

平成 29 年度卒業論文

映像からの簡便な人物行動検出と

閾値の自動決定法

宮崎大学 工学部 情報システム工学科

松本 亜沙美

指導教員 椋木雅之

## 目次

1.	はじめに .....	1
2.	人物の行動検出.....	2
2.1	行動検出処理の流れ.....	2
2.2	対象行動 .....	2
2.3	人物検出 .....	4
2.3.1	HOG .....	4
2.3.2	SSD.....	5
2.3.3	比較実験.....	5
2.4	対象行動の検出.....	9
2.4.1	しゃがむ.....	9
2.4.2	転倒.....	12
2.4.3	停止.....	14
3.	閾値の自動設定.....	15
4.	実験 .....	18
5.	おわりに .....	22

## 1. はじめに

近年社会の高齢化が進む中、老人ホームなどで高齢者を支える存在である介護福祉士の数が不足しており、それによって介護福祉士 1 人 1 人の負担が大きくなっている。老人ホーム等施設内に設置したカメラの映像から入居者の転倒などの行動を検出し、介護福祉士に知らせることで定期的な廊下の見回りなどの負担を軽減できると考えられる。この見守りシステムを開発するための実験段階として、廊下を奥行き方向に向けてカメラで撮影した映像から、人物の行動検出を行う手法の開発を目標とする。

見守りシステムについての研究はこれまでも行われている。[1]や[2]は赤外線センサーやベッドセンサーのような特殊なセンサーを使用して人物の行動を観測している。本研究では廊下を実験場所とするため、より広範囲をカバーできるカメラを採用する。[3]では機械学習を用いて異常行動を検出している。機械学習は実際に検出を行うために大量のデータが必要であり簡便な手法とは言えない。そこで、より簡便な手法を用いて人物の行動検出を行うため、1 台のカメラで撮影した映像のみを用いて行動検出を行う手法を提案する。本研究では、自動検出した人物が一定フレーム間に移動した距離と傾きに閾値を設定することで、動画内の人物の行動検出を行う。この提案手法では、カメラの設置位置や角度によって、行動検出を行うための閾値が変化する。それを解決するため、カメラの設置状況に合わせた閾値の自動設定方法の開発も同時に行う。

## 2. 人物の行動検出

### 2.1 行動検出処理の流れ

提案する行動検出処理では、まず動画をフレームごとに分割し、10 フレームごとに人物検出を行う。検出した人物領域の重心の座標を取得し、フレーム間での座標の距離と傾きを用いて行動検出を行う。

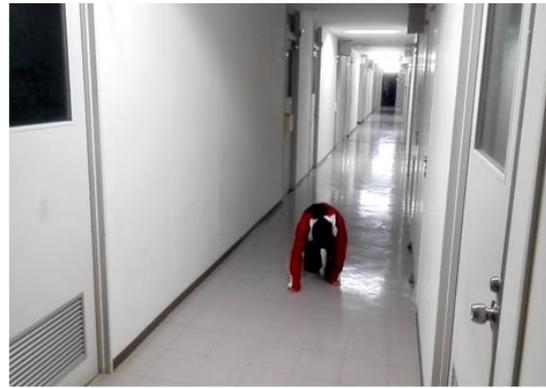
以下に行動検出の対象行動と各行動の判定基準について説明する。

### 2.2 対象行動

本研究において、主に検出したい対象行動は「転倒」(図 1 (b))である。これに加え、転倒に似た動作である「しゃがむ」(図 1 (a))と、動画の撮影場所である廊下でおこる頻度が高い「停止」(図 1 (c) (d))の3つを対象行動とした。本研究で使用する動画は、三脚を用いて廊下を奥行き方向に撮影する。また、人物は奥から手前に向けて歩行し、その途中で上記のいずれかの行動を行っている。



(a)しゃがむ



(b)転倒



(c)停止(立ち止まる)



(d)停止(座る)

図 1 : 検出対象行動

## 2.3 人物検出

動画内の人物を検出するための手法として本研究では HOG 特徴量を用いた人物検出[4] (図 2 (a)) と SSD を用いた人物検出[5] (図 2 (b)) の 2 手法で人物の検出精度についての比較実験を行った。



図 2 : 人物検出

### 2.3.1 HOG

HOG (Histograms of Oriented Gradients) とは N.Dalal 等[4]により 2005 年に発表された特徴量であり、局所領域 (セル) の輝度の勾配方向をヒストグラム化したものである。

HOG 特徴量は画像をグレイスケールで読み込み、各 pixel の輝度から勾配強度と勾配方向を求め、これをもとにセル領域 (5×5[pixel]など) ごとにヒストグラムを作成し、ブロック (3×3[セル]など) ごとに正規化することで抽出できる。本研究では、SVM (Support Vector Machine) を用いて HOG 特徴量を学習した人物検出器を使用して人物検出を行った。実装には OpenCV を用いた。

## 2.3.2 SSD

SSD (Single Shot MultiBox Detector) とは Liu.Weii 等[5]により 2016 年に発表された物体検出手法であり、畳み込みニューラルネットワークという深層学習を用いた検出手法である。この手法は、フィーチャーマップと呼ばれる特徴量をいくつかの階層で求め、 $8 \times 8$  や  $4 \times 4$  などの異なるスケールで評価することで物体の検出を行う。本研究では PASCAL VOC[6]というデータセットを学習させたモデルを用いて、画像内から人物らしさが一定値以上の箇所を取得することで人物検出を行った。

## 2.3.3 比較実験

実験には、人物が廊下を歩いている様子を撮影した動画を用いた。実験に用いたデータの詳細を表 1 に示す。動画は人数が 1~3 人の 3 パターン、カメラアングルが正面、斜め、俯瞰の 3 パターンの組み合わせで 9 パターン撮影した (図 3)。9 パターンの動画をフレームごとに分割し、HOG 特徴量を用いた手法と SSD を用いた手法でそれぞれ人物検出を行い、検出精度を比較した。検出精度は、

$$\frac{\text{検出成功した人数}}{\text{フレームに映っている人物の合計}}$$

として計算する。また、人物ではない場所を人物と検出した誤検出率

$$\frac{\text{誤検出ボックス数}}{\text{フレーム数}}$$

についても計算を行った。実験結果を表 2 と表 3 に示す。



図 3 : 実験使用データ

表 1 : 実験使用データ

人数	カメラアングル	時間	フレーム数	fps	サイズ
1 人	正面	6.9s	201	29	720 × 1280
	斜め	7.8s	228		
	俯瞰	7.7s	223		
2 人	正面	7.9s	228		
	斜め	7.5s	217		
	俯瞰	11.5s	334		
3 人	正面	9.1s	273		
	斜め	9.3s	271		
	俯瞰	10.6s	306		

表 2 : 実験結果 (HOG)

人数	カメラアングル	処理時間	検出精度	誤検出率	fps
1 人	正面	28s	56%	33%	7.24
	斜め	31s	66%	7%	
	俯瞰	31s	72%	34%	
2 人	正面	32s	58%	2%	
	斜め	30s	71%	34%	
	俯瞰	46s	42%	49%	
3 人	正面	38s	52%	26%	
	斜め	37s	36%	11%	
	俯瞰	42s	25%	27%	

表 3 : 実験結果 (SSD)

人数	カメラアングル	処理時間	検出精度	誤検出率	fps
1 人	正面	197s	91%	0%	0.99
	斜め	230s	100%	0%	
	俯瞰	242s	98%	0%	
2 人	正面	227s	97%	0%	
	斜め	217s	99%	1%	
	俯瞰	334s	90%	0%	
3 人	正面	273s	91%	3%	
	斜め	270s	90%	15%	
	俯瞰	307s	90%	23%	

実験の結果、HOG 特徴量を用いた場合の検出精度は平均 53%、SSD を用いた場合の検出精度は平均 94%であった。誤検出率については HOG の方が圧倒的に悪い結果であった。SSD に関しては人数が増えると誤検出率が増加する結果となった。HOG 特徴量を用いた時、歩行者が 3 人の場合に誤検出率が低いのは、そもそも人物としての検出が少なかった事が原因だと考えられる。処理速度に関しては、SSD の方が悪い結果であるが、本研究では行動検出のために人物の検出領域を用いるため、検出精度を優先する。よって、9 割以上の優れた人物検出精度であった SSD を本研究では用いることとする。

## 2.4 対象行動の検出

人物検出で切り出した検出領域（矩形）の情報を用いて、対象行動を検出する。以下、各対象行動の検出方法を述べる。

### 2.4.1 シャガむ

本研究では、人物が歩行状態からうずくまる様な行動を「シャガむ」と定義する（図4）。「シャガむ」という行動について、歩行状態からしゃがんだ場合、人物の検出領域の高さが歩行状態時よりも短くなることによって、検出領域の重心が大きく下方に変化するという特徴が挙げられる。そのため重心座標間の傾きと距離が大きくなることを利用して「シャガむ」という行動の検出を行う。



(a) 歩行時

(b) しゃがんだ状態

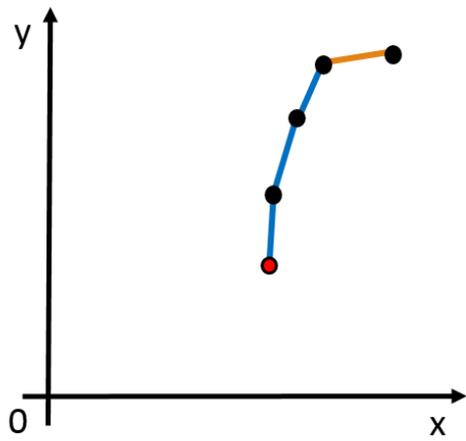
図4: 「シャガむ」時の検出領域の変化

具体的には、『重心座標間の距離と傾きが設定した閾値を連続して超えた場合』しゃがんでいると判定する。距離の閾値については、行動速度による距離の変化を考慮して短距離閾値 $[D_1]$ 、長距離閾値 $[D_2]$ の2つの閾値を設定する。傾きの閾値は1.5とする。また、カメラとの距離によって行動の大きさも離れた場所では小さく、近い場所では大きく変化する。それにあわせて重心座標間距離もカメラから離れた場所では変化するため判定基準を以下のように設定する。

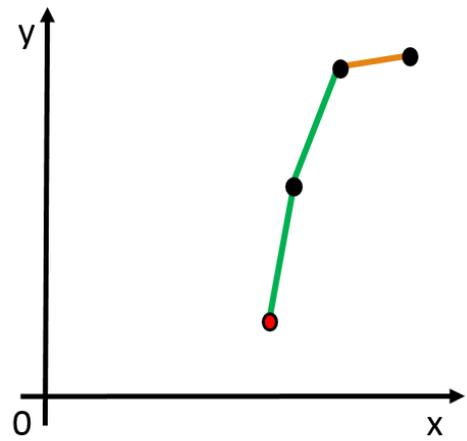
- ・3回連続で重心座標間距離と傾きが短距離閾値 $[D_1]$ と1.5をそれぞれ上回った場合「しゃがんだ」と判定する。(図5(a))

- ・2回連続で重心座標間距離と傾きが長距離閾値 $[D_2]$ と1.5をそれぞれ上回った場合「しゃがんだ」と判定する。(図5(b))

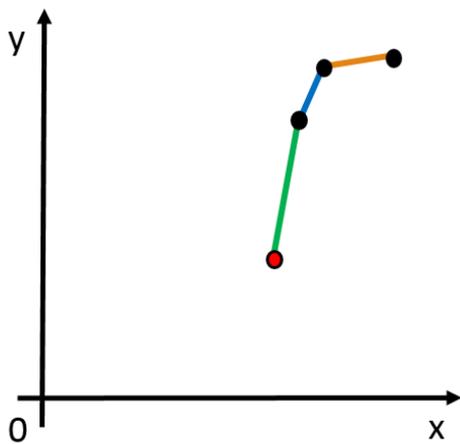
- ・重心座標間距離と傾きが長距離閾値 $[D_2]$ と1.5を上回っており、かつ前後どちらかの重心座標間距離と傾きが短距離閾値 $[D_1]$ と1.5を上回った場合「しゃがんだ」と判定する。(図5(c)(d))



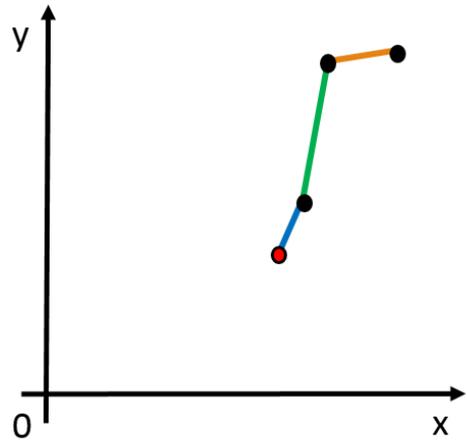
(a)



(b)



(c)



(d)

図 5 : 「しゃがむ」 の判定

青線 : 傾き  $> 1.5$  かつ距離  $> [D_1]$ 、緑線 : 傾き  $> 1.5$  かつ距離  $> [D_2]$

## 2.4.2 転倒

本研究では、人物が床に倒れる等した後自力で立ち上がれない状態を「転倒」と定義する（図 6）。「転倒」という行動について、「しゃがむ」という行動同様人物の検出領域の重心座標が大きく下方に変化するという特徴があるため、本研究では「転倒」という行動を「しゃがむ」という行動の延長上の行動として考える。

具体的な判定方法としては、『「しゃがんだ」と判定された後、一定時間内に立ち上がらなかった場合』転倒したと判定する。立ち上がったかどうかの判定は以下の手法で行う。

○立ち上がりの基準:

- ・しゃがんだと判定された際の重心座標間（図 7 青線）の中間地点の y 座標（図 7 赤線）を判定ラインとして設定する。

- ・しゃがんだと判定されてから 200 フレーム経過するまでに人物検出領域の重心の y 座標が判定ラインより上になった場合、立ち上がったと判定する。（図 7 水色線）

「転倒」に関しては、重心座標間距離を用いる方法とは別に、人物の検出領域を用いた判定方法も設定する。これは、図 6 のように転倒時の人物の検出領域の「高さ」と「幅」の比率が歩行時と大きく異なることを利用した判定方法である。

具体的には『検出領域の「高さ/幅」の値が設定した閾値[R]を下回った場合』転倒したと判定する。



(a) 歩行時の検出領域



(b) 転倒時の検出領域

図 6 : 「転倒」 時の検出領域の変化

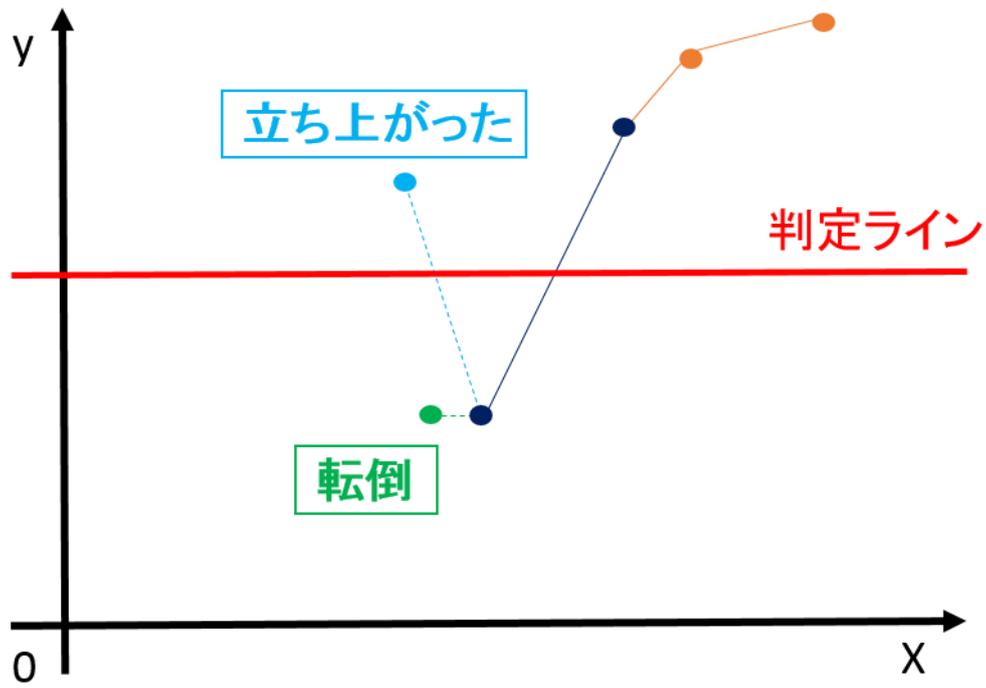


図 7 : 立ち上がり判定

### 2.4.3 停止

本研究では、人物が1か所にとどまっている状態を「停止」と定義する。「停止」という行動については、人物検出領域の重心座標があまり変化しないということが特徴として挙げられる。したがって重心座標間の距離も非常に短くなる。これを利用して「停止」という行動の検出を行う。

具体的には、『重心座標間の距離が連続して設定した閾値[S]以下になった場合』停止していると判定する。(図8)

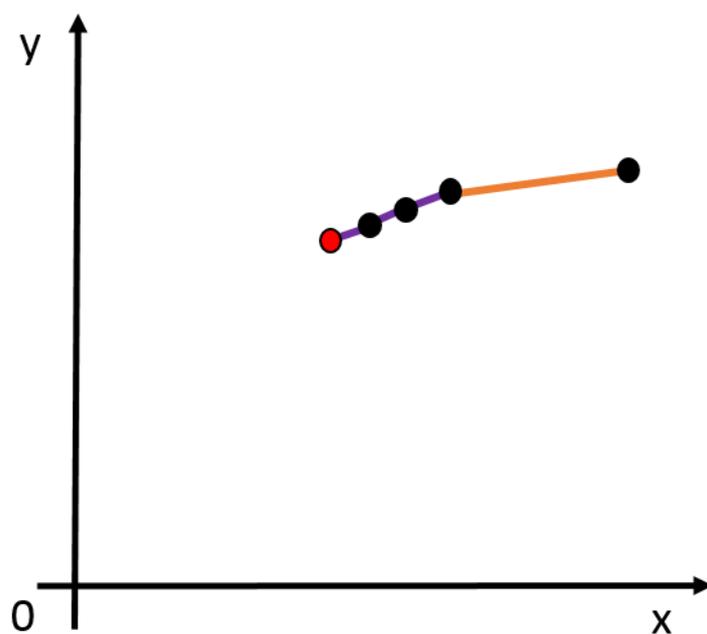


図8：「停止」の判定

紫線：距離 $\leq$ [S]

### 3. 閾値の自動設定

2 節で述べた手法で行動検出を行うにあたって設定しなければならない閾値は、

- ・「しゃがむ」の短距離閾値[D<sub>1</sub>]
- ・「しゃがむ」の長距離閾値[D<sub>2</sub>]
- ・「転倒」の閾値[R]
- ・「停止」の距離閾値[S]

である。

これらの閾値はカメラの設置角度や高さなどの違いによって変化する。また、重心座標間の距離がカメラと検出対象人物との距離によって変化するため、それに合わせて変化させる必要もある。

この問題を解決するため、カメラの設置後、廊下を歩いている様子を撮影し、その動画をもとにカメラの設置状況に対応した検出閾値の設定を自動で行う手法を提案する。

閾値設定のために撮影した動画に対して、行動検出を行うときと同じように人物検出と検出領域の重心座標の取得を行う。動画内の人物は奥から手前に向けて移動しているため奥を始点、手前を終点とする。得られた重心座標の軌跡の始点と終点の y 座標間の距離を水平方向に 10 分割し、各領域を始点に近い方から順に Sp1-Sp10 とする (図 9)。Sp1 より上の領域は Sp1、Sp10 より下の領域は Sp10 とする。

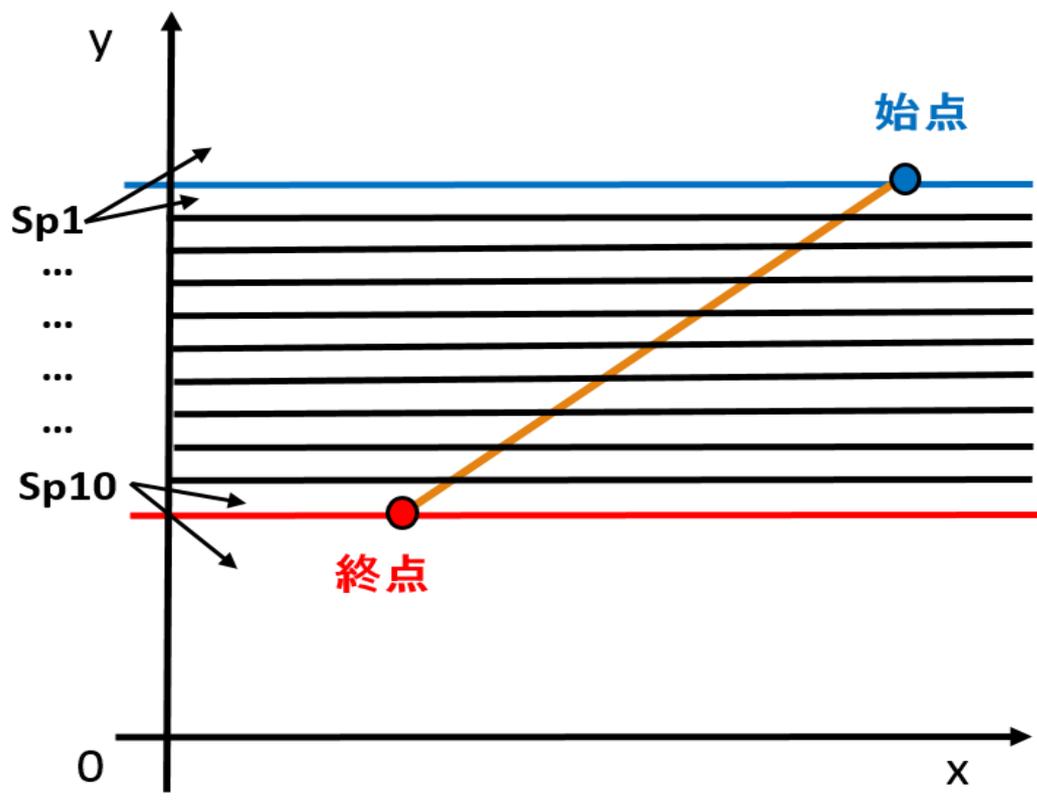


図 9 : 閾値設定

その後、各領域内に存在する重心座標間の距離の平均値を領域ごとに求める。この平均値を用いて閾値の自動設定を行う。閾値の計算法は以下の通りである。

○Sp1-Sp3:

- ・  $[D_1] = \text{平均値}$
- ・  $[D_2] = \text{平均値} \times 4$
- ・  $[S] = \text{平均値} / 4$

○Sp4-Sp10:

- ・  $[D_1] = \text{平均値} / 1.5$
- ・  $[D_2] = \text{平均値} \times 3$
- ・  $[S] = \text{平均値} / 3.5$

$[S]$ については重心座標間距離の0以上の最低値が0.5であるため、 $[S]$ が0.5を下回った場合、「停止」の条件を満たすのは重心座標間距離が0の時のみとなる。人物検出時の多少のズレや検出領域の大きさの変化でも重心座標間距離は変化するため、 $[S] < 0.5$ と設定するとほとんど検出されなくなる。そのため、 $[S]$ の最小値は0.5とした。Sp1-Sp3はカメラから遠く動きが小さいため、 $[S]$ を小さく設定する。また、本研究のカメラのセッティング方法では、人物がカメラから遠い場合傾きが大きくなりやすく、傾きの閾値を満たすことが多いので誤検出を出さないように $[D_2]$ に関しては大きく設定する。 $[D_1]$ は大きく設定しすぎると「しゃがむ」の検出自体ができなくなってしまうため、この設定とした。

閾値 $[R]$ に関しては、人物検出領域の「高さ/幅」の平均値を用いて閾値の自動設定を行う。閾値の計算法は以下の通りである。

- ・  $[R] = (\text{平均値} - 0.5) / 2$

## 4. 実験

3 節で述べた閾値の自動決定法を用いた場合での、対象行動の検出精度を調査した。「しゃがむ」を行っている動画 5 本・「転倒」を行っている動画 5 本・「停止」を行っている動画 6 本の計 16 本の動画、および閾値設定用の歩行動画 1 本を 1 つのデータセットとする。実験ではカメラの設置状況（廊下を撮影するカメラ角度、カメラの高さ）が異なる 2 種類のデータセット（データセット A、データセット B）を使用する（図 10）。各データセットの詳細については表 4 と表 5 に示す。

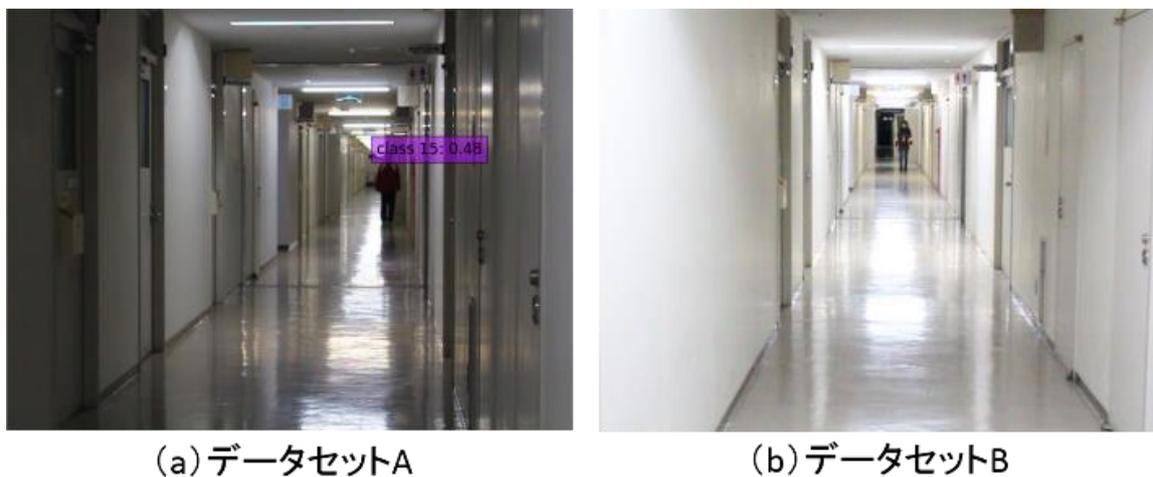


図 10：実験に用いたデータ

表 4 : データセット A

行動	時間	フレーム数	fps	サイズ
しゃがむ	29s	1639	59	480 × 640
	28s	1549		
	28s	1396		
	27s	1503		
	26s	1426		
転倒	23s	1779		
	26s	1683		
	24s	1690		
	22s	1652		
	22s	1576		
停止	38s	2282		
	35s	2109		
	34s	2123		
	34s	2095		
	32s	1963		
	30s	1843		

表 5 : データセット B

行動	時間	フレーム数	fps	サイズ
しゃがむ	26s	1586	59	480 × 640
	27s	1622		
	28s	1698		
	27s	1629		
	24s	1492		
転倒	30s	1845		
	30s	1807		
	32s	1930		
	27s	1666		
	29s	1795		
停止	31s	1880		
	34s	2040		
	36s	2193		
	39s	2341		
	35s	2141		
	36s	2203		

閾値の設定と行動検出の精度の算出は、データセットごとに行う。データセット A での閾値を閾値 A、データセット B での閾値を閾値 B とする。実験では、異なるデータセットの閾値を用いた場合との比較も行った。行動検出の精度については、動画内で対象行動を行っているフレーム区間内で、その対象行動の検出に成功しているフレームの割合が 30% を超え、かつ「転倒」と「しゃがむ」に関しては立ち上がり判定に成功した動画を行動検出成功動画とし、行動検出成功率

行動検出成功動画数

16

として求める。結果の一例を図 11 に示す。実験結果を表 6 に示す。また、しゃがんでいない場所でしゃがんだと判定するなど行動を誤検出したフレーム数を表 7 に示す。

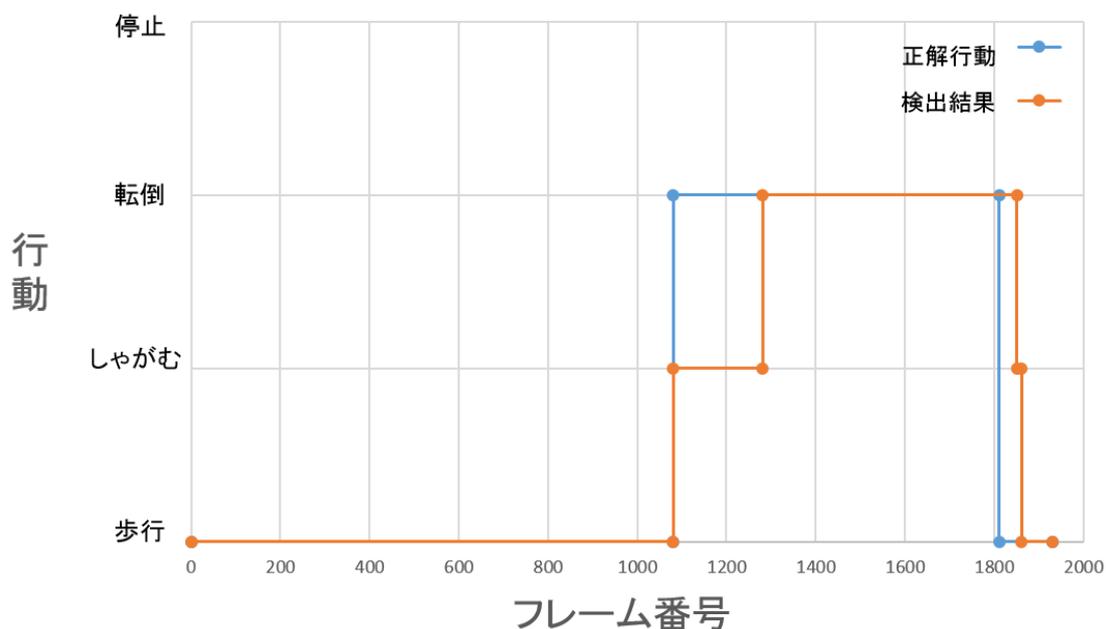


図 11：検出結果の一例  
(データセット B 転倒動画 3)

表 6：行動検出成功率

	閾値 A	閾値 B
データセット A	94%	88%
データセット B	63%	94%

表 7:行動誤検出フレーム数

	閾値 A	閾値 B
データセット A	27	206
データセット B	67	43

同じデータセットから自動設定した閾値を使用した方が正解率は高かった。また、異なるデータセットの閾値を使用した場合は、停止していない場所で停止したと検出されるなどの誤検出も多かった。特にデータセット B にデータセット A の閾値を用いた場合、しゃがんだ後立ち上がりの判定が行われないことがあり、これにより転倒していないのに転倒したという誤検出があった。これは [D<sub>1</sub>] の値が短すぎたことが原因であると考えられる。本研究で一番検出したい行動は「転倒」であり、この誤検出は非常に良くない結果である。これらのことから、閾値の自動設定は有効に機能していると言える。行動検出できなかったものに関しては、人物の検出自体がうまくできていなかった。

## 5. おわりに

本研究では、1台のカメラの映像から人物の行動検出を行う手法と、システムの設置を簡略化するための閾値の自動設定方法を提案した。実験により、自動設定した閾値に基づく簡便な処理で、行動検出が精度よく行えることを示した。行動検出のさらなる精度向上には、人物検出精度の向上が必要である。また今回は「三脚を使用し廊下を奥行き方向に撮影する」、「人物はカメラの奥から手前に向けて行動する」という制限を加えての実験だったため、今後はこの制限の緩和と人物検出の精度向上が課題である。

## 謝辞

卒業研究を行うにあたり、指導教員である棕木雅之教授にはお忙しい中指導していただき大変感謝しております。本研究は、受託研究「歩行者の転倒するリスク等の数値化及び転倒を検出するカメラの位置・角度などの設置指針の研究」の助成を受けて実施いたしました。また、機能強化経費「ロコモティブシンドローム（ロコモ）の病態解明・対策」の一環としても実施しており、両プロジェクトの関係者の方々にも感謝しております。そして棕木研究室の皆様をはじめ、他の研究室の方々、様々なアドバイスや実験への参加をしていただきありがとうございました。最後に、これまでご指導、協力していただいた皆様に心より感謝申し上げます。

## 参考文献

- [1] 田中仁,中内靖,“ユビキタスセンサによる独居高齢者見守りシステム”,日本機械学会論文集 C 編,75 卷 760 号,(2009), pp.3244-3252.
- [2] 前川泰子,他,“居住空間のスマート化に向けた高齢者見守りシステム開発の取り組み”,ヒューマンケア研究学会誌,5 卷 2 号,(2014),pp.51-54.
- [3] 関弘和,堀洋一,“高齢者モニタリングのためのカメラ画像を用いた異常動作検出”,IEEEJ Transactions on Industry Applications,Volume 122 Issue2, (2002), pp.182-188.
- [4] N.Dalal,B.Triggs, “ Histograms of Oriented Gradients for Human Detection” ,IEEE CVPR, (2005), pp.886-893.
- [5] Liu.Wei, et al, “SSD:Single Shot MultiBox Detector” ,ECCV, (2016), pp.21-37.
- [6] Pascal VOC, <http://host.robots.ox.ac.uk/pascal/VOC/> (2018/02/12 アクセス)