

# リカレント型ニューラルネットワークを用いた人体のモデルフィッティング手法の提案

田中豪<sup>†</sup> 直江健介<sup>†</sup> 武藤佳恭<sup>††</sup>

本研究では人体のシルエット画像を対象としたモデルフィッティング手法の提案を行う。フィッティングアルゴリズムにはリカレント型ニューラルネットワークを用い、各関節の再帰的・逐次的な最適化処理を行った。これにより関節間の距離・角度・位置を考慮したフィードバック評価を実現した。結果として、従来あった関節を評価する順序への依存性を最小限に抑えたモデルフィッティングを可能になった。本研究はモデルフィッティングの精度向上を図り、人間の姿勢・状態の高速判定への貢献を目指す。

## Model Fitting Methodology for Human Body with Recurrent Neural Network

Go Tanaka<sup>†</sup>, Kensuke Naoe<sup>†</sup> and Yoshiyasu Takefuji<sup>††</sup>

We propose the model fitting methodology for silhouette picture of human body. We employ Recurrent Neural Network in that algorithm, which can measure the joints recursively and sequentially. This enables the feedback measurement with considerations about the distances, angles and locations between the joints. As a result, we have achieved the model fitting without dependence of measurement order. Finally we aim to improve the accuracy of model fitting and to contribute high-speed processing for human posture and state.

## 1. イントロダクション

シルエット画像に対し多関節物体のモデルフィッティングにおいて課題となる1つに、各関節を評価する順序がある。体幹に近い関節は末端の関節の評価に影響を及ぼすため、体幹に近い関節から評価を始めることが望ましい。しかしながら、他方で体幹に近い関節もまた単独では一意に評価できず末端の関節に比較的弱いながらも依存性を持っている。そのため、一方に体幹に近い関節から評価するのではなく、関節間でのフィードバックのしくみが必要となる。[1]こういった従来よりも関節を評価する順序に依存の少ないモデルフィッティング手法を提案し、認識精度の向上が求められている。

本研究では、このような体幹に近い関節と末端の関節の間で双方向的な依存性に起因する関節を評価する順序課題に対しリカレント型ニューラルネットワーク（以下、RNN）による方法を提案する。人体を多関節モデルとみなし人間のシルエット画像を対象にモデルフィッティングを行う。その際、多関節モデルとシルエット画像のモデルフィッティングを最適化問題と捉え、照合度を評価するアルゴリズムに RNN を用いる。RNN は人間の脳が持つニューロンネットワークを簡易的に模したものであり、学習的な側面を持ち合わせた問題解決手法である。RNN の学習的な性質により、多関節モデルとシルエット画像のフィッティングという非線形空間における、体幹に近い関節と末端の関節の再帰的・逐次的な評価が可能となる。RNN のこの特性を利用することで、関節を評価する順序に影響されないモデルフィッティングが可能になる。

## 2. 提案手法

### 2.1 概要と手順

人体を多関節モデルとみなし人間のシルエット画像を対象にモデルフィッティングを行うとき、どの関節から評価を始めるかという課題がある。これは体幹に近い関節が末端の関節に影響を与えるが、体幹に近い関節を決定するためには単独では評価できず末端の関節も評価に関わるという逆説的な問題に起因している。この課題に対し我々は RNN によって解くことを提案する。本提案手法ではシルエット画像のピクセルをニューロンとし、すべてのニューロンに対し各関節の評価関数を元にどの関節であるかを決定する。RNN のアルゴリズムが持つ学習的な側面によって、非線形空間における体幹に近い関節と末端の関節の再帰的・逐次的な評価が可能になる。これにより各関節の並列的な再帰評価により、関節を評価する順序の影響を最小限にとどめ

<sup>†</sup> 慶應義塾大学 政策・メディア研究科  
Graduate School of Media and Governance, Keio University

<sup>††</sup> 慶應義塾大学 環境情報学部  
Faculty of Environment and Information Studies, Keio University

たモデルフィッティングを提案する。

本研究ではモデルフィッティングを島田ら[1]によってまとめられている関節物体の最適解探索問題として解く。まず背景画像を用いて抜き出した人間のシルエット画像を画像特徴  $F_i$  として用いる。背景差分の他にも数々の前景画像抜き出しの手法はあるが、本研究では手順を簡略化するため、背景差分を用いる。また人体の多関節モデル特徴  $F_M(\theta)$  ( $\theta$  はパラメータ) は図1のように表わす。この多関節モデルは実際の人体に基づいた制約を保持しており、具体的には関節間の距離や関節の可動角がその制約に当たる。上腕や大腿などは剛体として認識できるため、各剛体は距離の理想値を保持している。同様に各関節は固有の可動角を保持している。このように表現された多関節モデル  $F_M(\theta)$  と画像特徴  $F_i$  の照合度を表す評価関数  $S(F_M(\theta), F_i)$  を定義する。

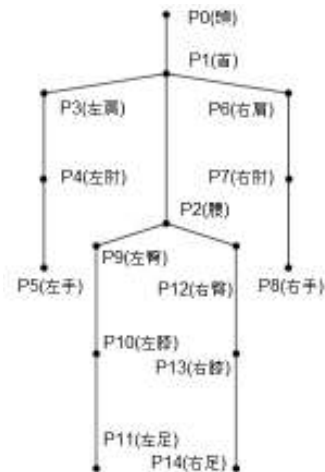


図1 人体の多関節モデル

ここで多関節モデル特徴  $F_M(\theta)$  は、各関節特徴  $F_m(\theta_k)$  ( $k=0,1,2,\dots,14$ ) によって構成されている。そのため、評価関数  $S(F_M(\theta), F_i)$  は数式1のように表現することができる。

$$S(F_M(\theta), F_i) = \sum_{k=0}^{15} S(F_m(\theta_k), F_i)$$

数式1 評価関数

$$\begin{aligned} S(F_m(\theta_0), F_i) &= D(\theta_1) + A(\theta_1, \theta_2) + L(\theta_1) \\ S(F_m(\theta_1), F_i) &= D(\theta_0) + D(\theta_2) + D(\theta_3) + D(\theta_6) + L(\theta_0) + L(\theta_2) \\ S(F_m(\theta_2), F_i) &= D(\theta_1) + D(\theta_9) + D(\theta_{12}) + A(\theta_1, \theta_0) + L(\theta_9) + L(\theta_{12}) \\ S(F_m(\theta_3), F_i) &= D(\theta_1) + D(\theta_4) + A(\theta_1, \theta_6) + L(\theta_1) \\ S(F_m(\theta_4), F_i) &= D(\theta_3) + D(\theta_5) + A(\theta_3, \theta_1) \\ S(F_m(\theta_5), F_i) &= D(\theta_4) + A(\theta_4, \theta_3) + L(\theta_3) \\ S(F_m(\theta_6), F_i) &= D(\theta_1) + D(\theta_7) + A(\theta_1, \theta_3) + L(\theta_1) \\ S(F_m(\theta_7), F_i) &= D(\theta_6) + D(\theta_8) + A(\theta_6, \theta_1) \\ S(F_m(\theta_8), F_i) &= D(\theta_7) + A(\theta_7, \theta_6) + L(\theta_6) \\ S(F_m(\theta_9), F_i) &= D(\theta_2) + D(\theta_{10}) + A(\theta_2, \theta_1) + A(\theta_2, \theta_{12}) + L(\theta_{12}) \\ S(F_m(\theta_{10}), F_i) &= D(\theta_9) + D(\theta_{11}) + A(\theta_9, \theta_2) + L(\theta_{13}) \\ S(F_m(\theta_{11}), F_i) &= D(\theta_{10}) + A(\theta_{10}, \theta_9) + L(\theta_2) \\ S(F_m(\theta_{12}), F_i) &= D(\theta_2) + D(\theta_{13}) + A(\theta_2, \theta_1) + A(\theta_2, \theta_9) + L(\theta_9) \\ S(F_m(\theta_{13}), F_i) &= D(\theta_{12}) + D(\theta_{14}) + A(\theta_{12}, \theta_2) + L(\theta_{10}) \\ S(F_m(\theta_{14}), F_i) &= D(\theta_{13}) + A(\theta_{13}, \theta_{12}) + L(\theta_2) \end{aligned}$$

数式2 関節ごとの評価関数

また各関節特徴と画像特徴の評価関数  $S(F_m(\theta_k), F_i)$  ( $k=0,1,2,\dots,14$ ) は関節ごとに数式2に定義する。ここで評価対象関節  $P_i$  とおくと、 $P_i$  と関節  $P_k$  が理想距離からのどれほど離れているかを評価する関数を  $D(\theta_k)$ 、関節  $P_k$  と  $P_i$  において  $\angle P_i P_k P_i$  が理想角からどれほど離れているかを評価する関数を  $A(\theta_k, \theta_i)$ 、 $P_i$  は関節  $P_k$  との上下・左右の位置関係が適切かを評価する関数を  $L(\theta_k)$ 、とそれぞれ表わしている。

## 2.2 ニューラル表現と期待される効果

RNNによって各関節の評価関数を最小化するような  $\theta$  を求めるために、課題をRNNの問題形式に置き換えるの必要があり、これをニューラル表現[2][3]と呼ぶ。主にニューロン、制約条件、動作式[a]を決定・定義する作業のことである。画像特徴  $F_i$  は座標情報を持ったピクセルの集合であり、このピクセルをニューロンとみなす。各ニューロンにおいて、制約条件である関節間の距離、角度、位置関係に基づいた動作式(数式

a) 動作式とは、ニューロンが課題解決を目指して発火するような式のこと。図1に基づきニューロン間の接続を決定している。この場合、評価関数と同義である。

2) により値  $U$  を算出する。値  $U$  は理想値から離れるほど大きな値を返すため、値が小さいほど照合度が高いことを意味する。よって、最小値を返すニューロンを発火させる。発火ニューロン群の重心座標を算出し、代表発火ニューロンを決定する。そして、直前に発火ニューロンから代表発火ニューロンに基づいて発火ニューロン座標を更新する。このような手順によりモデルフィッティングをピクセルをニューロンとした、提案した制約条件を満たす最適化問題として捉える。

この処理を全関節の評価関数に適用すると、ある関節として発火したニューロンの更新が次の関節の評価に逐一反映されるというメリットを得ることが可能になる。また、1 ステップ[b]のみの評価ではなく再帰的にステップを重ねることで学習効果を得ることが可能であり、より一層評価する順序に影響されないモデルフィッティングが可能になる。

### 3. 実験

#### 3.1 実験 1：関節を評価する順序に対する依存度の検証

##### 3.1.1 方法と評価

実験 1 では関節を評価する順序を入れ替えたいいくつかのケースを用意し、テスト画像を画像特徴とした結果を考察する。評価する順序のケースを下記に示す。

- ケース 1：P0→P1→P2→P3→P4→P5→P6→P7→P8→P9→P10→P11→P12→P13→P14
- ケース 2：P2→P1→P3→P9→P12→P6→P4→P10→P13→P7→P5→P11→P14→P8→P0
- ケース 3：P0→P8→P14→P11→P5→P7→P13→P10→P4→P6→P12→P9→P3→P1→P2
- ケース 4：P8→P1→P13→P4→P7→P0→P10→P6→P5→P11→P12→P3→P2→P9→P14

ケース 1 は関節の番号順であり腕・脚など部位単位の順序で評価するという特徴がある。ケース 2 は体幹に近い関節から末端への順序での評価、対照的にケース 3 は末端から体幹に近い関節への順序での評価。ケース 4 はランダムな順序での評価とした。評価方法はケースごとに結果に大きな違いがないか相対的な評価を行う。この相対的な相違の大小により、関節を評価する順序の依存度を評価する。

##### 3.1.2 結果

図 2 のテスト画像を対象に収束するまでフィッティングを行った結果を表したのが図 3 である。ケース 1 から 3 まではいずれも 40 ステップほどで収束し、精度としてはほぼ予想通りの妥当性が見受けられた。しかしながら、ケース 4 に関してのみ 65 ステップほどまで収束せず、結果画像も他よりも大きく精度を欠くものとなった。

b) 全関節を一通り評価することを 1 ステップと数える。

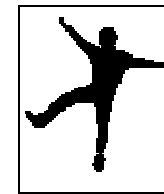


図 2 画像特徴

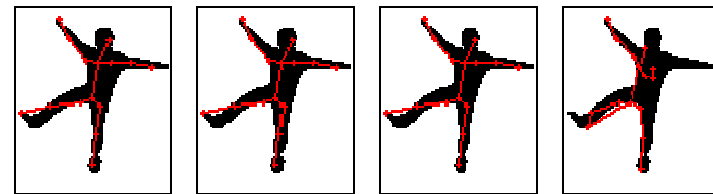


図 3 ケース別結果画像（左から順にケース 1, 2, 3, 4）

#### 3.2 実験 2：提案手法によるフィッティング検証

##### 3.2.1 方法と評価

実験 2 では画像タイプ別に提案手法によるモデルフィッティングを行い、それぞれ手や足において左右のどちらが上にあるかを正しくフィッティングできるかを検証する。用いる画像特徴を図 4 に示す。

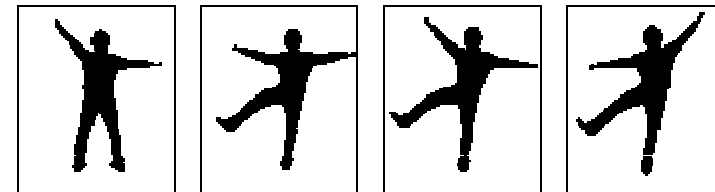


図 4 状態タイプ別画像特徴（左から順にタイプ 1, 2, 3, 4）

表 1 画像タイプ別手・足の識別目標（-は検証しない）

画像タイプ	左手	右手	左足	右足
タイプ1	上	下	-	-
タイプ2	-	-	上	下
タイプ3	上	下	上	下
タイプ4	下	上	上	下

タイプ別実験例としては、タイプ 1 については左手が右手より上にあることを識別すること、あるいはタイプ 3 では手についてはタイプ 1 と同様、かつ左足が右足より上にあることを識別できるかを検証する。

### 3.2.2 結果

結果画像を図 5 に示し、下記に画像タイプ別に結果を記述する。

#### ● タイプ 1

左手と右手の座標(X,Y)がそれぞれ(18,7), (69,28)であり、左手のほうが右手より Y 座標が小さい。よって、左手が右手より上にあることを識別した。(21 ステップで収束した。)

#### ● タイプ 2

左足と右足の座標(X,Y)がそれぞれ(11,57), (40,80)であり、左足のほうが右足より Y 座標が小さい。よって、左足が右足より上にあることを識別した。(16 ステップで収束した。)

#### ● タイプ 3

左手と右手の座標(X,Y)がそれぞれ(21,5), (68,29)であり、左手のほうが右手より Y 座標が小さい。よって、左手が右手より上にあることを識別した。また、左足と右足の座標(X,Y)がそれぞれ(2,52), (37,78)であり、左足のほうが右足より Y 座標が小さい。よって、左足が右足より上にあることを識別した。(37 ステップで収束した。)

#### ● タイプ 4

左手と右手の座標(X,Y)がそれぞれ(11,30), (47,29)であり、右手のほうが左手より Y 座標が小さい。よって、右手が左手より上にあることを識別した。また、左足と右足の座標(X,Y)がそれぞれ(4,56), (39,79)であり、左足のほうが右足より Y 座標が小さい。よって、左足が右足より上にあることを識別した。(27 ステップで収束した。)

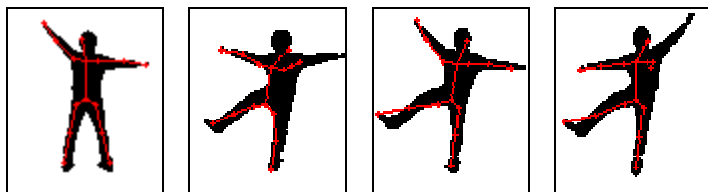


図 5 状態タイプ別結果画像 (左から順にタイプ 1, 2, 3, 4)

すべての画像タイプにおいて、37 ステップ以下で手と足ともに上下の識別について成功した。さらにタイプ 1, 3 においてはフィッティング結果に一定の精度が認められる。しかしながら、タイプ 4 における手の識別はわずか 1 ピクセル差での識別成功で

あり、精度が十分でない面も見受けられる。

## 4. まとめ

実験 1 ではケース 4 が大きく認識精度を欠くものであるが、他のすべてケースにおいて一定の精度を持った結果を導くことができた。これは、提案手法がすべてのケースにおいて評価する順序に対しての依存性の排除を保証するものではないが、ある規則性を持って評価した場合、依存性を排除できると言える。言い換えると、提案手法である再帰的・逐次的な最適化処理を行うことで、評価する順序のある規則性を持たせた条件下での精度向上を達成できたと考えられる。また実験 2 では設定した識別目標はわずか 37 ステップ以内ですべて達成し、さらには予想通りの認識結果を出力した画像タイプもあり、評価する順序に対しての依存性の排除が速度・精度向上に貢献したものであると予想できる。これは RNN を用いた提案手法が一定の速度と精度を保ち、モデルフィッティングを行えることを示唆している。今後の課題としては、精度向上に貢献する関節を評価する順序の規則性を解明すること、あるいはフィッティング精度が思わしくない結果も少なからずあり、精度の質向上がとして挙げられる。実験を重ねることで評価関数を洗練させていくことが求められる。また、実験 2 での識別問題は手・足の上下という大枠での識別を設定したが、より細かい人間のコンテキストを識別できるためには、画像タイプを詳細に切り分ける必要がある。今後の発展として人間の状態・姿勢の細かいコンテキストを抜き出せるようになった後、複数人モデルフィッティングを可能にすることで人間の状態判定付き人数カウンタなどへの応用を考えている。

**謝辞** 実験画像を撮影するにあたり協力していただいた浅井まり江氏、栗村大輝氏、山内裕己氏に深く感謝いたします。

## 参考文献

- 1) 島田伸, 有田大, 玉木徹: 関節物体のモデルフィッティング, 情報処理学会研究報告. CVIM, Vol.2006, No.51, pp. 375-392 (2006).
- 2) 武藤佳恭: ニューラルコンピューティング, コロナ社, 1996.
- 3) 武藤佳恭, 斎藤孝之: 応用事例ハンドブック ニューラルコンピューティング, 共立出版株式会社, 2001.