

プロブ特徴を用いた実時間歩行者数推定

吉永 諭史^{†1} 島田 敬士^{†1} 谷口 倫一郎^{†1}

本稿では「画像上における歩行者の位置と人数」を実時間で推定できるシステムを提案する。提案手法では、まず観測画像に対し、照明変動に頑健な動的背景モデルを用いた背景差分を施すことで、移動物体を前景として検出する。次に、前景領域として検出された各プロブから特徴ベクトルを抽出する。オフライン処理として、特徴ベクトルと歩行者数の対応関係を手動で与えることで、学習データを作成し、ニューラルネットワークに学習させておく。オンライン処理では、各プロブから抽出した特徴ベクトルをもとに、プロブが何人の歩行者で構成されているかを学習済みニューラルネットワークを用いて推定する。このとき、歩行者数が1人以上と推定されたプロブの画素に対しては、背景モデルの更新を行わない。これによって、歩行者領域の消滅や分裂を防ぐことができる。

Real-time Pedestrian Estimation Using Blob Features

SATOSHI YOSHINAGA,^{†1} ATSUSHI SHIMADA^{†1}
and RIN-ICHIRO TANIGUCHI^{†1}

We propose a system which estimates “how many and where pedestrians are in video sequences” in real-time by following procedures. In the preprocess, we let the neural network learn training data which we generated by giving a correspondence between a set of blob features and the number of pedestrians. In the online process, first, candidate regions are segmented into blobs according to background subtraction based on a background model. Second, a set of features is extracted from each blob and a neural network estimates the number of pedestrians corresponding to each set of features. Background model in the blob in which there is at least one person is not updated to prevent the human region from being embedded in the background.

1. はじめに

計算機の飛躍的な発達と共に画像処理技術は発達し続けている。また、小型監視カメラの安価化が進んだことで、防犯目的や調査目的など様々な目的のために、社会の至る所にカメラが設置されている。これらのカメラをセキュリティや市場調査、マネージメントなどに利用するためには、監視員を配置するなどしてカメラからの映像を監視する必要がある。カメラ画像の監視方法としては、各所に配置された映像を一ヶ所に集約させ、モニタに映し出された映像を監視員が目視によって状況を把握するという方法が一般的である。この方法を用いれば、同時に広範囲を監視することが可能であるが、同時に次のような問題も発生する。

- カメラの台数が比較的小さい場合でも、常時監視にかかるコストは非常に高い。
- カメラの台数が増えるにつれ、映像の監視員にかかる疲労とコストが増大する。
- 長時間の目視による監視は、監視員の疲労や怠惰による見落としが発生する原因となる。カメラの映像を録画しておき後から解析すれば、監視員の疲労による見落としを軽減できる。しかし、防犯目的でカメラを設置している場合、実時間で解析が求められているため、この方法では不十分である。そこで、画像処理技術を用いた異常検出システムが必要とされている。

画像処理技術を用いることで、歩行者に関する情報を提供できれば、監視システムの自動化や、監視員にかかる負担とコストの軽減に繋がると期待される。歩行者に関する情報は、多くのアプリケーションに利用可能な情報である。歩行者に関する情報として、入力画像中に映る歩行者数を推定する手法はいくつも提案されている。既存の手法は、大きく次の3通りの方法に分けられる。

- (1) 歩行者領域を検出する手法
- (2) 歩行者の軌跡を分類する手法
- (3) 特徴量抽出による手法

精度を考慮した場合、オクルージョンや背景の微小な変化に対応可能な(3)特徴量抽出による方法が有効である。しかし、特徴量抽出を利用した既存の手法は、実時間での処理が不可能であったり、画像全体での歩行者数を推定は可能であるが歩行者の位置情報までは分からないといった問題がある。画像全体での歩行者数が分かるだけでは、情報の応用先が限られ

^{†1} 九州大学
kyushu University

てしまう。しかし、歩行者の人数と共に歩行者の位置情報が得られれば、応用先が広がると考えられる。例えば、駅や空港などではセキュリティシステムに利用可能であるし、デパートやショッピングモールなどではマネージメントや市場調査に利用可能である。そこで本稿では、「画像上における歩行者の位置と人数」を実時間で推定可能なシステムを提案する。

第2節では、歩行者数計測の関連研究について述べる。第3章では、提案手法であるプロブ特徴を利用した歩行者数の計測について述べる。第4章では、実験内容と結果を述べ、考察する。最後に第5章では、本研究の結論と今後の課題について述べる。

2. 関連研究

画像処理技術を用いた歩行者数計測とは、観測画像中に映る歩行者数を計算機を用いて計測することである。画像処理技術を用いた歩行者数の計測にはいくつかの手法が提案されている。それらは、主に次の3通りの手法に分けることができる。

1つ目は歩行者領域を検出する手法で、画像中の全ての歩行者領域を検出することにより、歩行者数を推定するというものである。Violaらは、歩行者の外見と動きに関する特徴を学習させた検出器を用いて歩行者を検出する手法を提案している¹⁾。また、Zhaoらはベイジアンモデルを基にした領域分割による手法を提案している²⁾。しかし、この手法では、画像中の全ての歩行者領域を検出することで歩行者数を推定するため、混雑した状況を観察した場合、オクルージョンの発生などによってアルゴリズムが破綻してしまう。

2つ目は歩行者の軌跡を分類する手法で、追跡アルゴリズムによって得られた歩行者の軌跡をクラスタリングし、最終的なクラスタの数を歩行者の人数として推定するというものである。Antoniniらは、追跡アルゴリズムによって得られた軌跡を、軌跡の長さと同士の空間的近さに基づいてクラスタリングすることで歩行者数を推定している³⁾。しかし、この手法では、歩行者の軌跡を用いて歩行者数を推定しているので、一定時間内に通過した歩行者の数を推定することは可能であるが、実時間で歩行者数を推定することは不可能である。

3つ目は特徴量抽出による手法で、まず、入力画像に対し背景差分などの領域分割処理を施すことで得られた前景領域から特徴量を抽出する。その後、抽出した特徴量と歩行者数との間に成立する対応関係を回帰関数を用いて推定するというものである。Kongらは、背景差分により検出された前景領域とエッジ方向に関するヒストグラムを特徴量として、ニューラルネットワークにより歩行者数を推定している⁴⁾。Chanらは、領域分割に *mixture of dynamic texture*⁵⁾ を用いることで、プライバシーを保護しながら検出した前景領域から28次元の特徴ベクトルを抽出することで、歩行者数を推定する手法を提案している⁶⁾。

これまでに紹介したように、歩行者数計測の手法はいくつも提案されているが、アルゴリズムの性質上実時間での推定が不可能なものや、画像全体での歩行者数を推定するものである。例えば、Kongらが提案する手法では、入力画像から1つの特徴ベクトルを抽出するため、画像全体で何人の歩行者がいるか推定することは可能であるが、何処にいるかまでは分からない。さらに、Kongらは特徴ベクトルとして2つのヒストグラムを作成しているため、特徴量抽出に時間がかかってしまうと思われる。また、Chanらが提案する手法では、領域分割に *mixture of dynamic texture* を用いているため、過去と未来それぞれ数フレームの画像が必要となり、実時間で歩行者数を推定することは、そもそもアルゴリズム的に不可能である。さらに、Chanらは、フラクタル次元や、Homogeneity, Energy, Entropy などといった複雑な特徴量を含んだ28次元に及ぶ特徴ベクトルを作成しているため、特徴量抽出にも時間がかかってしまうと思われる。

そこで、本稿では入力画像に対し背景差分と影除去処理を施すことで得られる前景領域の各プロブが何人の歩行者で構成されているか推定することによって、「画像上における歩行者の位置と人数」を実時間で推定する手法を提案する。提案手法では、「特徴量の簡略化」と「領域分割の高速化」を行うことで、実時間での処理を可能としている。

特徴量の簡略化：単純且つ有効な特徴量のみを利用することで、特徴量抽出を高速に行う。
領域分割の高速化：高速なParzen推定を用いた動的背景モデルにより、背景でないかと判断された画素に対して、影除去処理を施すことで、前景領域として移動物体のみを精度よく高速に検出する。

3. プロブ特徴を利用した実時間歩行者数計測

「画像上における歩行者の位置と人数」を実時間で推定できるシステムを提案する。提案手法の概要を図1に示す。提案手法では、まず領域分割により移動物体を前景として検出する。次に、前景領域として検出された各プロブから特徴ベクトルを抽出する。また、各プロブから抽出した特徴ベクトルとそのプロブを構成する歩行者数の対応関係を手動で与えることで、事前に学習データを作成しておき、ニューラルネットワークにその学習データを学習させておく。その後、各プロブから抽出した特徴ベクトルをもとに、そのプロブ何人の歩行者で構成されているかを学習済みニューラルネットワークを用いて推定する。歩行者数を精度良く検出するためには、歩行者領域をできるだけ正確に検出する必要がある。提案手法では、歩行者数の推定結果を背景モデルにフィードバックすることで、背景差分における前景領域の検出精度を向上させている。

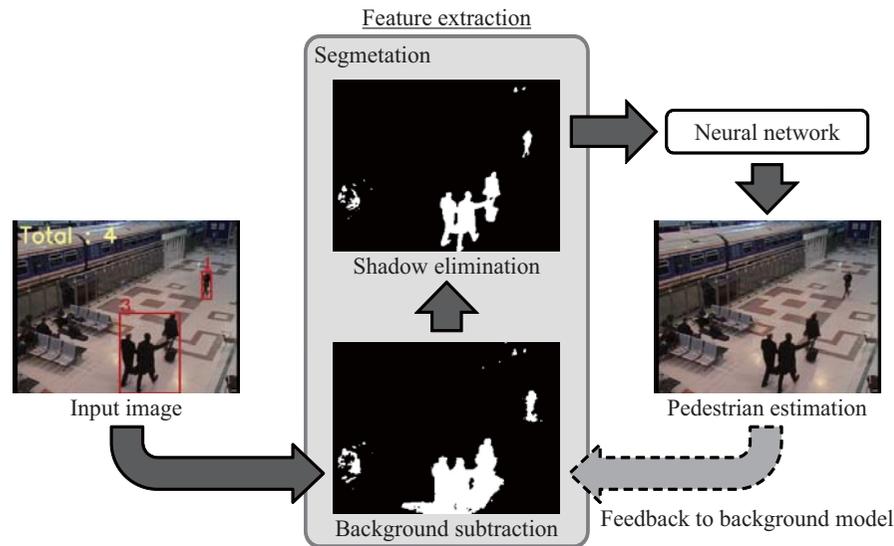


図 1 提案システムの概要

提案手法は、歩行者領域分割、特徴量抽出、特徴量正規化、ニューラルネットワークによる歩行者数の推定、動的背景モデルの更新の 5 つの処理で構成される。本章では、これらの処理について詳しく述べる。

3.1 領域分割

入力画像に映る背景以外の物体を検出するために背景差分を用いる。背景差分を用いることで、物体に関する事前知識を必要とせずに物体領域を抽出することができる。提案手法では、背景の微小な変化による影響を抑えつつ、物体領域のみを精度良く高速に検出するため、背景差分に Parzen 推定を用いた動的背景モデル⁷⁾を用いる。これは、ノンパラメトリックな確率密度の推定法である Parzen 推定を用いて、過去に観測された画素値を基に背景モデルを推定する手法である。しかし、背景差分では、図 1 における *Background subtraction* のように物体によってできる影まで前景として検出してしまう。日の傾きや、照明との位置関係などが変化することで、物体によってできる影の面積も変化する。これら影領域は、システムに悪影響を及ぼす可能性があるため、背景差分で得られた前景領域から影領域を除去する必要がある。影領域の除去に YUV 表色系を用いる手法⁸⁾がある。この手法では、影領域の画素値は背景の画素値に比べ明度、彩度のみが小さくなり、色相は変化しないという性

質を用いている。

図 1 における *Shadow detection* のように、背景差分によって検出された物体領域から影領域を除去した領域を、領域分割によって得られる前景領域とする。このとき、前景領域は複数のプロブで構成されることになる。

3.2 特徴量抽出

理想的環境下で歩行者数の計測を行った場合、領域分割によって得られる前景領域と歩行者数は線形関係を持つ。しかし、実際に観測される映像では、歩行者同士が重なり合っていることで発生するオクルージョンや、ノイズなど様々な要因が影響し、前景領域と歩行者数は非線形関係を持つ。この非線形関係をモデル化するために、領域分割処理で前景領域として検出された各プロブからそれぞれ以下に示す 6 次元の特徴ベクトルを抽出する。

プロブ特徴：領域分割により前景領域として検出された各プロブの大きさや形状に関する特徴量。

- 面積：プロブ内に含まれる画素数。
- 周囲長：プロブの輪郭を構成する画素数。
- 複雑度 (周囲長/面積)：プロブの面積と周囲長の比。この比によってプロブの形状の複雑さが表現できる。

内部エッジ特徴：前景領域として検出された各プロブ内部に存在するエッジに関する特徴量。プロブ内部に存在する物体が何であるかに関する強い手掛りとなる。Canny エッジ検出器を用いて入力画像全体から抽出したエッジ画像と領域分割によって得られる前景領域の共通部分をとった画像から抽出される。

- エッジ画素数：プロブ内部のエッジを構成する画素数。

プロブを包含する矩形特徴：前景領域として検出された各プロブを包含する矩形に関する特徴量。

- アスペクト比：矩形の高さと横幅の比。
- 画面外判定：歩行者の全身が画像中に映りきっているか判定する。具体的には、歩行者の上半身の一部が映りきれていない場合・下半身の一部が映りきれていない場合・右または左半身の一部が映りきれていない場合を判定する。画面外判定を行うことで、上記のそれぞれの場合における特徴量の変化を予測できる。

3.3 特徴量正規化

歩行者をカメラで撮影した場合、歩行者が立っている平面とカメラの投影面は平行でない

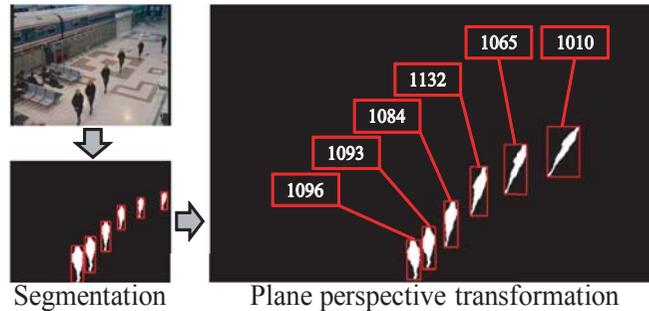


図 2 同一人物に対する領域分割 (手動) + 平面射影変換

ことが多いため遠近法により、画像上にはカメラに近い物体は大きく、カメラから離れた物体は小さく映る。このため、同一人物でも画像中の何処に映っているかで、その人が画像上で占める面積が変化する。画像上での歩行者のスケール変化と共に、特徴量が変化してしまうため、特徴量を正規化する必要がある。本節では、2通りの正規化法を提案する。

正規化法 1: 平面射影変換前後のプロブの面積比を用いた方法

正規化法 1 では、まず、平面射影変換を用いて、画像中に映る人物の占める面積が位置に非依存になるように入力画像を変換する。平面射影変換とは、歩行者が立っている平面 (基面) と画像上の注目領域 (ROI) 間の対応関係から homography と呼ばれる 3×3 行列 H を求めることで、ROI の点 (u, v) と基面上の点 (x, y) を対応付けることである。行列 H の要素は、4 点以上の対応関係をとることで式 (1) によって求められる。

$$u(h_{31}x + h_{32}y + h_{33}) = h_{11}x + h_{12}y + h_{13} \quad (1)$$

$$v(h_{31}x + h_{32}y + h_{33}) = h_{21}x + h_{22}y + h_{23}$$

平面射影変換によって、図 2 のように、同一人物であれば画像中の位置にかかわらず、同程度の面積を占めるような画像に変換出来る。しかし、歩行者は地面に対し垂直に立っているため、平面射影変換では、歩行者が左右に傾いた画像に変換されてしまう。このため、変換後の画像においても依然として、プロブの周囲長やエッジに関する特徴量は画像中の位置に依存してしまう。そこで、平面射影変換前のプロブの面積 (S_a) と、平面射影変換後のプロブの面積 (S_b) の比 (S_b/S_a) を用いて特徴量を正規化する。プロブの周囲長やエッジに関する特徴量は、変換前の画像から抽出した特徴量に対し、平面射影変換前後のプロブの面積比の平方根 ($\sqrt{S_b/S_a}$) を乗じることで正規化する。プロブの複雑度に関しては、正規化後の面積と周囲長を用いて計算することで正規化する。

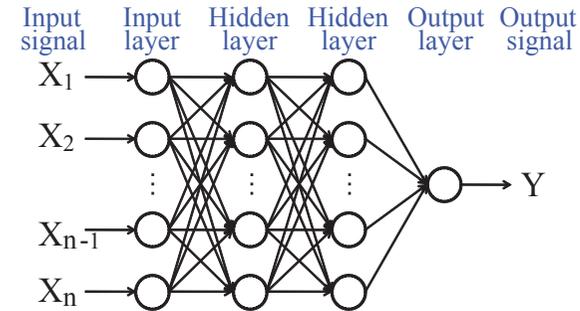


図 3 提案手法で用いるニューラルネットワーク

正規化法 2: スケールパラメータを用いる方法

正規化法 2 では、スケールパラメータを用いて特徴量を正規化する。具体的には、プロブを包含する矩形の位置情報をスケールパラメータとして特徴ベクトルに組み込む。この特徴ベクトルをニューラルネットワークに学習させることで、画像上での歩行者のスケール変化も一緒に推定させることができる。

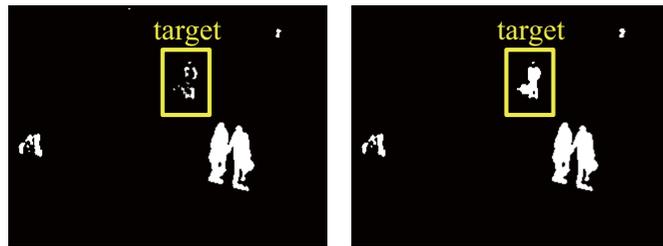
3.4 ニューラルネットワークを用いた歩行者数の推定

特徴ベクトルと歩行者数との間の非線形関係に対応付けるために、ニューラルネットワークを用いた。図 3 に、提案手法で用いたニューラルネットワークである、 n 入力、1 出力、2 層の中間層にそれぞれ n 個のニューロンを持つ 4 層パーセプトロンを示す。ここで、 n は特徴ベクトルの次元数であり、入力層と 2 層の中間層には、特徴ベクトルの次元数と同じ数のニューロンを配置している。特徴量の正規化として正規化法 2 を用いた場合は $n = 6$ 、正規化法 3 を用いた場合は $n = 7$ となる。出力は歩行者数である。また、活性化関数にはシグモイド関数を、学習アルゴリズムには RPROP アルゴリズムを用いている。

まず、各プロブから抽出した特徴ベクトルとそのプロブを構成する歩行者数の対応関係を手動で与えることで、学習データを作成する。作成した学習データをニューラルネットワークに学習させておくことで、特徴ベクトルと学習済みニューラルネットワークを用いて歩行者数を推定することができる。

3.5 動的背景モデルの更新

提案手法では、前景領域を検出するのに動的背景モデルを用いた背景差分法を用いている。通常、動的背景モデルを用いた背景差分法では、毎フレーム全ての画素の背景モデルを



(a) 歩行者情報の利用無し (b) 歩行者情報の利用有り

図 4 歩行者情報の利用の有無による歩行者領域の検出精度

更新している。しかし、毎フレーム背景モデルを更新すると、歩行者が同じ場所に居続けた場合、歩行者も背景として学習されてしまい、数フレーム後には歩行者を正確に検出できなくなってしまう。提案手法では、背景差分により得られるプロブから抽出した特徴ベクトルを用いて歩行者数を推定しているため、精度良く歩行者数を推定するには、歩行者が同じ場所に居続けても前景として検出する必要がある。

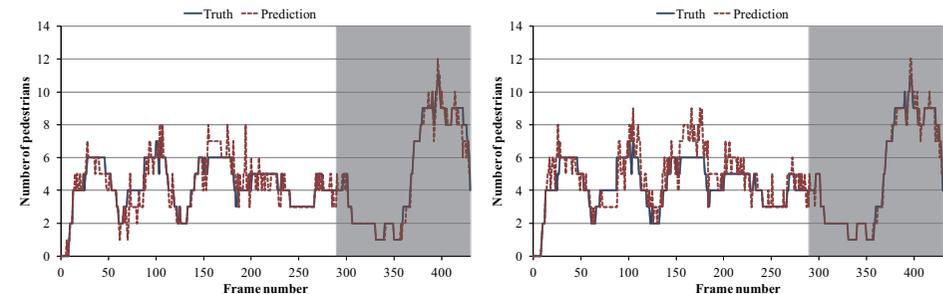
そこで提案手法では、歩行者数が1人以上と推定されたプロブの画素に対しては、背景モデルの更新を行わない。これによって、歩行者が同じ場所に居続けた場合でも、背景差分により歩行者領域を前景として検出することができる。歩行者情報を利用しなかった場合と、利用した場合における領域分割処理での歩行者領域の検出精度の違いを図4に示す。

4. 実験

実験1では、3章で述べた各正規化法を用いた場合の提案手法を、歩行者数の計測精度、計算時間、必要とされる学習データの点から評価する。実験2では、提案手法の有効性を検証するために、実験1とは別のシーンのデータを用いて、提案手法をクロスバリデーションにより評価する。

4.1 実験1：提案手法に関する実験

実験1では、PETS2006のデータセット⁹⁾を解像度320×240にリサイズしたものをを用いた。このデータセットは3020フレームの画像で構成されており、このうち2035~3020フレームの画像を用いて学習データを作成した。また、最初のフレームから7フレーム飛ばしに画像を用いて、合計431フレームで構成されるテストデータを作成した。本節では、このテストデータを用いて実験を行っている。誤差や精度の評価には、テストデータのうち



(a) 正規化法 1 (b) 正規化法 2

図 5 各正規化法を用いた場合のフレーム毎の歩行者検出数

学習データと重複しない最初の290フレーム分のみを用いている。

4.1.1 精度の評価

3.3節での述べた2種類の正規化法を用いた場合における提案手法の精度を作成したテストデータを用いて比較した。プロブベースで精度を比較した結果を表1に示す。ここで、False-PositiveとFalse-Negativeは次のように定義している。また、各正規化法を用いた場合におけるフレーム毎の歩行者検出数を図5に示す。図5において、網掛けされている部分は、学習データと重複するフレームである。

False-Positive: 歩行者以外を歩行者として検出してしまった、または、プロブを構成する歩行者の数を真値よりも多く判断した数。

False-Negative: 検出すべき歩行者を検出できなかった、または、プロブを構成する歩行者の数を真値より少なく判断した数

表1より、False-Positiveの値は、正規化法1を用いた場合に最小となり、False-Negativeの値は、正規化法2を用いた場合に最小となった。正規化法2においてFalse-Negativeの値が最小となったのは、位置情報によりスケール変化を学習しているため、ある場所では歩行者が障害物に隠れてしまうことで体の一部しか映らないなどといった局所的な情報まで学習できたからだと考えられる。

4.1.2 計算時間の評価

3.3節での述べた各正規化法を用いた場合の提案手法の計算時間を評価する。なお、実験に用いた計算機は、Pentium4 3.2GHzのCPUおよび2GBのメモリを搭載している。1フレー

表 1 各正規化法を用いた場合の精度

	正規化法 1	正規化法 2
Confirmed Pedestrians	1249	1249
False-Positive	77	199
False-Negative	84	78
Total-Error	161	277
Accuracy (%)	87.1	77.8



図 6 偏りのある学習データ

表 2 各正規化法を用いた場合の 1 フレームあたりの平均処理時間

	正規化法 1	正規化法 2
Segmentation(ms)	49.1	49.1
Pedestrian Estimation(ms)	39.2	3.9
Total(ms)	88.3	53.0

ムあたりの平均処理時間を表 2 に示す．ここで，Segmentation，Pedestrian Estimation，Total は次のように定義している．

Segmentation： 入力画像が観測されてから，入力画像に対する背景差分と影除去処理を完了するまでにかかる時間

Pedestrian Estimation： 前景領域が計算されてから，歩行者数推定までにかかる時間

Total： 入力画像が観測されてから，歩行者数を推定するまでにかかる時間

表 2 より，どちらの正規化法を用いた場合も，提案手法では 10fps 以上と高速に歩行者数の計測を行えていることがわかる．これは，実時間性を考えた場合，十分に処理速度と言える．また，正規化法 2 を用いた場合，正規化法 1 を用いた場合と比べ，かなり高速な処理が行えていることがわかる．正規化法 2 では，プロブの位置情報を組み込んだ特徴ベクトルをニューラルネットワークに学習させることで，歩行者のスケール変化に対応している．これにより，正規化法 2 では，特徴量の正規化に特別な処理を必要としないため，特徴量抽出から歩行者数の推定までが，わずか 4msec 足らずで行えている．これに対し，正規化法 1 では，特徴量の正規化のために入力画像に対し，平面射影変換を適応する必要があるため，処理時間が増大している．

4.1.3 必要とされる学習データの比較

正規化法 1 では特徴量自体を正規化しているのに対し，正規化法 2 ではプロブの位置情

報を組み込んだ特徴ベクトルをニューラルネットワークに学習させることで，画像上での歩行者のスケール変化に対応しているという違いがある．そこで本節では，得られる学習データが変化した場合，この正規化方法の違いが精度に影響を及ぼすかどうかを実験する．得られる学習データが変化する場合として，「歩行者が観測される領域に偏りがある場合」と「得られる学習データの量が変化した場合」を想定して実験を行う．

歩行者が観測される領域に偏りがある場合：

歩行者が観測される領域に偏りがあると，学習時に歩行者が全く通らなかった場所を，観測時になって歩行者が通るといったことが考えられる．このような状況において，正規化法 1 と正規化法 2 を用いた歩行者数計測では，精度に差が出るかどうかを実験する．歩行者が観測される領域に偏りがある場合として，図 6 のように画像下から数十パーセントの範囲に含まれる学習データのみを使って歩行者数の計測を行う．

得られる学習データの量が変化した場合：

学習データが少なく十分に学習できなかった場合，正規化法 1 と正規化法 2 を用いた歩行者数計測では，精度に差が出るかどうかを実験する．得られる学習データの量が変化する場合として，約 1000 枚の学習データを 100 枚ずつ段階的に増やしながら歩行者数の計測を行う．

「歩行者が観測される領域に偏りがある場合」において，正規化法 1 と正規化法 2 を用いた場合の歩行者数計測の精度を比較した結果を図 7(a) に示す．また，「得られる学習データの量が変化した場合」において，正規化法 1 と正規化法 2 を用いた場合の歩行者数計測の精度を比較した結果を図 7(b) に示す．

図 7(a) より，正規化法 1 を用いた歩行者数計測では，50% 以上の範囲内に含まれる学習データが得られれば，学習データに偏りがある場合でも 75% 程の精度が得られていることが分かる．このことから，特徴ベクトルが画像中の位置に非依存になるように正規化されていることが確認できる．しかし，正規化法 2 を用いた場合では，学習データに偏りがあると，精度が著しく低くなっている．これは，位置情報により歩行者のスケール変化を学習しているため，学習出来ていない範囲にある未知の情報に対する予測の精度が非常に低くなることが原因と考えられる．

図 7(b) より，正規化法 1 と正規化法 2 を用いた歩行者数計測では，得られる学習データの量が変化しても，両者に精度の差があまり見られないことがわかる．このことから，正規化法 1 のようにあらかじめ正規化した特徴量をニューラルネットワークに学習させる方法

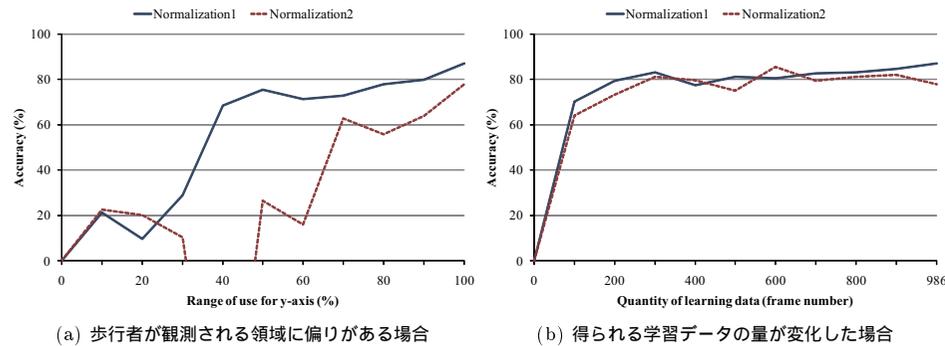


図 7 学習データの変化に対する精度比較

表 3 UCSD のデータセットに対するクロスバリデーション

	data1	data2	data3	data4
Confirmed Pedestrians	7492	6465	6860	7734
False-Positive	108	412	596	522
False-Negative	1216	365	203	809
Total-Error	1324	777	799	1331
Accuracy	82.3	88.0	88.4	82.8

と、正規化法 2 のようにスケールパラメータを用いることで、ニューラルネットワークにより内部的に特徴量を正規化する方法では、結果的に同じ処理が特徴ベクトルに対し行われていると考えられる。また、図 7(b) より正規化法 2 を用いた場合の提案手法では、全学習データを用いた場合よりも、全体の 6 割程度の学習データのみを用いた方が精度がかなり高くなっていることがわかる。これは、過学習が起こったためと思われる。このことは、図 5(b) で、正規化法 2 を用いた場合の歩行者数計測では学習データと重複する部分において、正規化法 1 を用いた場合と比べても、かなり高い精度で歩行者数を推定出来ていることから確認できる。以上より、正規化法 2 は、学習データへの依存度が高いことがわかる。正規化法 2 を用いるためには、学習データとして用いる映像に注意を払う必要がある。

4.2 実験 2: 提案手法の有効性

実験 1 より、精度と必要とされる学習データの両面から考慮すると、正規化法 1 を用いた歩行者数計測が有効であることが分かった。そこで実験 2 では、正規化法 2 を用いた場合

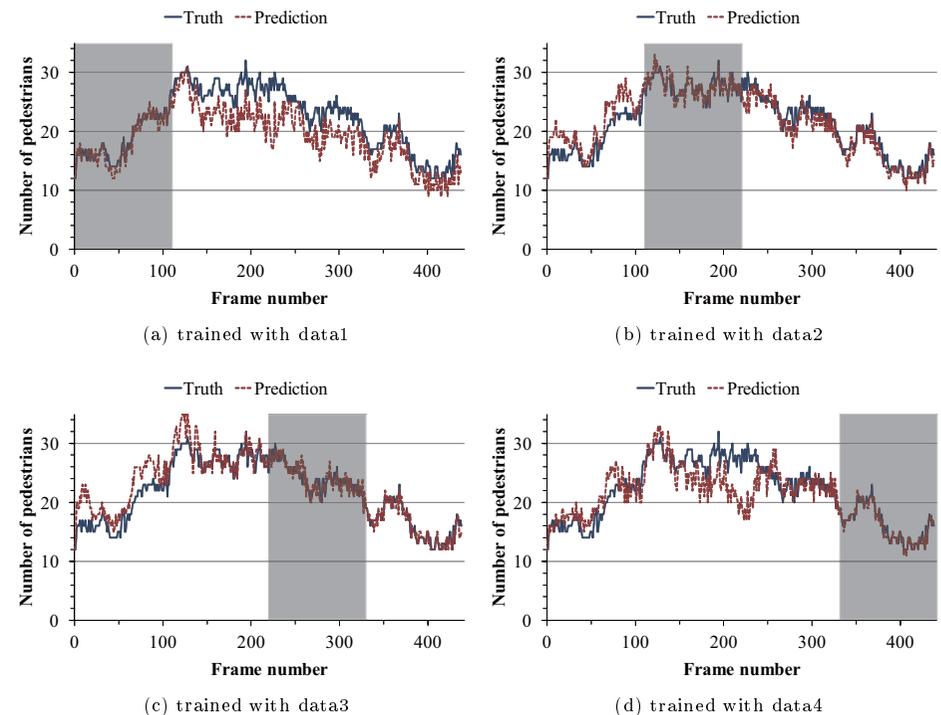


図 8 data1 ~ data4 を学習データとして用いた場合におけるフレーム毎の歩行者検出数

における提案手法の有効性を調べるため、実験 1 とは別のシーンの映像を用いて実験を行う。本節では、Chan らが用いている UCSD(University of California, San Diego) のデータセットを解像度 320×240 にリサイズしたものをを用いた。このデータセットから 440 フレーム分の画像を使用し、学習データを作成した。本節では、作成した学習データを 110 枚ずつ 4 つのデータセット (data1 ~ data4) に分割し、これらのデータセットを用いてクロスバリデーションを行うことで提案手法の有効性を検証する。プロベースで精度を比較した結果を表 3 に示す。また、フレーム毎の歩行者検出数を図 8 に示す。図 8 において、網掛けされている部分は、学習データと重複するフレームである。

表 3 より、分割した 4 つのデータセットのうちのどれを用いても、80% 以上の精度で歩行

者数を推定出来ていることが確認でき、提案手法は歩行者数の推定に有効であると言える。

また、表3と図8より、data2, data3を学習データとして用いた場合、精度が高くなっていることがわかる。これは、学習データとして用いた映像中に、単独で歩く歩行者から大人数の集団となって歩く歩行者まで映っていたことで、様々な情報が学習できたため、テストデータに対しても精度良く推定が行えたからである。

逆に、data1, data4を学習データとして用いた場合、100~300フレームにおいて歩行者の検出精度が特に悪くなっていることがわかる。これは、data1, data4の映像には大人数の集団となって歩く歩行者が映っておらず、十分な学習ができなかったことで、100~300フレームの映像で観測される大人数の集団となって歩く歩行者数を正しく推定できなかったと考えられる。

どのような映像を学習データとして用いるかで、歩行者数の計測精度に差が出てくる場合がある。このため、学習データとしてどの映像を用いるが十分に考慮する必要がある。

5. おわりに

本稿では、「画像上における歩行者の位置と人数」を実時間で推定する手法を提案した。異なる2シーンで実験を行った結果、提案手法では、80%以上の精度かつ10fps以上の処理速度で「画像上における歩行者の位置と人数」を推定できることが確認できた。

今後の課題としては、最適な特徴量の選択、背景差分の改良、平面射影変換に関する指標の提示、学習データの再利用（汎用的な学習データの作成）などが挙げられる。

最適な特徴量の選択：歩行者数計測に最適な特徴量を用いて精度が上がる可能性がある。

このため、各プロブから抽出する特徴ベクトルとして、幾通りもの特徴量の組み合わせを試し、歩行者数の計測に最適な特徴ベクトルを探す必要がある。

背景差分の改良：提案手法では、「画像中の何処に何人の歩行者がいるか」を推定するのに背景差分で検出されたプロブを利用している。背景差分では、歩行者が同じ場所に居続くと背景に取り込まれてしまうという問題があり、歩行者のプロブが分裂したり、異常に小さくなってしまふ。分裂した同一歩行者のプロブをそれぞれ別の歩行者としてカウントしたり、小さくなったプロブを歩行者として検出できなくなることで、誤差が大きくなるという問題がある。プロブを構成する歩行者数が1人以上と推定された場合、そのプロブの背景モデルを更新しないようにすれば、この問題を解決できると思われる。

平面射影変換に関する指標の提示：正規化法1を用いるには、入力画像を平面射影変換しなくてはならない。平面射影変換に必要なhomographyは実験的に求める必要がある。

図2のような同一人物に対する領域分割画像とROIの4頂点を与えることで、人物領域の面積が同程度になる基底面のとり方を示せるなど、homographyを求める際に何らかの指標を示すとこの作業にかかる手間が軽減できる。

学習データの再利用（汎用的な学習データの作成）：提案手法では、観測するシーン毎にそれぞれ学習データを用意しなくてはならない。学習データを作成する作業は、大変な手間がかかるため、一度作成した学習データを別のシーンに対しても転用可能なように変換したり、シーンの変化に対し不変な特徴量を探し、その特徴量を用いて汎用的な学習データを作成する必要がある。

参考文献

- 1) Viola, P., Jones, M. and Snow, D.: "Detecting pedestrians using patterns of motion and appearance", *IJCV*, Vol.63(2), pp.153-161 (2005).
- 2) Zhao, T. and Nevatia, R.: "Bayesian human segmentation in crowded situations", *CVPR*, Vol.2, pp.495-466 (2003).
- 3) Antonini, G. and Thiran, J.P.: "Counting Pedestrians in Video Sequences Using Trajectory Clustering", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.16, pp.1008-1020 (2006).
- 4) Kong, D., Gray, D. and Tao, H.: "Counting pedestrians in crowds using viewpoint invariant training", *British Machine Vision Conf.* (2005).
- 5) Chan, A.B. and Vasconcelos, N.: "Mixtures of Dynamic Textures", *IEEE International Conference on Computer Vision* (2005).
- 6) Chan, A.B., Liang, Z.-S.J. and Vasconcelos, N.: "Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking", *IEEE Conf. on Computer Vision and Pattern Recognition(CVPR)* (2008).
- 7) 田中達也, 島田敬士, 有田大作, 谷口倫一郎, 富浦洋一: "高速なParzen推定を用いた動的背景モデルによる映像からの物体検出", *映像情報メディア学会誌*, Vol.62, pp.(141)2045-(148)2052 (2008).
- 8) Schreer, O., Feldmann, I., Golz, U. and Kauff, P.: "FAST AND ROBUST SHADOW DETECTION IN VIDEOCONFERENCE APPLICATION", *4th IEEE Intern. Symposium on Video Proces. and Multimedia Comm.*, pp.371-375 (2002).
- 9) with IEEE Conference on Computer Vision, I.C. and 2006, P.R.: "Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2006)", <http://ftp.pets.rdg.ac.uk/PETS2006/> ((accessed 2008/01/15)).