能動的物体認識のための視点計画法

〇中里 一幾, 諸岡 健一, 辻 徳生, 倉爪 亮 (九州大学)

Viewpoint Planning for Active Object Recognition

O Kazuki NAKAZATO, Ken'ichi MOROOKA, Tokuo TSUJI, and Ryo KURAZUME (Kyuhu University)

Abstract: This paper presents a new viewpoint planning method for an active object recognition. Using 3D shape information, the proposed method determine the viewpoint to recognize a given object efficiently and robustly. The determination is formulated as an entropy minimization, and the next view is chosen so that the entropy is reduced when the information obtained from the view.

1. 緒言

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.ロボットは認識対象物を直接扱えるため,認識時にロボットの視線方向や対象物の姿勢を変えながら,様々な方向から対象物を観測することができる.この特徴を生かして,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機能を備えることで,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やすい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、単一ロボットに取り付けるセンサ数を増やすと、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.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. 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.

我々は、1台の距離センサを有するロボットを使って、形状が未知の日用品に対しても認識可能な、視点計画に基づく能動的物体認識法を提案した $^{5)}$. この手法では、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。2つのカテゴリの物体認識問題を扱った実験結果から、ランダムで視点を選ぶ方法と比べ、我々の手法が効率的に物体認識ができることを確認した。

本稿では、我々の手法を、3カテゴリ以上の物体認識問題に適用できるよう拡張した新たな能動的物体認識法を提案する.また、提案手法を用いた物体認識実験を行った結果について報告する.

2. 能動的物体認識法 5)

能動的物体認識の流れを、Fig.1 に示す。本手法では、形状幾何特徴量として、Clustered Viewpoint Feature Histogram (CVFH)⁶⁾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 は以下の手順で求める。まず、各点ごとに、全点群データでの平均法線ベクト

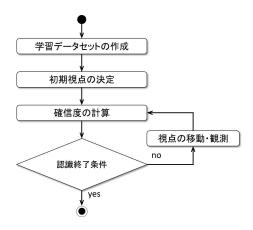


Fig.1 Our active object recognition

ルに対する、その点での法線ベクトルの傾きを求める、次に、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。

2.1 学習データセットの作成

1つの日用品のカテゴリごとに、複数の物体の3次元形状モデルを用意する。各モデルを80個の視点から計測したときの点群データを仮想的に生成し、視点ごとにCVFHを求める。80個の視点は、物体を覆う正80面体の各パッチの重心を用いる(Fig.2)。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、モデルの番号と視点の組み合わせからなる。

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可能性が高い。この問題を解決するために、本手法では、Farshidi らの確率モデル 4)を導入する。k 番目の視点 $v_k(k=1,2,\cdots,80)$ から、カテゴリ $c_i(i=1,2,\cdots,N_c;N_c;N_c:$ カテゴリ数)の日用品の D 次元 CVFH g_{ik} が得られる確率 N_a を

$$\mathcal{N}_{g}(g_{ik}, \mu_{ik}, \sigma_{ik}^{2}) = \frac{1}{(2\pi)^{\frac{D}{2}} \sigma_{ik}} \exp\left\{-\frac{|g_{ik} - \mu_{ik}|^{2}}{2\sigma_{ik}^{2}}\right\} (1)$$

とする. カテゴリ c_i に含まれる $N_m^{(i)}$ 個のモデルを $M_{ij}(j=1,2,\cdots,N_m^{(i)})$ と記述すると,式 (1) の平均ベクトル μ_{ik} ,分

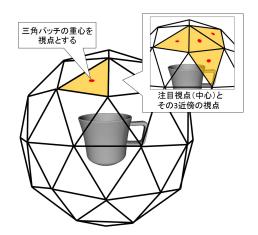


Fig. 2 80 viewpoints used to generate a CVFH feature.

散 σ_{ik}^2 は,注目視点 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:

$$\mu_{ik} = \frac{1}{4N_m^{(i)}} \sum_{j=1}^{N_m^{(i)}} \sum_{k' \in L_k} g_{ijk'};$$
 (2)

$$\sigma_{ik}^2 = \frac{1}{4N_m^{(i)}} \sum_{j=1}^{N_m^{(i)}} \sum_{k' \in L_k} |g_{ijk'} - \mu_{ik}|^2.$$
 (3)

 g_{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 CVFH 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.

2.2 初期視点の決定

初期視点から物体を観測し、CVFH $g^{(0)}$ が得られた時、各カテゴリから、 $g^{(0)}$ 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 $g^{(0)}$ 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形状位置合わせを行い、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。選ばれた視点を、 c_i の初期視点 $\hat{v}_i^{(0)}$ とする。同様に、全カテゴリから、初期視点を決定し、初期視点を使って視点計画を行う。

2.3 確信度の計算

観測をn回行った時、それまでの視点 $\mathbf{v}^{(0)},\cdots,\mathbf{v}^{(n-1)}$ と、その時に得られた特徴量 $\mathbf{g}^{(0)},\cdots,\mathbf{g}^{(n-1)}$ 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 $\mathbf{c}^{(n)}$ である確率は、ベイズ推定により事後確率密度関数 $p_{post}(\mathbf{c}^{(n)})$ で表される:

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

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

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

本手法では、式 (5) の事後確率を、物体の確信度とみなす。 ここで、式 (4) の分母の確率密度関数は、

$$p(\mathbf{g}^{n}|\mathbf{v}^{(n)},\mathbf{g}^{(n-1)},\mathbf{v}^{(n-1)},\cdots,\mathbf{g}^{(0)},\mathbf{v}^{(0)})$$

$$=\sum_{i=1}^{N_{c}}p(\mathbf{g}^{(n)}|c_{i}^{(n)},\mathbf{v}^{(n)})P_{pri}(c_{i}^{(n)}) \quad (7)$$

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、

$$P_{pri}(c^{(n)}) = P_{post}(c^{(n-1)})$$
 (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 v_k から観測した時, CVFH g_{ik} が得られる尤度 $p(g_k|c,v_k)$ は、次の式で表される.

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

2.4 視点の決定

カテゴリ c_i において、式(5)の確信度が高い場合は、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確信度に差がない場合、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,2.2 節で選んだ各カテゴリ c_i の代表モデル \hat{M}_i を使って,次の視点候補を選択する。80 個の各視点 v_k から得られる CVFH $g^{(k)}$ を, \hat{M}_i の CVFH \tilde{g}_{ik} として式 (4) に代入し,仮の事後確率を視点ごとに求める。次に, $g^{(n)}$ が得られた時の情報量のエントロピーを

$$H(P(X)) = -\sum_{X} P(X) \log P(X) \tag{10}$$

と定義する. このエントロピーは、物体の不確実さを表し、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{g}_{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:

$$I(c^{(n)}; \tilde{\mathbf{g}}_{ik}) = H(P_{pri}(c^{(n)})) - H(P_{post}(c^{(n)}))$$
(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:

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

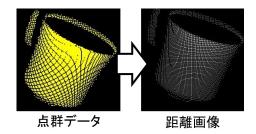


Fig.3 Depth map

3. 形状の類似性を用いた確信度の計算

CVFH のみで 3 カテゴリ以上の物体認識問題を扱うと、異なるカテゴリ内に類似した特徴が含まれる可能性がある。そのため、複数あるカテゴリから真のカテゴリを特定するのに、多くの視点移動を要していた.この問題を解決するために、既に取得している点群データを利用する確信度計算法を提案する.この計算法では、取得した点群データと、データベース内にある点群データの類似性を評価することで確信度を求める.

具体的には、それぞれの点群データから作成した距離画像間の距離を求めることで行う。距離画像は、各視点から得られる点群データから作成する(Fig.3)。作成した距離画像は、点群の奥行き方向の座標値とグレースケールの輝度値(0-255)の割合を対応させている。カメラに近い部分ほど、輝度値は255に近い値としている。

形状の類似性を用いた確信度の計算は,2回目以降の視点の決定から適用する.確信度の計算は,以下の手順で行う.まず,カテゴリごとに学習モデルの中から,形状が平均的であるモデルを平均モデル \bar{M}_i として選択する.次に,その平均モデルの k 番目の視点 v_k までの点群データと,モデル M_{ij} の点群データとの距離画像間の距離 $d_{(v_0,\dots,v_k)ij}$ を求める.l(k>l) 番目の視点 v_l における,平均モデルとモデル M_{ij} の距離画像間の距離を d_{v_lij} とすると,

$$d_{(\mathbf{v}_0,\dots,\mathbf{v}_k)ij} = \frac{1}{k+1} \sum_{l=0}^{k} d_{\mathbf{v}_l ij}$$
 (13)

によって求められる。また,k 番目の視点 v_k における,平均モデルとモデル M_{ij} の距離画像間の距離の平均 μ_{dik} と分散 σ_{dik}^2 は次の式で表される.

$$\mu_{dik} = \frac{1}{N_m} \sum_{i=1}^{N_m} d_{(\boldsymbol{v}_0, \dots, \boldsymbol{v}_k)ij}$$
 (14)

$$\sigma_{dik}^2 = \frac{1}{N_m} \sum_{j=1}^{N_m} (d_{(\mathbf{v}_0, \dots, \mathbf{v}_k)ij} - \mu_{dik})^2$$
 (15)

そして、取得した点群データとデータベース内の点群データとの距離画像間の距離を求める.

観測をn回行った時、それまでの視点 $\mathbf{v}^{(0)},\cdots,\mathbf{v}^{(n-1)}$ と、その時に得られた形状の類似性を表す距離 $d^{(1)},\dots,d^{(n-1)}$ 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 $\mathbf{v}^{(n)}$ から形状の類似性を表す距離 $d^{(1)},\dots,d^{(n-1)}$ が得られた時、それがカテゴリ $\mathbf{c}^{(n)}$ である確率を求める。ベイ

ズ推定により、2.3 節式 (4) の事後確率密度関数 $p_{post}(c^{(n)})$ は次のように書き換えられる:

$$p_{post}(c^{(n)}) = \frac{p(\boldsymbol{g}^{(n)}, d^{(n)}|c^{(n)}, \boldsymbol{v}^{(n)})p_{pri}(c^{(n)})}{p(\boldsymbol{g}^{(n)}, d^{(n)})}$$
(16)

$$p(\boldsymbol{g}^{(n)}, d^{(n)})$$

$$= p(\boldsymbol{g}^{(n)}, d^{(n)} | \boldsymbol{v}^{(n)}, \boldsymbol{g}^{(n-1)}, d^{(n-1)}, \boldsymbol{v}^{(n-1)},$$

$$\cdots, \boldsymbol{g}^{(1)}, d^{(1)}, \boldsymbol{v}^{(1)}, \boldsymbol{g}^{(0)}, \boldsymbol{v}^{(0)}) \quad (17)$$

$$P_{post}(c^{(n)}) = P(c^{(n)}|\boldsymbol{g}^{(n)}, d^{(n)}, \boldsymbol{v}^{(n)}, \boldsymbol{g}^{(n-1)}, d^{(n-1)}, \boldsymbol{v}^{(n-1)}, \dots, \boldsymbol{g}^{(1)}, d^{(1)}, \boldsymbol{v}^{(1)}, \boldsymbol{g}^{(0)}, \boldsymbol{v}^{(0)}) \quad (18)$$

$$P_{pri}(c^{(n)}) = P(c^{(n)}|\boldsymbol{g}^{(n-1)}, d^{(n-1)}, \boldsymbol{v}^{(n-1)}, \dots, \boldsymbol{g}^{(1)}, d^{(1)}, \boldsymbol{v}^{(1)}, \boldsymbol{g}^{(0)}, \boldsymbol{v}^{(0)}) \quad (19)$$

ここで、式(16)の分母の確率密度関数は、

$$p(\mathbf{g}^{(n)}, d^{(n)}) = p(\mathbf{g}^{(n)})p(d^{(n)})$$
 (20)

と分解でき,

$$p(\boldsymbol{g}^{(n)}) = p(\boldsymbol{g}^{n}|\boldsymbol{v}^{(n)}, \boldsymbol{g}^{(n-1)}, d^{(n-1)}, \boldsymbol{v}^{(n-1)}, \\ \cdots, \boldsymbol{g}^{(1)}, d^{(1)}, \boldsymbol{v}^{(1)}, \boldsymbol{g}^{(0)}, \boldsymbol{v}^{(0)}) \\ = \sum_{i=1}^{N_c} p(\boldsymbol{g}^{(n)}|c_i^{(n)}, \boldsymbol{v}^{(n)}) P_{pri}(c_i^{(n)}) \quad (21)$$

$$p(d^{(n)}) = p(d^{(n)}|\boldsymbol{v}^{(n)}, \boldsymbol{g}^{(n-1)}, d^{(n-1)}, \boldsymbol{v}^{(n-1)}, \dots, \boldsymbol{g}^{(n-1)}, d^{(n)}, \boldsymbol{v}^{(n)}, \boldsymbol{g}^{(n)}, \boldsymbol{v}^{(n)})$$

$$= \sum_{i=1}^{N_c} p(d^{(n)}|c_i^{(n)}, \boldsymbol{v}^{(n)}) P_{pri}(c_i^{(n)}) \quad (22)$$

と表される. 式(16)の尤度は,

$$p(\mathbf{g}^{(n)}, d^{(n)}|c^{(n)}, \mathbf{v}^{(n)})$$

$$= p(\mathbf{g}^{(n)}|c^{(n)}, \mathbf{v}^{(n)})p(d^{(n)}|c^{(n)}, \mathbf{v}^{(n)}) \quad (23)$$

のように分解できる. k 番目の視点 $m{v}_k$ から,平均モデル $ar{M}_i$ との形状の類似性を表す距離 $ilde{d}_{(m{v}_0,\dots,m{v}_k)i}$ が得られる確率 \mathcal{N}_d を

$$\mathcal{N}_{d}(\tilde{d}_{(\boldsymbol{v}_{0},...,\boldsymbol{v}_{k})}, \mu_{dik}, \sigma_{dik}^{2})$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_{dik}} \exp\left\{-\frac{(\tilde{d}_{(\boldsymbol{v}_{0}},...,\boldsymbol{v}_{k}) - \mu_{dik})^{2}}{2\sigma_{dik}^{2}}\right\} \quad (24)$$

とする. 式 (16) \sim (20), 式 (22) \sim (23) を用いると, 尤度 $p(\tilde{d}_{(\boldsymbol{v}_0,\ldots,\boldsymbol{v}_k)i}|c_i,\boldsymbol{v}_k)$ は次の式で表される:

$$p(\tilde{d}_{(\boldsymbol{v}_0,\dots,\boldsymbol{v}_k)i}|c_i,\boldsymbol{v}_k) = \mathcal{N}_d(\tilde{d}_{(\boldsymbol{v}_0,\dots,\boldsymbol{v}_k)},\mu_{dik},\sigma_{dik}^2). \quad (25)$$

ここで求めた確信度は、データ取得-物体認識-視点計画のループの中で、物体認識のときにのみ適用する。視点決定における相互情報量(式(11))での事前確率 $P_{pri}(c^{(n)})$ は、式 (19) から求める。また、事後確率 $P_{post}(c^{(n)})$ は、CVFHのみの式(5)で行う。データ取得-物体認識-視点計画の処理ループは、任意のカテゴリの確信度(式(18))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。

4. 実験

提案手法の有効性を検証するため、6つのカテゴリを分類する実験を行った。Fig.4は、学習した6つのカテゴリと、学習モデルの一例を示す。

実験は、学習データに含まれていないテスト用モデルを用いて、提案手法と、CVFH のみで確信度を求めた従来法で認識実験を行った。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した。次に、式 (5) を用いてカテゴリ毎の確信度を計算した。認識終了条件は、いづれか1つのカテゴリの確信度が0.90以上とし、認識終了条件を満たすまで、視点計画-データ取得-確信度の計算の処理ループを繰り返した。本実験では、各カテゴリで代表モデルを2つ選び、モデル1つに対して10個のテスト用データを作成した。テスト用データを入力し、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、Fig.5 は、カテゴリがカップのテスト用データを入力した際の、本手法と従来法におけるカップの確信度の変化を示す.Table1 は、本手法と、従来法で視点を決定した場合の平均視点移動回数と、最大視点移動回数を示す.Table1 から明らかなように、従来法に比べて、本手法の方が、視点移動回数および最大視点移動回数共に少ない回数で認識ができている.しかし、初期視点の決定の際に、実際とは異なる視点を初期視点とした場合、誤認識や視点移動回数が増加する問題も発生した.



Fig.4 Objects that was used in the experiment

5. 結言

本稿では、提案手法を、3カテゴリ以上の物体認識問題に適用できるよう拡張した手法を提案した、提案手法は、CVFHのみで物体認識問題を扱うのではなく、既に取得している点群データを利用した確信度計算法である。この手法を使って、6カテゴリの物体認識を行った結果、従来法よりも、認識終了までの視点移動回数を少なくすることができた。しかし、初期視点の決定の際に、実際とは異なる視点を

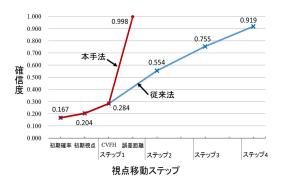


Fig.5 "cup"'s probabilities

Table 1 Average and maximum number of viewpoint movement

	平均視点移動回数		最大視点移動回数	
テスト	本手法	従来法	本手法	従来法
カテゴリ				
バナナ	2.32	7.76	6	13
ペットボトル	1.61	5.52	3	9
カップ	1.50	4.12	4	9
ミニカー	1.37	2.72	3	5
ハンマー	1.45	4.42	5	13
牛乳パック	2.37	6.89	6	10

選んだ場合は、誤認識や視点移動数増加の原因となることも確認できた. 今後は、この問題に対して、対策を行っていく予定である. また、最終的には実機での実験を予定しているため、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も対処していく必要がある.

参考文献

- [1] 諸岡健一, 査紅彬, 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 Roya, Santanu Chaudhuryb, Subhashis Banerjeec: "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] 中里一幾, 諸岡健一, 辻徳生, 倉爪亮: "幾何特徴量を用いた視点計画に基づく物体認識", 第 32 回日本ロボット学会学術講演会予稿集, 1J1-5, 2014.
- [6] Aitor Aldoma, Markus Vincze: "CAD-Model Recognition and 6DOF Pose Estimation Using 3D Cues", 2011 IEEE International Conference on Computer Vision Workshops, 978-1-467:3-0063-6, 2011.