

混合分布推定に基づく単一カメラによる動画像の人物検出に関する研究 分布パラメータの比較による人物の増減判定

○小関 英剛[†] 森下 壮一郎[‡] 浅間 一[‡]

[†]東京大学 工学系研究科 [‡]東京大学 人工物工学研究センター

A Study for Multiple Person Detection from a Monocular Image Sequence Based on Normal Mixture Model

Discrimination of Decrease or Increase the Number of Persons Comparing Distribution Parameters

○Hidetaka Koseki[†], Soichiro Morishita[‡], and Hajime Asama[‡]

[†]School of Engineering, The University of Tokyo

[‡]RACE, The University of Tokyo

Abstract: Detection and analysis of trajectories are effective for providing services to pedestrians in public spaces. We have proposed a method to detect moving objects from a monocular image sequence with a normal mixture model. However, the method cannot accommodate a changing number of moving objects. As described herein, to overcome this shortcoming, we proposed two algorithms. One algorithm detects a increase of pedestrians, the other detects a decrease of pedestrian. We did an experiment using image sequences from an actual environment to verify the availability of the proposed method.

1 緒言

公共空間等の広い場所における歩行中の人物へサービス提供を行うには、人物の移動軌跡を取得し、適切な解析をすることが有効である [1]。歩行者の位置検出に関する研究は数多く行われているが、精度は高いもののコストが高いセンサを必要としたり、一般的な環境への導入が困難であったりすることが多い。実用化のコストを考慮すると、既に防犯カメラとして普及している単一カメラを用いた検出手法が有用である。今までに我々は単一カメラを用いた混合分布推定による複数人物の検出手法を提案している [2]。これは人数が既知のとき移動物体領域からいくつかの点をサンプリングして、これらを混合正規分布に従うものとみなし、混合分布推定を用いることで人物に対応する領域を決定するものであった。しかしながらこれは人数が未知のときには適用できなかった。本稿では単一カメラ画像の人数変化を検出することで画像中の人数を計数する手法を提案する。さらに実環境で取得した動画像に提案手法を適用する実験を行い、その有効性を検証する。

2 従来研究

単一カメラによる移動物体計測としては、(1) テンプレートマッチングを用いるもの (2) 時空間情報を用いるものが挙げられる。

テンプレートマッチングを用いるものとしては次のようなものが挙げられる。Curioらは車道を横断する歩行者の歩容をテンプレートとする手法 [3] を提案した。また広野

らは移動している人物をフレーム間差分で検出し、静止人物を移動人物より毎フレーム更新したテンプレートにより検出する手法 [4] を提案した。しかしながらこれらの手法はテンプレートによっては画面内の人物がカメラに対して向いている方向について頑健ではない。

時空間情報を用いるものとしては、Markov Random Field モデルを時空間モデルとして拡張した時空間 MRF モデルに基づく手法 [5] がある。この手法は移動物体の向きについては頑健であるものの、計算コストが高くリアルタイム処理を行うには高価な計算機を必要とする。

以上のことから本研究では人物の向きにより検出率が変動せず、背景差分画像を用いた計算コストの比較的低いアルゴリズムを提案する。

3 先行研究

本節では前述した先行研究について述べる。まず屋内公共空間等で移動物体のほとんどが人物であるような状況を仮定する。移動物体領域の抽出には背景差分法 [6] を用いる。そして抽出した移動物体領域の重心と分散共分散行列から移動物体領域を覆う楕円を得て、その楕円の長軸の先端を足元とみなす [7]。

動画像中の人数が複数のとき、人数分の重心と分散を求めるには、EM アルゴリズムによる混合分布推定 [2] を行う。以下にその手続きを述べる。移動物体領域の画素から取得したサンプル点の総数を N とし、 n 番目のサンプル点の座標を $x_n = (x_n, y_n)$ とする。EM アルゴリズムの E

ステップにおいて、適当な初期の混合度 $\hat{\alpha}$, 重心 $\hat{\mu}$, 分散共分散行列 $\hat{\Sigma}$ が与えられたとき、各サンプル点 x_n について、それが合計 K 個中の k 番目の混合正規分布によって生成される事後確率 $Q_n(k)$ を推定する。

次にEMアルゴリズムのMステップから、 $Q_n(k)$ を用いてもっともらしい α , μ , Σ を得る [8]。得られたパラメータを元に各画素が正規分布に属する確率 $p(x_n; \mu_k, \Sigma_k)$ を次の式 (1) より求める。

$$p(x_n; \mu_k, \Sigma_k) = \frac{\exp(-(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)/2)}{2k\pi |\Sigma_k|^{1/2}} \quad (1)$$

また事後確率 $Q_n(k)$ は以下の式から求める。

$$Q_n(k) = \frac{p(x_n; \hat{\alpha}_k, \hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{i=1}^K p(x_n; \hat{\alpha}_i, \hat{\mu}_i, \hat{\Sigma}_i)} \quad (2)$$

サンプル点が K 個の混合正規分布に従うとき、サンプル点のクラス数は K 個であると定義する。また本稿では k 番目の混合正規分布による $p(x_n; \mu_k, \Sigma_k)$ のクラスを M_k と定義する。

先行研究においては人数は既知として、クラスの追加と削除を行った。しかし動画像の人数の初期値が既知で、かつフレーム間での人数の増減が検出できれば、1フレームごとに人数の計数を行わなくても、複数人物の位置同定に対応できる。次節からその手法について述べる。

4 提案手法

本研究で用いた動画像は、人物の増減に対してフレームレートは十分大きい。そこでフレームごとに画面内に入り出す人数は、最大1人ずつであると仮定した。人数の増減に対応したクラスの追加と削除を、以下の手法を組み合わせることで複数人物の検出を行う。

4.1 人数増加判定

1つのクラスにおける人数判定

画面内の人数とクラス数 K が異なると、適切な分散共分散行列が得られず、楕円領域の大きさ及び傾きが適切に人物同定している状態と大きく異なる。森下らは、あらかじめ人物が1人の画像を取得して、1人のときの分散共分散行列を計算した。本研究ではこの計算結果を推定値と定義する。そして実際に計測した人物領域から求めた重心座標と分散共分散行列との比較を行い、楕円領域内の人物の人数判定を行った [9]。本稿ではクラス M_k に関する人数判定を、以下の式 (3) を用いて評価する。

$$\beta(M_k) = \begin{cases} S_1 S_2 / S'_1 S'_2 & (S_1 S_2 < S'_1 S'_2) \\ S'_1 S'_2 / S_1 S_2 & (S'_1 S'_2 \leq S_1 S_2) \end{cases} \quad (3)$$



Fig. 1 複数のクラスが割り当てられた失敗例

式 (3) において、実測値から得た楕円の長軸短軸長をそれぞれ S_1, S_2 , 推定値による楕円の長軸短軸長をそれぞれ S'_1, S'_2 とする。本研究では $\beta(M_k)$ を適合度と定義する。

画面内の人数が増えたとき、クラスを追加しないと、 $\beta(M_k)$ は大きく低下する。一方で人数が減少すると、Fig.1 のように1人の人物領域に複数のクラスが割り当てられる。このとき適合度 $\beta(M_k)$ は画面内の人数が増えたときの適合度ほど低下しない。またオクルージョンによる適合度 $\beta(M_k)$ の低下との判別が難しい。そこで、適合度 $\beta(M_k)$ は人数が増えたときの判定のみに用いる。楕円領域内の人数が1人か複数かを判断する閾値を θ_1 とする。 $\beta(M_k) > \theta_1$ のとき実測値の楕円領域は1人の人物とみなす。

クラス追加領域の設定

人数が変化していないときに、実測値の楕円領域の適合度が $\beta(M_k) < \theta_1$ となる状況は次の2つが考えられる。(1) 人物が画面端に移動した場合、(2) 画面内にノイズが生じた場合である。(1) においては動画像に全身が映らない状態で分散共分散行列を求めるので、適切な適合度が得られない。そこで画面端の全身が映らない領域では、適合度の判定は行わない。本稿では動画像よりひと回り小さな矩形領域を設定し、矩形領域内に適合度の低いクラスの重心がある場合のみクラスの追加を行う。(2) に関して、本稿で用いる背景差分画像のノイズは適合度に影響の無い程度に小さい。しかしノイズの無い動画像であっても適合度は常に変動する [9]。本稿ではスパイク状のノイズへ頑健性を高めるため動画像の直近 γ フレーム間のクラス M_k の適合度の最大値を $\beta(M_k)$ とする。

クラスの追加

動画の t フレーム目に、あるクラス $M_m(t)$ ($m = 1, \dots, K$) の適合度が $\beta(M_m) < \theta_1$ のとき、次のフレームで $M_m(t+1)$ に従うサンプル点を楕円短軸で分割し、2つのクラス $M'_m(t+1), M'_{K+1}(t+1)$ にそれぞれ従うものとする。 M_m のサンプル点の座標ベクトルを $x = (x_m, y_m)$, 重心ベクトルを $\mu = (\mu_x, \mu_y)$, 楕円の傾きを ϕ と仮定するとき、楕円短軸の直線は以下の式で示せる。

$$f(x) = \tan \phi (x - \mu_x) + \mu_y \quad (4)$$

式 (4) を用いて、サンプル点は以下のように分配する。

$$M_m = \begin{cases} M'_m & = \{x | f(x_m) \leq y_m\} \\ M'_{K+1} & = \{x | f(x_m) > y_m\} \end{cases} \quad (5)$$

ただし $\phi = 90$ [deg] のときは式 (6) のように分配する。

$$M_m = \begin{cases} M'_m & = \{x | x_m \leq \mu_x\} \\ M'_{K+1} & = \{x | x_m > \mu_x\} \end{cases} \quad (6)$$

クラスを追加する前後の概念図を Fig.2 に示す。

クラス追加前後のクラス対応付け

適切な移動軌跡を取得するために M_m が、 M'_m と M'_{K+1} のどちらに対応するか判定する。Fig.2 のフレームのみからクラス M_m が、 M'_m と M'_{K+1} のどちらに対応するかの判別は困難である。しかし $t-\gamma$ フレームは適合度が下がりはじめた直後のフレームであり、画面内に新たな人物領域の一部が入ったフレームと判断できる。すなわち $t-\gamma$ フレームでの M_m の重心座標は画面内に既にある人物の重心座標に近いと判断できる。そこで、 t フレーム目でクラスを分割した後、 $t-\gamma$ フレーム目まで時間をさかのぼって人物同定を行う。 $M'_m(t-\gamma)$ と $M'_{K+1}(t-\gamma)$ の重心座標のうち、 $M'_m(t-\gamma)$ の重心座標に近いクラスを t フレーム以降の M_m と定義する。

4.2 クラスの削除

画面内の人数が減少したとき、クラス数 K を削除しないと Fig.1 のように 1 人の人物領域のサンプル点に複数のクラスが割り当てられる。このとき適合度 $\beta(M_m)$ は、あるクラス M_m の楕円領域の人数が 1 人か複数かを判断する指標であるため、人数の減少の判定は行えない。以下に人数の減少を判定するための手法を述べる。

クラス M_m で同定している人物が画面内にいるとき、その人物領域のサンプル点の事後確率 $Q_n(m)$ は m 番目以外のクラス M_k による事後確率 $Q_n(k)$ より大きくなる。クラス M_m が同定している人物が画面外へ出ると、全ての

サンプル点のクラス M_m による事後確率 $Q_n(m)$ はその他全てのクラス M_k による事後確率 $Q_n(k)$ より小さくなる。そこで式 (2) の事後確率を用いて人数の減少を判定する手法を提案する。画面内に K 人の歩行者を混合分布推定で適切に位置同定しているときを考える。クラス M_m によって位置同定をしている人物が、 t フレーム目で画面外へ完全に出ると t フレーム目の事後確率 $Q_n(m)$ は N 個全てのサンプル点で

$$Q_n(m) < Q_n(k)$$

$$(n = 1, \dots, N; k = 1, \dots, m-1, m+1, \dots, K) \quad (7)$$

が成り立つ。 M_m で同定している人物が外に出るとき、サンプル点がわずかに画面内に残ることも考えられる。そのため人物減少を判別する閾値 θ_2 を設定し、式 (7) が成り立つサンプル点が θ_2 個以上のとき人数が減少したと判断し、クラス M_m を削除する。

4.3 提案手法の流れ

画面内に人がいない状態から検出を行う。毎フレーム EM アルゴリズムを適用し、クラス M_m について以下の条件に従いクラスの増減を行う。

1. $\theta_1 > \beta(M_m)$ かつ μ_m が矩形領域内に存在する場合
→クラス数を 1 追加
2. 式 (7) を満たすサンプル点の数が θ_2 個以上の場合
→クラス数を 1 削除

5 実験

5.1 実験環境

提案手法の有効性を検証するために、東京大学柏図書館の一階入口上部にカメラを設置して得られた映像に提案手法を適用する実験を行った。

5.2 実験方法

背景画像は M-estimation による適応的背景推定手法 [6] を用いて作成した。また実験で用いたパラメータは経験的に Table.1 のように設定した。また本稿で使用した動画内の人数は最大 2 人である。

5.3 実験結果

画面内人数が最大 2 人の動画像に提案手法を適用した結果を Fig.3 に示す。過去 5 フレームの適合度が $\theta_1 < \beta(M_1)$ となった次フレームの Fig.3(b) においてクラス追加が行われている。また Fig.3(c) において、人物が 1 人画面外へ出ると、クラスの削除が適切に行われている。このとき、2 人の人物領域の重心軌道を Fig.4 に示す。Fig.4 のグラフ中の数字は位置同定したときのフレーム番号をあらわす。

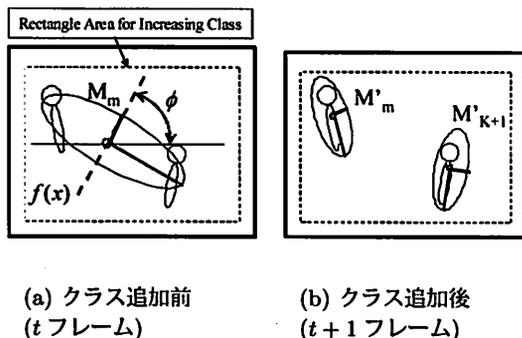
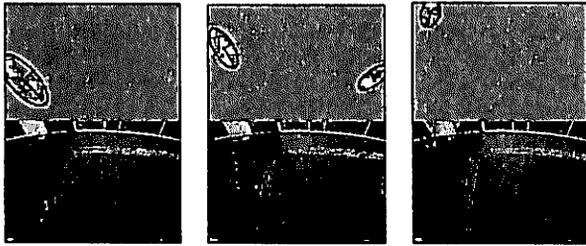


Fig. 2 クラス追加の概念図



(a) 1人を同定 (b) クラス追加 (c) クラス削除

Fig. 3 位置同定の成功例

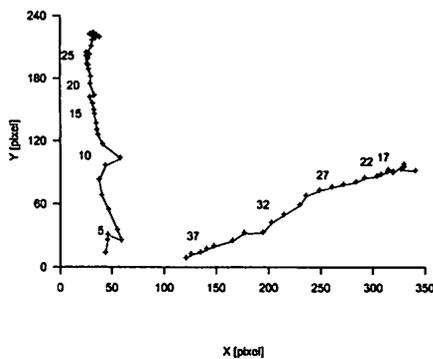


Fig. 4 2人の重心軌跡

次に人数が最大2人のFig.3を含む4つの動画に提案手法を適用した。全フレームを目視で評価し、Fig.3のように各楕円が各人物領域をそれぞれ覆うときに正しく判別されたものとした。結果、全フレーム272フレーム中に成功したフレームは223フレームで、判別率は82.0%であった。単一カメラを用いてかつ背景差分で移動物体検出を行ったことを考慮すれば良好な結果が得られた。

6 結言

本稿では単一カメラによる動画中の複数人物の自動検出方法を提案した。人数の増加に対応するために、移動物体領域の分散共分散行列の推定値と実測値を比較し、人物の増加の判定を行った。そして1つのクラスに複数の人物が含まれると判断した場合、クラスの分割を行った。また人数の減少に対応するため、各サンプル点の事後確率を観

Table 1 実験で用いたパラメータ

フレームレート	15 [fps]
画像サイズ	352 × 240 [pixel]
サンプリング数 n	230
人数判定の適合度 $\beta(M_k)$ の閾値 θ_1	0.3
$\beta(M_k)$ を保持するフレーム数 γ	5
矩形領域の左下, 右上座標	(30,30),(322,200)
人数減少判別の閾値 θ_2	225

測することで、人数の減少を判定した。さらに提案手法を実環境で取得した動画に適用し、有効性を確認した。今後の展望として多数の歩行者がいる動画を用了有効性の確認や、非線形モデルを用いて、人数増減の判別率を高める手法の検討などが挙げられる。

謝辞

本研究は、次世代ロボット知能化技術開発プロジェクト(独立行政法人 新エネルギー・産業技術総合開発機構)における「ロボストに作業を実行するための作業知能モジュール群の開発」の成果の一部である。

参考文献

- [1] 西村 他: “隠れマルコフモデルによる歩行パターンからの目的地推定 —推定結果評価手法の検討—”, 日本機械学会, ROBOMEC07, Vol.7, No.2, pp. 2P1-C10(1)–2P1-C10(3), 2007.
- [2] 広瀬 他: “混合分布推定による動画からの複数人物の位置検出”, 計測自動制御学会, 第8回 SICE SI 部門講演会講演論文集, 3L2-2, 2007.
- [3] C. Curio et al.: “Walking Pedestrian Recognition”, IEEE Transactions on intelligent transportation systems, Vol.01, pp.155–163, 2000.
- [4] 広野淳之 他: “オフィス内人分布計測システム”, (特集 松下電工 東京本社ビル), 松下電工技報 02855054, 松下電工マーケティング部, 2003.
- [5] S. Kamijo et al.: “Traffic Monitoring and Accident Detection at Intersections”, IEEE TRANSACTION on ITS, Vol.1, No.2, pp.108–118, Jun., 2000.
- [6] 上端 他: “動画における移動物体抽出のための相関分析による影領域判定手法とその評価に関する研究”, 日本機械学会, ROBOMEC07, Vol.7, No.2, pp. 2P1-C09(1)–2P1-C09(4), 2007.
- [7] 森下 他: “固定カメラ画像における人物領域の重心からの足元位置同定に関する研究”, 計測自動制御学会, 第7回 SICE SI 部門講演会講演論文集, pp. 1380–1381, 2006.
- [8] A. P. Dempster et al.: “Maximum likelihood from incomplete data via the em algorithm”, Journal of Royal Statistical Society, Series B, Vol.39, No.1, pp.1–38, 1977.
- [9] 森下 他: “動画中の人数判定のための移動物体領域の分布パラメータ推定”, 計測自動制御学会, 第8回 SI 部門講演会講演論文集, 3L2-1, 2007.