

(d) 情報

# 画像の学習による姿勢推定手法の学習サンプル数による精度変化の検討

## Accuracy Change According to the Number of Learning Samples for View-Based Pose Estimation

奥川 裕之<sup>†</sup> 原田 健吾<sup>††</sup> 玉木 徹<sup>†</sup> 天野 敏之<sup>†††</sup> 金田 和文<sup>†</sup>Hiroyuki Okugawa<sup>†</sup> Kengo Harada<sup>††</sup> Toru Tamaki<sup>†</sup> Toshiyuki Amano<sup>†††</sup> Kazufumi Kaneda<sup>†</sup><sup>†</sup> 広島大学大学院工学研究科 <sup>††</sup> 広島大学工学部 <sup>†††</sup> 奈良先端科学技術大学院大学情報科学研究科

### 1 背景

本研究では、ある物体の見え方が違う画像を学習することで、その物体の姿勢を推定する手法の推定精度を検討する。物体の見え方が違う画像とは、物体を図1のように様々な方向から撮影した画像である。その画像を学習サンプルとし、推定したい画像と学習サンプルを比較することで、物体の姿勢を推定する。

画像の学習による姿勢推定手法の精度は、学習サンプル数に依存している。よって、推定する際に重要となってくることは、どのぐらいの枚数を学習サンプルとするかである。学習サンプルが多いほど精度が良くなると考えられるが、様々な姿勢を推定するとなると、膨大な学習サンプルが必要である。そのため、なるべく学習サンプルは少なくしたい。しかし、どのくらいまで学習サンプルを減らして、精度良く姿勢推定できるかはまだわかっていない。よって、学習サンプル数と推定精度の関係を明らかにし、学習サンプル数の基準を作ることを、この研究の目的とする。

推定精度の関係を明らかにするために、今回、学習サンプル数を変化させて精度検討をした。その結果、物体により精度変化が異なり、物体によっては学習サンプル数が多い場合より、少ない場合の方が精度が良くなった。その原因を探るため学習サンプル数をより詳細に検討、また画像の相関の検討を行った。その結果、学習サンプル数が多い場合より、少ない場合の方が精度が良くなった原因は、画像の相関が高い画像同士が学習されていないためであるということがわかった。

まず、姿勢推定手法として用いた EbC ( Estimation-by-Completion ) 法 [1] について 2 節で説明し、3 節で学習サンプル画像と学習セットについて説明する。4 節で学習サンプル数を变化させた精度の検討、5 節で画像の相関の検討、6 節で今後の課題について述べる。

### 2 EbC 法

画像を学習して姿勢推定する手法として、EbC 法を用いる。この手法は、姿勢推定を、画像に埋め込まれた情報トラックの復元という問題としてとらえ、固有空間法による学習結果をもとに画像を補間することで情報トラックを復元し、パラメータ推定する方法である ( 図 2 )。

まず、各画素  $x_1, \dots, x_N$  を要素にもつ  $N$  次元の画素

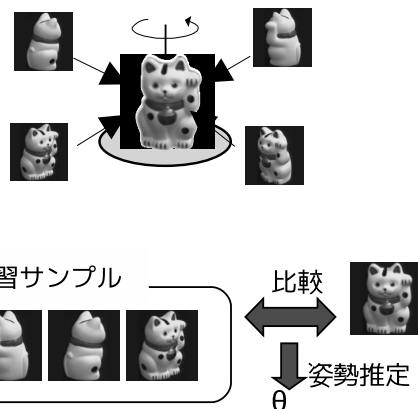


図 1: 画像の学習に基づく姿勢推定

ベクトル  $x$  に情報トラック  $p$  を付加した行列を画像  $n$  枚分並べた行列

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix} \quad (1)$$

を用意する。この情報トラックベクトル  $p$  は、物体のパラメータ  $\theta$  と一意に対応するもので、正弦波を用い、

$$p_j = \begin{bmatrix} \cos(0 - \theta_j) \\ \cos\left(\frac{2\pi}{w} - \theta_j\right) \\ \cos\left(2\frac{2\pi}{w} - \theta_j\right) \\ \vdots \\ \cos\left(\frac{2(w-1)\pi}{w} - \theta_j\right) \end{bmatrix} \quad (2)$$

と表す。ここで  $w$  は画像の幅である。

次に、この行列  $X$  に対して特異値分解を用いて、固有ベクトル  $\{e_1, \dots, e_n\}$  を求める。それらを表す固有空間を

$$E = [e_1 \ e_2 \ \dots \ e_n] = \begin{bmatrix} E_{in} \\ E_{out} \end{bmatrix} \quad (3)$$

とおく。ここで  $E_{in}$  は固有空間の画像部分、 $E_{out}$  は固有空間の姿勢を表す情報トラック部分である。

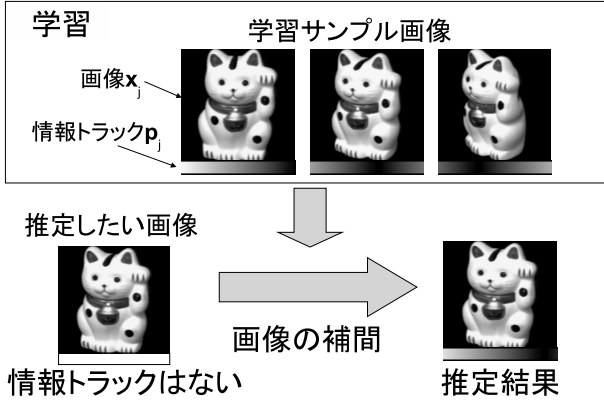


図 2: EbC 法の概略

最後に、姿勢推定を以下の式で情報トラック部分を補間することで行う。

$$\hat{\theta} = \tan^{-1} \left( \frac{\Omega_s^T x}{\Omega_c^T x} \right) \quad (4)$$

$$\Omega_s^T = \omega_s^T E_{\text{out}} (E_{\text{in}}^T E_{\text{in}})^T E_{\text{in}} \quad (5)$$

$$\Omega_c^T = \omega_c^T E_{\text{out}} (E_{\text{in}}^T E_{\text{in}})^T E_{\text{in}} \quad (6)$$

$$\omega_s = \left[ \sin 0, \sin \left( \frac{2\pi}{w} \right), \dots, \sin \left( \frac{2(w-1)\pi}{w} \right) \right]^T \quad (7)$$

$$\omega_c = \left[ \cos 0, \cos \left( \frac{2\pi}{w} \right), \dots, \cos \left( \frac{2(w-1)\pi}{w} \right) \right]^T \quad (8)$$

$\Omega_s$  と  $\Omega_c$  は、姿勢推定をする前に計算を行うことができる。そのことより、姿勢推定計算は、推定したい画像と  $\Omega_s$ 、 $\Omega_c$  との内積演算 2 回と三角演算 1 回で計算できる。よって、高速な姿勢推定を可能としている。

### 3 学習サンプル画像と学習セット

今回、学習サンプル画像として COIL-20[2] を用いた。この学習サンプルは、物体の鉛直軸回転の 1 自由度である (図 1)。また学習サンプルの間隔は 5[deg] 刻みで、サンプル数は一物体に対して 72 枚である。その学習サンプルを、 $x_i$  と定義する。例えば、 $\theta = 0$  の時、 $x_0$ 、 $\theta = 5$  の時、 $x_5$  である (図 4)。そして物体数は 20 物体である (図 3)。

今回、学習サンプル数を変化させる方法として、学習セットを次のように定義した。

$$S_{i,s} = \{x_{ik+s}\}, \quad k = 0, 1, \dots, n_i - 1 \quad (9)$$

ここで、 $i$  は学習サンプル間の角度 [deg]、 $s$  は始点とする角度 [deg]、そして、 $n_i$  は学習サンプル数であり  $n_i = \frac{360}{i}$  である。

例えば、学習サンプル間の角度  $i = 20$ 、始点とする角度  $s = 0$  ならば、学習セットは、 $S_{20,0} = \{x_0, x_{20}, x_{40}, \dots, x_{340}\}$  の 18 枚 ( $= \frac{360}{20} = n_{20}$ ) であり、 $i = 20, s = 5$  なら学習セットは、 $S_{20,5} = \{x_5, x_{25}, x_{45}, \dots, x_{345}\}$  の 18 ( $= \frac{360}{20} = n_{20}$ ) である

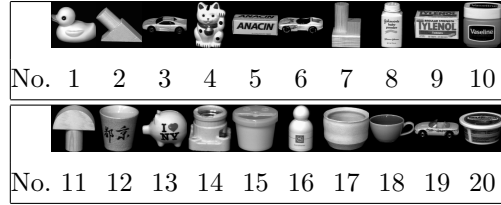


図 3: COIL-20 の 20 物体の画像。これらの画像はグレイスケール画像で、解像度は、128×128 である。

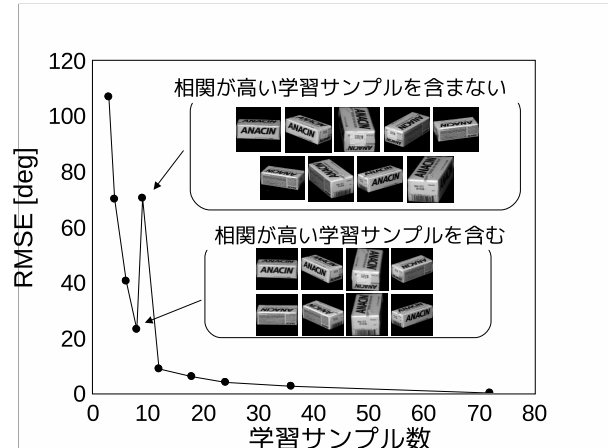


図 5: 学習サンプル数変化による精度変化 (obj5) (縦軸: RMSE[deg] 横軸: 学習サンプル数 [枚])

(図 4)。このように学習セットを定義することで、学習枚数が同じでも学習画像が違う場合を扱う。

### 4 学習サンプル数変化による精度検討

学習サンプル数が多いほど、推定精度は良くなると仮定できる。だが、実際どのように精度が変化するかは検討されていない。そこでサンプル数を減らし、実際の姿勢との推定誤差を計算することで、推定精度を調査する。

まず最初に、始点角  $s = 0$ 、学習サンプルの間隔  $i = 5k$  ( $k = 1, 2, 3, 4, 6, 8, 9, 12, 18, 24$ ) [deg] とし、サンプル数を  $n_i = 72, 36, 24, 18, 12, 9, 8, 6, 4, 3$  枚と減らして検討を行った。推定誤差の評価方法として以下の式を用いた。

$$\text{RMSE}_{i,s} = \sqrt{\frac{1}{72 - n_i} \sum_{x_j \notin S_{i,s}} (\hat{\theta}_j - \theta_j)^2}, \quad (10)$$

$\theta_j$  [deg] は実際の角度、 $\hat{\theta}_j$  [deg] は推定角度である。この RMSE の値が大きいくほど推定精度が悪くなる。

結果として、図 5 のような精度変化が、20 物体中 6 物体存在した。その精度変化は仮定と違い、学習サンプル数が多い場合より、少ない場合の方が精度が良くなっている。また、精度が悪くなっている学習サンプル間隔は 40 度の時だけである。そこで、精度が悪くなっ

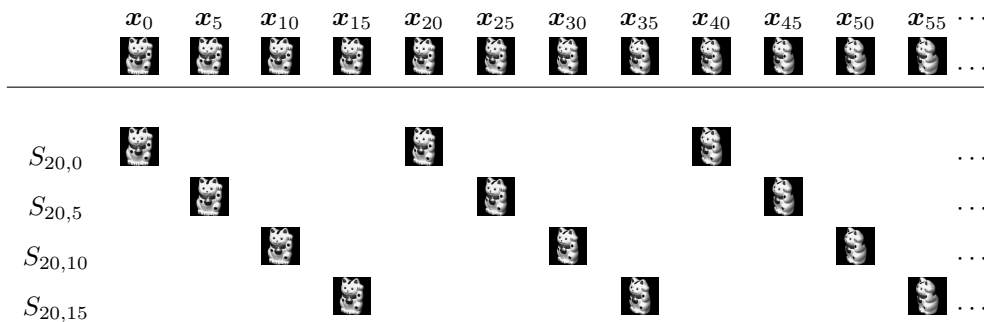


図 4: 学習セットの例 ( $S_{20,0}, S_{20,5}, S_{20,10}, S_{20,15}$ ). これらの学習セットは, 学習画像は違うが, 同じ枚数, 同じサンプル間隔 20 度である. ここでの  $x_j$  は,  $\theta_j = j$  の時の画像である.

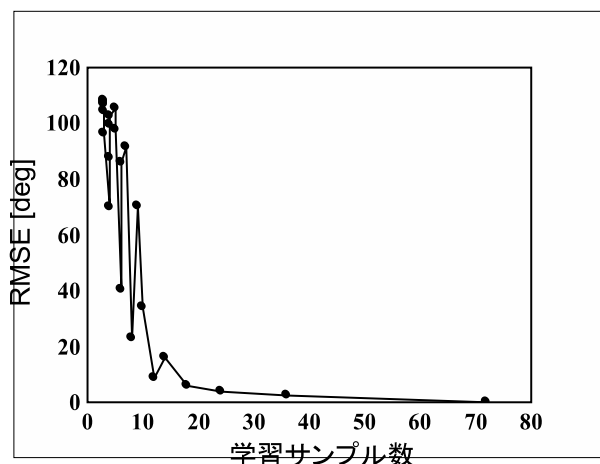


図 6: 学習サンプル数を細かくした場合の精度変化 (obj5) (縦軸: RMSE[deg] 横軸: 学習サンプル数 [枚])

ている学習サンプル間隔は 40 度以外にもないかどうかを見るために, 間隔  $i = 5k$  ( $k = 1, 2, \dots, 24$ ) [deg], 始点角  $s = 5m$  ( $m = 0, 1, \dots, k - 1$ ) と変化させた. 評価方法は, 各学習セットに対して RMSE を計算し, 同じ学習枚数に対して平均を求めることで行った. 結果として, 図 6 のような精度変化となり, 精度が悪くなっている学習サンプル間隔は 40 度だけではないことがわかる. だが, 精度悪くなる角度には一貫性はない. だが, 姿勢推定が学習画像の類似性に影響があると考えられるため, 次節で画像の相関の検討を行う.

## 5 画像の相関の検討

学習サンプル数が多い方が精度が悪くなる原因は, 学習した画像間の類似度が影響すると考えられる. そこで, 画像間の相関を計算した. 画像  $x_i$  と  $x_j$  の相関  $C_{i,j}$  を, 以下の式で計算した.

$$C_{i,j} = \frac{x_i^T x_j}{\|x_i\| \|x_j\|} \quad (11)$$

相関  $C_{i,j}$  は, 1 に近いほど相関が強くなり, 0 に近いほど相関が弱くなる.

結果として, obj5 の相関関係を表したものを, 図 7, 図 8 に示す. 図 7 は, 図 6 において精度が悪くなっていたサンプル数 9, つまり学習セット  $S_{40,0} = \{0, 40, 80, 120, 160, 200, 240, 280, 320\}$  の 0 度, 40 度, 200 度の画像と他の全ての画像との相関  $C_{0,j}, C_{40,j}, C_{200,j}$  を表したものである. また図 8 は, 図 6 において精度が良くなっていたサンプル数 8, つまり学習セット  $S_{45,0} = \{0, 45, 90, 135, 180, 225, 270, 315\}$  の相関  $C_{45,j}, C_{90,j}, C_{225,j}, C_{270,j}$  を表したものである. グラフの縦軸は, 相関値を, 横軸は角度  $j$  を示している. プロット点の違いは, 基準となる画像の違いを示しており, 例えば, 図 7 の三角のプロット点 ( $\triangle$ ) は, 0 度の画像と  $j$  度の画像との相関  $C_{0,j}$  を表している.

図 7, 図 8 を見ると, どの基準画像であっても, 基準画像付近以外に 180 度ずれた画像 (例えば, 図 7 では 200 度の画像に対して 20 度) の相関が高くなっている. つまり画像の姿勢推定に, 推定したい画像付近以外の 180 度ずれた画像が大きな影響を与えることになる. だが, サンプル数 8 の場合 (図 8) には, 180 度ずれた画像同士 (例えば 45 度に対して 225 度) がセットで学習されているため, 180 度ずれた画像に与える影響が小さくなっている. ところが, サンプル数 9 の場合 (図 7) には, 180 度ずれた画像同士が学習画像には含まれていない. よって, 180 度ずれた画像に与える影響が大きいため推定誤差が大きくなり (例えば, 20 度の画像を 200 度付近と推定 (図 7)), 精度が悪くなったと考えられる.

その考えは, 図 6 の精度変化にも当てはまっている. 図 6 の精度が良くなっている  $S_{20,s}, S_{30,s}, S_{45,s}, S_{60,s}, S_{90,s}$  では, 180 度ずれた画像が学習されている. 反対に, 図 6 の精度が悪くなっている  $S_{25,s}, S_{40,s}, S_{50,s}, S_{55,s}, S_{75,s}$  などでは, 180 度ずれた画像が学習されていない. つまり, 180 度ずれた画像を学習できているかどうかで精度の良し悪しが変わっていると考えられる. これが, 学習サンプル数が少ない場合より多い場合の方が精度が悪くなる原因となったと考えられる.

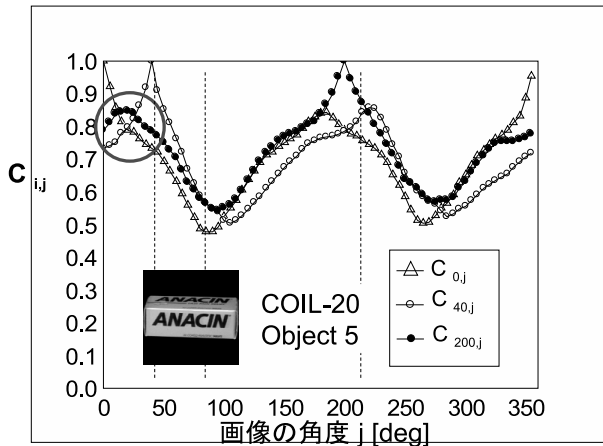


図 7: サンプル数 9 の場合の画像の相関変化 ( サンプル角 40 度  $i = 0, 40, 200$  )

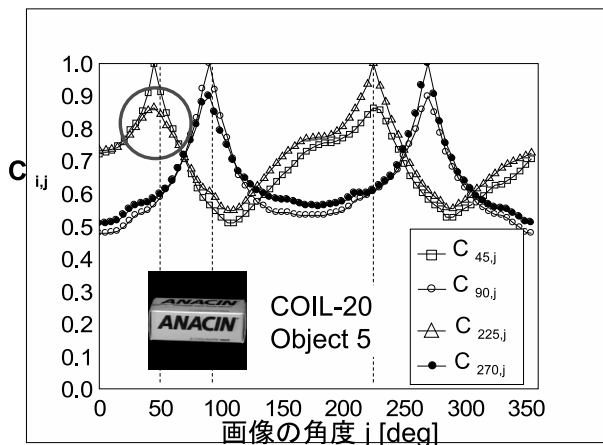


図 8: サンプル数 8 の場合の画像の相関変化 ( サンプル角 45 度  $i = 45, 90, 225, 270$  )

## 6 今後の課題

物体の回転は 3 次元であるため、自由度を 3 まで増加させて精度検討を行う。だが自由度を増加させると、様々な姿勢を学習するために、学習サンプルが膨大になるので、CG 画像を学習サンプルとして用いることを検討している。

### 参考文献

- [1] 天野敏之, 玉木徹: 「EbC 画像対を用いた三次元物体の高速線形姿勢推定」, 電子情報通信学会論文誌 D, Vol.J90-D, No.8, pp.2060-2069 (2007).
- [2] S. A. Nene, S. K. Nayar, H. Murase: "Columbia Object Image Library (COIL-20)", Technical report CUCS-005-96, Columbia University (1996).