

最適カメラパラメータ選択による能動的物体認識

○大津裕美, 出口光一郎 (東北大学)

Active Object Recognition by Optimal Camera Parameter Selection

○ Hiromi Ohtsu, Koichiro Deguchi (Tohoku University)

Abstract— This paper discuss the problem of recognizing an object from its appearance (i.e. image) using an actively controllable camera. In this case the performance of recognition can be improved by controlling the pose of the active camera and choosing a good viewpoint for recognizing each object. In this paper we propose two methods for choosing viewpoint that minimizes the ambiguity of the object hypotheses. One uses average information, the other uses average mutual information.

Key Words: Object recognition, Eigenspace method, Parameter selection,

1 はじめに

本研究では, 能動的に動かせるカメラを用いて, 見かけ (画像) から物体を認識する問題を扱う. このとき, 大変似通った特徴を持つ物体に対して, それらの特徴の違いが最も分かるカメラパラメータ (カメラの視点位置) を選択し, その位置から得た画像を使って認識を行うことを目的とした (Fig.1).

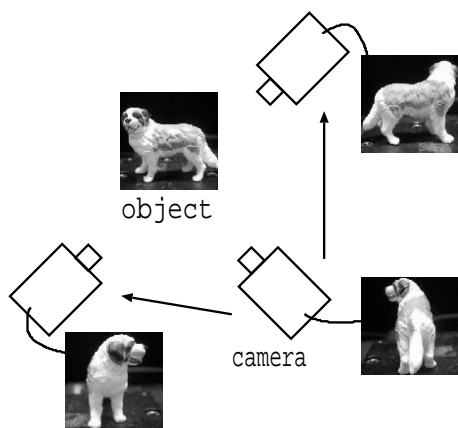


Fig. 1: Active vision

画像を用いた物体認識の研究では, 多くの手法が提案されている. 一般的に, これらの手法は, 得られた画像から物体を認識するのに有利な特徴を見つけ, それを利用して認識を行う. しかし, この方法では認識したい物体が大変似通った特徴を持つものである場合には認識が困難である. 例えば, Fig.2 のような場合である.



Fig. 2: Objects which look very similar except for its neck's barrel

このような問題に対して, 得られた画像だけでなく, 能動的に新たな画像を得ることで, 物体が何であるかと

いう認識率を向上させることが期待できる. そしてこの新たな物体の画像を得る際, 似通った特徴を持つ物体が最も異って見えるカメラパラメータを選択することが重要になってくる [1] [2].

本研究では, このカメラパラメータの選択に 2 つの方法を用いた. 1 つ目の方法では, 現時点で物体が何かという認識の不確かさ (平均情報量) と, あるカメラパラメータを選んで新たな画像を得てからの認識の不確かさを計算する. その差の期待値が最大になるようにカメラパラメータを選択する. 2 つ目の方法では, あるカメラパラメータを選ぶことで画像を得, その画像から物体が何であるかという認識の情報量を得る. この情報量の期待値 (平均相互情報量) を最大にするようにカメラパラメータを選ぶ.

このような認識を行うにあたり, まず物体の画像から特徴を抽出して数値的に表し, 物体の画像とその物体が何であるかということとの関係を確率的に表現する必要がある. そのために, 固有空間法を用いた.

2 固有空間における物体認識

物体認識を行うためには, 物体の画像から特徴を抽出して数値的に表さなければならない. 本研究では, 画像そのものをパターンとして扱い, 有名な手法である固有空間法を使って低次元の特徴量を構成し直し, 認識に用いた.

2.1 固有空間法

固有空間法は膨大な次元をもつ特徴ベクトルをより低次元で効果的な特徴ベクトルに作り直す方法である.

具体的には, ある画像の特徴をその画像の各々の画素における濃淡値として, それらを要素に持つベクトルを考える. つまり, n 画素の画像 S 枚を

$$\mathbf{x}_i = (x_1, x_2, \dots, x_n) \quad i = 1, 2, \dots, S \quad (1)$$

と表す. ここで, 変換行列 \mathbf{W} を用いると, m 次元の新たな特徴ベクトル \mathbf{y} は

$$\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i \quad (2)$$

と表せる. このとき, 変換行列 \mathbf{W} は n 行 m 列の行列とし, $m < n$ とする. 固有空間法においてこの \mathbf{W} は以

下のように求まる．まず分散共分散行列 \mathbf{S}_T を

$$\mathbf{S}_T = \sum_{k=1}^S (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T \quad (3)$$

とし， $\boldsymbol{\mu}$ はすべての \mathbf{x}_i の平均であり， n 次元を持つとする．このとき，新たな特徴ベクトル $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_S\}$ の分散 $\tilde{\mathbf{S}}_T$ は $\mathbf{W}^T \mathbf{S}_T \mathbf{W}$ となる．この \mathbf{S}_T の行列式を最大化するように \mathbf{W}_{opt} を選ぶ．これは

$$\begin{aligned} \mathbf{W}_{opt} &= \arg \max_{\mathbf{W}} |\tilde{\mathbf{S}}_T| \\ &= \arg \max_{\mathbf{W}} |\mathbf{W}^T \mathbf{S}_T \mathbf{W}| = (w_1 w_2 \dots w_m) \end{aligned} \quad (4)$$

と表される．ここで $\{\mathbf{w}_i | i = 1, 2, \dots, m\}$ は \mathbf{S}_T の固有値の大きい方から $1, 2, \dots, m$ 個に対応するそれぞれ n 次元の固有ベクトルである．

2.2 固有空間における確率分布

認識をするにあたり，あらかじめ与えられた情報を整理しておく必要がある．(1) 式におけるそれぞれの画像 \mathbf{x}_i ，また対応する特徴ベクトル \mathbf{y}_i はクラス $\{O_1, O_2, \dots, O_C\}$ のうちの 1 つに属している．ここで， C は物体の数であり，どの物体であるかを示している．ここで，行動 (どの位置から物体の画像を得るかの選択) を a とし，新たに得たクラスが未知の画像が固有空間において \mathbf{g} で与えられるとすると， $p(\mathbf{g}|O_i, a)$ の確率分布は一般的なモデルである正規分布と仮定することで学習用画像から学習することで得られる．このとき，ベイズの定理により，

$$p(O_i|\mathbf{g}, a) = \frac{p(\mathbf{g}|O_i, a)p(O_i|a)}{p(\mathbf{g}|a)} \quad (5)$$

を得る．(5) 式で得られた各状態の確率により，どの物体であるかの認識を行う．

3 物体認識

似通った特徴を持つ物体を認識することは困難である．そこで，このような物体の認識率を向上させるために，与えられた画像だけでなく，能動的に新たに得た物体の画像を用いて物体認識を行った．ここで，新たに画像を得る際に，似通った特徴を持つ物体が最も異って見えるような視点を選ぶことが大切である．

3.1 認識結果の逐次的な更新

能動的に次々に得た画像を用いて物体認識を行うということは，当然次々に認識結果が更新されることとなる．今，固有空間での特徴 \mathbf{g}_1 を持つ画像を 1 ステップ目として得た．このときの物体の推定は (5) 式より求まる．ただし， $p(O_i|a_1)$ は 1 ステップ目においては一樣とする．次のステップでも同様に求まるが，このとき，

$$p(O_i, a_2) = p(O_i|\mathbf{g}_1, a_1) \quad (6)$$

とする．よって， t ステップでは

$$p(O_i|\mathbf{g}_t, a_t) = \frac{p(\mathbf{g}_t|O_i, a_t)p(O_i|\mathbf{g}_{t-1}, \dots, \mathbf{g}_1)}{p(\mathbf{g}_t|a_t)} \quad (7)$$

となり，これにより t ステップ終了後に各クラスである確率が求まる．

3.2 最適なパラメータの選択 1

今回，次ステップに用いる画像を選択する方法として，2 つの方法を用いた．この節ではまず 1 つ目の方法について述べる．

Shannon のエントロピーを用いると，

$$H(\mathbf{p}_t) = - \sum_{i=1}^C p_t^i \log p_t^i \quad (8)$$

と表せる．ここで， $\mathbf{p}_t = (p_t^1, \dots, p_t^C)$ であり， p_t^i は t ステップ後に O_i である確率を示し， $p_t^i = p(O_i|\mathbf{g}_1, \dots, \mathbf{g}_t)$ である．これを用いて，不確かさ (エントロピー) の減少量の期待値が最大となる行動，つまり視点選択を行う計画をとる [2]．この計画では，物体が何であるかという現時点での認識の不確かさと，ある行動をとった時点での認識の不確かさを計算する．そしてこの差の期待値が最大となるような行動選択をする．つまり，この方法は各時点での認識の不確かさに注目した方法である．具体的には不確かさの減少量の期待値は，(8) 式を用いて，

$$\Delta H(a) = \sum_{i=1}^C p_t^i \{H(\mathbf{p}_t) - \sum_{\mathbf{g}_{t+1}} p(\mathbf{g}_{t+1}|O_i, a)H(\mathbf{p}_{t+1})\} \quad (9)$$

と表せる．これを最大にする最適な行動 a^*

$$a^* = \arg \max_a \Delta H(a) \quad (10)$$

を選んで認識を行う．

3.3 最適なパラメータの選択 2

2 つ目の方法として，平均相互情報量を最大にする行動をとる計画をとる [3]．これは画像を得たことによって現在の認識の不確かさをどれだけ減らせるかに着目したものである．つまり，ある行動に依存する画像が得られたことによって与えられる認識の不確かさの減少量を最大にするような計画をとるものである．この方法は物体が何であるかという認識とある行動により得られる画像との関係に注目した方法である．具体的にある行動に伴って得られる画像と認識との平均相互情報量 $I(O; \mathbf{g}_{t+1}|a)$ は

$$I(O; \mathbf{g}_{t+1}|a) = \sum_{i=1}^C \sum_{\mathbf{g}_{t+1}} p_t^i p(\mathbf{g}_{t+1}|O_i, a) \log \frac{p(\mathbf{g}_{t+1}|O_i, a)}{p(\mathbf{g}_{t+1}|a)} \quad (11)$$

で求まり，(10) 式と同様に，これが最大をなす行動 a^*

$$a^* = \arg \max_a I(O; \mathbf{g}_{t+1}|a) \quad (12)$$

を選んで認識を行う．

4 物体認識の予備実験

Fig.3に今回の実験で用いた各物体の正面から見た画像を示す。物体を乗せたターンテーブルを回転させて画像を得て、2章で述べた固有空間での特徴を利用して、どの物体であるかを認識する。つまり、ターンテーブルの回転角が行動 a (パラメータ選択) である。

最適な行動をとるためには、各物体と画像の関係を知る必要がある。このために学習用画像を用いた。学習用画像には、Fig.3のそれぞれのクラスについて、Fig.4に示すようにターンテーブルの回転角を30度ごとに360度回転させて得た画像をそれぞれ10枚ずつ用いた。なお、画像は100×100の白黒画像である。Fig.3の O_1 と O_2 , O_5 と O_6 , O_8 と O_9 はそれぞれ色や模様が違うだけの形状が同じ物体であり、 O_1 と O_3 は首の部分の樽以外は色も形状もほぼ同じ物体である。これらの学習用画像は2章の方法で整理しておく。

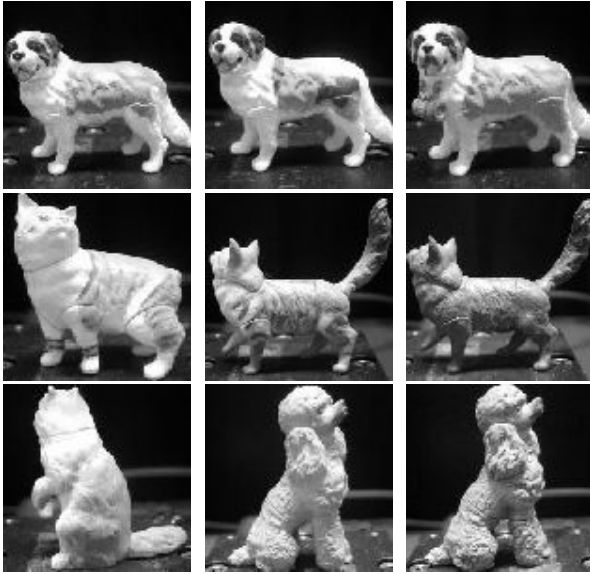


Fig. 3: Objects which use in recognition experiment. top row (left to right) : object O_1, \dots, O_3 , middle : O_4, \dots, O_6 , bottom : O_7, \dots, O_9 .

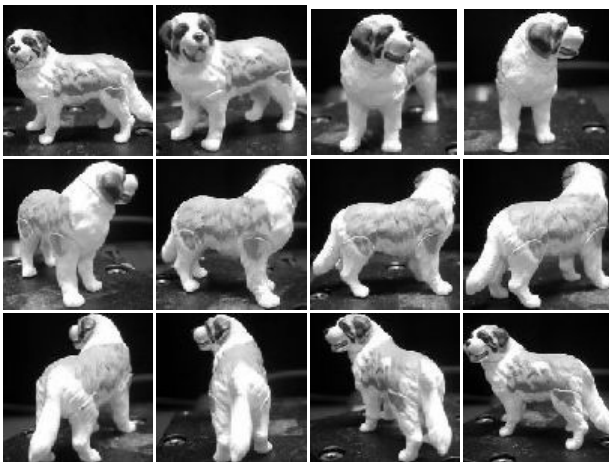


Fig. 4: Poses of object1. top row (left to right) : view a_1, \dots, a_4 , middle : a_5, \dots, a_8 , bottom : a_9, \dots, a_{12}

4.1 実験

カメラの正面に認識したい物体を置くものとする。このとき、任意の行動(カメラパラメータの選択)を一回だけ選択して行い、そこから得た物体の画像の情報を用いて、どのクラスであるかを認識する実験を行った。なお、(5)式で得られる値が最も大きくなったクラスをその物体のクラスとした。

4.2 認識結果

クラス O_1, O_2 のそれぞれに属する物体を認識した結果について Table1 に示す。行動 a_1-a_5, a_{12} では物体間の違いである樽が見えず、模様の差もほぼない画像が得られ、行動 a_6-a_{11} ではそれらが見える画像が得られる。認識結果の差は明らかである。このことから、認識する際に、物体の違いが目立つ画像がとれる行動を選択することの重要性が分かる。

Table 1: 物体 O_1, O_2 の画像の認識率 (%)

	a_1-a_5, a_{12}	a_6-a_{11}
object O_1	93	67
object O_2	80	50
average	87	58

5 最適なパラメータ選択を行う物体認識実験

4章の実験結果からも明らかのように、ある方向から見ると大変似通った特徴を持つような物体間での認識は、様々な方向から得た物体の画像を用いて行うのがよい。そこで、3章の方法を用いて能動的に行動を選択して徐々に画像を得ることで、認識を行う実験をした。また、比較のためにランダムに次のパラメータを選ぶ方法も実験した。

5.1 実験

カメラの正面に認識したい物体を置くものとする。手順を次に示す。

1. まず1ステップ目で用いる画像を任意の行動(カメラパラメータ)を選択して得る。
2. そこから得た物体の画像の情報を用いて、(7)式からどのクラスであるかの確率を求める。
3. そのとき、どれかのクラスの確率が0.9を越えれば認識終了。そうでない場合は次へ。
4. 次の行動の選択を行う。(9)式が最大になる行動、(11)式が最大になる行動、ランダムに行動を選ぶ方法で次の行動を選択し、画像を取得し、ステップ2へ。

5.2 認識結果

ある任意の視点から得た物体の画像から認識を開始し、認識が終了するまでに要したステップ数をカウントした。次にまた違う視点から得た物体の画像を与えて同様に行う、という試行を繰り返した。このときの各試行(横軸)ごとに認識終了までに要したステップ数

(縦軸)を Fig.5, Fig.6, Fig.7 に示す. それぞれ, ランダムにパラメータ選択を行ったときの結果, 3.2 節のパラメータ選択を行ったときの結果, 3.3 節パラメータ選択を行ったときの結果をである. 各試行ごとに与えた画像は3つの方法で同じである. 明らかに, ランダムにパラメータ選択を行ったときに比べて, 計画的にパラメータ選択を行ったときに要したステップ数が減っているのが分かる. このときの認識率はどの方法を用いたときも認識率の差はほぼなく, どれも 90% を越えていた.

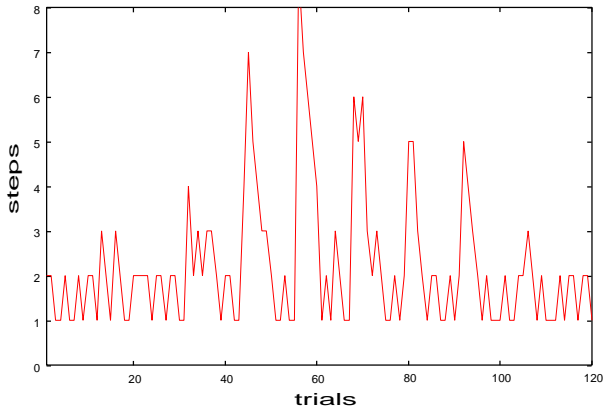


Fig. 5: Number of steps using random strategy

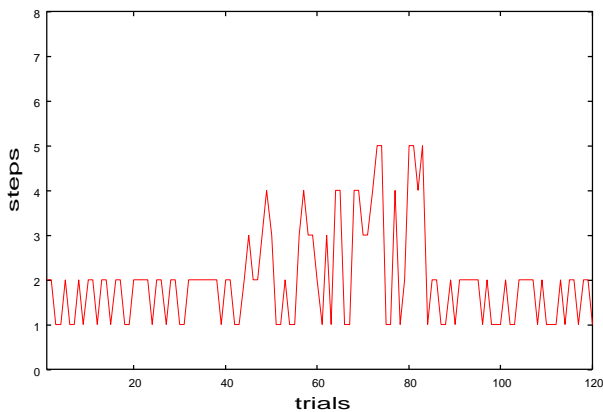


Fig. 6: Number of steps using optimal strategy (section 3.2)

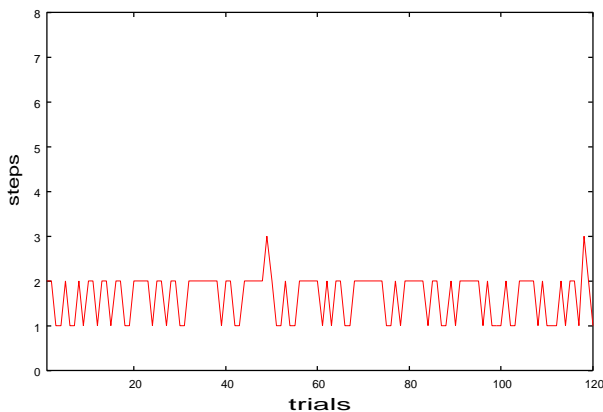


Fig. 7: Number of steps using optimal strategy (section 3.3)

6 まとめ

平均情報量および平均相互情報量を用いることにより, より効率のよい認識が行えることが分かった. よりよい認識を行うためには, 多くの学習用画像を用いる必要がある.

また, これらの方法は照明条件の変化を伴う物体の認識についても適用できることが考えられる.

参考文献

- [1] L.Paletta, A.Pinz, Active object recognition by view integration and reinforcement learning, *Robotics and Autonomous Systems* 31, pp.71-86, 2000
- [2] H.Borotschnig, L.Paletta, M,Prantl, A.Pinz, Appearance-based active object recognition, *Image and Vision Computing* 18, pp.715-727, 2000
- [3] 出口光一郎, 横断型科学としてのセンシングについての一考察'計測=情報の獲得という立場から', 第21回センシングフォーラム, pp.141-146, 2004