

$$p(\mathbf{g}_{ik} | c_i, \mathbf{v}_k) = \mathcal{N}(\mathbf{g}_{ik}, \boldsymbol{\mu}_{ik}, \sigma_{ik}^2) \quad (9)$$

4. 視点の決定

カテゴリ c_i において、式(5)の認識率が高い場合は、観測している物体が c_i に属する可能性が高いことを示す。一方、複数のカテゴリ間の認識率に差がない場合、これらの認識が曖昧であることを示す。そこで、これらの物体を区別するために、以下の手順で最適な視点を決定する。

まず、3.1節で選んだ各カテゴリ c_i の代表モデル \tilde{M}_i を使って、次の視点候補を選択する。80個の各視点 \mathbf{v}_k から得られる特徴ベクトル $\mathbf{g}^{(n)}$ を、 \tilde{M}_i のCVFH $\tilde{\mathbf{x}}_{ik}$ として式(4)に代入し、仮の事後確率を各視点ごとに

求める. 次に, $g^{(n)}$ が得られた時の情報量のエントロピーを

$$H(P(X)) = - \sum_X P(X) \log P(X) \quad (10)$$

と定義する. このエントロピーは, 物体の不確かさを表し, この値が小さいほど物体を特定できていることを示す. そこで, エントロピーが減少する, つまり, 物体を認識しやすい情報が得られる視点を選ぶ. 具体的には, \tilde{x}_{ik} によるエントロピーの減少量を, 式 (11) の相互情報量によって算出する:

$$I(c^{(n)}; \tilde{x}_{ik}) = H(P_{pri}(c^{(n)})) - H(P_{post}(c^{(n)})) \quad (11)$$

ここで, 事前確率 $P_{pri}(c^{(n)})$ は, 式 (6) より得る. 式 (11) を用いて, 移動可能な全ての視点の候補について相互情報量を計算し, 相互情報量が最大となる視点, すなわち, 観測後にいずれかのカテゴリの曖昧さを最小にする視点を, 次に移動する視点として決定する:

$$v_{k^*} = \arg \max_{k^*} \frac{1}{N_c} \sum_{i=1}^{N_c} I(c^{(n)}; \tilde{x}_{ik}) \quad (12)$$

データ取得-物体認識-視点計画の処理ループは, 任意のカテゴリの認識率 (式 (5)) が閾値以上となるまで繰り返す. 但し, 既に選ばれた視点は, 視点候補から除く.

5. 実験

提案手法の有効性を検証するため, 2つのカテゴリを分類する実験を行った. 表 1 に, 本実験に用いた5つの2カテゴリの組み合わせパターンと, 各パターンで学習に用いたモデル数を示す. また, 図 4 は, パターン 1 の学習用モデルを示す. 学習データに含まれていないテスト用モデルを用いて, 本手法で視点を選択した場合と, ランダムで選択した方法 (以後, ランダム法) で比較実験を行った. 実験は次の手順で行った. まず, 80 視点分のテスト用点群データから1つランダムに選択し, これを初期観測データとする. 次に, 式 (5) を用いてカテゴリ毎の認識率を計算する. この時, 一つのカテゴリの認識率が 0.95 以上となれば終了する. そうでなければ, 2.4 節で述べた方法で次の視点を決定し, その視点から物体を観測する. 本実験では, 各パターンで 30 個の初期観測データを選択し, 認識終了までの平均視点移動回数と最大視点移動回数を求めた.

実験結果の一例として, 図 5 は, パターン 1 における本手法とランダム法でそれぞれ視点を選択した場合のボトルの認識率の変化を示す. 図 6 は, 図 5 の本手法による結果において, 各ステップで入力された点群データを示す. 図 5 で見られるように, ランダム法と比べ, 提案手法は, 少ない観測回数で, ボトルの認識率が向上している. 表 2 は, 本手法と, ランダム法で視点を決定した場合の平均視点移動回数と, 最大移動回数を示す. 表 2 から明らかなように, ランダム法と比較すると, 本手法の視点移動回数および最大移動回数共に少なく, 効率的な視点移動ができています.

表 1 実験で使用したデータセット

データパターン	カテゴリ (テスト用)	カテゴリ (学習用)	学習モデル数
1	ボトル	ボトル	7
		カップ	8
2	リンゴ	リンゴ	7
		バナナ	6
3	牛乳パック	ボウル	8
		牛乳パック	7
4	ミニカー	ミニカー	7
		牛乳パック	8
5	ハンマー	ボトル	8
		ハンマー	7

ボトル 7種類



カップ 8種類



図 4 パターン 1 の学習用モデル

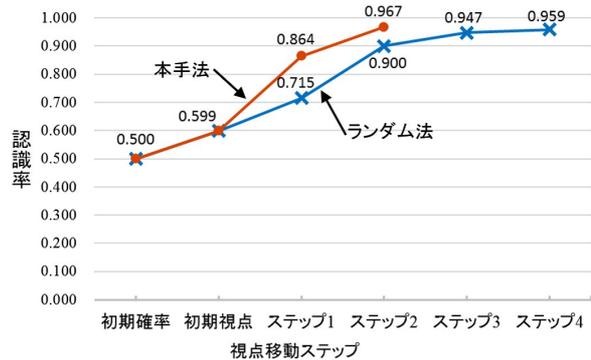


図 5 パターン 1 における本手法とランダム法の場合のボトルの認識率の変化

6. まとめ

本稿では, 距離センサから得られる形状幾何特徴量 CVFH を用いて, 能動的物体認識を実現するための視点計画法を提案した. 提案手法は, 対象物の不確かさをエントロピーで定量化し, そのエントロピーが最も減少する視点を選択する. シミュレーション実験を行い, 提案手法と, ランダム法で視点を選択した場合の結果を比較し, 提案手法により, 効率的な視点選択を行えていることが確認できた. 今後は, 実機での実験を行っていく予定であるが, 更なるノイズの増加やロボットハンドによるオクルージョンの問題に対処していく必要がある. これに対し, 新たな特徴量の選択や, その特徴量の性質にあった認識率の計算方法などの検討を行う必要がある.



図6 図5の本手法による結果において、各ステップで入力されたテスト用点群データ

表2 本手法およびランダム法で視点を決定した場合の平均視点移動回数と最大視点移動回数

データパターン	平均視点移動回数		最大視点移動回数	
	本手法	ランダム法	本手法	ランダム法
1	3.13	7.9	6	16
2	1.17	1.23	2	3
3	2.33	4.75	4	7
4	1.77	2.79	3	6
5	1.23	5.13	3	7

参考文献

- [1] 諸岡健一, 查紅彬, 長谷川勉: “複数の距離画像の統合による3次元物体モデル生成のための視点計画”, 電子情報通信学会 D-II, Vol.J82-D-II, No.3, pp.371-381, 1999.
- [2] Sumantra Dutta Roy, Santanu Chaudhury, Subhashis Banerjee: “Active recognition through next view planning: a survey”, Pattern Recognition, Volume 37, Issue 3, March 2004, Pages 429-446
- [3] 松ヶ野修功, 高橋友和, 井出一郎, 日加田慶人, 村瀬洋: “複数視点物体認識における視点配置の決定法” 電子情報通信学会, pp.19-pp.24, 2004.
- [4] F. Farshidi, S.Sirouspour, T.Kirubarajan: “Robust sequential view planning for object recognition using multiple cameras”, Image and Vision Computing 27 pp.1072-1082, 2009.
- [5] Aitor Aldoma, Markus Vincze: “CAD-Model Recognition and 6DOF Pose Estimation Using 3D Cues”, 2011 IEEE International Conference on Computer Vision Workshops, 978-1-4673-0063-6, 2011.