

2次元 LMedS クラスタリングによる複数直線の推定

北海道大学 ○乾健太郎 金子俊一 五十嵐悟

要旨

2次元パラメータ空間より独自に提案したクラスタリングアルゴリズムに基づき複数直線のそれぞれのパラメータを推定する。また、例外値と真値とを区分することにより、より多くの例外値を含む点群からの直線の推定や、それを応用し複数直線の推定を行った。

1. はじめに

近年ロバストな統計的推定手法の画像認識・理解への応用が行われている[1][2]。そこでは、個々には「ゆらぎ」をもつ多くの手掛かり（データ）を利用するための、従来の全データにわたる最小2乗評価に基づく算術平均処理に替わるロバストな手法が指向されている。我々は LMedS クラスタリングによって、例外値の影響を低減する手法を提案している[3]。多くの例外値を除去する特性は、ロバスト統計量である最小2乗 median 推定 (LMedS) の挙動によく似ている。

本論文では、1次元 LMedS 推定アルゴリズム (配置推定) を拡張した2次元 LMedS クラスタリングアルゴリズムを提案する。クラスタリングは教師なし学習とも呼ばれ、データ集合に関する事前知識が少ない場合のパターン認識手法として重要である。その応用例として、2次元点群への直線あてはめを扱う。また、例外値を多く含む直線推定と、それを応用し、複数直線構造の推定を行った。

2. 2次元 LMedS クラスタリングアルゴリズム

我々は点データの辞書式順序付けを併用した準厳密アルゴリズムを提案している[3]。ロバストクラスタリングは画像パターン認識においては非常に効果的な手法として期待できる。パターン識別アルゴリズムの基礎として、また複数構造の推定などへの拡張が可能であると考えられる。ここでは、その基本を述べる。

2次元に散布する点群の重心(u^*, v^*)を求めるとき、最小2乗 median 推定を用いる。誤差量 (距離) を次のチェス盤距離 (2乗値)

$$d^2 = \min_x \left\{ (u - u^*)^2, (v - v^*)^2 \right\}$$

で近似することにより、重心推定問題は過半数の点を含む最小周囲長の正方形 (通常は円あるいは楕円) の中心座標を求める探索問題となる。ここでは各座標値の統計的スケール (分散あるいは標準偏差) が一様な場合を扱う。このような条件は画像上のパターンを扱う場合にはよく成立すると考えられる。すなわち、 x, y 座標値に含まれる統計誤差は等価の場合が多い。

次にアルゴリズムを示す。

- (0) $\{T_i(x_i, y_i)\} (i=1, 2, \dots, n)$: xy 座標値をもつ総数 n 個の終点候補集合。 $h = \lfloor n/2 \rfloor + 1$, $hh = \lfloor n/2 \rfloor$. $S[i, j, v]$: T_i および $T_j (i < j)$ を含む、辺の長さ : $u_j - u_i (=W)$, 中心座標 : $(u_c = (u_i + u_j)/2, v_i + W/2 < v_c < v_j - W/2)$ の正方形領域。 /* 準備 */
- (1) $\{T_i\}$ を座標値 ($u_i = x_i, v_i = y_i$) の辞書式順序に整列する。 $u^* = 0$,

$v^* = 0, W^* = \max\{u_n - u_1, v_n - v_1\}$. /* 初期化 */

- (2) $i = 1$. /* 左境界点の初期化 */
- (3) $i < n - hh$ の間、以下を繰り返す。 $i = n - hh$ のとき、 [9] へ。
- (4) $j = i + hh$. /* 右境界点の初期化 */
- (5) $j < n + 1$ の間、以下を繰り返す。 $j = n + 1$ のとき、 [8] へ。
- (6) $S[i, j, v]$ を算出する。 $W > W^*$ であれば、 [8] へ。
- (7) $S[i, j, v]$ の何れかが h 個以上の点を含み、かつ $W < W^*$ の場合には、 $u^* = u_c, v^* = v_c, W^* = W$ とする。 /* この正方形領域の中心座標あるいは重心座標の LMedS の候補となるので、最小領域候補として登録する */
- (8) $j = j + 1$. [5] へ。 /* 右境界点の更新 */
- (9) $i = i + 1$. [3] へ。 /* 左境界点の更新 */
- (10) $\{T_i\}$ を座標値 ($u_i = y_i, v_i = x_i$) の辞書式順序に整列する。 [2] へ戻る。 /* 縦・横を交換して探索を続ける */
- (11) 終了時に、誤差評価量 $W^*/2$ を得る。

発見された最小正方形領域は余分の空白を含む。これはチェス盤距離を用いたことによる。処理後、トリミングを行い、最小長方形としてその中心を算出して用いることとする。

この提案アルゴリズムの計算オーダーは、[1]の作業で $O(n \log n)$, [2]~[5]の作業で最悪 $O(n^3)$, これに正方形の生成と検査で、 $O(n)$ である。したがって、すべての作業で $O(n^3)$ となり、これは整列せず、 n 点から3点を選択し、正方形を生成させる手法の $O(n^4)$ に対し、効率的である。

3. 直線検出

次に点群にあてはまる直線回帰問題にロバストクラスタリングを応用し、直線を検出する。ここでは図1に示すように、画像面上の点群の中から任意の2点 P_i, P_j を選び、これがあてはめべき回帰直線の1つの候補直線と考える..そして、この平面と平行に置かれた球面座標系のレンズ中心 O と点を結びベクトル V_i, V_j を作成する。この V_i, V_j の外積を球面座標系に変換する。この変換では画像面上の位置誤差は一様に球面座標上に投影される。変換式を次に示す。

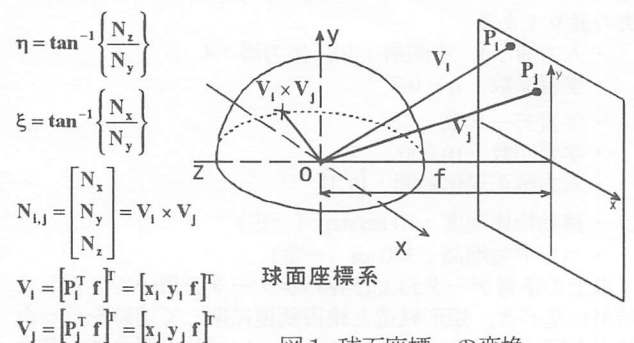


図1 球面座標への変換

この変換では、同じ直線上の点群ならば変換後、一点に集中する。そして、2章で示した LMedS クラスタリングを球面座標系において実行する。そこで直線のパラメータを得、結果を直線に戻すという手法である。この手法を用いてシミュレーション実験を行った。直線は、図3に示すように量子化幅 0.010mm に対し、x 方向、y 方向に標準偏差値 0.02mm となる点群に例外値を含ませている。図2は最小2乗法による結果、図3は LMedS クラスタリングによる結果を示している。

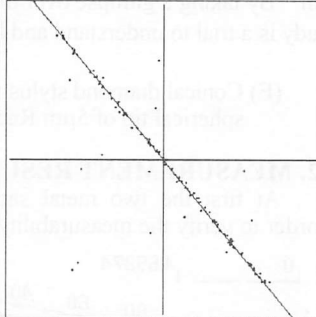
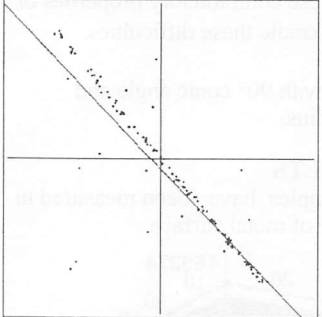
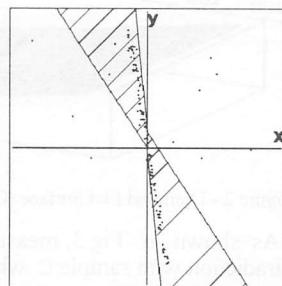
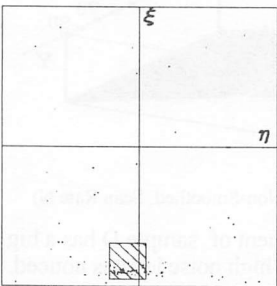


図2 最小2乗法による直線推定 (例外値 約15%の場合)

図3 本手法による直線推定 (例外値 約15%の場合)

4. 例外値の除去

3章では、例外値が比較的少ない場合での直線を推定した。しかし、理論的には、例外値の含有率が約30%を超える場合には、この手法を用いることができない。そこで、球面座標系でもとめられた最小領域(正方形)に対応する xy 空間での領域を求めた。図4は、このそれぞれに対応する領域を示している。例外値の含有率が30%を超える点群で直線を推定した場合、球面座標系で求めた最小領域には、まだ例外値を含んでいる可能性が高い。そこで最小領域外の点群を例外値とし、その最小領域に対応する xy 空間の領域外の点も例外値と考える。そしてその例外値を除去する。これを繰り返し例外値の含有率を減らすことにより、正確な直線を得ることが可能となった。



(1) ξ η 空間(球面座標系)

(2) xy 空間

図4 球面座標系の領域に対応する xy 平面の領域

5. 複数直線の推定

4章で述べた手法を応用し、複数直線を推定する手法を提案する。以下にその定式化したアルゴリズムを示す。また図6は推定の結果を示している。

- (0) k 本の直線を推定し、i 回例外値を除去した xy 平面座標系の点群(N 点)を $\{P_{i,j,k}\}(x_{i,j,k}, y_{i,j,k})$, LMedS クラスタリングの最小正方形の辺 W, 重心 (ξ_k, η_k) , 最小領域の $\xi_{max}, \xi_{min}, \eta_{max}, \eta_{min}$ /*初期値の設定*/
- (1) $i=0, n=N$

- (2) 点群 $\{P_{i,j,k}\}$ から球面座標系①に変換し、LMedS クラスタリング /*2章*/ を行う。
- (3) $W < T_{wide}$ ならば(10)へ /*直線の評価*/
- (4) $j=0, l=0$ ($j < n$ の間を繰り返す)
- (5) $x_{i,j,k} < 0$ ならば(7)へ /*例外値の除去*/
- (6) $-\frac{x_i}{\tan(\xi_{min})} + f \times \tan(\eta_{min}) \leq y_{i,j,k} \leq -\frac{x_i}{\tan(\xi_{max})} + f \times \tan(\eta_{max})$ ならば、例外値を除去した点群 $\{P_{i+1,j,k}\}, l=l+1$
- (7) $-\frac{x_i}{\tan(\xi_{max})} + f \times \tan(\eta_{min}) \leq y_{i,j,k} \leq -\frac{x_i}{\tan(\xi_{min})} + f \times \tan(\eta_{max})$ ならば、例外値を除去した点群 $\{P_{i+1,j,k}\}, l=l+1$
- (8) $j=j+1$ とし(5)へ
- (9) $\{P_{i,j,k}\} = \{P_{i+1,j,k}\}$ ならば終了 /*停止条件*/
- (10) $i=i+1, n=n-1$ とし(2)へ
- (11) $k+1$ 本目の直線のパラメータ (ξ_k, η_k) を得る
- (12) $j=0, l=0$ ($j < n$ の間を繰り返す)
- (13) $x_{0,j,k} < 0$ ならば(14)へ /*点群から $k+1$ 本目の直線上の点群を除去*/
- (14) $-\frac{x_i}{\tan(\xi_{min})} + f \times \tan(\eta_{min}) \geq y_{i,j,k}$ または $y_{i,j,k} \geq -\frac{x_i}{\tan(\xi_{max})} + f \times \tan(\eta_{max})$ ならば新たな点群 $\{P_{0,l,k+1}\}, l=l+1$
- (15) $-\frac{x_i}{\tan(\xi_{max})} + f \times \tan(\eta_{min}) \geq y_{i,j,k}$ または $y_{i,j,k} \geq -\frac{x_i}{\tan(\xi_{min})} + f \times \tan(\eta_{max})$ ならば新たな点群 $\{P_{0,l,k+1}\}, l=l+1$
- (16) $j=j+1$ とし(12)へ
- (17) $l < T$ ならば終了 /*停止条件*/
- (18) $k=k+1, N=N-1$ とし(1)へ

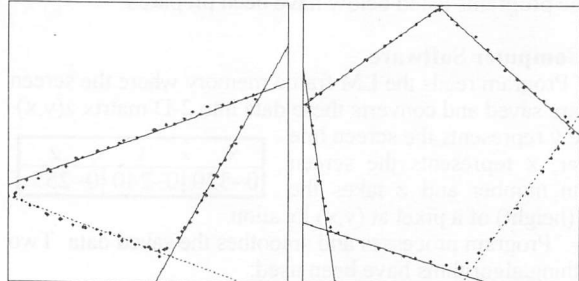


図6 複数直線の推定結果

6. まとめ

2次元 LMedS を導入したロバストクラスタリングアルゴリズムを提案し、画像面上の直線回帰問題へ応用することを試み、良好な結果を得た。また例外値を除去する手法を用い、例外値の含有率が大きい場合にも、直線を推定することができた。また、複数直線の推定に成功した。

参考文献

- [1] R.J.Rousseeuw and A.M.Leroy: Robust Regression and Outlier Detection, John Wiley & Sons, 1987.
- [2] B. Kamgar-Parsi, B. Kamgar-Parsi, and Nathan S. Netantahu: A Nonparametric Method for Fitting a Straight Line to a Noisy Image, IEEE Trans. of PAMI, vol.11, No.9, pp.998-1001, 1989.
- [3] 乾健太郎, 金子俊一, 五十嵐悟. “最小2乗メディアンクラスタリングによる直線検出” 第4回画像センシングシンポジウム講演論文集. pp.241-246.