

Procrustes Analysis を用いた姿勢推定

寺門和哉[†] BisserRaytchev[†] 玉木徹[†] 金田和文[†][†] 広島大学大学院工学研究科 〒739-8527 広島県東広島市 鏡山 1-4-1E-mail: [†]terakado@eml.hiroshima-u.ac.jp, ^{††}{bisser,tamaki,kin}@hiroshima-u.ac.jp

あらまし 本論文では新しいアピアランス空間の姿勢推定手法である Local Procrustes Regression (LPR) を提案する。LPR では、全ての学習サンプルとその姿勢パラメータ空間とのマップを学習するのではなく、テストサンプル近傍の学習サンプル同士の距離関係を学習に利用し、テストサンプルの姿勢パラメータを推定する。このとき Procrustes Analysis を用いてテストサンプル近傍の学習サンプルの低次元空間と、それに対応する姿勢パラメータ空間との位置合わせを行う。物体ごとに 2 つの回転軸において 5 度ずつサンプリングされた 703 枚の画像データセット Object Pose Estimation Database (OPED) を用いて実験を行った結果、提案手法の姿勢推定精度は半数以上の物体で Support Vector Regression より優れていることが示された。

キーワード 姿勢推定, 回帰, Multidimensional Scaling, Procrustes Analysis

1. はじめに

コンピュータビジョンにおける姿勢推定問題として、回帰に基づく姿勢推定問題がある。この問題は、ある画像が与えられたときに画像中の物体の傾きや方向などの姿勢パラメータを与える関数を求めるものである。姿勢推定は物体追跡や物体認識、仮想現実や拡張現実などに用いられており、これまでに姿勢推定問題に対して様々な手法が提案されている。それらはモデルベース手法 [1]、アピアランス空間手法 [2]、あるいはその 2 つを組み合わせたものの 3 つに大別できる。本研究では姿勢推定問題に対してアピアランス空間手法に着目する。この種の手法は主に学習サンプルと姿勢パラメータ空間とのマップを学習する。従来は全てのサンプルを用いてマップを学習する。本論文ではその代わりに局所的に選択した学習サンプルを用いてマップを学習する Local Procrustes Regression (LPR) を提案する。この手法では、学習サンプルとテストサンプル間の距離を用いて、未知のテストデータの姿勢情報を局所的に選択した学習サンプルから推定する。

LPR は線型代数に基づいているため、簡潔で高速に実行することが可能である。さらに、高次元の画像を効率的に扱うことが可能であるという利点がある。提案手法を評価するために、公開されている姿勢推定用データセット Object Pose Estimation Database (OPED) [7] を用いて実験を行い、提案手法はサポートベクトル回帰と比べて半数以上の物体で優れた結果が得られた。

本論文では 2 節で提案手法である LPR について述べる。3 節では今回用いた次元削減手法である Multidimensional Scaling (MDS) について、4 節では Procrustes Analysis について述べる。5 節では提案手法を用いた実験結果を示す。

2. Local Procrustes Regression (LPR)

本節では提案手法である LPR の概要を述べる。まず $X_T = \{x_1, \dots, x_N\}$ をある物体の N 枚の画像 (列ベクトル形式) 群とする。ここで x_i は角度が異なる物体の画像を表す。 X_T の画像の姿勢パラメータは既知であり、 $P_T = \{p_1, \dots, p_N\}$ で表されるとする。ここで p_i は x_i の姿勢パラメータのベクトルである。 X_T と P_T を対にして学習セット T を表す。姿勢推定においては、 X_T に含まれないテスト画像 x_0 の姿勢パラメータ p_0 を求める。

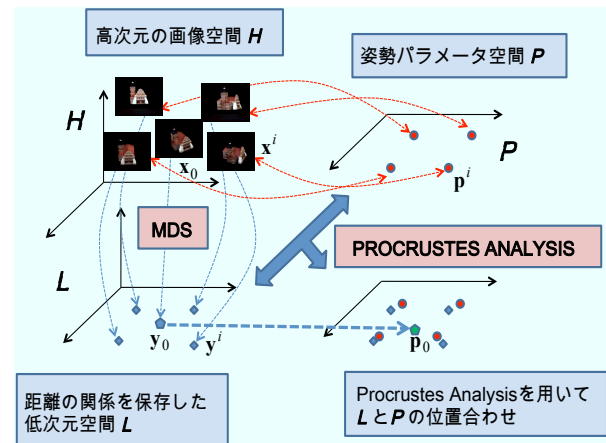


図 1 Local Procrustes Regression の概要

LPR の概略を図 1 に示す。LPR は 3 段階で姿勢推定を行う。まず X_T の中から x_0 の k 近傍となる x_i を求め、それらを用いて学習画像 $X^{(k)} = (x^1, \dots, x^k)^t$ とそれに対応するパラメータ $P^{(k)} = (p^1, \dots, p^k)^t$ を作成する。これにテスト画像ベクトル x_0 を追加し、 $X = (x_0, x^1, \dots, x^k)^t$ を作成する。次に MDS を用いて X を高次元の画像空間 H から低次元空間 L にマッピングする。このとき X の低

次元表現を $Y = (y_0, y^1, \dots, y^k)^t$ と表し, H, L 間のマッピングは距離が保存されるように行う. つまり H 内のベクトル $x_i - x_j$ 間と L 内のベクトル $y_i - y_j$ 間の距離関係を保つようにマッピングする. 最後に, $X - P^{(k)}$ 間, $Y - P^{(k)}$ 間の対応を用いて最適な変換を求める. このとき Procrustes Analysis [5] を用いて適合度検定の最小化を行う. そして, 求めた変換を用いて y_0 から p_0 を求める. MDS と Procrustes analysis については 3 節, 4 節で詳しく述べる.

3. Multidimensional Scaling (MDS)

Multidimensional Scaling [3] はデータの距離行列を解析することでデータ点群の低次元表現を得る手法の総称である. MDS には様々な手法があるが, 本論文では距離を保存して低次元にマッピングされた Y を得るために classical MDS [4] を用いる. まず, $(k+1) \times (k+1)$ の距離行列 $D = (d_{ij})$ を作成する. ここで (d_{ij}) は X 内の各画像同士のユークリッド距離である. 距離行列 D より, 次の行列を得る.

$$A = (a_{ij}), a_{ij} = -\frac{1}{2}d_{ij}^2 \quad (1)$$

次に doubly-centered 行列 $B = CAC$ を作成する. ここで C はセンタリング行列である.

$$C = I_n - n^{-1}J_n, J_n = \mathbf{1}_n \mathbf{1}_n^t \quad (2)$$

J_n は $(k+1) \times (k+1)$ のすべての要素が 1 の行列である. そして, B の m 個の最大固有値 λ_i に対応する固有ベクトル v_i を求める. m 次元の低次元マッピングは以下のように与えられる.

$$\begin{aligned} Y &= V\Lambda^{\frac{1}{2}} = (\sqrt{\lambda_1}v_1, \dots, \sqrt{\lambda_m}v_m) \\ &= (y_0, y^1, \dots, y^k)^t \end{aligned} \quad (3)$$

y^i は元の高次元のサンプル x^i の principal coordinate と呼ばれる. このとき principal coordinate 同士の距離は元の距離行列 D の対応する距離と等しくなる.

4. Procrustes Analysis

前節で学習サンプル X の低次元表現 Y が得られた. 次に, Y と姿勢パラメータ $P^{(k)}$ の位置合わせを行うための変換を求める. LPR では適合度検定を用いて $P^{(k)}$ の各要素 p^i に関して, Y のすべての y^i (ただし y_0 を除く) において二乗残差が最小となる値を求める.

$$R = \min_{A,b,s} \sum_{i=1}^k (p^i - sA^t y^i - b)^t (p^i - sA^t y^i - b) \quad (4)$$

ここで $(m \times m)$ の直交行列 A は m 次元の回転, 鏡映を表し, b は並進, s はスケールを表す. この 3 つの要素を用いて Y と $P^{(k)}$ の位置合わせを行う. Procrustes

Analysis では A, b, s は以下のように与えられる [5].

$$A = VU^t, Z = Y^t P^{(k)} = VLU^t \quad (5)$$

$$b = \bar{p} - sA^t \bar{y}, \bar{p} = \frac{1}{k} \sum_{i=1}^k p^i, \bar{y} = \frac{1}{k} \sum_{i=1}^k y^i \quad (6)$$

$$s = \text{trace}(L) / \text{trace}(YY^t) \quad (7)$$

式 (5) において, $Z = VLU^t$ は $Y^t P^{(k)}$ の特異値分解である. ここで V と U は $(m \times m)$ の直行列, L は特異値からなる対角行列である. Y と $P^{(k)}$ の位置合わせを評価する残差 R は以下のように表される.

$$\begin{aligned} R &= \text{trace}(PP^t) + s^2 \text{trace}(YY^t) \\ &\quad - 2s \text{trace}(P^t YY^t P)^{1/2} \end{aligned} \quad (8)$$

A, b, s が求められたら, テスト画像 x_0 の姿勢パラメータ p_0 の推定値は低次元表現 y_0 を用いて以下のように与えられる.

$$p_0 = sA^t y_0 + b \quad (9)$$



図 2 実験に用いたデータベースの一部

5. 実 験

ここでは典型的な姿勢推定課題に対して提案手法を用いた結果を述べる. 本論文では公開データセット Object Pose Estimation Database (OPED) [6][7] を用いた. このデータセットは 16 物体の画像群からなり, 2 つの回

転軸 (y 軸まわりの回転 pan: 0 度?180 度, x 軸まわりの回転 tilt: 0 度?90 度) において 5 度ずつサンプリングされている．そのためこのデータセットには各物体に 703 枚の画像が含まれる．各物体の例を図 2 に示す．

まず，各物体でランダムに選んだ 100 枚の画像をテストサンプルとし，残りの 603 枚を学習サンプルとする．そして各テストサンプルに対して姿勢パラメータを推定し，得られた推定誤差の平均を評価に用いる．表 1 に各物体の結果を示す．LPR においては k 近傍 ($k = 10$) を用いて実験を行った．今回比較手法として，学習サンプルと姿勢パラメータ空間との全体的なマップを学習する線形回帰 と Support Vector Regression (SVR) [8]，テストサンプルと最も近い学習サンプルを選ぶ nearest neighbor (NN) 法を用いた．評価を簡単にするために，2 つの姿勢パラメータ (pan, tilt) を球座標上のベクトルへと変換し，真値との誤差をベクトル同士の角度として表した．表 1 からわかるように，LPR は $k = 10$ において 16 物体中 7 つにおいて最も誤差が小さい．一方 SVR においては 16 物体中 6 つである．

表 1 実験結果 (単位: 度)

Object	NN	SVR	Linear	LPR
adapter	3.39	1.70	1.87	1.11
bay	5.70	1.65	1.72	2.43
cable	3.39	0.82	0.79	0.81
cap	3.37	0.82	1.13	3.39
clamp	3.72	2.15	1.58	1.23
fuse	3.57	1.12	0.83	1.41
goldcar	3.40	1.36	1.24	1.13
house	4.11	1.19	1.94	1.86
ipipe	4.76	1.18	1.70	3.79
redcar	3.46	1.40	1.77	1.14
socketin	3.70	0.84	1.34	2.17
socketout	3.70	0.84	1.01	1.83
tpipe	4.56	1.74	1.74	1.21
trap	3.67	0.92	0.82	1.76
whitecar	3.82	1.15	1.53	0.48
wood	3.41	2.12	1.09	0.90

図 3 は物体 “clamp” の場合に低次元の画像空間と姿勢パラメータ空間との位置合わせを行った結果の一例である．図 3 ではテストサンプルの 10 近傍の学習サンプルの値とその推定値も表示している．このとき，学習サンプルの真値の姿勢パラメータ (pan, tilt) の角度を赤丸，学習サンプルの低次元空間でマッピングされた位置を青星印で表している．テストサンプルにおいては正しい姿勢パラメータを黒十字 (135, 0)，推定された姿勢パラメータを緑十字 (135.9, -1.1) で表している．

MATLAB 上での LPR の実行時間は，1 枚のテストサンプルの姿勢パラメータを 603 枚の学習サンプルから推定するには 0.5 秒以下であった．そのうち最も時間がかかっているのは，学習サンプルからテストサンプルの k

近傍を求める計算であった．今回用いたデータセットにおいてはよい姿勢推定結果が得られた．しかし，より大きなデータセットにおいては近傍を求める際に計算時間を削減するためより効率の良い手法が必要となる．

6. おわりに

本論文では新しいアピアランス空間の姿勢推定である LPR を提案した．提案手法は，未知の姿勢情報をテストサンプルとその近傍サンプルとの距離を用いて局所的に推測し，Procrustes Analysis を用いてテストサンプルの近傍サンプルより作成された低次元の画像空間とそれに対応する姿勢パラメータ空間との位置合わせを行う．OPED データセットを用いて行った実験結果において，提案手法は良い結果が得られた．今後の予定は，計算時間の削減やより正確な姿勢推定を行うために，距離を保存して低次元マッピングを行う手法について MDS の代わりに多様体学習手法 [9-12] を用いることや，低次元画像空間と姿勢パラメータ空間の位置合わせを行うよりよい手法を考案することである．

文 献

- [1] V. Lepetit and P. Fua, “Monocular Model-Based 3D Tracking of Rigid Objects: A Survey,” *Foundations and Trends in Computer Graphics and Vision*, 1(1), pp. 1-89, 2005
- [2] H. Murase and S. K. Nayar, “Visual Learning and Recognition of 3-D Objects from Appearance,” *Int. Journal of Computer Vi-sion*, 14(1), pp. 5-24, 1995
- [3] T. Cox and M. Cox, *Multidimensional Scaling*. 2nd ed. Chapman & Hall/CRC, 2000
- [4] W. Torgeson, “Multidimensional Scaling: I. Theory and method,” *Psychometrika* 17, pp. 401-419, 1952
- [5] K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis*. Academic Press, 1979
- [6] F. Viksten, P. E. Forssen, B. Johansson, and A. Moe, “Comparison of Local Image Descriptors for Full 6 Degree-of-Freedom Pose Estimation,” *in IEEE International Conference on Robotics and Automation*, May 2009
- [7] <http://www.cvl.isy.liu.se/research/objrec/posedb/>
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006
- [9] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science* 290, pp. 2319-2323, 2000
- [10] M. Belkin, and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,” *Adv. NIPS* 15, Vancouver, Canada, 2001
- [11] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, “Graph Embedding and Extensions: A General Framework for Dimensionality Rediction,” *IEEE Trans. PAMI*, 29-1, pp. 40-51, 2007
- [12] S. T. Roweiss, and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” *Science* 290, pp. 2323-2327, 2000

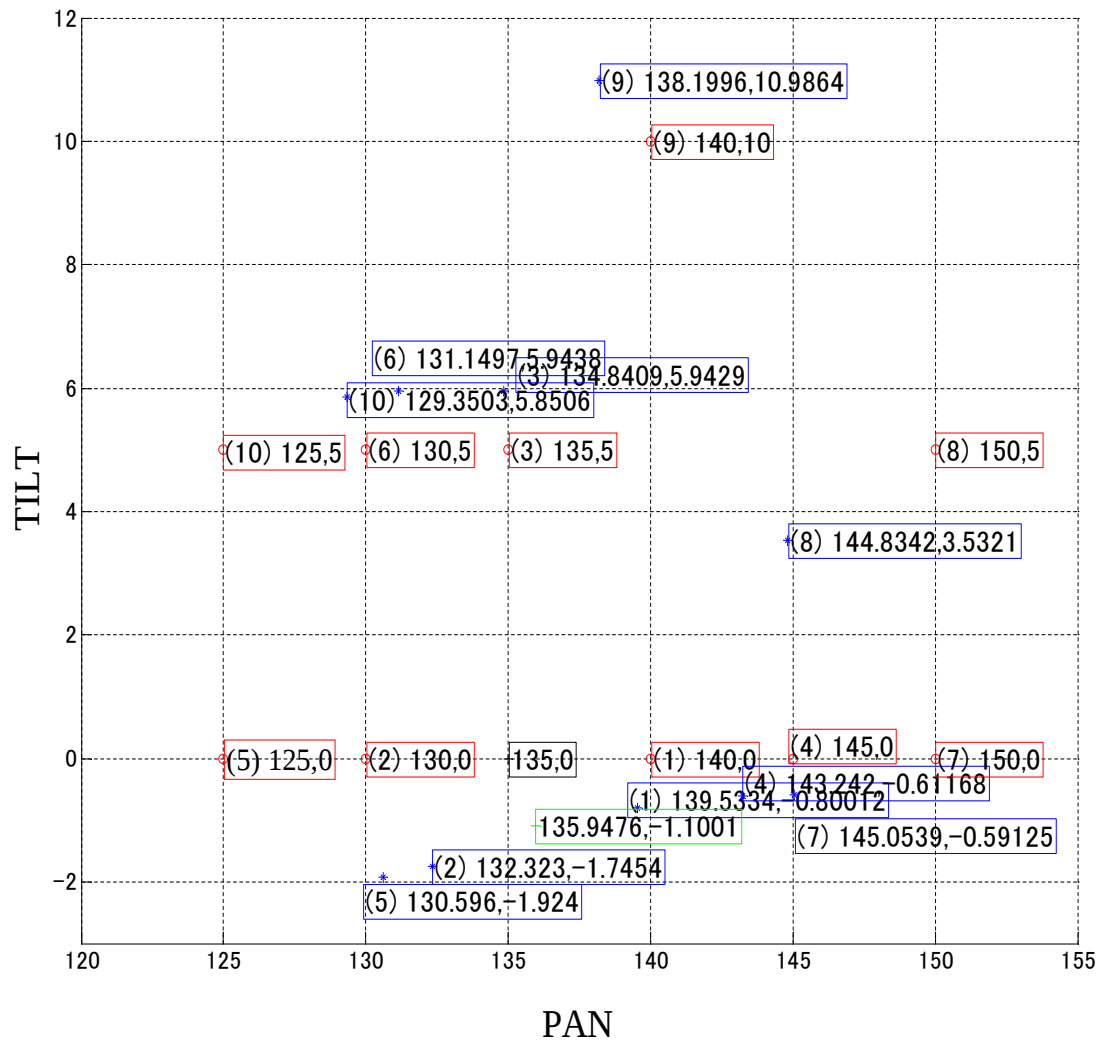


図 3 “clamp” のテストサンプルについて LPR を適用した際のマッピング結果の一例 (赤丸) 学習サンプルの真値の姿勢パラメータ (青星印) 学習サンプルの画像のマップ (黒十字) テストサンプルの真値の姿勢パラメータ (135, 0) (緑十字) 推定された姿勢パラメータ (135.9, -1.1) .