

平成26年度 修士論文

Deep Learningを用いた列車運行映像における
高精度な乗客検出に関する研究
Research on High-precision Passenger Detection
Using Deep Learning for Train Operation Video

指導教授 渡辺 裕 教授

2015年3月

早稲田大学大学院 国際情報通信研究科
国際情報通信学専攻

4213A146-8

張 傑

Jie ZHANG

目次

第1章 序論	1
1.1 研究の背景	1
1.2 本研究の目的	2
1.3 本論文の構成	3
第2章 列車運行の安全の現状	4
2.1 まえがき	4
2.2 鉄道運転事故	4
2.2.1 鉄道運転事故の種類	
2.2.2 鉄道運転事故の発生状況	
2.3 安全対策の実施状況	7
2.4 安全対策の問題点	8
2.5 むすび	8
第3章 画像処理による人物検出の関連研究	9
3.1 まえがき	9
3.2 固定カメラによる高速な人物検出	10
3.2.1 背景差分	
3.2.2 人物判定処理	
3.3 人物検出に利用される特徴量	11
3.3.1 HOG特徴量	
3.3.2 LBP特徴量	
3.4 人物検出に利用される機械学習	13
3.5 比較実験	16
3.5.1 実験条件	
3.5.2 実験結果	
3.6 むすび	18
第4章 Deep Learning を用いた高精度な人物検出	19
4.1 まえがき	19
4.2 Deep Learning の考え方	19

4.3 多層パーセプトロン	20
4.4 畳み込みニューラルネットワーク	22
4.4.1 多層パーセプトロンと畳み込みニューラルネットワーク	
4.4.2 畳み込みニューラルネットワークの特性	
4.4.3 畳み込みとマックスプーリングの実行例	
4.5 高精度な人物検出のための畳み込みニューラルネットワーク	24
4.5.1 Supported Layer Types	
4.5.2 Forward Propagation	
4.5.3 Back Propagation	
4.5.4 提案手法の畳み込みニューラルネットワークの構造	
4.6 むすび	29
第 5 章 提案手法の評価実験	30
5.1 まえがき	30
5.2 識別率調査	30
5.2.1 実験概要	
5.2.2 実験条件	
5.2.3 実験結果	
5.3 回転安定性調査	34
5.3.1 実験概要	
5.3.2 実験条件	
5.3.2 実験結果	
5.4 乗客検出の実行	34
5.4.1 学習画像の生成	
5.4.2 システム概要	
5.4.3 実行結果	
5.5 むすび	37
第 6 章 結論	38
6.1 総括	38
6.2 今後の課題	39
参考文献	40

第1章

序論

1.1 研究の背景

交通機関や重要施設等において、人身事故抑制の強化は大きな課題となっている。特に、交通機関の代表となる鉄道では、駅構内で多数の利用者が行き交うため、列車との接触危険性がある。列車運行中に人身事故が発生した場合には、多数の利用者に影響を与えることとなる。したがって、列車運行の安全を確保するための技術が益々必要となっている。

駅構内においては、監視カメラの設置台数を増やすことで、監視範囲を十分に広くとることができる。ただし、全てのカメラ映像を常時注視しているわけではないため、リアルタイムで異常を発見する事は困難である。実状では監視カメラの映像は、事故抑制ではなく、事故発生時の証拠取得としての側面が強い。そのため、目視監視を補助あるいは代行する機能を備えた画像処理を実現できれば、映像から異常を自動的に検知することで、リアルタイムに監視員に判定画像を提示し、問題発生の予防、予測などに役立てることができる。

映像解析システムの中で、基礎となっている部分が乗客検出である。目視と同様の精度をもつ検出機能を画像処理で実現することが人工知能の研究として長年に渡り検討されてきたが、その達成は未だ困難である。また、問題発生の予防、及び予測機能を達成するため、高精度であるだけでなく、処理速度が重要となる。従来提案されてきた様々な画像特徴量と学習手法の組み合わせを用いた検出手法では、高精度かつ高速な処理を実現することが困難である。そこで、近年ではディープラーニングというアプローチが提案され、人工知能、機械学習の分野で大変注目を集めている。ディープラーニングとは、多層ニューラルネットワーク構築手法の1つであり、人間の神経回路と似た深い構造を作り出し、学習により最適な特徴量を獲得可能であり、従来の特徴量を任意に設定する手法と比較して、精度の向上が可能である。しかし、この手法を用いた列車運行映像における乗客検出の研究はほとんどされていないのが現状である。

このような研究背景において、本研究では、ディープラーニングを用いた列車運行映像における高精度な乗客検出の実現を目的とする。

1.2 本研究の目的

本研究の目的は、列車運行映像における高精度かつ高速な乗客検出の実現である。従来の検出手法では、勾配、色、動き、距離などの情報に基づく手法が採用されてきた。これらの方法は次のような特徴がある。

- 低次元の特徴だけを抽出する
- 高次元の特徴を抽出できる特徴量の設計が難しい
- 高精度と高速の両立が難しい
- 過学習が起きやすい

具体的な内容を Table 1.1 に示す[1]。従来手法のこのような特徴は、複雑な背景と揺るぎを持つ列車運行映像における高精度な乗客検出に向かない。

そこで我々は、より高精度かつ高速な検出を行うため、ディープラーニングを採用する。しかしながら、ディープラーニングは複数のレイヤーを通じて処理を行うため、高い精度と処理速度を両立するディープラーニングの設計が困難である。そこで、本研究では、ディープラーニングを構成するレイヤーの構成を変更した多数の実験により、各レイヤーの作用と影響を明らかにする。

Table 1.1. 従来の検出手法とそれらの特徴

特徴量	ベース	長所	短所
EOH Edgelet HOG	勾配	人の特徴をよく表現できる	テクスチャの影響を受ける
LBP ICF CSS	色	エッジが明確でない画像からも有効な特徴が抽出できる	計算コストが大きい
前景尤度 HOF RDSF	動き	テクスチャの影響を受けない	連続した複数枚の画像が必要
距離差分	距離	テクスチャの影響を受けない 人の形状を明確に捉えられる	特殊なデバイスが必要 画角、距離に制限がある

1.3 本論文の構成

以下に本章以降の構成を示す。

第1章 本章であり、研究の背景およびその目的について述べている。

第2章 列車運行の安全の現状について述べる。まず、列車の運転中における事故の発生状況を述べる。次に、これらの事故に対し、相応な安全対策の実施状況を述べる。特に、鉄道人身障害事故と踏切障害事故の防止対策を述べる。最後に、これらの安全対策の問題点を示す。

第3章 画像処理による人物検出の関連研究について述べる。3.2 節では、移動中の人物を検出する観点から固定カメラによる高速な人物検出方法を紹介する。3.3 節では、人物検出に利用される代表的な特徴量を紹介する。3.4 節では、人物検出に利用される代表的な機械学習を紹介する。3.5 節では、代表的な特徴量と学習手法を使った2つの実験を行う。

第4章 Deep Learning を用いた高精度な人物検出について述べる。まず、Deep Learning の考え方を述べる。次に、多層パーセプトロンを紹介する。そして、畳み込みニューラルネットワークの基礎について述べる。最後に、提案手法である高精度な人物検出のための多層畳み込みニューラルネットワークの構造や演算などについて述べる。

第5章 第4章で提案した、畳み込みニューラルネットワークを用いて、評価実験を行う。まず、提案手法の識別率調査を行う。次に、提案手法の回転安定性調査を行う。最後に、列車運行映像における乗客検出を実行する。

第6章 本研究の総括と今後の課題について述べる。

第2章

列車運行の安全の現状

2.1 まえがき

本章では列車運行の安全の現状について述べる。まず、列車の運転中における事故の発生状況を述べる。次に、これらの事故に対し、相応な安全対策の実施状況を述べる。特に、鉄道人身障害事故と踏切障害事故の防止対策を述べる。最後に、これらの安全対策の問題点を示す。

2.2 鉄道運転事故

本節では、列車の運行中における鉄道人身事故、踏切障害事故等の発生状況を述べる。都営地下鉄、JR東日本、西武鉄道、三つの企業から鉄道運転事故の発生状況を調査する。

2.2.1 鉄道運転事故の種類

鉄道運転事故は、列車の運転中における事故である。鉄道事業法及び鉄道事故等報告規則で、鉄道運転事故は「列車衝突事故」「列車脱線事故」「列車火災事故」「踏切障害事故」「道路障害事故」「鉄道人身事故」「鉄道物損事故」の7種類に分けられる。具体的な事故及びその内容は Table 2.1 のように分類されている[2]。

Table 2.1. 事故の分類と内容

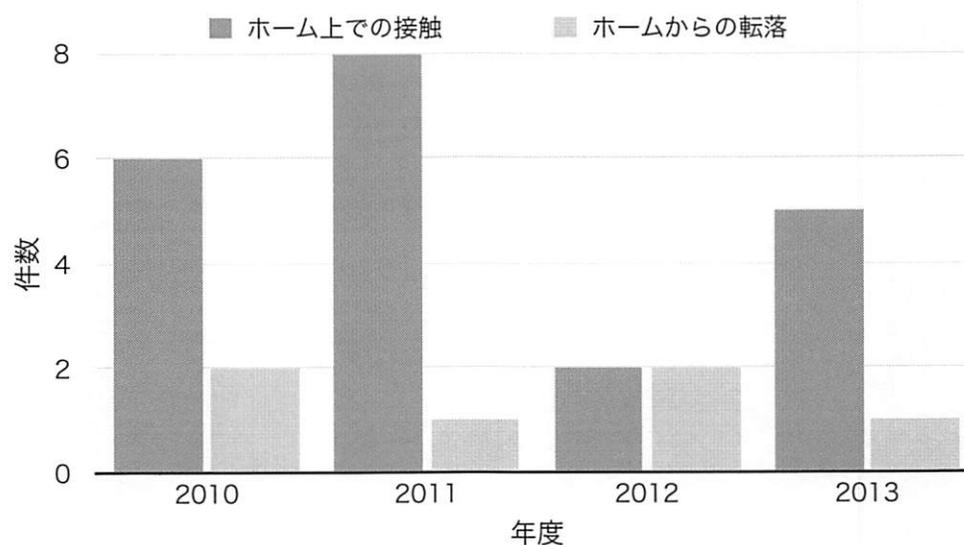
分類		内容
鉄道運転事故	列車衝突事故	列車が他の列車又は車両と衝突し, 又は接触した事故
	列車脱線事故	列車が脱線した事故
	列車火災事故	列車に火災が生じた事故
	踏切障害事故	踏切道において列車又は車両が道路を通行する人又は車両等と衝突し, 又は接触した事故
	道路障害事故	踏切道以外の道路において, 列車又は車両が道路を通行する人又は車両等と衝突し, 又は接触した事故
	鉄道人身事故	列車又は車両の運転により人の死傷を生じた事故(上記の事故に伴うものを除く)
	鉄道物損事故	列車又は車両の運転により500万円以上の物損を生じた事故(上記の事故に伴うものを除く)

2.2.2 鉄道運転事故の発生状況

(1) 都営地下鉄鉄道運転事故

平成25年, 鉄道人身障害事故が6件発生し, 内訳はホーム上での接触が5件, ホームからの転落が1件である。また, 過去4年の鉄道運転事故件数は Fig. 2.1 に示す[3]。

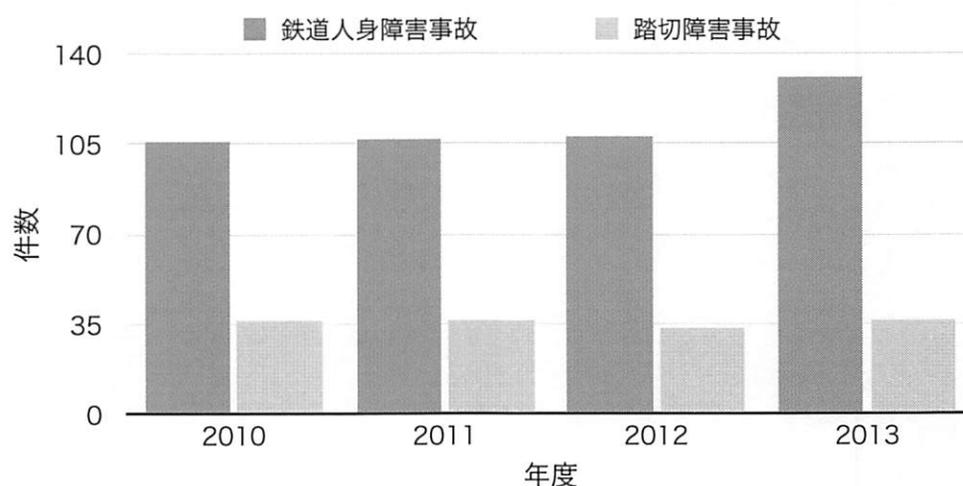
Fig. 2.1. 都営地下鉄鉄道運転事故件数



(2) JR東日本鉄道運転事故

平成25年、鉄道運転事故が171件発生した。このうち、鉄道人身障害事故が全体の約77%を占めている。また、過去4年の鉄道運転事故件数は Fig. 2.2 に示す[4]。

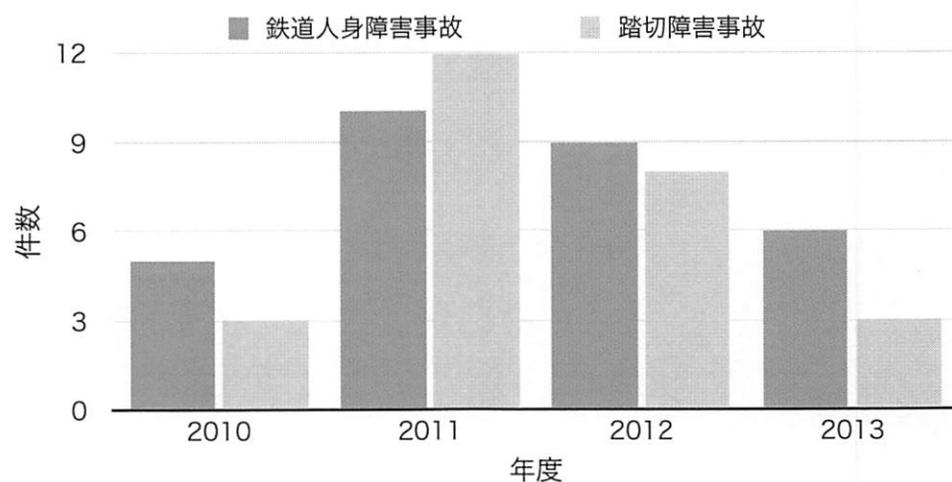
Fig. 2.2. JR東日本鉄道運転事故件数



(3) 西武鉄道運転事故

平成25年、鉄道運転事故が9件発生した。このうち、鉄道人身障害事故が6件、踏切障害事故が3件である。また、過去4年の鉄道運転事故件数は Fig. 2.3 に示す[5]。

Fig. 2.3. 西武鉄道運転事故件数



2.3 安全対策の実施状況

本節では、鉄道運転事故における人身障害事故及び踏切障害事故を防ぐための対策を述べる。各企業の対策を統合し、現在実施している対策は大きく以下に示す 14 種類である [3,4,5].

01) 列車緊急停止スイッチ・ボタン

乗客がホームから転落した場合などに、駅付近の列車を自動的に非常停止させるスイッチ・ボタンである。

02) ステップ

乗客がホームから転落した場合にホームに上がりやすくするためのステップである。

03) ホームの内側を表示する内方線

目の不自由な乗客が、何方がホームの内側なのか、足で踏むことで判別できるような線状突起である。

04) ホーム端転落防止柵

ホーム両端部に設置し、乗客の転落を防止するための柵である。

05) 転落感知マット

転落が起きると、スイッチが入るといったマットである。

06) ITV モニター装置

列車がホームを出発する際や停車中に、乗務員が乗客の乗り降りやホーム上の安全を確認するための装置である。

07) ホームドアの整備

乗り降りの時だけ開ける自動ドアである。

08) 防災監視盤

駅務室に設置し、ITV モニターや放送装置などにより、乗客の避難誘導に迅速に対応できるような総合防災装置である。

09) 非常通報器

列車内で異常事態が発生した場合に、乗客から乗務員又は指令所に通報できるようなスイッチ・ボタンである。

10) 非常発報装置

乗務員が緊急時に架線の送電を止めるため、司令所に異常を通報する非常発報ボタンである。

11) ATC (自動列車制御装置)

列車が先行列車との間隔及び進路の条件によって決定される運行速度を超えると、自動的ブレーキをかけて減速又は停止させることにより、列車の速度を制御するものである。

12) ATS (自動列車停止装置)

運転士が停止信号を誤認して出発した時などに、自動的にブレーキをかけ、列車を減速又は停止させる装置である。

13) 運転士異常時列車停止装置

運転士の異常時に列車を自動的に停止させる装置である。

14) 列車防護無線装置

緊急時に列車から電波を発信し、付近を走行する列車に停止を求める装置である。

2.4 安全対策の問題点

前節では、人身障害事故及び踏切障害事故を防ぐための対策を述べた。しかし、これらの対策において、下記の問題点がある。

上記の安全対策はホームを中心に行っている。事故は全線路において何処でも発生する可能性があるため、理想的には線路上の全区間において安全装置を設置する必要がある。また、安全装置の設置を含めた総合的な対策費用の問題がある。東京都交通局安全報告書によると、平成 25 年度の設備投資は約 221 億円を投入した[3]。このような膨大な資金が企業に対して大きい負担となっている。

一方、事故の本質的な問題として、発生予測が困難であるという点が挙げられる。前節で紹介した 14 種類の安全対策の内、殆どは事故発生時の対応方法である。事故抑制の目的を達成するため、事故予測の技術が益々必要となっている。現状、事故の予測には人的な判断が必要である。ITV モニター装置や防災監視盤などの安全対策は駅務員の常時注視が必要である。また、緊急停止ボタン・スイッチや非常通報器や非常発報装置などの対策も人の操作が必要である。ただし全ての異常をリアルタイムで発見することは困難である。

2.5 むすび

本章では列車運行の安全の現状について述べた。まず、列車の運転中における事故の発生状況を述べた。次に、鉄道運転事故に対し、相応な安全対策の実施状況を述べた。特に、鉄道人身障害事故と踏切障害事故の防止対策を述べた。最後に、これらの安全対策の問題点を示した。

2.4 節に示した問題点を解決するため、知的な画像解析技術が必要である。第3章では、最も重要な画像解析技術、すなわち画像処理による人物検出の関連研究について述べる。

第3章

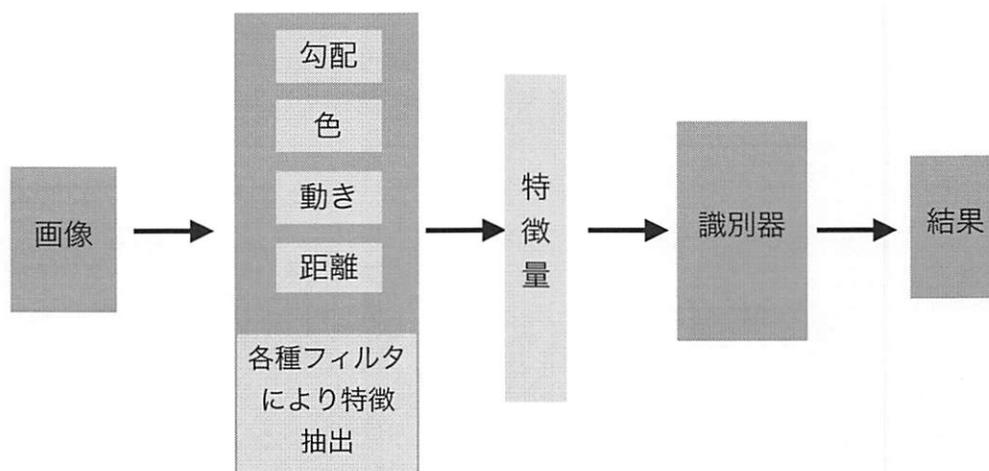
画像処理による人物検出の関連研究

3.1 まえがき

ITV モニター装置や防災監視盤などの設備の中で、自動化できていない部分が駅務員による目視機能である。このような目視機能を画像処理で実現することが人工知能の研究として長年に渡り検討されてきたが、未だに困難な課題である。映像から移動中の人物を検出する観点から、従来の代表的な画像処理を概観すると、以下の Fig. 3.1 に示すような関係が描ける。従来の代表的な画像処理は、様々な種類のフィルタによる特徴抽出と識別器による特徴量を識別の2段階と見なすことができる。

人物検出の画像処理では、人物の検出精度並びに処理速度が重要となる。そこで、本章の3.2 節では、移動中の人物を検出する観点から固定カメラによる高速な人物検出方法を紹介する。3.3 節では、人物検出に利用される代表的な特徴量を紹介する。3.4 節では、人物検出に利用される代表的な機械学習を紹介する。3.5 節では、代表的な特徴量と学習手法を使った2つの実験を行う。

Fig. 3.1. 従来の人物検出手法の流れ



3.2 固定カメラによる高速な人物検出

本節では、固定カメラ画像から人物を高速に検出する手法を紹介する。まず、背景差分について述べる。次に、移動体候補領域内の人物判定処理を述べる。

3.2.1 背景差分

固定カメラで撮影した映像の静止背景を人物が移動する場合、背景差分やフレーム間差分により検出した画像内の变化領域が歩行者などの移動体候補領域となる。例として、以下のような最も一般的な背景差分による变化領域の検出方法を紹介する[6]。

時刻 t における画像 I_t 内の座標 (x, y) での画素値を $I_t(x, y)$ とする。各座標の画素値を一定時間平均した画像を A とし、例えば時刻 $t-1$ まで平均した画像を $A_{t-1}(x, y)$ とすると、背景差分による变化領域 $B_t(x, y)$ は式 (3.1) により求められる。ここで、 T_t は閾値である。

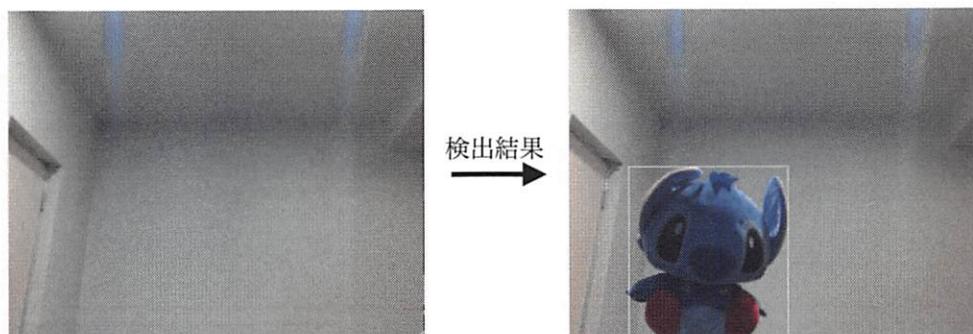
$$B_t(x, y) = \begin{cases} 1 & \text{if } |I_t(x, y) - A_{t-1}(x, y)| \geq T_t \\ 0 & \text{else} \end{cases} \quad (3.1)$$

閾値 T_t は背景画像などの標準偏差 $\sigma_{t-1}(x, y)$ を使い、式 (3.2) により定義される。 α は時間的な経過を反映するための重み係数である。

$$T_t(x, y) = \alpha \times \sigma_{t-1}(x, y) \quad (3.2)$$

照明変化のない室内で動く物体が人物だけであれば、差分処理で検出した变化領域が人物領域となる。ただし、実状では、動く物体は人物に限らない。例えば、Fig. 3.2 はぬいぐるみが検出された例である。従って、人物のみを検出する場合には、背景差分以外の処理が必要である。

Fig. 3.2. 背景差分法による動物体の検出



3.2.2 人物判定処理

前節で紹介した背景差分処理で検出した変化領域に人物がいるかどうかを判断するため、パターン認識により人物画像と非人物画像を区別する人物判定処理を行う。本節では、人物画像と人物以外の二種類の画像を学習データとして Real AdaBoost に与える実験を簡単に説明する。特徴量と機械学習について詳しい解説は次節で行う。

Real AdaBoost は機械学習の一つの手法である。Real AdaBoost での人物判定関数 $H(x)$ は式 (3.3) で定義される。ここで $h_i(x)$ は i 番目の弱識別器の出力を示す。判定関数 $H(x)$ が閾値より高ければ人、低ければ非人と判定する。

$$H(x) = \text{sign} \left[\sum_{i=1}^n h_i(x) - \lambda \right] \quad (3.3)$$

3.3 人物検出に利用される特徴量

本節では、人物検出に利用される代表的な特徴量を紹介する。3.3.1 節では、HOG特徴量を述べる。また 3.3.2 節では、LBP特徴量を紹介する。

3.3.1 HOG特徴量

HOG は “Histograms of Oriented Gradients” の略称で、画像の局所領域から輝度勾配、及び輝度強度を取り出す特徴量である。幾何学的変換に強く、また照明の変動に頑健ため、歩行者や人工物などの検出によく用いられる特徴量となっている[7,8,9,10]。

HOG では、輝度勾配ヒストグラムの対応する要素に輝度強度を累積する。さらに、正規化処理によりヒストグラムの形状を整える。具体的に以下4つのステップがある。

- 1) 局所領域をセル・ブロックに分割。
- 2) ブロックを移動し、セルの中における輝度勾配ヒストグラムを計算する。
- 3) ブロックごとに正規化する。
- 4) 正規化したヒストグラムを統合して特徴量とする。

輝度の勾配方向と強度の算出

勾配方向は式 (3.4) で計算される。

$$\theta(x,y) = \tan^{-1} [f_y(x,y) \div f_x(x,y)] \quad (3.4)$$

勾配強度は式 (3.5) で定義される.

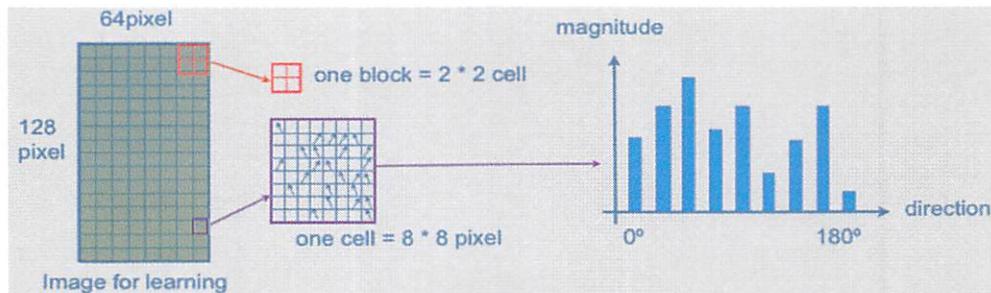
$$m(x,y) = [f_y(x,y)^2 + f_x(x,y)^2]^{0.5} \quad (3.5)$$

$f_x(x,y)$ は画素 (x,y) の右画素 $(x+1,y)$ と左画素 $(x-1,y)$ の差である. $f_y(x,y)$ は画素 (x,y) の下画素 $(x,y+1)$ と上画素 $(x,y-1)$ の差である.

ヒストグラムの作成

ヒストグラムは Fig. 3.3 に示すように作成できる. ここでは, 学習画像のサイズを $64 * 128$ pixel に統一し, セルの大きさを $8 * 8$ pixel に規定する. また, ブロックの大きさを $2 * 2$ セルとする. さらに, 勾配方向を 0° から 180° まで, 20° ずつ 9 方向に分割する. この場合, 特徴量は 3780 次元となる.

Fig. 3.3. ヒストグラムの作成



ブロックごとの正規化

ヒストグラムの形状を整えるため, ブロックごとの正規化を行う. ある n 番目の HOG 特徴量の正規化は式 (3.6) により行う. ここで, q はブロックのサイズである. N は勾配方向数である. ε は分母が 0 にならないための定数である.

$$V(n) = V(n) \div \left(\sum_{k=1}^{q \times q \times N} V(k)^2 + \varepsilon \right)^{0.5} \quad (3.6)$$

3.3.2 LBP特徴量

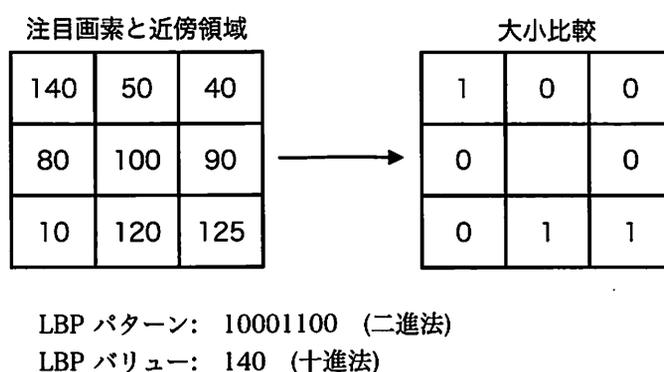
LBPは“Local Binary Pattern”の略称で、画像の認識や分類に用いられる代表的な特徴量の1つである[10,11]。LBPは通常3*3の画素領域で計算され、局所的な特徴を抽出する。LBPは照明変化の影響を受けにくく、高速に計算できるという特徴から近年注目されている。特徴量はラスタスキャンにより、注目画素の周囲の半径をR、近傍領域の画素数をPとして以下式(3.7)のように算出される。

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, s(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad (3.7)$$

ここで、 g_c は注目画素の画素値を示し、 g_p は注目画像を中心に参照点の画素値を表している。近傍領域はR=1の時、近傍領域は3*3であり、Pの最大は8となる。R=2の時、近傍領域は5*5となり、Pの最大は16となる。LBPは注目画素と近傍領域の画素値の大小比較によりパターン $LBP_{P,R}$ が算出され、その種類は 2^P となる。LBPはこのパターンの頻度を記述したヒストグラムの特徴量である。

LBPは理論上PとRの値に制限はないが、特徴量の次元数が爆発するため、比較実験の節で用いるLBPはR=1、P=8に設定する。この時、LBPはFig. 3.4のように3*3の近傍領域の画素で算出され、 $LBP_{8,1}$ の次元数は256となる。

Fig. 3.4. LBP特徴量の計算



3.4 人物検出に利用される機械学習

本節では、人物検出に利用される代表的な機械学習を説明する。機械学習は、教師なし学習と教師あり学習の2タイプに分けられる。教師なし学習は、ベクトル量子化あるいはクラスタリングなどにより、パターンの分布を幾つかのクラスに分ける手法である。一方、教師

あり学習はサポートベクターマシンや AdaBoost などのように学習パターンとその正解クラスを与え、パターンを正しく分類する関数を作成する手法である。

本研究では、検出対象の人物画像が収集できる前提で、教師あり学習手法を紹介する。まず、サポートベクターマシンについて述べる。次いで、AdaBoost と Real AdaBoost について解説する。

3.4.1 サポートベクターマシンの概要

SVM は線型の識別器であるが Mercer カーネルを使うことで非線型に拡張され、近年画像認識で高い識別率を上げているパターン認識手法である[12]。

SVM は2クラス分類器であり、学習データ $x_i \in R^d$ とそのクラス $y_i \in \{-1, 1\}$ で、事前にラベル付けられた学習データ $\{x_i, y_i\}$, $i = 1, \dots, n$ を使い、式 (3.8) の形式で表記される識別関数を求める手法である。

$$F(x) = \text{sign}(w \cdot x + b) \quad (3.8)$$

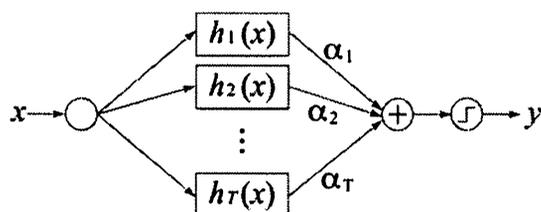
ここで、 sign は -1 と 1 を返す符号関数であり、「 \cdot 」は内積を表す。 b はバイアス項と呼ばれるパラメーターである。識別面から各クラスまでの距離をマージンと呼び、SVM ではマージンを最大化するように識別面を決めることで誤認識を最小化している。

全ての学習データが分離できれば、 $y_i = 1$ の時 $(w \cdot x_i + b) \geq 1$, $y_i = -1$ の時 $(w \cdot x_i + b) \leq -1$ となり、 $y_i(w \cdot x_i + b) \geq 1$ と表現できる。この時のマージンの大きさは $1 / \|w\|$ と表され、このマージンを最大化する問題は $\min 0.5 \times \|w\|^2$ を求めることと等価となる。

3.4.2 AdaBoostの概要

AdaBoost は逐次に学習データの重みを変化させながら異なる識別器を作り、これら複数の識別器の重み付き多数決により最終的な識別関数を与える手法である[13,14]。個々の識別器は弱識別器と呼ばれ、それらを組み合わせたものは強識別器と呼ばれる。AdaBoost により得られる識別器の構造を Fig. 3.5 に示す。弱識別器は入力パターン x に対し、そのパターンが属するクラスのラベル y を出力する。AdaBoost では、 T 個の弱識別器 $h_t(x)$ を信頼度 α_t で重み付けて結合することにより、強識別器を構成する。

Fig. 3.5. AdaBoost により得られる識別器



AdaBoost の学習アルゴリズムは以下の手順で行う[13].

- 1) 各サンプルの重みを初期化
- 2) For $t = 1, \dots, T$:
 - A: サンプル分布において, 弱識別器を学習する
学習サンプルに対するエラー率が最小となる
弱識別器を選ぶ
 - B: エラー率から信頼度 α を計算する
 - C: サンプルの重みを更新する
 - D: サンプルの重みの和が1になるように正規化する
- 3) 強識別器を構築

AdaBoost の学習結果は SVM と同じように学習サンプルを2分類している. T 個の弱識別器 $h_t(x)$ と信頼度 α_t の積を合計し, その値が閾値より高ければ, 目的とした対象であると判定する.

3.4.3 Real AdaBoostの概要

Real AdaBoost は AdaBoost からの派生である. 弱識別器の出力が特徴量の分布に応じて実数化されるので, 効果的な重みの更新が可能である. 単純な AdaBoost より, 学習の収束が速いという利点がある.

Real AdaBoost の学習アルゴリズムは以下の手順で行う[13].

- 1) 各サンプルの重みを初期化
- 2) For $t = 1, \dots, T$:
 - A: 弱識別器の確率密度関数を作成する
 - B: 弱識別器候補の評価値を算出する
 - C: 弱識別器候補から最適な弱識別器を選択する
 - D: サンプルの重みを更新する
 - E: サンプルの重みの和が1になるように正規化する
- 3) 強識別器を構築

Real AdaBoost の学習結果は AdaBoost と同じように学習サンプルを2分類している。しかし、AdaBoost のように信頼度 α は必要ない。T個の弱識別器 $h_t(x)$ の出力を合計し、その値が閾値より高ければ、目的とした対象であると判定する。

3.5 比較実験

本節では、HOG 特徴量と LBP 特徴量の有効性を調査するため、2つの実験を行う。

- 1) HOG 特徴量と Real AdaBoost を使った人物認識
- 2) LBP 特徴量と Real AdaBoost を使った人物認識

3.5.1 実験条件

本実験では、人物認識を対象とする。実験には INRIA Person Dataset を使用し、学習画像を 3500 枚、評価サンプルを 1000 枚とした。HOG 特徴量の抽出に用いるパラメータを Table 3.1 に示す。学習画像のサイズを $64 * 128$ pixel に統一し、セルの大きさを $8 * 8$ pixel に規定する。また、ブロックの大きさが $2 * 2$ セルとする。さらに、勾配方向を 0° から 180° まで、 20° ずつ9方向に分割する。結果として、HOG 特徴量の次元数は 3780 次元

となる。

LBP 特徴量を抽出するときに対応するパラメータを Table 3.2 に示す。学習画像のサイズを 64×128 pixel に統一し、ブロックの大きさを 16×16 pixel に規定する。また、 $R=1$ 、 $P=8$ に設定する。この時、LBP 3×3 の近傍領域の画素で算出され、ブロックごとの次元数は 256 となる。LBP 特徴量の次元数は 8192 次元となる。

識別器には Real AdaBoost を用いる。確率密度分布のビン数を 64 に、学習回数を 300 回に設定する。

Table 3.1. HOG 特徴量を抽出するときに対応するパラメータ

学習画像のサイズ	64*128 (pixel)
セルの大きさ	8*8 (pixel)
ブロックの大きさ	2*2 (セル)
勾配方向	9
特徴量次元数	3780

Table 3.2. LBP 特徴量を抽出するときに対応するパラメータ

学習画像のサイズ	64*128 (pixel)
セルの大きさ	3*3 (pixel)
ブロックの大きさ	16*16 (pixel)
ブロックごとの次元数	256
特徴量次元数	8192

3.5.2 実験結果

評価サンプルの 1000 枚の画像に対し、Real AdaBoost 識別器における HOG 特徴量と LBP 特徴量の認識結果をそれぞれ Fig. 3.6 と Fig. 3.7 に示す。293 枚の人物画像に対し、HOG+Real AdaBoost による識別された数は 256 枚であり、LBP+Real AdaBoost による識別された数は 258 枚である。識別率に対する大きな差が見られなかった。

Fig. 3.6. HOG+Real AdaBoost の認識結果

```
*****
Human Images = 293 , Nonhuman Images = 707
*****
                Human      Nonhuman
Human           256         1
Nonhuman        37         706
Recognition Rate 0.87372    0.998586
Total Recognition Rate = 0.962
*****
```

Fig. 3.7. LBP+Real AdaBoost の認識結果

```
*****
Human Images = 293 , Nonhuman Images = 707
*****
                Human      Nonhuman
Human           258         4
Nonhuman        35         703
Recognition Rate 0.88055    0.99434
Total Recognition Rate = 0.961
*****
```

3.6 むすび

本章では画像処理による人物検出の関連技術について述べた。まず、移動中の人物を検出する観点から固定カメラによる高速な人物検出方法を紹介した。次に、人物検出に利用される HOG 特徴量と LBP 特徴量を紹介した。そして、人物検出に利用される代表的な機械学習の中のサポートベクターマシン, AdaBoost, 及び Real AdaBoost, 3種類の識別器を紹介した。最後に、HOG 特徴量と LBP 特徴量の有効性を調査するための2つの実験を実施し、結果を示した。

第4章

Deep Learning を用いた高精度な人物検出

4.1 まえがき

前章では、従来の人物検出手法として、特徴量と識別器の組み合わせによる手法を紹介した。従来の特徴量と識別器の組み合わせによる人物検出では、勾配や色などに基づく数値を特徴量とするため、画像の状況に合わせて、人が適切なフィルタを使用する必要がある。また、ここで得られる特徴量は局所領域から抽出された勾配情報や色情報に基づくものであり、表現可能な特徴に限られる。したがって、多クラス分類に適さない。さらに、2クラス分類においても、入力画像の変動が大きい場合などには識別率が上がらないという問題がある。

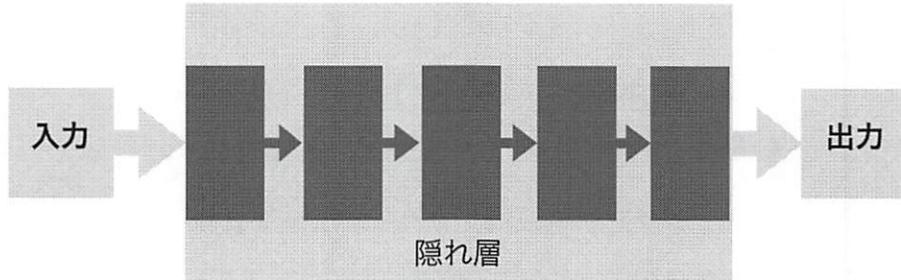
一方で、近年 Deep Learning という技術が人工知能、機械学習の分野で大変注目を集めている[15]。Deep Learning とは、多層ニューラルネットワークの1つであり、人間の神経回路の様な深い構造を作り出し、最適化された特徴量そのものを学習により獲得することで、精度を大幅に向上させる手法である。従来の特徴量は1つの情報から抽出された特徴の連結であり、Deep Learning は複数の情報から抽出された特徴の組み合わせである。従って、Deep Learning では高次元の特徴を表現でき、かつ特徴量自身も学習により獲得するため、フィルタ設計も不要である。

本章では、Deep Learning を用いた高精度な人物検出について述べる。まず、Deep Learning の考え方を述べる。次に、多層パーセプトロンを紹介する。そして、畳み込みニューラルネットワークの基礎について述べる。最後に、提案手法である高精度な人物検出のための多層畳み込みニューラルネットワークの構造や演算などについて述べる。本章で述べる内容に基づき第5章において、多層畳み込みニューラルネットワークの評価実験を行う。

4.2 Deep Learning の考え方

本節では、Deep Learning の考え方を述べる。例えば、Fig. 4.1 の様なシステムがある。入力層と出力層の間に、 n 層の隠れ層がある。もし出力層と入力層の情報が同じであれば、これは隠れ層での情報伝達に損失がないことを意味する。即ち、隠れ層の任意の一層は入

Fig. 4.1. 情報伝達システム

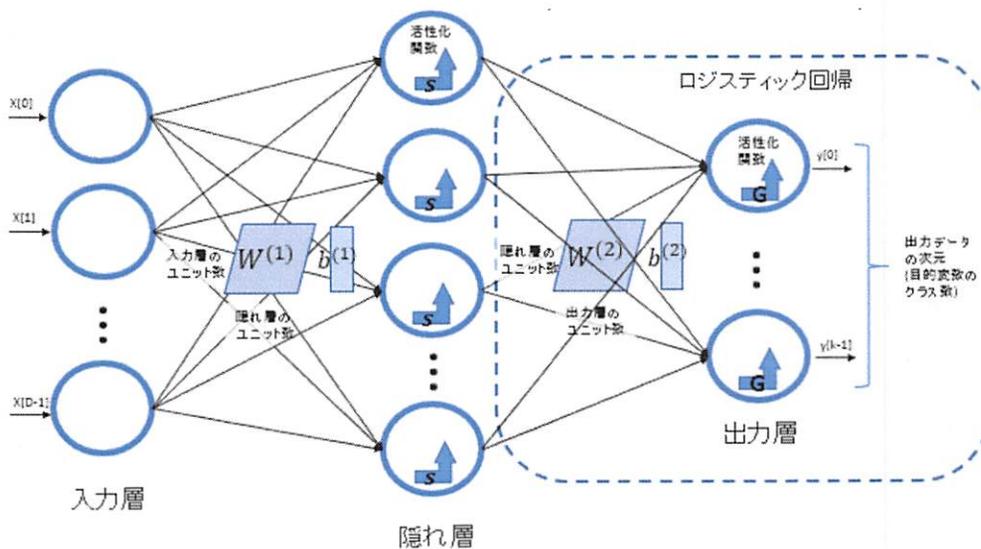


力層の別の表現である。しかし、情報理論によると、情報がレイヤーごとに損失する。一方、もし損失した情報が無駄な情報であれば、または有用な情報の損失が十分小さければ、出力層と入力層の意味は変わらない。Deep Learning は多層隠れ層による前のレイヤーの出力を次のレイヤーの入力とし、情報の段階的表現を実現する方法である。

4.3 多層パーセプトロン

本節では、多層パーセプトロンを述べる。多層パーセプトロンとは、複数のレイヤーにニューロンを配置させたフィードフォワード型ニューラルネットワークである。多層パーセプトロンは、主に Fig. 4.2 に示される入力層 / 隠れ層 / 出力層 の3層で構成される。

Fig. 4.2. 多層パーセプトロンの構造



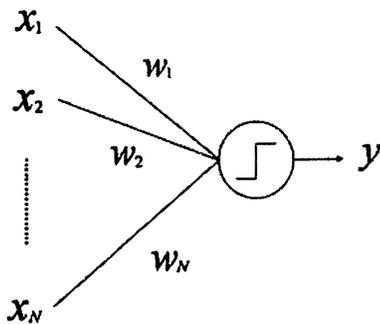
ニューロン

実際の生物のニューロンは、自分に繋がっているニューロンから受け取った電気信号の総和がある閾値を超えると、他のニューロンへ電気信号を発する。これを模したニューロンのモデルを Fig. 4.3 に示す。ここで、注目されるニューロンは、自身に繋がっているN個のニューロンから信号を受け取る。この信号を、ニューロンへの入力と呼ぶ。また、シナプスの伝達効率に相当する重みを結合重みと呼ぶ。

生物のニューロンの電気信号を発する性質は式 (4.1) で表現される。この関数 $f(s)$ は不連続であるため、微分できない。即ち誤差逆伝播法により重みの更新ができない。従って、実際にはニューロンの関数にはシグモイド関数が用いられることが一般である。

$$\begin{cases} f(s) = 1 & (s \geq \theta) \\ f(s) = 0 & (s < \theta) \end{cases} \quad s = \sum_{i=1}^N x_i w_i, \theta = \text{threshold} \quad (4.1)$$

Fig. 4.3. ニューロンのモデル



バックプロパゲーション

バックプロパゲーション法は、1986年に Rumelhart, Hinton, Williams によって発表された学習方法である。バックプロパゲーション法は、誤差関数をパラメータで微分することによってパラメータの修正方向を決めて、パラメータの修正を繰り返すという勾配法を多層パーセプトロン用に拡張したものである。誤差を出力層から前の層へ伝搬していくことから“Back Propagation”と名づけられている[16]。

4.4 畳み込みニューラルネットワーク

本節では、2次元の物体認識においてよく用いられる Deep Learning 及び畳み込みニューラルネットワークの基礎を述べる。

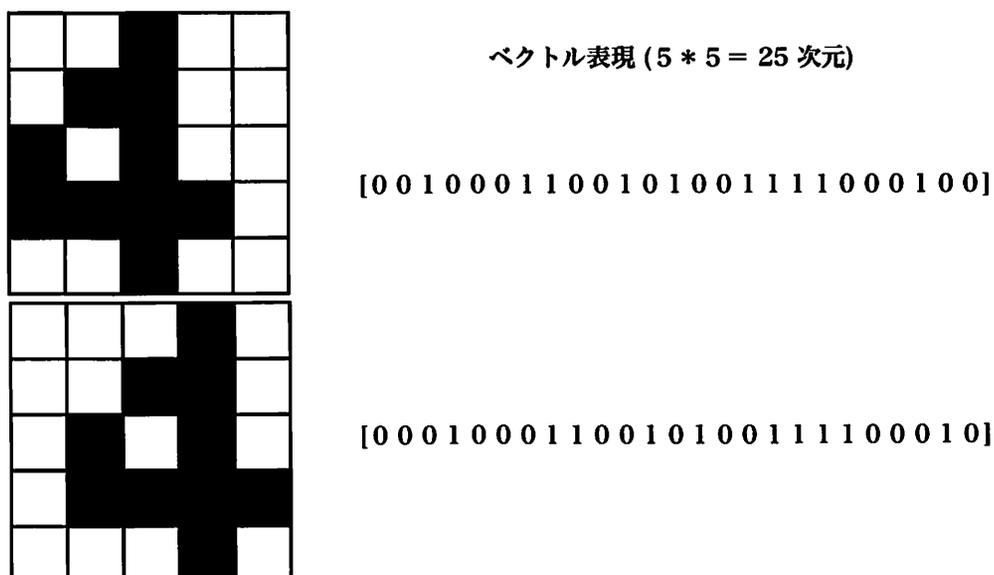
4.4.1 多層パーセプトロンと畳み込みニューラルネットワーク

従来の多層パーセプトロンでは、2次元の画像を1次元のベクトルとして扱う。例えば、数字認識の MNIST データセットでは、画像のサイズが 28×28 pixel なので、1レコードは $28 \times 28 = 784$ 次元のベクトルとして処理される。そのため、Fig. 4.4 のように入力画像の形状が同一であっても、その位置がずれていると全く別のデータに見えてしまう。結果、このような入力に対しては誤判別、或いは過学習が起きることがある。

畳み込みニューラルネットワークでは、入力を2次元のまま扱って特徴抽出を行い、最終的に1次元化して多層パーセプトロンに渡す。つまり、多層パーセプトロンに以下のような前処理を加えたものとなる。これは人間の視覚野の動きを模倣したものと考えることができる。

- 1) 入力をいくつかの領域に分割する (受容野に対応)
 - ・単純な受容野では、分割した領域から局所特徴を抽出する (e.g. エッジ)
 - ・複雑な受容野では、物体の位置の移動などを吸収する
- 2) これらを統合して入力の特徴をとらえ、多層パーセプトロンで判別する

Fig. 4.4. 平行移動した数字のベクトル表現



4.4.2 畳み込みニューラルネットワークの特性

画像への畳み込み

画像への畳み込みは、適当な大きさの領域に含まれる各値を重み付けして足し合わせることになる。重み行列はフィルタの役割となる。畳み込みにより、画像のエッジ検出や平滑化などの効果を得ることができる[17]。

今回の畳み込みニューラルネットワークでは、重み行列の値を直接指定することはない。学習画像に対して、全体のエラー率が最小となるように重み行列の値が更新される。

疎な結合

多層パーセプトロンでは、Fig. 4.4 のような平行移動した画像が入力されると、そのずれは誤差として学習されてしまう。これは、入力層 / 隠れ層 / 出力層それぞれで、お互いの全ユニットが接続されているために誤差が逆伝播することに起因する。畳み込みニューラルネットワークでは、このような誤差が逆伝播することを防ぐため、ニューラルネットワークを疎結合の形で構成する。実際には、後述するマックスプーリング層が疎結合に対応する。入力される情報の一部を捨てることで、誤差となる情報を除外する。

重み共有

$m-1$ 層の特徴マップから m 層の特徴マップへの変換は、畳み込み処理に対応した重み行列によって行われる。この重み行列は、入力の各領域ごとに変更せず、1つの重み行列を入力全体に対して使うようにする。重み独立の多層パーセプトロンと比べると、重みの数を大幅に削減し、高速な学習が可能となる。

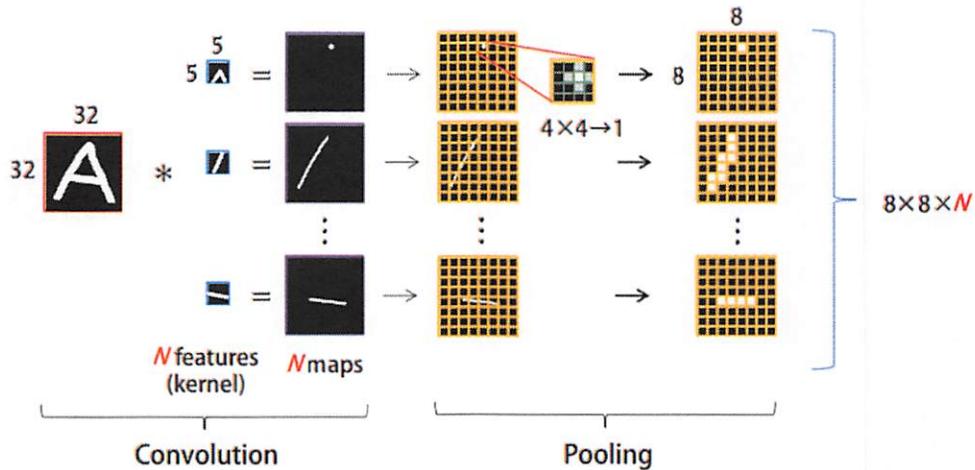
特徴マップ

入力から抽出できる特徴は、1通りとは限らない。入力から抽出される複数の特徴を特徴マップと呼ぶ。1つの特徴マップは、畳み込み+後処理 (e.g. 正規化, 整流 etc.) で生成される。

4.4.3 畳み込みとマックスプーリングの実行例

畳み込みとマックスプーリング処理で特徴マップを生成する例を Fig. 4.5 に示す。

Fig. 4.5. 畳み込みとマックスプーリングの実行例



4.5 高精度な人物検出のための畳み込みニューラルネットワーク

本節では、高精度な人物検出のための畳み込みニューラルネットワークを述べる。まず、構成できるレイヤーの種類を紹介する。次に、前方伝搬処理と後方伝搬処理を述べる。最後に、提案手法の畳み込みニューラルネットワークの構造を示す。

4.5.1 Supported Layer Types

一つの畳み込みニューラルネットワークは多段接続された複数のレイヤーを通じて処理される。ここで、サポートされるレイヤーは全部で7種類ある。

入力層

入力層は畳み込みニューラルネットワークの最初のレイヤーである。ニューロンの出力は2次元画像の画素値である。

畳み込み層

畳み込み層は全構造の唯一の重み行列で生成されるレイヤーである。畳み込み処理は大脳皮質一次視覚野の単純型細胞に相当する情報処理を行い、特定方向の線分に反応する方位選択性を持つ。

整流層

整流層は特徴マップを生成する前に、整流という後処理により生成される典型的な1つのオプショナルレイヤーである。整流は、畳み込み層のアウトプットに対し、負数を除去する関数である。

マックスプーリング層

マックスプーリング層は疎結合に対応するレイヤーである。入力の一部を捨てることにより、特徴マップの次元数を減らすだけでなく、誤差の発生を抑制する。マックスプーリングとは、局所領域内におけるニューロンの反応の最大値を抽出し、要約する操作である。

正規化層

正規化層は特徴マップを生成する前に、正規化という後処理により生成される典型的なもう1つのオプショナルレイヤーである。正規化とは、入力データを一定の基準に従って変形し、汎化させる処理である。正規化処理は式 (4.2) に従って行う。

$$b_{x,y}^i = a_{x,y}^i \div \left(k + \alpha \times \sum_{j=\max(0, i-n+2)}^{\min(N-1, i+n+2)} (a_{x,y}^j)^2 \right)^\beta \quad (4.2)$$

全結合層

全結合層は 4.3 節で紹介した多層パーセプトロンの隠れ層と同様である。多層パーセプトロンでは、隠れ層のニューロンが入力層と出力層の全てのニューロンと連結している。

出力層

目的に応じた結果を示すレイヤーである。例えば、物体認識の場合、出力層のニューロンが対応のクラスを代表する。また、物体検出の場合、出力層のニューロンが物体の位置を意味する。

4.5.2 Forward Propagation

入力層から出力層までのプロセスは“Forward Propagation”と呼ばれている。以下に提案手法の“Forward Propagation”の演算について述べる[17,18,19,20]。

任意層 → 畳み込み層 / 全結合層

畳み込み層 / 全結合層を生成するための演算は式 (4.3) に従って行う。畳み込み層 / 全結合層のニューロンの出力は、このニューロンと連結している前のレイヤーの出力を重み付けして足し合わせ、さらにシグモイド関数を用いた結果である。本研究でのシグモイド関数は双曲線関数を用いた。

$$y_n^i = \sum_{c=0}^{C_{n-1}} x_{n-1}^{ic} w_n^c, \quad x_n^i = f(y_n^i) \quad (4.3)$$

畳み込み層 → 整流層

整流層を生成するための演算は式 (4.4) に従って行う。整流層のニューロンの出力は、前の畳み込み層の出力を処理した結果となる。

$$x_n^i = \max(0, x_{n-1}^i) \quad (4.4)$$

任意層 → 正規化層

正規化層を生成するための演算は式 (4.5) に従って行う。正規化層のニューロンの出力は、このニューロンと連結している前のレイヤーの局所領域のニューロンの出力を処理した結果となる。

$$x_n^i = x_{n-1}^i \div \left(1 + 0.2 \times \sum_{c=0}^{C_{n-1}} (x_{n-1}^{ic})^2 \right)^{0.75} \quad (4.5)$$

任意層 → マックスプーリング層

マックスプーリング層を生成するための演算は式 (4.6) に従って行う。マックスプーリング層のニューロンの出力は、このニューロンと連結している前のレイヤーの局所領域のニューロンの出力の最大値である。

$$x_n^i = \max(x_{n-1}^{i0}, x_{n-1}^{i1}, x_{n-1}^{i2}, \dots, x_{n-1}^{iC_{n-1}}) \quad (4.6)$$

4.5.3 Back Propagation

出力層から入力層までのプロセスは“Back Propagation”と呼ばれる。“Back Propagation”のプロセスは出力層の誤差を無くすため、この誤差を前の各レイヤーに伝搬し、各レイヤーで修正できるよう、畳み込み層の結合重みを修正する過程である。以下に提案手法の“Back Propagation”の演算について述べる。演算は出力層での誤差を式 (4.7) により求めることから開始する[17,18,19,20]。

$$E_n^p = 0.5 \times \sum_{i=0}^{I-1} (x_n^i - T_n^i)^2 \quad (4.7)$$

畳み込み層 / 全結合層 → 任意層

仮に畳み込み層 / 全結合層での誤差を E_n とする。畳み込み層 / 全結合層での“Back Propagation”は、この誤差を使用し、以下のように演算する。式 (4.8) は畳み込み層 / 全結合層のニューロンの出力の誤差を入力層の誤差に転換することを意味する。ここで、 $f'(x_n^i)$ はシグモイド関数の導関数である。式 (4.9) は畳み込み層 / 全結合層のニューロンの入力層の誤差を前のレイヤーのニューロンの出力の誤差に転換することを意味する。式 (4.10) は結合重みの更新である。

$$\frac{dE_n}{dy_n^i} = f'(x_n^i) \cdot \frac{dE_n}{dx_n^i} \quad (4.8)$$

$$\frac{dE_{n-1}}{dx_{n-1}^k} = \sum_i w_n^{ik} \cdot \frac{dE_n}{dy_n^i} \quad (4.9)$$

$$\frac{dE_n}{dw_n^{ij}} = x_{n-1}^j \cdot \frac{dE_n}{dy_n^i}, \quad (w_n^{ij})_{new} = (w_n^{ij})_{old} - \eta \cdot \frac{dE_n}{dw_n^{ij}} \quad (4.10)$$

整流層 → 畳み込み層

Deep Learning において、整流層を生成する時、結合重みがないため、整流層での“Back Propagation”は誤差を畳み込み層の誤差に転換するだけである。演算は式 (4.11) に従って行う。

$$\frac{dE_{n-1}}{dx_{n-1}^i} = x_{n-1}^i > 0 ? \frac{dE_n}{dx_n^i} : 0 \quad (4.11)$$

正規化層 → 任意層

整流層と同様に、正規化層での“Back Propagation”も誤差を前のレイヤーの誤差に転換するだけである。演算は式 (4.12) に従って行う。ここで、 $g'(x_{n-1}^i)$ は正規化関数の導関数である。

$$\frac{dE_{n-1}}{dx_{n-1}^i} = \frac{dE_n}{dx_n^i} \cdot g'(x_{n-1}^i) \quad (4.12)$$

マックスプーリング層 → 任意層

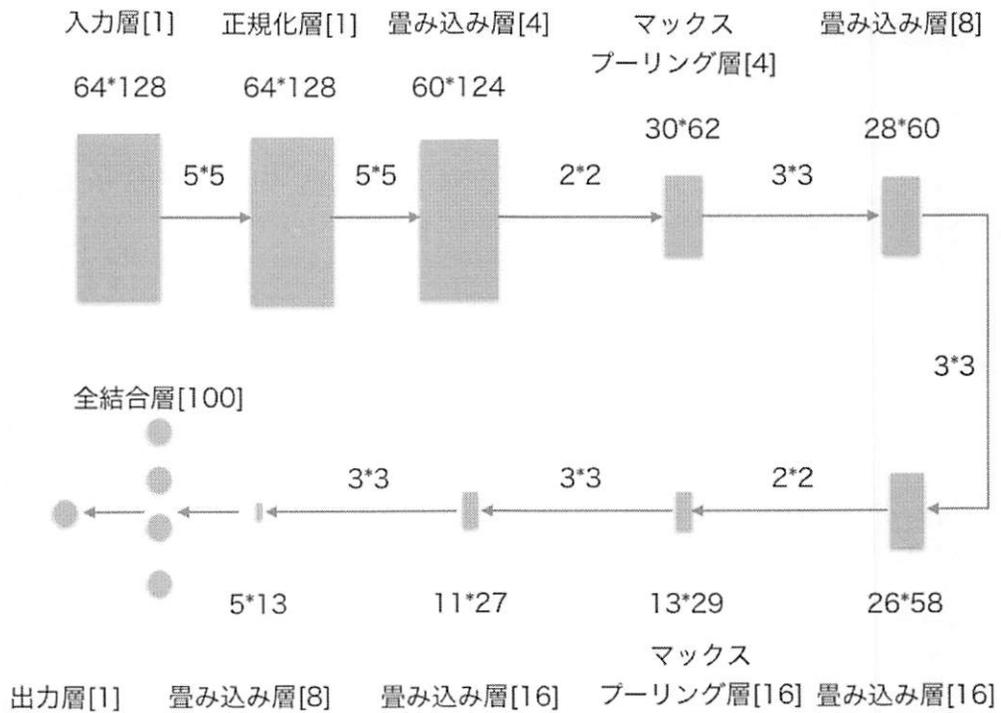
マックスプーリング層での“Back Propagation”は式 (4.13) に従って行う。

$$\frac{dE_{n-1}}{d\max(x_{n-1}^{i0}, x_{n-1}^{i1}, x_{n-1}^{i2}, \dots, x_{n-1}^{iC_{n-1}})} = \frac{dE_n}{dx_n^i} \quad (4.13)$$

4.5.4 提案手法の畳み込みニューラルネットワークの構造

処理コストを考慮した高精度な人物検出のための畳み込みニューラルネットワークの構造を Fig. 4.6 に示す。後述の識別率調査の実験5の構造に対応する。

Fig. 4.6. 提案手法の畳み込みニューラルネットワークの構造



4.6 むすび

本章では、Deep Learning を用いた高精度な人物検出について述べた。まず、Deep Learning の考え方を述べた。次に、多層パーセプトロンを紹介した。そして、畳み込みニューラルネットワークの基礎について述べた。最後に、提案手法の高精度な人物検出のための多層畳み込みニューラルネットワークの構造や演算などについて述べた。

第5章では、本章の内容に基づき、多層畳み込みニューラルネットワークの評価実験を行う。

第5章

提案手法の評価実験

5.1 まえがき

本章では、前章で提案した畳み込みニューラルネットワークを用いて、評価実験を行う。まず、提案手法の識別率調査を行う。次に、提案手法の回転安定性調査を行う。最後に、列車運行映像における乗客検出を実行する。

5.2 識別率調査

提案手法の有効性を明らかにするために識別率調査の比較実験を行う。以下に示す5種類の組み合わせについて畳み込みニューラルネットワークを構築し、精度を調査する。

- 1) 畳み込み層だけで構成されるニューラルネットワーク
- 2) 畳み込み層と整流層で構成されるニューラルネットワーク
- 3) 畳み込み層とマックスプーリング層で構成されるニューラルネットワーク
- 4) 畳み込み層と整流層とマックスプーリング層で構成されるニューラルネットワーク
- 5) 正規化層と畳み込み層とマックスプーリング層で構成されるニューラルネットワーク

5.2.1 実験概要

本実験では、第3章の比較実験と同様な学習画像と評価サンプルを使用する。同じく人物認識を対象とする。データセットには INRIA Person Dataset を使用する。学習画像は3500枚、評価サンプルに1000枚を使用する。

5.2.2 実験条件

実験1～5に対応するパラメータを Table 5.1 ～ 5.5 に示す。

Table 5.1. 実験1に対応するパラメータ

	横*縦*枚数	フィルター	ストライド
入力層	64*128*1	—	—
畳み込み層1	60*124*4	5*5	1*1
畳み込み層2	29*61*8	4*4	2*2
畳み込み層3	27*59*16	3*3	1*1
畳み込み層4	13*29*16	3*3	2*2
畳み込み層5	11*27*16	3*3	1*1
畳み込み層6	5*13*8	3*3	2*2
全結合層	1*1*100	—	—
出力層	1*1*1	—	—

Table 5.2. 実験2に対応するパラメータ

	横*縦*枚数	フィルター	ストライド
入力層	64*128*1	—	—
畳み込み層1	60*124*4	5*5	1*1
整流層1	60*124*4	—	—
畳み込み層2	29*61*8	4*4	2*2
整流層2	29*61*8	—	—
畳み込み層3	27*59*16	3*3	1*1
整流層3	27*59*16	—	—
畳み込み層4	13*29*16	3*3	2*2
整流層4	13*29*16	—	—
畳み込み層5	11*27*16	3*3	1*1
整流層5	11*27*16	—	—
畳み込み層6	5*13*8	3*3	2*2
全結合層	1*1*100	—	—
出力層	1*1*1	—	—

Table 5.3. 実験3に対応するパラメータ

	横*縦*枚数	フィルター	ストライド
入力層	64*128*1	—	—
畳み込み層1	60*124*4	5*5	1*1
マックスプーリング層1	30*62*4	—	2*2
畳み込み層2	28*60*8	3*3	1*1
畳み込み層3	26*58*16	3*3	1*1
マックスプーリング層2	13*29*16	—	2*2
畳み込み層4	11*27*16	3*3	1*1
畳み込み層5	5*13*8	3*3	1*1
全結合層	1*1*100	—	—
出力層	1*1*1	—	—

Table 5.4. 実験4に対応するパラメータ

	横*縦*枚数	フィルター	ストライド
入力層	64*128*1	—	—
畳み込み層1	60*124*4	5*5	1*1
整流層1	60*124*4	—	—
マックスプーリング層1	30*62*4	—	2*2
畳み込み層2	28*60*8	3*3	1*1
整流層2	28*60*8	—	—
畳み込み層3	26*58*16	3*3	1*1
整流層3	26*58*16	—	—
マックスプーリング層2	13*29*16	—	2*2
畳み込み層4	11*27*16	3*3	1*1
整流層4	11*27*16	—	—
畳み込み層5	5*13*8	3*3	1*1
全結合層	1*1*100	—	—
出力層	1*1*1	—	—

Table 5.5. 実験5に対応するパラメータ

	横*縦*枚数	フィルター	ストライド
入力層	64*128*1	—	—
正規化層	64*128*1	—	1*1
畳み込み層1	60*124*4	5*5	1*1
マックスプーリング層1	30*62*4	—	2*2
畳み込み層2	28*60*8	3*3	1*1
畳み込み層3	26*58*16	3*3	1*1
マックスプーリング層2	13*29*16	—	2*2
畳み込み層4	11*27*16	3*3	1*1
畳み込み層5	5*13*8	3*3	1*1
全結合層	1*1*100	—	—
出力層	1*1*1	—	—

5.2.3 実験結果

各実験の結果を Table 5.6 に示す。今回の実験から、全て5つの畳み込みニューラルネットワークの識別率が第3章で行った従来手法の比較実験の結果より優れている。また、提案手法が5つの畳み込みニューラルネットワークの内一番良い結果となっている。

Table 5.6. 識別結果

	True Positive	True Negative	Average
実験1	0.952	0.972	0.966
実験2	0.949	0.976	0.968
実験3	0.962	0.986	0.979
実験4	0.969	0.984	0.980
実験5 (提案手法)	0.969	0.99	0.984

5.3 回転安定性調査

提案手法の回転安定性を調査するため、従来手法との比較実験を行う。以下に示す3つの手法を用いて、回転画像に対する人物認識を行う。

- 1) HOG + Real AdaBoost
- 2) LBP + Real AdaBoost
- 3) 提案手法

5.3.1 実験概要

本実験では、前節の比較実験と同様な学習画像と評価サンプルを使用する。評価サンプルを -20 度から 20 度まで、 5 度ずつ回転し、1つの手法に対して9回の識別を行う。

5.3.2 実験条件

実験1と実験2の条件は第3章の比較実験の実験条件と同様、実験3の条件は前節の実験5の条件と同様である。

5.3.2 実験結果

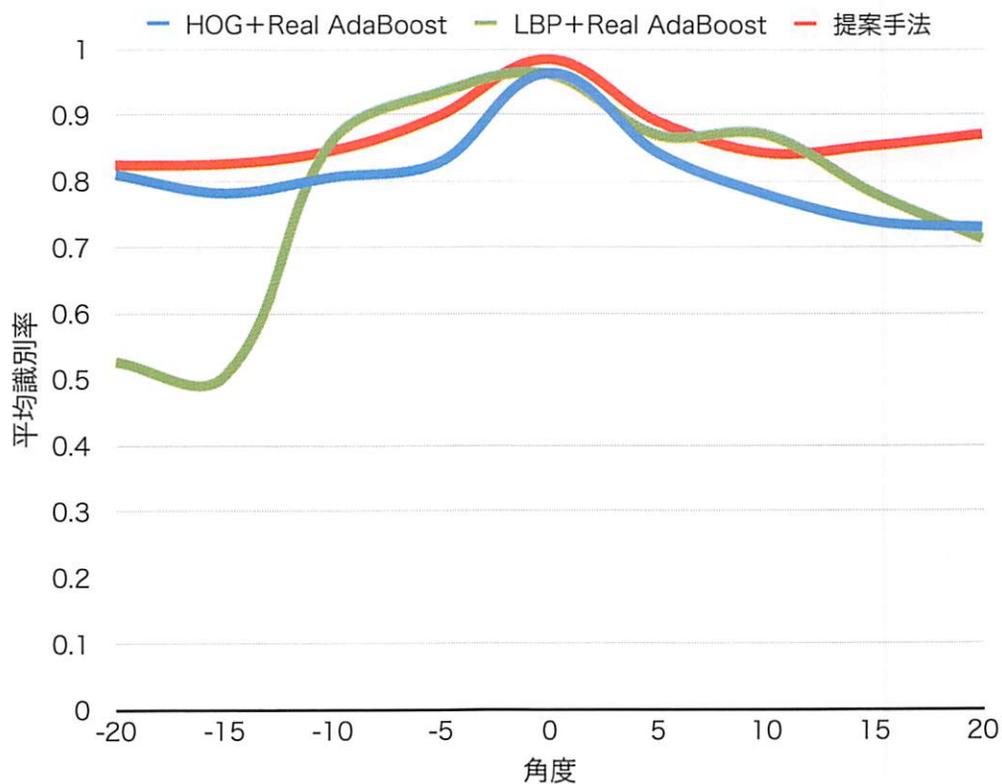
3つの実験結果を Fig. 5.1 に統合して示す。各曲線は対応手法の回転角度に対する平均識別率変化を表す。9回の識別に対し、提案手法が6回の最高識別率を得ている。また、提案手法の線型が3本の内一番緩やかであり、然も対称性を持っている。この実験から、提案手法は従来手法より安定していることが確認された。

5.4 乗客検出の実行

提案手法を用いて、実際の列車運行映像における乗客検出を行う。以下に示す3つの機能を持つシステムを構築し、乗客検出を実行する。

- 1) 検出領域の特定
- 2) 検出領域での乗客検出
- 3) 誤検出の再学習

Fig. 5.1. 回転安定性調査の結果



5.4.1 学習画像の生成

学習画像の生成は手作業による切り出しと自動生成の2方式で行う。まず、幾つの列車運行映像から乗客が存在するフレームを保存する。次に、手作業で乗客を切り出し、ポジティブ画像のデータセットを生成する。最後に、切り出した乗客領域を除いて、残る領域からネガティブ画像のデータセットを自動生成する。結果、ポジティブ画像は1350枚、ネガティブ画像は5740枚を生成した。

5.4.2 システム概要

本システムは、高精度な乗客検出を目的としている。ここでは検出速度については取り扱わないが、現状のシステム上ではリアルタイム検出は不可能である。検出処理はキーボードでスペースキーを押してから始まる。本システムでは、特定した領域から乗客検出を行う。また、誤検出に対しては、これを再学習することができる。

5.4.3 実行結果

システムの実行結果を Fig. 5.2 と Fig. 5.3 に示す. Fig. 5.2 は1回目の検出結果である. 5つの検出領域に対し, 1つの誤検出(右下の赤枠)が発見された. 再学習したい領域を青枠で表す. 右下の赤枠領域を青枠領域に変更し, 誤検出を再学習してからの乗客検出結果は Fig. 5.3 である. 同じフレームに対し, 誤検出が無くなった. 100 フレームに対する平均正解率が 95% である.

Fig. 5.2. 1回目の検出結果

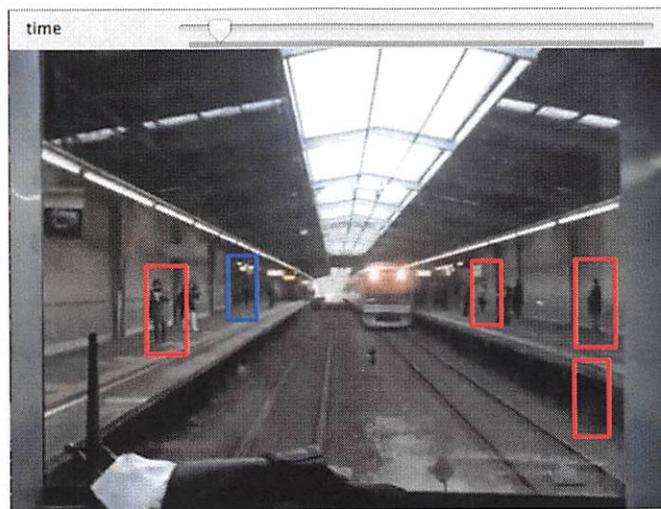


Fig. 5.3. 誤検出を再学習してからの検出結果



5.5 むすび

本章では、前章で提案した畳み込みニューラルネットワークを用いて、評価実験を行った。まず、提案手法の識別率調査を行った。次に、提案手法の回転安定性調査を行った。最後に、列車運行映像における乗客検出を実行した。

識別率調査から、全て5つの畳み込みニューラルネットワークの識別率が第3章で行った従来手法の比較実験の結果より優れている。また、提案手法が5つの畳み込みニューラルネットワークの内一番良い結果となっている。また、回転安定性調査から、提案手法は従来手法より安定していることが確認された。更に、提案手法を用いた列車運行映像における乗客検出を実行し、平均正解率が95%という結果を得た。

第6章

結論

6.1 総括

第1章では、本研究の背景と目的、及び本論文の構成について述べた。

現在、列車運行の安全を確保するための技術が益々必要となっている。乗客検出はその内の1つである。目視と同様の精度をもつ検出機能を画像処理で実現することが人工知能の研究として長年に渡り検討されてきたが、その達成は未だ困難である。また、問題発生予防、及び予測機能を達成するため、高精度であるだけでなく、処理速度が重要となる。従来提案されてきた様々な画像特徴量と学習手法の組み合わせを用いた検出手法では、高精度かつ高速な処理を実現することが困難である。そこで、近年ではディープラーニングというアプローチが提案され、人工知能、機械学習の分野で大変注目を集めている。従来の特徴量を任意に設定する手法と比較して、精度の向上が可能である。しかし、この手法を用いた列車運行映像における乗客検出の研究はほとんどされていないのが現状である。ディープラーニングを用いた高精度な乗客検出に関して社会的な要求があることについて述べた。

第2章では、列車運行の安全の現状として、列車の運転中における事故の発生状況と、鉄道運転事故に対する相応な安全対策の実施状況とこれらの安全対策の問題点について述べた。

都営地下鉄、JR東日本、西武鉄道、三つの企業から鉄道運転事故の発生状況を調査した結果、鉄道人身障害事故と踏切障害事故は毎年数多く発生していることが分かった。また、事故に対して、企業側から細かく対策を施してきたが、殆どは事故発生時の対応方法であるため、事故率を収めることが困難である。知的な画像解析技術による事故抑方法の重要性があることを示した。

第3章では、画像処理による人物検出の関連技術について述べた。まず、移動中の人物を検出する観点から固定カメラによる高速な人物検出方法を紹介した。次に、人物検出に利用される HOG 特徴量と LBP 特徴量を紹介した。そして、人物検出に利用される代表的な機械学習の中のサポートベクターマシン、AdaBoost、及び Real AdaBoost、3種類の識別器を紹介した。最後に、HOG 特徴量と LBP 特徴量の有効性を調査するための2つの実験を実施し、結果を示した。

第4章では、列車運行映像における高精度な乗客検出を実現するために、Deep Learning を用いた高精度な人物検出手法を提案した。

はじめに、提案手法の Deep Learning の考え方を述べた。次に、多層パーセプトロンを紹介した。そして、畳み込みニューラルネットワークの基礎について述べた。最後に、提案手法の高精度な乗客検出のための多層畳み込みニューラルネットワークの構造や演算などについて述べた。

第5章では、第4章で提案した畳み込みニューラルネットワークを用いて、識別率と回転安定性の調査を行った。また、列車運行映像における乗客検出を実行した。

識別率調査から、全て5つの畳み込みニューラルネットワークの識別率が第3章で行った従来手法の比較実験の結果より優れている。また、提案手法が5つの畳み込みニューラルネットワークの内一番良い結果となっている。また、回転安定性調査から、提案手法は従来手法より安定していることが確認された。更に、提案手法を用いた列車運行映像における乗客検出を実行し、平均正解率が95%という結果を得た。

第6章は結論であり、本論文の総括および今後の課題について述べている。

6.2 今後の課題

今後の課題として以下がある。

乗客検出率向上の検討

本研究では、ディープラーニングを用いた列車運行映像における高精度な乗客検出の実現を目的とした。第5章で実行した乗客検出システムの平均正解率が95%であるが、検出率がそれほど高くない。運行映像の低解像度と乗客の集まりが低検出率の原因だと考えられる。そこで、乗客検出率を向上するための検討が必要である。

リアルタイム検出の検討

本研究では、検出速度については取り扱わないが、現状のシステム上ではリアルタイム検出は不可能である。提案した畳み込みニューラルネットワークの構造は消費時間を配慮して設計した。また、計算処理も並列化した。しかし、リアルタイムの検出をできるように、新たな検討が必要である。

参考文献

- [1] 山内悠嗣 他, “画像からの統計的手法に基づく人検出,” 信学論, vol.J96-D, no.9, pp.2017-2040, 2013.
- [2] “鉄道事故の種類,” <http://hunken.rtri.or.jp/PDF/cdroms1/0004/2005/20000405070401.pdf>, 2015.
- [3] “安全報告書,” http://www.kotsu.metro.tokyo.jp/information/safety/pdf/railwaybusiness_20.pdf, 2015.
- [4] “安全報告書,” https://www.jreast.co.jp/safe/pdf/report2014/report2014_all.pdf, 2015
- [5] “安全報告書,” <http://www.seibu-group.co.jp/railways/company/hokoku/2014/index.html>, 2015
- [6] “背景差分による移動物体の検出に関する基礎的な検討,” <http://www.cc.kyoto-su.ac.jp/~kano/pdf/study/student/OkabePresen.pdf>, 2015
- [7] Q. Zhu, et al., “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients,” CVPR, pp.1491-1498, 2006.
- [8] 後藤雄飛 他, “色の類似性に基づいた形状特徴量 CS-HOG の提案,” SSII, 2012.
- [9] N. Dalal, et al., “Histograms of Oriented Gradients for Human Detection,” CVPR, vol.1, pp.886-893, 2005.
- [10] X. Wang, et al., “An HOG-LBP Human Detector with Partial Occlusion Handling,” ICCV, 2009.
- [11] Y.D. Mu, et al., “Discriminative local binary patterns for human detection in personal album,” CVPR, pp.1-8, 2008.
- [12] “サポートベクターマシン入門,” <http://home.hiroshima-u.ac.jp/tkurita/lecture/svm.pdf>, 2015.
- [13] 山内悠嗣 他, “Boosting に基づく特徴量の共起表現による人検出,” 信学論, vol.J92-D, no.8, pp.1125-1134, 2009.
- [14] P. Viola, et al., “Rapid Object Detection Using a Boosted Cascade of Simple Features,” CVPR, pp. 511-518, 2001.
- [15] “Deep Learning に関する記事,” <http://wired.jp/list/tags/deep-learning/>, 2015.
- [16] “多層パーセプトロンによる関数近似,” <http://aidiary.hatenablog.com/entry/20140122/1390395760>, 2015.
- [17] A. Krizhevsky.: “Convolutional Deep Belief Networks on CIFAR-10,” <http://www.cs.utoronto.ca/~kriz/conv-cifar10-aug2010.pdf>.
- [18] D. Ciresan, U. Meier, and J. Schmidhuber.: “Multi-column Deep Neural Networks for Image Classification,” <http://people.idsia.ch/~juergen/cvpr2012.pdf>.
- [19] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, L.D. Jackel, et al.: “Handwritten Digit Recognition with a Back-Propagation Network,” <http://yann.lecun.com/exdb/publis/pdf/lecun-90c.pdf>.
- [20] Y. LeCun, K. Kavukcuoglu, and C. Fawcett.: “Convolutional Networks and Applications in Vision,” In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253-256, May 30 2010 - June 2 2010.