

混雑した環境における歩行者数推定 のための疎な解を持つ学習法の研究

増田 慎也^{†1} 下坂 正倫^{†1}
森 武俊^{†1} 佐藤 知正^{†1}

混雑環境において人数推定を行うことは、都市環境における安全面や商業面の観点から役立つ。近年、人周りの領域の面積などといった、画像全体から得られる特徴量を用いて、回帰により人数推定を行う手法が目立っている。この手法の推定精度は、用いる人周りの領域抽出手法・特徴量に依存する。そこで本研究は、適切な領域抽出手法・特徴量を自動的に選出する学習法を提案する。提案手法は、特徴量選出の2段階のアルゴリズムから成り立ち、それぞれの段階に疎な解を持つことでしられる l_1 ノルム正則化を用いる。Lars という正則化パスを効率的に求める考え方により、提案手法はハイパーパラメータ決定を含む学習時間が他の手法より短い。本研究では、公開データセットを用いた実験を通し、提案手法の有用性を検証した。

Efficient Sparse Learning on Crowd People Counting

SHINYA MASUDA,^{†1} MASAMICHI SHIMOSAKA,^{†1}
TAKETOSHI MORI^{†1} and TOMOMASA SATO^{†1}

Counting people in crowded environments is valuable for applications such as traffic analysis, safety in urban or advertising. Recently, number of people in the scenes can be estimated using holistic image features. The performance of this approach depends on the combination of motion segmentations and image features. Therefore, we propose a learning method which selects only useful features and segmentations from a large number of those automatically. Our method mainly consists of two stages. Our method contains l_1 -norm regularization, which is known as a shrinkage and selection technique, in two stages. Thanks to using Lars methods, which calculate a regularization path efficiently, the proposed approach can determine a hyper-parameter under less training time. Experimental results show our sufficient estimation accuracy and reduction of required features within less training time.

1. 緒 論

混雑した環境において、存在する人数を推定することは、混雑環境における安全面や交通整備などの観点から重要であり、さらに商業的にも役立つ。従来、画像による人数推定は、個人の発見¹⁾²⁾、追跡³⁾に基づき行われていた。しかしこれらの手法をそのまま混雑環境に適用することは、人々の重なりや計算時間の人数依存性の問題から困難である。そこで近年、対象人物まわりを領域抽出(セグメンテーション)し、対象領域内の面積、エッジ量といった画像特徴量から、回帰により人数推定を行う手法⁴⁾が提案されており、近年の様々な研究の基礎となっている⁵⁾⁶⁾⁷⁾。この手法は、人数に対して非依存な画像全体から得られる特徴量を利用し回帰により推定を行うのため、人々の重なりにも強く、かつ、計算時間が人数に非依存である特徴をもつ。これらは、混雑環境で人数推定を行うにあたり、大きな利点となる。また、この手法の推定精度は、利用する画像特徴量の組み合わせ方に依存し、画像特徴量がセグメンテーション領域から抽出されるために、利用する領域抽出手法にも依存する。精度向上のためには、現在提案されている様々なセグメンテーション手法(前処理)や画像特徴量をすべてを用いることが考えられるが、有用でない入力を引き起こす過学習と呼ばれる現象により推定精度は必ずしも向上しない。そこで、最適な組み合わせの前処理や特徴量のみを利用することが重要となる。また、有用な入力のみを利用することは、過学習を防ぐだけでなく、前処理や特徴量抽出の計算時間を削ることもつながる。このように、適切に利用する前処理、特徴量を選択することが重要であるが、人の手により適切な組み合わせを見つけることは、明らかに容易ではない。そこで我々は、最適な組み合わせを自動的に学習する手法を提案する。ここで、組み合わせ最適解を見つける問題は、対象数が多いとき、学習時間が膨大になってしまうことが多いことに注目し、本研究では効率的な選択を可能とする手法を考える。

有用な前処理、特徴量を選択するにあたり、最も単純な手法は、すべての組み合わせを探索する手法である。この手法は、最も適した組み合わせを見つけることが可能であるかもしれないが、探索には膨大な時間を要する。例えば、 g 個の前処理が存在し、前処理の組み合わせ選択を行うとする。このとき、前処理だけでも探索パターンが $2^g - 1$ となってしま

^{†1} 東京大学
University of Tokyo

う。これは g の増加に伴い、明らかに計算時間が破綻する。同様の組み合わせ探索手法として、前向き選択、後ろ向き選択があるが、これらも探索すべきパターンが $\frac{1}{2}g(g+1)$ となり、 g^2 に計算時間が比例するため、 g が大きくなると計算時間の観点から適用が難しい。以上の手法と異なるアプローチとして正則化項を用いた評価関数の最小化から、有用な入力を見つける Lasso⁸⁾ という手法も存在する。これは、前処理から得られるすべての特徴量を列挙、 l_1 正則化を利用し得られた解から、有用な前処理のみを判断する手法である。基本的な Lasso は特徴量選出しが行わないが、Lasso をグループ選択機能に拡張した Group Lasso⁹⁾ も存在する。しかし、GroupLasso は特徴量に対する選出を行わない。そこで、前処理と特徴量両方に疎な解を得るため、前処理と特徴量に対する選出方法を組み合わせた手法が必要である。Lasso の拡張により、1 つ基準で前処理と特徴量の両方に疎な解を得る SparseGroupLasso¹⁰⁾ が提案されているが、これにはハイパーパラメータが2つ存在する。2変数のハイパーパラメータの決定は、1変数の場合に比べて格段に難しくなる。そこで本研究では、効率的にハイパーパラメータ決定ができる Lars (Least angle regression)¹¹⁾ に基づく手法である Lars-Lasso¹¹⁾, GroupLars⁹⁾ の比較的共通した性質を持つ手法を2段階に組み合わせた手法を提案する。1段階目で、各前処理の有用性に基づき、探索範囲を $2^g - 1$ 通りから g 通りに減らす。2段階目で、各候補に疎な学習を行い、特徴量に対して疎な解を求める。Lars という考え方をを用いることで、正則化パスを高速に求め、交差検定によるハイパーパラメータ決定を容易に行う。そのため、効率的に解を求めることが可能となる。

最後に、この論文の構成は以下の通りである。2章で提案手法について詳細を述べ、3章で人数推定の概要を述べる。そして、4章に実験結果、5章に結論を示す。

2. 提案する疎な解を持つ学習法

混雑環境において人数推定を行うにあたり、線形回帰モデル

$$\hat{y} = \sum_{j=1}^g \sum_{i=1}^{d_j} w_{ji} \phi_{ji}(x), \quad (1)$$

を用いる。ここで入力ベクトル $\phi_j(x) = [\phi_{j1}(x), \dots, \phi_{jd_j}(x)]^T$ は、 j グループに属する特徴量を表し、 $w_j(x) = [w_{j1}, \dots, w_{jd_j}]^T$ は、入力ベクトル $\phi_j(x)$ に対応する重みベクトルを表す。 d_j は j グループに属する特徴量の数であり、 \hat{y} が推定出力、 x が入力画像である。

2.1 提案手法の概要

1節で述べたように、本研究の目的は大きく2つに分かれる。1つ目は、過学習を防ぐため利用する前処理、特徴量を選択することであり、2つ目は、短い時間で学習を行うこと

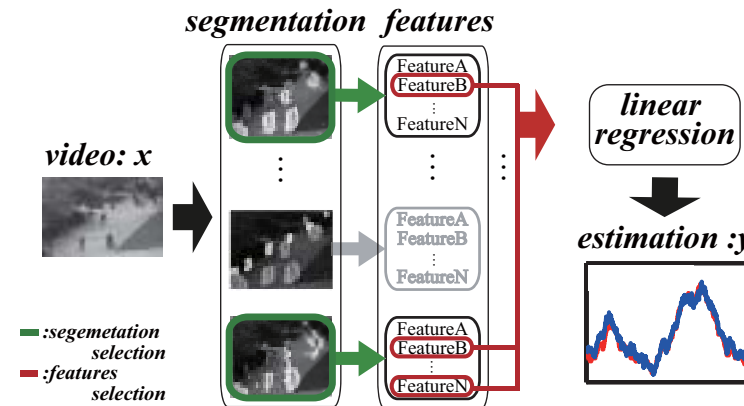


図1 人数推定のための学習の流れ
Fig. 1 learning overview for people counting

である。まず、1つ目の目的のために、グループと個々に対して疎な解をもつ学習が必要となる。理由は以下の通りである。線形回帰モデルにおいて、 w_{ji} が0であれば、対応する $\phi_{ij}(x)$ は、出力 \hat{y} に影響を与えない。すなわちこれは、 $\phi_{ij}(x)$ が不要であることを意味する。さらに、 w_j が零ベクトルであれば、対応する $\phi_j(x)$ もまた同様に不要であると判断可能である。このように、 $w_j = 0$ や $w_{ji} = 0$ を多く持つ疎な解を得られると、 $w_j \neq 0$ に対応する $\phi_j(x)$ や $w_{ji} \neq 0$ に対応する ϕ_{ij} のみを有用と判断できるのである。よって、 $w = [w_1^T, \dots, w_g^T]^T$ に対するグループごとの疎な解が、前処理選択として機能し、 $w_k \neq 0$ に対する $w_k = [w_{k1}, \dots, w_{kd_k}]^T$ の疎な解が、特徴量選択機能として働く。

N 個の学習データ $y = [y_1, \dots, y_N]^T$ が与えられたとき、重みベクトルは一般に

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_i (y_i - \hat{y}_i)^2 + \lambda \Omega(w), \quad (2)$$

により計算される。 $\Omega(w)$ は正則化項、 λ は正則化項の影響力を決めるハイパーパラメータであり、 $\Omega(w)$ によって、重みベクトルのもつ性質が決定される。本研究では、適切な正則項を用いることで、グループと個々に疎な解を持つ重みベクトルを求める。本研究における疎な解は、グループに対する疎な解の学習と、個々の特徴量に対する疎な解の学習の2段階に分けることが可能である。そこで本研究では2段階のアルゴリズムを提案する。1つ目の段階で、グループの組み合わせ候補を選出し、2つ目の段階で、選ばれた候補に対して、

**Algorithm 1 Proposed Learning method
using Group Selection and Feature Selection**

Input:

$\Phi(\mathbf{x})$: features for N frames

\mathbf{y} : a vector of N observations

K : upper limit number of used segmentation

{Group Selection Stage}

while $|S_k| \leq K$ do

 search a new group as λ decreases from ∞

$S_k \leftarrow$ selected groups

$|S_k| \leftarrow$ number of selected groups ($|S_k| = |S_{k-1}| + 1$)

$k = k + 1$

end while

{Feature Selection Stage}

for all S_k do

$\hat{\mathbf{w}}_k \leftarrow$ sparse learning

 (hyper parameter is determined with cross-validation)

$score_k \leftarrow$ MSE scored by cross-validation

end for

Output:

$\mathbf{w}_{\hat{k}} : \hat{k} = \underset{k}{\operatorname{argmin}} score_k$

個々に対して疎な解を持つ学習を行う。そしてこれらにより、全体としてグループと個々の両方に疎な解を持つ学習を行う。提案手法では、各段階で疎な解を導く正則化項である l_1 正則化を用いる。そのため、各段階でハイパーパラメータ λ の決定が必要であるが、Lars と呼ばれる正則化パスを高速に求める考え方を用いることで、容易に交差検定が可能となり、効率的に適切なハイパーパラメータが決定できる。それゆえ、提案手法は、2 つ目の目的である短い学習時間を実現しながら、疎な解の決定を可能としている。

2.2 有用グループ選出

一つ目のグループ選出を行う段階では、グループに対して疎な解を持つ適切な正則項が必

要となる。そこで、グループに対して l_1 ノルム正則化を (2) に用いた、

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i \in N} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^g \left(\sum_{i=1}^{d_j} w_{ji}^2 \right)^{\frac{1}{2}}. \quad (3)$$

を適用する。これは GroupLasso⁹⁾ として知られており、グループに対して 1 次ノルム正則化を、個々に対しては 2 次ノルム正則化用いている。そのため、得られる解はグループにのみ疎であり、個々に対しては疎ではない解となる。したがって、(3) を適用するだけでは、本研究の全要求機能を満たさないため、本研究は 2 段階のアルゴリズムを用いるのである。そこで本研究は、一つ目の段階では選択されるグループの情報のみを用いる。選出されるグループの組み合わせは、ハイパーパラメータ λ に依存し、 λ を ∞ から 0 にしていくと、選出されるグループは 1 つずつ増加していく。そのため、選択されるグループの変化点は g 点存在し、GroupLasso を用いる事で得られる組み合わせパターンは最大で g 通りとなる。よって、この変化点を求めれば、選出されるグループの組み合わせの候補が得られる。変化点を求める手法には Lars¹¹⁾ の考えを利用した GroupLars⁹⁾ が存在し、選択されるグループの組み合わせ候補を容易に求めることが可能となる。ここでは、 $\mathbf{x} = [x_1, \dots, x_N]^T$ が訓練データ、 N が訓練データ数を表し、 $\Phi_j = [\phi_j(x_1), \dots, \phi_j(x_N)]^T$ は j 番目のグループに対応する $N \times d_j$ の行列、全体の計画行列を $\Phi = [\Phi_1, \dots, \Phi_g]$ で示すとする。GroupLars を用いるには、それぞれのグループの計画行列が、直交性 ($\Phi_j^T \Phi_j = \mathbf{I}_{d_j}$) をもつことが必要となる。ここで、 \mathbf{I}_{d_j} は大きさ d_j の単位行列を示す。しかし、一般に、入力直交性を持たないことが多い。そこで、GroupLars を適用する前に、各行列を特異値分解する方法がある¹⁰⁾。特異値分解は $\Phi_j = \mathbf{U}_j \mathbf{D}_j \mathbf{V}_j^T$ で示され、 \mathbf{U}_j が直交性を持つ。 $\Phi_j \mathbf{w}_j$ は j 番目のグループの、推定値 $\hat{\mathbf{y}} = [\hat{y}_1, \dots, \hat{y}_N]$ への影響を表すが、 $\mathbf{U}_j = \Phi_j \mathbf{V} \mathbf{D}^{-1}$ と $\mathbf{w}_j^* = \mathbf{D}$ を用いることで、 $\Phi_j \mathbf{w}_j$ は

$$\Phi_j \mathbf{w}_j = \mathbf{U}_j \mathbf{D} \mathbf{V}^T \mathbf{w}_j = \mathbf{U}_j \mathbf{w}_j^*. \quad (4)$$

と表す。 \mathbf{U}_j を入力と見なせば、GroupLars は解 \mathbf{w}^* を求めることが可能であるが、一般に \mathbf{w}_j^* から本来求めたい解 \mathbf{w}_j を求めることはできない。しかし、本研究では選ばれるグループに関する情報のみが必要であるため、 \mathbf{w}^* の解から、グループの選出情報を用いればよい。以上の方法により、グループ選出の段階で、全パターン $2^g - 1$ 通りに対して、候補となる組み合わせを g 通りまで削減できる。以後、 k 個目の変化点で選出されたグループ番号の集合を $S_k (k = 1, \dots, g)$ で表す。

2.3 個々の特徴量に関する疎な学習

2.2 節により, 選出されるグループの組み合わせが g 通りまで絞られる. 2 つ目の段階では個々に対して疎な解を得るため, l_1 ノルムを個々のレベルに対して適用する. すると (2) は,

$$\hat{w}_k = \operatorname{argmin}_w \sum_{i \in N} (y_i - \hat{y}_i)^2 + \lambda_k \sum_{j \in S_k} \sum_{i=1}^{d_j} |w_{ji}|, \quad (5)$$

となる. この (5) は Lasso⁸⁾ として知られ, 正則化項 $\Omega(x)$ が個に対して l_1 ノルムとなっており, 個に対して疎な解が得られる. それゆえ, 入力に 2.2 節でグループに対して疎である $S_k (k = 1, \dots, g)$ を適用することで, グループと個々の両方に対して疎な解を持つことが可能となる. (5) により得られる解 w_k はハイパーパラメータ λ_k に依存するが, Lasso の正則化パスは区分線形性をもつことで知られている. そこで, 変化点のみをとらえることで, 区分線形性から全体の正則化パスを効率的に求める手法が Lars¹¹⁾ である. 一旦, Lars の考え方をういて正則化パスが求められると交差検定が容易に可能となり, 最適なパラメータ λ_k と対応する解 \hat{w}_k が効率的に求まる.

(5) は, すべての $S_k (k = 1, \dots, g)$ に適応され, 交差検定による評価値 (E_k) を比較, 最高の評価値を持つ解 ($w = \hat{w}_{\hat{k}} | \hat{k} = \operatorname{argmax}_k E_k$) が選ばれる. 以上のように, S_k に対する個々に疎な学習を通して, グループと個々の両方に対して疎な解が得られる.

2.4 利用グループ数制限への応用

提案手法は, グループ選択の段階で, $2^g - 1$ 通りから g 通りの組み合わせ候補に絞り込む. この段階を通して提案手法は, 利用グループ数の制限に容易に対応可能となっている. 2.2 節で示したように, 選出されるグループは λ の増加に対して, 1 つずつ増加する ($|S_k|$ が S_k の要素数を表すとき, $|S_{k+1}| = |S_k| + 1$). したがって, 利用グループ数に対する制限 ($\max |S_k| \leq K$) が存在するとき, $|S_k| = K$ になった時点で計算を打ち切れれば, 利用グループに対する制限内で解が得られる.

2.5 人工データを用いた実験

我々は, 人工データを生成し, 提案手法を用いた実験を行った. 人工データにはそれぞれ 10 個の特徴量をもつ 20 グループ, 計 200 の特徴量から生成される 300 の訓練データを用意した. 正解データの重みベクトル w は, 半分の 10 グループが全て零ベクトル, 残り 10 グループがそれぞれ (10, 9, 8, 7, 6, 5, 4, 3, 2, 1) 個の 0 でない要素を持つ重みベクトルで構成されている. 重み係数, 特徴量は, 共に標準正規分布に従い生成され, 観測データは標

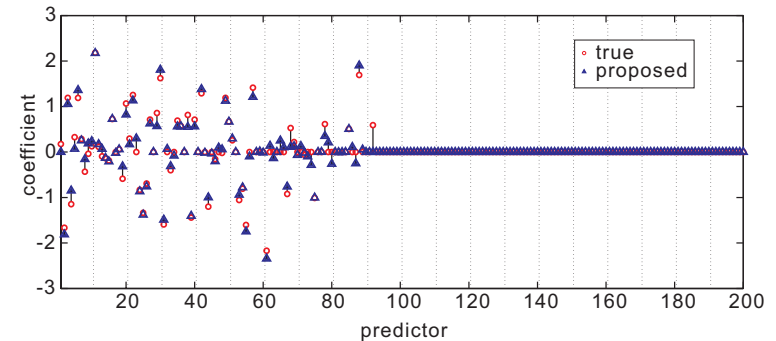


図 2 生成した重みベクトル (正解) と学習により得られた重みベクトル
Fig. 2 generated weights and learning result with the proposed method

表 1 人工データによる実験結果

Table 1 preliminary experimental result

	MSE	training time (s)	used group
Proposed	5.67	279	1,2,3,4,5,6,7,8,9
Forward	5.35	2029	1,2,3,4,5,6,7,8,9,10
Backward	5.35	2484	1,2,3,4,5,6,7,8,9,10
Ridge	9.22	11	all

準偏差 2.0 のガウシアンノイズを加えて, 生成した. 図 2 は生成したデータの重みと提案手法により得られた重みを表している. また, 表 2.5 に, 他の比較手法を用いたときの結果と, 提案手法の結果の比較を示した. 比較手法には, 前向き, 後ろ向き選択アルゴリズムを用意し, 参考手法にリッジ回帰を用いた. 以下に各手法の詳細を記す.

• 前向き選択アルゴリズム

この手法は, $|S_0| = 0$ から, 徐々に選択されるグループを増やしていく手法である. 加えるグループは, 実際にグループを加え (S_k に対して $g - k + 1$ の候補), 各組み合わせパターンに対して学習, 推定精度のよくなるグループを選択する. 本研究では, 各パターンに対しての計算に Lasso⁸⁾ を用いており, ハイパーパラメータ λ 決定に交差検定を用いた. 計算が行われる組み合わせは $\frac{1}{2}g(g+1)$ 通りとなる.

• 後ろ向き選択アルゴリズム

この手法は, $|S_0| = g$ から, 徐々に選択されるグループを減らしていく手法である. 減

らずグループは、実際にグループを減らし (S_k に対して k の候補), 各組み合わせパターンに対して学習, 推定精度のよくなるグループを選択する. 本研究では, 各パターンに対しての計算に Lasso⁸⁾ を用いており, ハイパーパラメータ λ 決定に交差検定を用いた. 計算が行われるグループ組み合わせは $\frac{1}{2}g(g+1)$ 通りとなる.

● リッジ回帰

この手法は, l_2 -ノルムを用いた正則項をもつ過学習を防ぐための一般的な手法である. l_2 -ノルムを用いているため, 解は疎とならず, グループや特徴量の選択は行われない. また, Lasso と同様にハイパーパラメータ λ をもつため, 交差検定により値を決定した. 各手法に対して, 同様のデータから生成された 1000 のテストデータを用いて MSE (Mean Squared Error) を計算した. また, 訓練時間は交差検定を行う時間を含めて評価した. 実験結果から我々の手法が比較的高い精度を保っていることがわかり, さらに, グループ選択を行う手法のうちで, 訓練時間が最も短いことがわかる. これはグループ選択の段階で, 他の手法が $\frac{1}{2}g(g+1)$ 通りの組み合わせを探索するのに比べ, g 通りの組み合わせまで, 候補を削減できていることにより, 基本的に $2/g$ 倍程度の計算時間で済むことによると考えられる. また, 利用グループに関して, 人工データは, グループ 1 がもっとも影響力が大きく, 10 に向かうにつれ影響力が小さくなり, 11 以降で影響力がなくなるよう生成されている. これに対し, 提案手法におけるグループ選出の解では, 全 20 グループから影響力の強いグループが選ばれていることが確認できた.

3. 混雑環境における人数推定

本研究では, 動画からセグメンテーションにより対象領域を切り出し, 画像特徴量を抽出する. そして, 補正された画像特徴量が人数推定のための入力になる. この入力ベクトルを求める処理は大きく 3 段階に分けられる. まず始めに, 対象領域を切り出すためにセグメンテーションを行い, 次に対象領域から画像特徴量を抽出する. そして最後に, 画像の位置情報に基づき特徴量の補正を行う. これらの手法は A.B.Chan らの方法⁴⁾ を参考としており, この章では, それぞれについて詳細を述べる.

3.1 対象領域の切り出し

人数推定のための入力を求めるにあたり, まず始めに違った方向に動く群衆に対し, 対象とする人間周りの領域を切り出す. 本研究では, 既存の他の研究⁴⁾⁶⁾ と異なり, 領域切り出し処理 (前処理) に 8 つの手法を用いている. これには同様の手法でハイパーパラメータのみを変化させたものを含む. 利用したセグメンテーション手法を, mixture of dynamics

textures¹²⁾, オプティカルフロー, 背景差分法の 3 種類に分け, 以下で詳細を述べる.

● mixture of dynamics textures

この手法は先行研究⁴⁾⁵⁾ で利用された手法である. この手法はいくつかの異なる動きを混合モデルで扱い, 初めに各モデルを学習する. その後, 学習した各動きモデルに対して尤度を計算, 最も尤度の高いモデルの割当てにより, セグメンテーションを行う. クラス数に関するハイパーパラメータを持っており, 本研究では対象とするクラスの違いによる 2 種類の前処理を利用した.

● オプティカルフロー

オプティカルフローは画像の輝度変化を用いて動きをとらえる手法である. セグメンテーションには, 得られた速度情報を用いる. オプティカルフローには様々な手法が提案されているが, 本研究では 3 種類の手法を扱った. Horn-Schunck の手法¹³⁾, Lucas-Kanade の手法¹⁴⁾, 時間的局所最適化法¹⁵⁾ の 3 つである. 本研究ではハイパーパラメータの違いによる 2 種類の Horn-Schunck の手法, 1 種類の Lucas-Kanade の手法, 1 種類の時間的局所最適化法を利用した.

● 背景差分法¹⁶⁾

背景差分法は, 背景画像と対象画像の差分を利用する手法である. 閾値以上の差分を取り出す処理であるため, 特定の動きを対象とした領域切り出しは行えないが, 人がいる領域は切り出せると考えられる. 本研究では閾値を適当に設定し, 1 種類の前処理を利用した.

実験には, 上述した 7 種類のセグメンテーション手法とランダムに領域を切り出した方法の計 8 種類を利用した.

3.2 特徴量抽出

次に, 前処理領域から様々な画像特徴量を抽出する. 本研究では, 先行研究⁵⁾⁴⁾ に基づき, 抽出する特徴量を計算した. 以下に利用した特徴量を 3 種類に分類して述べる.

● 対象領域自体の特徴:

これらの特徴量は, 対象領域の形, 大きさに基づく特徴量である. 具体的には, 対象領域の面積, 周囲長, ブロック数, 面積と周囲長の比, 周囲の方向 (4 方向) のヒストグラム, の計 8 種類を利用した.

● 対象領域内のエッジに基づく特徴:

これらの特徴量は, 対象領域に含まれるエッジの量を用いる. 混雑環境におけるエッジ量は, 人数推定に有用であることが知られている¹⁷⁾¹⁸⁾. 具体的に用いた特徴量は, エッ

ジ量, エッジ方向 (4 方向) のヒストグラム, の 5 種類の特徴量である .

● 対象領域のテクスチャ情報に基づく特徴 :

これらの特徴量は, 対象領域内のテクスチャ情報に基づく . テクスチャ情報は混雑状況をクラス分けした研究¹⁹⁾ で用いられた GLCM (gray-level co-occurrence matrix) に基づく情報を利用する . 具体的には, 4 方向に対応した 4 種類の GLCM から計算される, 同様性, エネルギー, エントロピー, の計 12 種類を利用した .

以上, 本研究では 25 種類の特徴量を用いており, 同一の前処理から抽出された特徴量を同一グループに存在するとして扱う .

3.3 位置情報に基づく特徴量補正

抽出された特徴量は補正の後, 回帰予測の入力 $\phi_{ij}(x)$ となる . この補正は, 画像の位置情報に基づき行う . これはカメラに近いほど, 画像上で大きく映る性質を補正するためである . 各位置による補正量は, 参照人物を決め, 2 枚の画像の参照人物を比較, Chan らの方法⁴⁾ に従い計算することで求める . 補正量は 2 次元の情報をもとに得られるため, 1 次元の特徴量に対しては根をとった値を適用する .

4. 実 験

本研究では, 混雑した環境で歩行者を扱った公開データセット, UCSD pedestrian database⁴⁾ を用いて提案手法に対する実験・評価を行った . この章では, 行った実験・結果について詳細を述べる .

4.1 実験基準

提案手法の有用性を 4 つの基準をもとに評価する . 1) 推定精度, 2) 選出されたグループ数, 3) 選出された特徴量の数, 4) 訓練時間 の 4 つである . 1) 推定精度の評価には MSE を用いる . これは Chan ら⁴⁾ と同様の評価指標である . 2) 選出されたグループ数は $w_j \neq 0$ の数により評価する . 選ばれたグループ数が少ないことは, 過学習を防ぐ観点から有用である . また, 一般にセグメンテーション計算は計算時間が長いので, 利用グループが少ないことは, 実行時間の短縮にもつながる . 3) 選出された特徴量の数は $w_{ji} \neq 0$ の数により評価する . 選出されたグループ数と同様に, 過学習を防ぐ観点から利用特徴量が少なく優れていると考えられる . 4) 訓練時間は, 学習を行う時間のことであり, これにはハイパーパラメータ決定の時間を含む . ハイパーパラメータ決定の時間は, 交差検定もしくは最尤推定を用いる時間である . また, 本研究においては, 訓練時間には前処理や特徴量抽出を行う時間は含まない . 本研究では, 4 つの評価基準を通して, 提案手法を他の手法と比較する .



図 3 データセットの例と対象領域

Fig. 3 Example frame of dataset and ROI (Region of Interesting)

4.2 対象データ

我々は, 公開データセット UCSD pedestrian database⁴⁾ を用いて実験を行った . このデータセットは 1 フレームあたり 11-46 の人が存在する 2000 フレームから成る . データセットには, 2 種類の方向に移動する人数の真値が含まれている . また, 動画は 10fps, 238×158 ピクセルのグレースケール画像から構成されている . フレームの例と人数推定を行う対象領域を図 3 に示す . この実験において, 対象人物はカメラから遠ざかる方向に動く人物とする . 対象の人数の推移は, 推定結果とともに図 4 に示す . データセット 2000 フレームに対し, 600 から 1399 番目の 800 フレームを訓練データとし, 残りの 1200 フレームをテストフレームとして利用する . これは Chan ら⁴⁾ の利用法に従っている . また, 我々は, 3.1 節に示した 8 種類のセグメンテーションと 3.2 節に示した 25 種類の特徴量の計 $25 \times 8 = 200$ を入力として用いる . また, 入力はそれぞれに対して標準化 (平均 0, 分散 1) とした状態で利用する .

4.3 比較手法

本研究では, 実験においていくつかの手法と比較を行う . 比較手法には, 一般によく知られる手法, グループ選出を行う手法, 先行研究⁴⁾ で用いられた手法, を用いた . 入力提案

手法と同様のものを用いており、実装は Matlab²⁰⁾ を用いた。以下に詳細を記す。

- リッジ回帰
この手法は、2.5 節において、説明した。
- Group Lasso⁹⁾
この手法は、2.2 節に登場した手法である。グループに対して l_1 ノルム正則化を、個々に対して l_2 ノルム正則化を行うため、グループに対してのみ疎な解を持つ。よって、有用なグループの選択は可能であるが、個々に対して疎な解を持たず、特徴量選択は行われない。解は cvx²¹⁾ を利用して求めた。また、この手法はハイパーパラメータ λ をもつが、交差検定を用いて λ を決定した。
- 前向き選択アルゴリズム
この手法は、2.5 節において、説明した。
- 後ろ向き選択アルゴリズム
この手法は、2.5 節において、説明した。
- GPR²²⁾
この手法は、先行研究⁴⁾ で利用されたノンパラメトリックな推定手法である。GPR はいくつかのカーネル関数を持ち、各カーネル関数はハイパーパラメータを持つ。これらのハイパーパラメータは、訓練データに対する尤度最大化を行う事で決定した。この尤度最大化は、局所解の存在により初期値依存性を持つ。そこで本研究では、適当な初期値を用いて何度か尤度最大化を行い、尤度が最も高くなるパラメータを利用した。

4.4 実験結果

提案手法を用いた推定結果を図 4 に示す。表 4.4 は提案手法と比較手法を用いた時の結果を示している。実験は、Intel Xeon E5335 2.00GHz 上の Matlab によって行った。提案手法の推定精度 (MSE) は 3.64 となっており、全手法の中で 2 番目に良い結果となっている。また、利用する前処理数は提案手法では、全 8 つの内 4 つが選ばれており、半分まで利用前処理を減らせていることが分かる。利用特徴量数においても、全 200 の内、13 個のみを利用しており、6.5%まで減らすことが可能となっている。これは、提案手法が前処理、特徴量に関する選出を行えていることを示しており、実行時間についても大幅な削減を可能とすることを表す。図 5 は提案手法による学習から得られた重みベクトルを示しているが、グループと特徴量に対して疎な解が得られている様子が分かる。また、最も用いられた特徴量は、エッジ量であり、エッジ量が人数推定において有用な手がかりとなっていることがわかる。訓練時間の観点からは、グループ選出を行わないリッジ回帰の手法を除き、最も短い

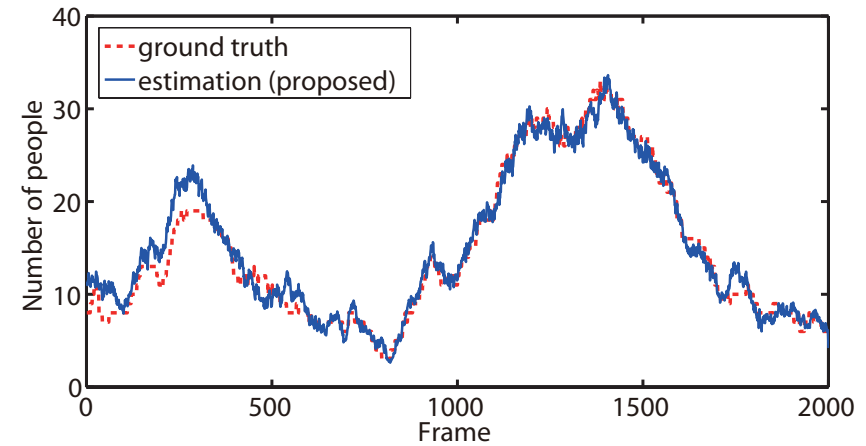


図 4 推定結果
Fig. 4 estimate result

訓練時間となっている。これは提案手法が、グループの組み合わせ候補を g 通り (今回は $g = 8$) まで減らせていることによる。したがって、次に短い前向き選択手法の 4 分の 1 程度まで、大幅に訓練時間を減らすことが可能となっている。ここまで述べたように、提案手法は他の手法と比較し、どの分野においてもよい結果が得られている。すなわち、提案手法はバランスのとれた手法であり、十分な推定精度、訓練時間の削減、前処理と特徴量に対する選出機能から本研究の目的を果たす有用な学習法であるといえる。

5. 結 論

本研究では、グループと個々の両方に対して疎な解を持ち、効率的に計算を行うことが可能である、2 段階に Lars を適用する学習法を提案した。提案手法は、効率的にグループ選出を行えるため、訓練時間を短縮できる長所も持っている。2 段階の各段階にハイパーパラメータが存在するが、Lars の考え方により正則化パスが容易に計算できるため、交差検定が容易となり、訓練時間の短縮が可能となる。我々は、提案手法を混雑環境において多数の画像特徴量から人数推定を行う手法に適用した。実験により、提案手法が過学習を防ぎながら、大幅に利用する前処理、特徴量を選択できるという結果が得られた。さらに、グループ選出を行う他の手法と比較し短い時間で学習を行うことが可能であり、効率的な学習ができ

表 2 比較実験の結果
Table 2 experimental result

	MSE	num. of segmentation	num. of features	training time (s)
Proposed	3.64	4/8	13/200	104
Ridge	4.97	8/8	200/200	15
G-Lasso	6.16	4/8	100/200	2068
Forward	4.16	3/8	10/200	426
Backward	4.11	2/8	9/200	567
GPR	3.49	8/8	200/200	17026

太文字：各評価に対する上位 2 位を示す
bold font: first and second in each section

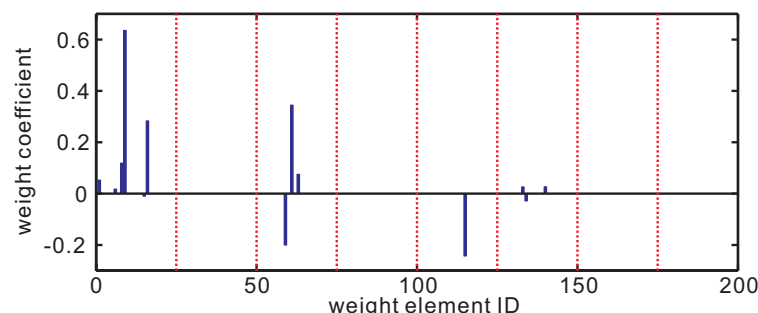


図 5 提案手法により計算された重み: 赤線の間が同じ前処理を示し、左から 3.1 節で説明したものである。
Fig.5 calculated weight with proposed method: A section between red lines means one group. The groups are listed in 3.1 from left.

ることを確認した。

参考文献

- 1) Viola, P., Jones, M. and Snow, D.: Detecting pedestrians using patterns of motion and appearance, *Proc. ICCV*, Vol.1, pp.734–741 (2003).
- 2) Zhao, T. and Nevatia, R.: Bayesian human segmentation in crowded situations, *Proc. of CVPR*, Vol.2 (2003).
- 3) Rabaud, V. and Belongie, S.: Counting crowded moving objects, *Proc. of CVPR*, Vol.1, pp.705–711 (2006).
- 4) Chan, A.B., Liang, Z.J. and N.Vasconcelos: Privacy preserving crowd monitoring:

- Counting people without people models or tracking, *Proc. of CVPR*, pp.1–7 (2008).
- 5) Chan, A. and N.Vasconcelos: Bayesian poisson regression for crowd counting, *Proc. ICCV*, pp.545–551 (2010).
- 6) Ryan, D., Denman, S., Fookes, C. and Sridharan, S.: Crowd counting using multiple local features, *Proc. DICTA*, pp.81–88 (2009).
- 7) Ryan, D., Denman, S., Fookes, C. and Sridharan, S.: Crowd counting using group tracking and local features, *Proc. of AVSS*, pp.218–224 (2010).
- 8) Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso, *J. R. Statist. Soc. B*, Vol.58, pp.267–288 (1996).
- 9) Yuan, M. and Lin, Y.: Model selection and estimation in regression with grouped variables, *J. R. Statist. Soc. B*, Vol.68, pp.49–67 (2006).
- 10) Friedman, J., Hastie, T. and Tibshirani, R.: A note on the group lasso and a sparse group lasso, *Technical report, Department of Statistics, Stanford University* (2010).
- 11) Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R.: Least angle regression, *Ann. Statist.*, Vol.32, pp.407–499 (2006).
- 12) Chan, A.B. and Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures, *Trans. of PAMI*, Vol.30, No.5, pp.909–926 (2008).
- 13) Horn, B. K.P. and Schunck, B.G.: Determining optical flow, *Artificial Intelligence*, Vol.17, pp.185–203 (1981).
- 14) Lucas, B. and Kanade, T.: An iterative image registration technique with an application to stereo vision, *Proc. IJCAI*, pp.674–679 (1981).
- 15) Kearney, J.K., Thompson, W.B. and Boley, D.L.: Optical flow estimation: an error analysis of gradient-based methods with local optimization, *Trans. of PAMI*, Vol.9, pp.229–244 (online), DOI:http://dx.doi.org/10.1109/TPAMI.1987.4767897 (1987).
- 16) Morita, S., Yamazawa, K. and Yokoya, N.: Networked telepresence and remote surveillance systems using as omnidirectional image sensor, *IEIC Technical Report*, Vol.102, No.318, pp.19–24 (2002).
- 17) Davies, A., Yin, J. and Velastin, S.: Crowd monitoring using image processing, *Electron. Commun. Eng. J.*, Vol.7, No.1, pp.37–47 (2002).
- 18) Kong, D., Gray, D. and Tao, H.: A viewpoint invariant approach for crowd counting, *Proc. of ICPR*, Vol.3, pp.1187–1190 (2006).
- 19) Marana, A., Costa, L., Lotufo, R. and Velastin, S.: On the efficacy of texture analysis for crowd monitoring, *Proc. of SIBGRAPI 1998*, pp.354–361 (1998).
- 20) MATLAB: version 7.4.0 (R2007a) (2007).
- 21) Grant, M. and Boyd, S.: CVX: Matlab Software for Disciplined Convex Programming, version 1.21, http://cvxr.com/cvx (2011).
- 22) Rasmussen, C.: Gaussian processes in machine learning, *Advanced Lectures on Machine Learning*, pp.63–71 (2004).