

# 姿勢情報を用いた深層学習による 人物動作認識のための自動クレンジング手法の改善

亀田明男<sup>1</sup> 磯貝愛<sup>1</sup> 志水信哉<sup>1</sup> 木全英明<sup>1</sup>

**概要:** 映像中の人物から、姿勢推定技術により得られる姿勢情報（関節位置）に基づき、機械学習や深層学習の技術を用いて動作認識を行う際に、姿勢推定の誤りにより間違った関節位置が出力されると、動作認識においても誤った分類を行うことがある。正しい動作認識を行うためには、このような場合、目視によるクレンジングを行うことが考えられるが、特に監視用途を想定した転倒や不審動作などの異常状態の認識を行う場合、リアルタイム性を確保できない問題がある。本稿では、姿勢推定技術として OpenPose を用いて、関節位置とともに出力される信頼度を活用し、目視時の考え方を反映して自動クレンジング（除去）する手法[4]を改善した。具体的には、重視する複数の関節点の閾値を、従来の一定値から重視する度合いに応じて可変とすることにより、誤って過剰に自動除去していたケース（過除去）で改善効果を確認できたので報告する。

**キーワード:** 深層学習, 姿勢情報, 自動クレンジング, 動作認識

## Improvement of Automatic Human Pose Information Cleansing Method for Human Motion Recognition Based on Deep Learning

AKIO KAMEDA<sup>1</sup> MEGUMI ISOGAI<sup>1</sup>  
SHINYA SHIMIZU<sup>1</sup> HIDEAKI KIMATA<sup>1</sup>

**Abstract:** In the process of human motion recognition using deep learning from pose information, if wrong pose information is output, motion recognition may also be wrong. In this case, although cleansing by human may be considered, processing cannot be performed in real time. In this paper, we propose an improvement method of automatic pose information cleansing technique [4] that reflects the way of thinking by human, using confidence score of each keypoint from OpenPose. Specifically, the thresholds of important joint position were variable according to each importance while those were fixed in the conventional method.

**Keywords:** Deep learning, Pose information, Automatic cleansing, Motion recognition

### 1. はじめに

画像から人の姿勢や物を使った動作など、人物の様々な状態や行動を認識することで、監視や防犯、スポーツなどの様々な分野での応用が期待される。近年、OpenPose[1]のように、高速に人物の画像から各関節位置（姿勢情報）を推定することが可能となり、姿勢情報を用いて機械学習や深層学習により転倒や不審動作等の動作認識を行う手法が提案されている[2][3]。

しかし、姿勢推定時に誤った関節位置を推定する「姿勢推定誤り」が発生すると、動作認識においても誤分類を行う「動作認識誤り」となる場合がある。この場合、誤った結果よりも、認識不可を通知することが重要と考え、このようなポリシーの下では、「動作認識誤り」の発生確率が高いと思われる姿勢情報については、予め認識不可としてクレンジング（除去）されることが望ましい。人が目視で除去する方法も考えられるが、リアルタイム性を確保できない問題がある。

そこで筆者らは、姿勢推定技術として OpenPose を用い、各関節位置とともに出力される信頼度を活用し、目視時の考え方を反映した自動クレンジング手法[4]を提案した。監視事例として、大型スーパーやショッピングモールといった商業施設の顧客の異常を検出する観点から、「倒れている」、「しゃがんでいる」状態の映像を評価対象とし、OpenPose により出力される姿勢情報に対して、目視時の考え方を反映した手法により、自動除去を行った。目視による除去を正解として、自動除去できた割合（自動除去率）に関して 78%~90%の精度を示したが、一部のシーンで、誤って過剰に自動除去してしまうケース（過除去）が発生していた。

このケースでは、正しいと思われる姿勢情報にも関わらず認識不可となり、誤った通知が頻繁に発生する。例えば、監視事例では、このような誤った通知により、監視員や警備員の確認作業の増大や複雑化を招くことが想定され、システムとしては、誤った通知の削減が求められる。

本稿では、従来手法[4]を発展させ、自動除去時に、誤って過剰に除去してしまう過除去にも対応可能な改善手法を提案する。さらに、目視によるクレンジング結果を正解と

<sup>1</sup> 日本電信電話（株）メディアインテリジェンス研究所  
NTT Media Intelligence Laboratories, Japan.

して比較し、本手法の評価を行ったので報告する。

## 2. 動作認識処理における処理誤りの原因

### 2.1 処理の概要

姿勢情報を用いた深層学習による人物動作認識の処理の概要を図1に示す。

まず、画像を入力として姿勢推定部から2次元(2D)の各関節位置の座標である姿勢情報を出力し、さらに、これを入力として、深層学習を用いた動作認識部にて、人の状態や動作を分類する。図1の例では、監視事例として商業施設での顧客の異常検出を想定し、「倒れている」、「しゃがんでいる」、「その他」の3つの状態に分類を行い、「その他」という状態情報を出力している。

### 2.2 姿勢推定誤りによる動作認識誤りの原因

異常検出を例として、図2に姿勢推定誤りによる動作認識誤りの例を示す。縦軸は図1で示した3つの状態情報、横軸は frame である。青い線は人が与えた正解で、赤い線は筆者らが独自に作成した学習用データ(約50万 frame)で学習した4層の多層パーセプトロン(MLP)により動作認識した分類結果を示している。

ここで、黄色の枠線で示す部分では、正解は「しゃがんでいる」であるが、図中の画像2例のように、肩付近を腰と誤り、さらに腕を足と誤ることで足が曲がっておらず、腰の位置を高く姿勢推定した「姿勢推定誤り」が発生し、これに起因して、動作認識部で「その他」と判定される「動作認識誤り」の問題が発生している。

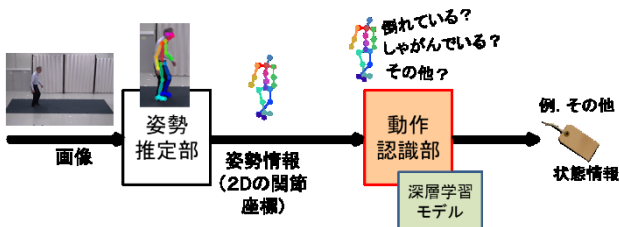


図1. 処理の概要

## 3. 姿勢情報の自動クレンジング手法

### 3.1 対処方針

このような「動作認識誤り」に対しては、[4]と同じく、異常検知という作業の中では、姿勢推定誤りによる動作認識誤りを防ぐことがシステムのポリシーとして考えられる。

より正確な動作認識を行うため、目視による姿勢情報のクレンジングが考えられるが、リアルタイム性を確保できない。そこで、誤った姿勢情報を、目視によるクレンジング時の考え方を反映して、自動的にクレンジングする手法を提案する。

### 3.2 従来手法の詳細

自動クレンジングのため、人が姿勢情報から動作認識する際の判断基準としている関節位置は重視するよう、閾値を高く設定し、それ以外の関節は、判断基準としている関節位置との関連度を関節間の距離と定義し、閾値の重みづけを設定する。今回は、閾値との比較対象に、OpenPoseによる各関節座標の信頼度(0~1の間を取り確かなほど値が高い)を用い、閾値以下の関節位置を含む姿勢情報は除去する。

具体的には、図3で示すように、異常検知の対象となる「倒れている」や「しゃがんでいる」際に重要と考えられる背骨(首と腰付近)と足首の位置を重視し、以下の手順で閾値を設定する。

- 1)重視の関節点(①, ⑤, ⑧)に、判定基準とする信頼度(閾値  $c_{req}$ : 図3では0.4)を手動で設定
- 2)端の関節点(④)に、最低の信頼度(閾値  $c_{min}$ : 図3では0.01)を手動で設定
- 3)任意の関節点(②, ③)は、線形補間により計算。ただし、重視の関節点の間(⑥, ⑦)は、これらの中央に最低の信頼度を設定して補間

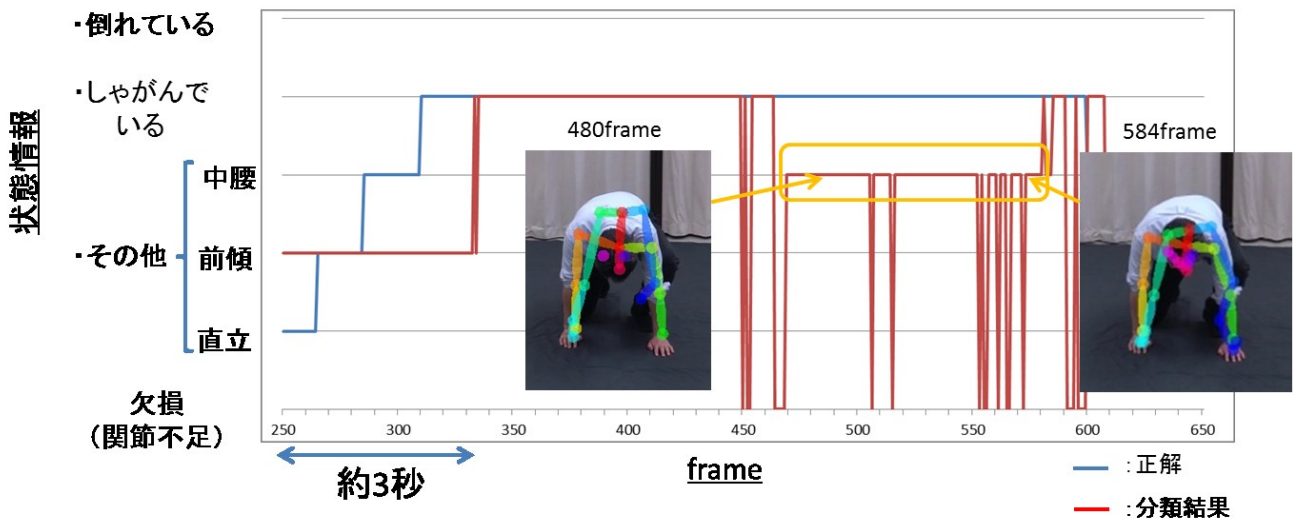


図2. 姿勢推定誤りによる動作認識誤りの例

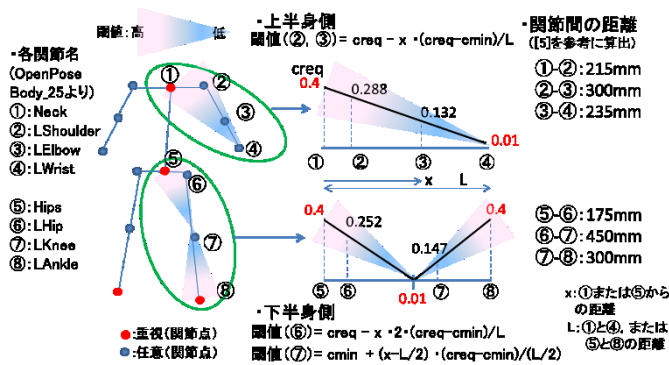


図3. 従来手法による閾値設定方法(creq=0.4, cmin=0.01)

なお、図3の例では、左腕、左足に閾値を設定しているが、右腕、右足にも同じ設定を行う。

#### 4. 従来手法の問題点と提案方式

##### 4.1 従来手法の問題点

従来手法では、上下および左右方向に、歩行動作から倒れる、しゃがむ動作を2名(演者 m,n)で行い、「倒れている」と「しゃがんでいる」シーンを抽出して評価映像とし、さらに、目視による除去(目視除去)を正解として、本手法で自動除去した割合(図4のc/a)を自動除去率とし評価している。

ここで、表1, 2に、cmin=0.01と固定した上で、creq=0.4および0.5の2パターンを対象にして、「倒れている」シーンを例に、図4で示す各ケースの件数と自動除去率を示す。なお、OpenPoseが全ての関節位置を出力したフレームを有効フレーム数とし、これらを実評価対象としている。

表1, 表2の左右方向/演者nに着目すると、正解となる目視除去(a)の94件に対して、従来手法により自動除去で誤って除去した「過除去」(e)が135件および246件と非常に多くなっている。

これは、図5の画像例のように、演者の足首部分の色と背景色が黒色で類似しており、OpenPoseによる足首の信頼度が、重視の関節点の閾値creqよりも低くなった(一例として、図5の画像にOpenPoseを適用した図6の左足首の信頼度は0.293で、この関節点の閾値creqよりも低い)ためである。

##### 4.2 提案方式の詳細

そこで、提案方式では、人が動作認識において判断基準としている関節位置は重視するよう、閾値を高く設定する際に、重視する度合いに応じて、閾値を可変とする。なお、重視以外の関節点は、従来手法と同様に、判断基準としている関節位置との関連度を関節間の距離と定義し、閾値の重みづけを設定する。

具体的には、図7で示すように、異常検知の対象となる「倒れている」や「しゃがんでいる」際に、重要と考え

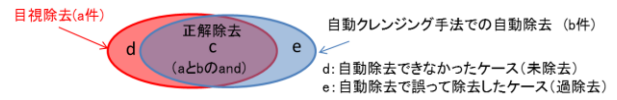


図4. クレンジング時の各ケース

表1. 「倒れている」シーンの評価結果(creq=0.4)

方向	演者	目視除去 a	自動除去 b	正解除去 c	未除去 d	過除去 e	有効フレーム数	全フレーム数	自動除去率 c/a
上下	m	281	275	275	6	0	281	1013	0.979
	n	39	20	14	25	6	241	567	0.359
左右	m	290	130	119	171	11	325	878	0.410
	n	94	229	94	0	135	471	366	1.000
全体		704	654	502	202	152	1318	2824	0.713

表2. 「倒れている」シーンの評価結果(creq=0.5)

方向	演者	目視除去 a	自動除去 b	正解除去 c	未除去 d	過除去 e	有効フレーム数	全フレーム数	自動除去率 c/a
上下	m	281	281	281	0	0	281	1013	1.000
	n	39	33	25	14	8	241	567	0.641
左右	m	290	175	148	142	27	325	878	0.510
	n	94	340	94	0	246	471	366	1.000
全体		704	829	548	156	281	1318	2824	0.778

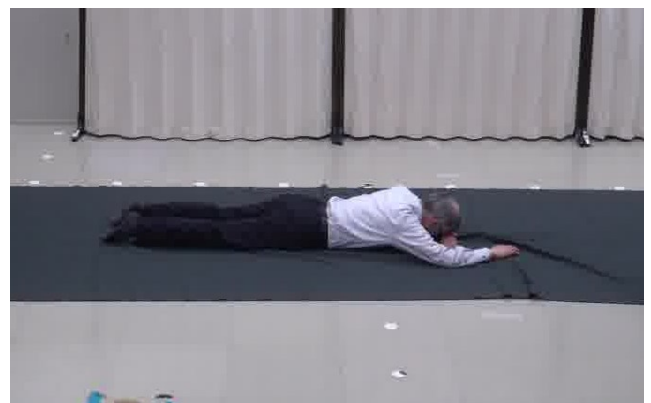


図5. 過除去した姿勢情報の画像例

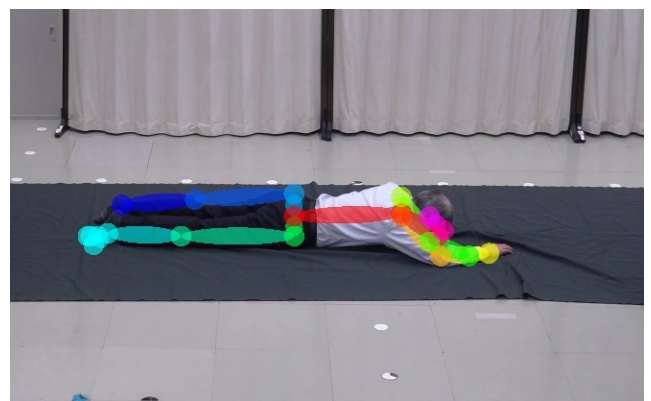


図6. 過除去した画像へのOpenPose適用例

られる背骨を構成する首と腰付近を最も重視(重視1)し、次に足首の位置を重視(重視2)し、以下の手順で閾値を設定する。なお、上半身側は従来手法と同じ手順となるため、以下では差分となる下半身側についてのみ示す。

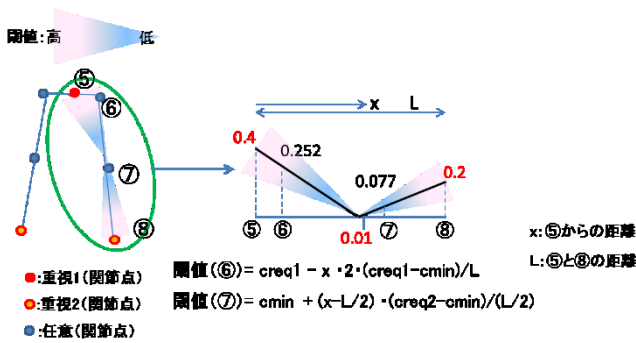


図7. 提案方式による閾値設定例 (下半身側のみ)  
( $creq1=0.4, creq2=0.2, cmin=0.01$ )

- 1) 重視1の関節点(⑤)に、判定基準とする最も高い信頼度(閾値  $creq1$ : 図7では0.4)を手動で設定
- 2) 重視2の関節点(⑧)に、判定基準とする次に高い信頼度(閾値  $creq2$ : 図7では0.2)を手動で設定
- 3) 重視1または重視2の関節点の間(⑥, ⑦)は、これらの中央に最低の信頼度(閾値  $cmin$ : 図7では0.01)を設定して補間

## 5. 評価

### 5.1 評価方法

従来手法の評価時に使用した評価映像から、従来手法の問題点となった左右方向に演者  $n$  が「倒れている」シーンを抽出し、これを今回の評価映像とした。

さらに、目視による除去を正解とし、従来手法および提案方式で正解を自動除去した正解除去の件数(図4のc)と、自動除去で誤って除去した過除去の件数(図4のe)の比( $c/e$ )を、過除去に対するスコアとし評価する。本スコアが高いほど、過除去を抑止しながら、自動除去で効率的に正解を除去していると考えられる。

### 5.2 評価結果

表3, 4に、 $cmin=0.01$ と固定した上で、 $creq1=0.4$ および0.5の2パターンを対象にして、 $creq2=0.2, 0.25, 0.3$ とした場合の、図4で示した各ケースの件数、自動除去率( $c/a$ )と過除去に対するスコア( $c/e$ )を示す。なお、比較のため、従来手法における各値を、 $creq2$ に「従来手法」と表記した行に併記した。また、従来手法からのスコアの改善率(従来手法のスコアに対する提案方式のスコアの倍率)も合わせて示す。

全体の傾向としては、 $creq2$ を低く設定すると、自動除去の件数(b)とともに、過除去の件数(e)が減少する。また、提案方式は従来手法に比べて、正解除去の件数(c)が減少するが、過除去に対するスコア( $c/e$ )が改善し、約1.9~3.3倍となることが分かった。

表3. 「倒れている」シーンの評価結果( $creq1=0.4$ )

creq2	目視除去	自動除去	正解除去	未除去	過除去	自動除去率	スコア	スコアの改善率
	a	b	c	d	e	c/a	c/e	
0.2	94	44	25	69	19	0.266	1.316	1.89
0.25	94	109	76	18	33	0.809	2.303	3.31
0.3	94	155	88	6	87	0.936	1.313	1.89
従来手法	94	229	94	0	135	1	0.696	-

表4. 「倒れている」シーンの評価結果( $creq1=0.5$ )

creq2	目視除去	自動除去	正解除去	未除去	過除去	自動除去率	スコア	スコアの改善率
	a	b	c	d	e	c/a	c/e	
0.2	94	141	61	33	80	0.649	0.763	2.00
0.25	94	166	82	12	84	0.872	0.976	2.55
0.3	94	190	88	6	102	0.936	0.863	2.26
従来手法	94	340	94	0	246	1.000	0.382	-

さらに、提案方式では  $creq2=0.25$  でスコアが最大となり、この時の自動除去率は、 $creq1=0.4$  で約81%、 $creq1=0.5$  で約87%となった。

## 6. 考察

全体の傾向として、閾値  $creq2$  を導入した提案方式において、従来手法の過除去に対するスコアが約1.9~3.3倍と改善し、さらに、スコア最大時の自動除去率は表1, 2の全体の平均を上回っており、本手法の有効性が示唆された。ただし、他のシーン、背景および演者で、同様な結果が得られるかは不明であり、今後、実験のバリエーションを増やすことが必要である。

また、提案方式において  $creq1=0.5$  とした場合、自動除去率は  $creq1=0.4$  の場合よりも高くなるものの、過除去に対するスコアは低くなり1を下回ることから、自動除去した件数のうち、半数以上を誤って過剰に除去している。このため、自動除去率や過除去に対するスコア等を含めた、総合的に判断可能な指標について、今後、検討する必要があると考える。

## 7. まとめ

本稿では、姿勢情報の誤りに対処するため、関節位置の信頼度を利用した自動クレンジング手法について、重視する関節位置の閾値を、従来では一定としていたが、重視する度合いに応じて可変とする改善手法について検討した。特に、従来手法の問題点である過除去に対し、提案方式では、過除去に対するスコアが改善することが確認された。

今後、他のシーンでの検証に加えて、今回の過除去に対するスコア以外に、自動除去率や未除去に対するスコアを含めた総合的なスコアの導入や、手動で設定している閾値の、自動的な決定アルゴリズムの検討が課題である。

## 参考文献

- [1] “OpenPose”. <https://github.com/CMU-Perceptual-Computing-Lab/openpose>, (参照 2019-07-22).

- [2] 一原賢吾, 竹内健, 金井謙治, 甲藤二郎: “姿勢推定技術に基づく機械学習を活用した人物の不審動作認識手法の精度評価”, 電子情報通信学会技術研究報告, 2018, vol.117, no. 483, pp89-94
- [3] Anahita Shojaei-Hashemi, Panos Nasiopoulos1, James J. Little, “Video-based Human Fall Detection in Smart Homes Using Deep Learning”, 2018 IEEE International Symposium on Circuits and Systems (ISCAS) , pp1-5
- [4] 亀田明男, 磯貝愛, 志水信哉, 木全英明: “深層学習による人物動作認識のための姿勢情報の自動クレンジングに関する一検討”, 映像情報メディア学会全国大会, 2019
- [5] “モノサシ -人体寸法編-” .  
[https://note.mu/takayoshi\\_kano/n/na8d6c6da969a](https://note.mu/takayoshi_kano/n/na8d6c6da969a), (参照 2019-07-22).