

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7537700号  
(P7537700)

(45)発行日 令和6年8月21日(2024.8.21)

(24)登録日 令和6年8月13日(2024.8.13)

|                 |                             |          |                                                          |         |
|-----------------|-----------------------------|----------|----------------------------------------------------------|---------|
| (51)国際特許分類      |                             | F I      |                                                          |         |
| G 0 6 T         | 7/20 (2017.01)              | G 0 6 T  | 7/20                                                     | 3 0 0 Z |
| G 0 6 T         | 7/00 (2017.01)              | G 0 6 T  | 7/00                                                     | 3 5 0 C |
| H 0 4 N         | 7/18 (2006.01)              | H 0 4 N  | 7/18                                                     | D       |
| G 0 6 N         | 20/00 (2019.01)             | H 0 4 N  | 7/18                                                     | K       |
|                 |                             | G 0 6 N  | 20/00                                                    |         |
| 請求項の数 11 (全29頁) |                             |          |                                                          |         |
| (21)出願番号        | 特願2020-203258(P2020-203258) | (73)特許権者 | 000000011<br>株式会社アイシン<br>愛知県刈谷市朝日町2丁目1番地                 |         |
| (22)出願日         | 令和2年12月8日(2020.12.8)        | (73)特許権者 | 504174135<br>国立大学法人九州工業大学<br>福岡県北九州市戸畑区仙水町1番1号           |         |
| (65)公開番号        | 特開2022-90760(P2022-90760A)  | (74)代理人  | 110004244<br>弁理士法人仲野・川井国際特許事務所                           |         |
| (43)公開日         | 令和4年6月20日(2022.6.20)        | (74)代理人  | 100096655<br>弁理士 川井 隆                                    |         |
| 審査請求日           | 令和5年9月29日(2023.9.29)        | (74)代理人  | 100091225<br>弁理士 仲野 均                                    |         |
|                 |                             | (72)発明者  | 山田 英夫<br>東京都千代田区外神田1丁目18番13号 秋葉原ダイビル7階 株式会社エクォ<br>最終頁に続く |         |

(54)【発明の名称】 学習装置、学習データの作成方法、学習方法、及び学習プログラム

(57)【特許請求の範囲】

【請求項1】

歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークの学習装置であって、前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報jと、当該完全骨格情報jによる歩行状態sとからなる複数の完全学習データ(j、s)を取得する完全学習データ取得手段と、前記取得した各完全学習データに対して、その完全骨格情報jから前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報cjと歩行状態sとからなる複数の遮蔽学習データ(cj、s)を作成する遮蔽学習データ作成手段と、前記取得した複数の完全学習データ(j、s)と前記作成した複数の遮蔽学習データ(cj、s)を使用して前記ニューラルネットワークの学習を行う学習手段と、を具備することを特徴とする学習装置。

【請求項2】

前記歩行状態sは、前記被写体の周期的な歩行動作における歩行段階と、前記被写体の歩行する方向を構成要素としている、ことを特徴とする請求項1に記載の学習装置。

【請求項3】

前記遮蔽学習データ作成手段は、鼻、首、及び、腰部中心に設定されたジョイントを前

記主要ジョイントとして削除対象から除外する、  
ことを特徴とする請求項 1、又は請求項 2 に記載の学習装置。

【請求項 4】

前記遮蔽学習データ作成手段は、前記被写体の右手、左手、右足、左足に対応する少なくとも 1 つの先端側から前記主要ジョイントに向けて繋がるジョイントを削除対象とする、  
ことを特徴とする請求項 1、請求項 2、又は請求項 3 に記載の学習装置。

【請求項 5】

前記遮蔽学習データ作成手段は、前記被写体に対して設定された全てのジョイントの 80 % 以下のジョイントを削除対象とする、  
ことを特徴とする請求項 1 ～請求項 4 のうちの 1 の請求項に記載の学習装置。

【請求項 6】

前記遮蔽学習データ作成手段は、前記主要ジョイントに加え、前記主要ジョイントに直結している 1 次ジョイントの少なくとも 2 以上を残して削除する、  
ことを特徴とする請求項 1 ～請求項 4 のうちの 1 の請求項に記載の学習装置。

【請求項 7】

前記遮蔽学習データ作成手段は、全ての完全学習データに対する遮蔽学習データを予め作成し、

前記学習手段は、完全学習データと予め作成した遮蔽学習データを使用して学習を行う、  
ことを特徴とする請求項 1 から請求項 6 のうちのいずれか 1 の請求項に記載の学習装置。

【請求項 8】

前記遮蔽学習データ作成手段は、1 の完全学習データに対する複数の遮蔽学習データを作成し、当該 1 の完全学習データと作成した複数の遮蔽学習データによる学習の後に、次の完全学習データに対する遮蔽学習データを作成する、  
ことを特徴とする請求項 1 から請求項 6 のうちのいずれか 1 の請求項に記載の学習装置。

【請求項 9】

歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークにおける学習データの作成方法であって、

前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報  $j$  と、当該完全骨格情報  $j$  による歩行状態  $s$  とからなる複数の完全学習データ ( $j$ 、 $s$ ) を取得する完全学習データ取得ステップと、

前記取得した各完全学習データに対して、その完全骨格情報  $j$  から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報  $c_j$  と歩行状態  $s$  とからなる複数の遮蔽学習データ ( $c_j$ 、 $s$ ) を作成する遮蔽学習データ作成ステップと、

を具備することを特徴とする学習データの作成方法。

【請求項 10】

歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークの学習方法であって、

前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報  $j$  と、当該完全骨格情報  $j$  による歩行状態  $s$  とからなる複数の完全学習データ ( $j$ 、 $s$ ) を取得する完全学習データ取得ステップと、

前記取得した各完全学習データに対して、その完全骨格情報  $j$  から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報  $c_j$  と歩行状態  $s$  とからなる複数の遮蔽学習データ ( $c_j$ 、 $s$ ) を作成する遮蔽学習データ作成ステップと、

前記取得した複数の完全学習データ ( $j$ 、 $s$ ) と前記作成した複数の遮蔽学習データ ( $c_j$ 、 $s$ ) を使用して前記ニューラルネットワークの学習を行う学習ステップと、  
を具備することを特徴とする学習方法。

【請求項 11】

10

20

30

40

50

歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークの学習プログラムであって、

前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報  $j$  と、当該完全骨格情報  $j$  による歩行状態  $s$  とからなる複数の完全学習データ ( $j$ 、 $s$ ) を取得する完全学習データ取得機能と、

前記取得した各完全学習データに対して、その完全骨格情報  $j$  から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報  $cj$  と歩行状態  $s$  とからなる複数の遮蔽学習データ ( $cj$ 、 $s$ ) を作成する遮蔽学習データ作成機能と、

前記取得した複数の完全学習データ ( $j$ 、 $s$ ) と前記作成した複数の遮蔽学習データ ( $cj$ 、 $s$ ) を使用して前記ニューラルネットワークの学習を行う学習機能と、をコンピュータに実現させることを特徴とする学習プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、学習装置、学習データの作成方法、学習方法、及び学習プログラムに関し、例えば、動画上で歩行者を追跡する歩行追跡装置で使用するニューラルネットワークの学習に関する。

【背景技術】

【0002】

歩行中の歩行者をカメラで動画撮影し、その動画上で歩行者を同定して追跡する技術の需要が高まっている。

例えば、特許文献1の技術では、外観による特徴や座標、及び歩行の方向などを手がかりにして複数のカメラで撮影した人物を追跡しつつ、アバターに変換している。

【0003】

しかし、一般的に、カメラによる歩行者の追跡は、監視カメラ設置エリアなど多数の人々が行き交う場所で行われる。

このような集団で歩行するようなシーンで歩行者を追跡する場合、外見の似た他者が近くを同じ方向に歩いていたり、歩行者の身体の一部が他の歩行者の陰に隠れたり、あるいは、服装のデザインの影響で角度によって歩行者の外見の見え方が変化したり、など、同様な情報が取得できるため、個々の歩行者を分離して追跡することが困難な場合があった。

【0004】

また、カメラによる歩行者の追跡において、追跡中の歩行者の一部が柱や荷物などの遮蔽物によって遮蔽されることで、歩行者の認識が困難になる場合がある。

そこで、特許文献2では、カメラによる対象物の骨格から歩行者を認識する場合に、遮蔽のない学習データから人工的に骨格の一部を欠損させたデータを用いて補間器を作成し、補間器と推定器を用いて対象物の骨格を認識する技術が提案されている。

【0005】

しかし、特許文献2記載技術では、補間器と推定器の2つを学習・推論させる必要があり、学習・推論にかかる処理時間が多い。

また、データを欠損させる際に現実空間でありえないような欠損を行うことは、逆にノイズとなる可能性がある。

【先行技術文献】

【特許文献】

【0006】

【文献】特開2019-9752号公報

【文献】特開2020-123105号公報

【発明の概要】

【発明が解決しようとする課題】

10

20

30

40

50

## 【0007】

本発明は、ニューラルネットワークによる歩行者追跡の頑健性を高めることを目的とする。

## 【課題を解決するための手段】

## 【0008】

(1) 請求項1に記載の発明では、歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークの学習装置であって、前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報 $j$ と、当該完全骨格情報 $j$ による歩行状態 $s$ とからなる複数の完全学習データ( $j$ 、 $s$ )を取得する完全学習データ取得手段と、前記取得した各完全学習データに対して、その完全骨格情報 $j$ から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報 $c_j$ と歩行状態 $s$ とからなる複数の遮蔽学習データ( $c_j$ 、 $s$ )を作成する遮蔽学習データ作成手段と、前記取得した複数の完全学習データ( $j$ 、 $s$ )と前記作成した複数の遮蔽学習データ( $c_j$ 、 $s$ )を使用して前記ニューラルネットワークの学習を行う学習手段と、を具備することを特徴とする学習装置を提供する。

10

(2) 請求項2に記載の発明では、前記歩行状態 $s$ は、前記被写体の周期的な歩行動作における歩行段階と、前記被写体の歩行する方向を構成要素としている、ことを特徴とする請求項1に記載の学習装置を提供する。

(3) 請求項3に記載の発明では、前記遮蔽学習データ作成手段は、鼻、首、及び、腰部中心に設定されたジョイントを前記主要ジョイントとして削除対象から除外する、ことを特徴とする請求項1、又は請求項2に記載の学習装置を提供する。

20

(4) 請求項4に記載の発明では、前記遮蔽学習データ作成手段は、前記被写体の右手、左手、右足、左足に対応する少なくとも1つの先端側から前記主要ジョイントに向けて繋がるジョイントを削除対象とする、ことを特徴とする請求項1、請求項2、又は請求項3に記載の学習装置を提供する。

(5) 請求項5に記載の発明では、前記遮蔽学習データ作成手段は、前記被写体に対して設定された全てのジョイントの80%以下のジョイントを削除対象とする、ことを特徴とする請求項1～請求項4のうちの1の請求項に記載の学習装置を提供する。

(6) 請求項6に記載の発明では、前記遮蔽学習データ作成手段は、前記主要ジョイントに加え、前記主要ジョイントに直結している1次ジョイントの少なくとも2以上を残して削除する、ことを特徴とする請求項1～請求項4のうちの1の請求項に記載の学習装置を提供する。

30

(7) 請求項7に記載の発明では、前記遮蔽学習データ作成手段は、全ての完全学習データに対する遮蔽学習データを予め作成し、前記学習手段は、完全学習データと予め作成した遮蔽学習データを使用して学習を行う、ことを特徴とする請求項1から請求項6のうちのいずれか1の請求項に記載の学習装置を提供する。

(8) 請求項8に記載の発明では、前記遮蔽学習データ作成手段は、1の完全学習データに対する複数の遮蔽学習データを作成し、当該1の完全学習データと作成した複数の遮蔽学習データによる学習の後に、次の完全学習データに対する遮蔽学習データを作成する、ことを特徴とする請求項1から請求項6のうちのいずれか1の請求項に記載の学習装置を提供する。

40

(9) 請求項9に記載の発明では、歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークにおける学習データの作成方法であって、前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報 $j$ と、当該完全骨格情報 $j$ による歩行状態 $s$ とからなる複数の完全学習データ( $j$ 、 $s$ )を取得する完全学習データ取得ステップと、前記取得した各完全学習データに対して、その完全骨格情報 $j$ から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報 $c_j$ と歩行状態 $s$ とからなる複数の遮蔽学習データ( $c_j$ 、 $s$ )を作

50

成する遮蔽学習データ作成ステップと、を具備することを特徴とする学習データの作成方法を提供する。

(10) 請求項10に記載の発明では、歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークの学習方法であって、前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報 $j$ と、当該完全骨格情報 $j$ による歩行状態 $s$ とからなる複数の完全学習データ( $j$ 、 $s$ )を取得する完全学習データ取得ステップと、前記取得した各完全学習データに対して、その完全骨格情報 $j$ から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報 $c_j$ と歩行状態 $s$ とからなる複数の遮蔽学習データ( $c_j$ 、 $s$ )を作成する遮蔽学習データ作成ステップと、前記取得した複数の完全学習データ( $j$ 、 $s$ )と前記作成した複数の遮蔽学習データ( $c_j$ 、 $s$ )を使用して前記ニューラルネットワークの学習を行う学習ステップと、を具備することを特徴とする学習方法を提供する。

10

(11) 請求項11に記載の発明では、歩行している被写体を撮影した動画のフレーム画像から取得した前記被写体の骨格情報を入力し、当該被写体の歩行状態を出力するニューラルネットワークの学習プログラムであって、前記被写体に骨格情報として設定されている複数のジョイントの位置情報が全て揃った完全骨格情報 $j$ と、当該完全骨格情報 $j$ による歩行状態 $s$ とからなる複数の完全学習データ( $j$ 、 $s$ )を取得する完全学習データ取得機能と、前記取得した各完全学習データに対して、その完全骨格情報 $j$ から前記被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報 $c_j$ と歩行状態 $s$ とからなる複数の遮蔽学習データ( $c_j$ 、 $s$ )を作成する遮蔽学習データ作成機能と、前記取得した複数の完全学習データ( $j$ 、 $s$ )と前記作成した複数の遮蔽学習データ( $c_j$ 、 $s$ )を使用して前記ニューラルネットワークの学習を行う学習機能と、をコンピュータに実現させることを特徴とする学習プログラムを提供する。

20

【発明の効果】

【0009】

本発明によれば、取得した各完全学習データに対して、その完全骨格情報 $j$ から被写体の中心線に沿って設定された主要ジョイントを除くジョイントの位置情報を削除した遮蔽骨格情報 $c_j$ と歩行状態 $s$ とからなる複数の遮蔽学習データ( $c_j$ 、 $s$ )を作成するので、ニューラルネットワークによる歩行者追跡の頑健性を高めることができる。

30

【図面の簡単な説明】

【0010】

【図1】歩行追跡装置のハードウェア的な構成を示した図である。

【図2】歩行状態の検出手順を説明するための図である。

【図3】骨格情報を説明するための図である。

【図4】歩行方向を説明するための図である。

【図5】歩行段階を説明するための図である。

【図6】歩行状態を説明するための図である。

【図7】MLPを説明するための図である。

【図8】歩行状態を用いた人物同定方法を説明するための図である。

40

【図9】実験結果を説明するための図である。

【図10】歩行追跡処理を説明するためのフローチャートである。

【図11】歩行状態計算処理を説明するためのフローチャートである。

【図12】隠れを有する学習データを説明するための図である。

【図13】部分領域を説明するための図である。

【図14】MLPの全体構成を説明するための図である。

【図15】歩行状態計算処理を説明するためのフローチャートである。

【図16】学習モデルの精度を比較した実験結果を説明するための図である。

【図17】歩行状態による対応付け手順を説明するためのフローチャートである。

【図18】隠れを有する学習データを使用したMLPの隠れ学習処理を表したフローチャ

50

ートである。

【発明を実施するための形態】

【0011】

(1) 実施形態の概要

歩行追跡装置1(図1)は、動画の異なるフレーム画像に写った歩行者9が同一人物であることを同定して追跡する。

同定は、歩行者9の外見情報による同定と歩行状態による同定を相補的に組み合わせて行い、これによって頑健性を高めている。外見情報による同定は、一般的に利用されている技術を用いる。

【0012】

歩行状態による同定は今回新たに開発した技術であって、歩行者9が連続的な歩行動作を行う際の姿勢に着目したものである。

歩行者は、歩行する際に周期 $2\pi$ の周期的な歩行動作を行うが、この歩行動作の各歩行段階を位相角 $\phi$ で表し、更に、歩行方向を $\theta$ で表し、 $\theta$ と $\phi$ の組み合わせによって歩行状態を定義する。

【0013】

類似した外見情報と歩行状態を有する複数の歩行者が近接して歩行していることは希であるため、外見情報と歩行状態を組み合わせることにより、頑健に歩行者9を識別して同定することができる。

また、歩行状態自体も、歩行方向 $\theta$ と歩行段階 $\phi$ を相補的に組み合わせてあるため、多少歩行方向がぶれたり、歩調が乱れたりしても、頑健に歩行者9の歩行状態を検出することができる。

【0014】

歩行状態については、25個のジョイントで定義した歩行者9の骨格情報15を抽出し、骨格情報による深層学習済みのMLP16(多層パーセプトロン)に入力することで、その出力値から推定する。

本実施形態のMLP16に対しては、フレーム画像から抽出した骨格情報15が遮蔽などによって一部欠損してても、より高い精度で歩行状態を推定することが可能な学習を済ませてある。

すなわち、歩行状態の学習の際に、歩行者と判断するために重要な骨格(顔の中心や胴など)は保持し、遮蔽が発生しやすい手やつま先等の骨格の末端から一定の割合で削除した学習データを使用してMLP16の学習を行う。

学習データは、全てを備えた骨格情報15とその歩行状態 $s$ (教師信号)からなる完全学習データと、当該骨格情報15から末端側のジョイントの位置情報を欠如(削除)した遮蔽骨格情報と同一の歩行情報 $s$ (教師信号)とからなる遮蔽学習データ、を使用する。遮蔽骨格情報と教師信号については、予め作成しておく場合と、学習の際に作成する場合のいずれでもよい。

【0015】

(2) 実施形態の詳細

カメラで歩行者を動画撮影し、2つのフレーム画像に写った歩行者を対応させて同一人物であると同定する場合、同じ動画データの時間的に前後するフレーム画像間で歩行者を対応させて同定する場合と、複数台のカメラで撮影した異なる動画データのフレーム画像間で歩行者を対応させて同定する場合がある。

以下、第1～第3実施形態で前者について説明し、第4実施形態で後者について説明するが、歩行状態を用いて2つのフレーム画像に写っている歩行者を同定する技術は共通である。

【0016】

[第1実施形態]

図1(a)は、第1実施形態に係る歩行追跡装置1のハードウェア的な構成を示した図である。

10

20

30

40

50

歩行追跡装置 1 は、CPU (Central Processing Unit) 2、ROM (Read Only Memory) 3、RAM (Random Access Memory) 4、表示部 5、入力部 6、通信制御部 7、カメラ 8、記憶部 10 などから構成されている。

【0017】

歩行追跡装置 1 は、歩行者 9 を動画撮影して、動画内で歩行者 9 を追跡する。骨格を有する対象であれば、歩行状態を検出して歩行を追跡できるが、本実施形態では、追跡対象を人とし、歩行追跡装置として利用する。

【0018】

図では、単一の歩行者 9 を図示しているが、歩行追跡装置 1 は、集団で歩行している複数の歩行者 9、9、・・・を個別に追跡することができる。

10

歩行追跡装置 1 は、歩行者ごとにばらばらである歩行状態を用いて個々の歩行者 9 を識別するため、集団的な歩行で特に顕著な頑健性を示す。

歩行追跡装置 1 は、例えば、自律走行ロボットに搭載されて省人化を行ったり、監視カメラ映像を用いて防犯対策やマーケティング分析などを行うのに用いられる。

【0019】

CPU 2 は、記憶部 10 や ROM 3 などに記憶されたプログラムに従って、各種の情報処理や制御を行う中央処理装置である。

本実施形態では、歩行追跡プログラムに従って情報処理することにより、動画データを構成する 2 つのフレーム画像に写った歩行者 9 を同定し、この同定処理を繰り返すことによって歩行者 9 を動画内で追跡する。

20

【0020】

ROM 3 は、読み取り専用メモリであって、歩行追跡装置 1 を動作させるための基本的なプログラムやパラメータなどを記憶している。

RAM 4 は、読み書きが可能なメモリであって、CPU 2 が動作する際のワーキングメモリを提供する。

本実施形態では、動画を構成するフレーム画像（1コマの静止画像）の画像データを展開して記憶したり、計算結果を記憶したりすることにより、CPU 2 が、歩行者 9 を追跡するのを支援する。

【0021】

30

表示部 5 は、液晶画面などの表示デバイスを用いて構成されており、歩行追跡装置 1 の操作画面や動画の表示などを行う。

入力部 6 は、キーボードやマウスなどの入力デバイスを用いて構成されており、操作画面への入力などを受け付ける。

【0022】

通信制御部 7 は、インターネットや LAN (Local Area Network) などの通信ネットワーク 11 を介して、外部の装置との通信を行う。

本実施形態では、例えば、遠隔地のイベント会場などの外部のカメラで撮影した動画データを通信ネットワーク 11 経由で受信して、この動画について歩行追跡処理を行うことができる。

40

【0023】

カメラ 8 は、汎用の動画撮影カメラであって、レンズで構成された光学系と、これによって結像した像を電気信号に変換する画像素子を用いて構成されている。

カメラ 8 は、歩行者 9 が歩行している場所を所定のフレームレートで撮影し、これら連続するフレーム画像で構成された動画を RGB 信号などの所定の画像信号によって出力する。

このように歩行追跡装置 1 は、歩行している被写体（歩行者 9）を撮影した動画をカメラ 8 や通信ネットワーク 11 経由で外部の装置などから取得する動画取得手段を備えている。

【0024】

50

記憶部10は、ハードディスクやEEPROM (Electrically Erasable Programmable Read-Only Memory) などの記憶媒体を用いて構成されており、歩行追跡プログラムやその他のプログラム、及びデータ（解析に必要なパラメータや解析によって得た歩行者9の歩行経路や歩行状態の時間的推移など）を記憶している。

【0025】

歩行追跡装置1は、ニューラルネットワークの一種であるMLP (Multilayer perceptron) による深層学習を用いて歩行者9の歩行状態を解析するが、これを高速化するためにGPGPU (General-Purpose computing on Graphics Processing Units) を用いたり、あるいは、CPU、GPU、メモリなどを一体にモジュール化した小型コンピュータを用いて小型化することもできる。この小型コンピュータは、自律走行ロボットに搭載して人物を自動追尾するのに適している。

【0026】

図1(b)の式(1)は、歩行追跡装置1が歩行者追跡のために行う計算式の一例を示している。

式(1)は、2つのフレーム画像に写っている歩行者9を対応させて同定するのに用いる式であって、第1実施形態～第3実施形態では、単一のカメラ8で撮影した時間的に前後するフレーム画像間で歩行者9を同定するのに用い、第4実施形態では、複数台のカメラ8、8、・・・で同時刻に撮影した2つのフレーム画像間で歩行者9を同定するのに用いる。

このように、動画撮影したカメラ8の異同にかかわらず、式(1)を用いて2つのフレーム画像間で歩行者9を同定することができる。

【0027】

歩行追跡装置1は、動画フレームで歩行者9を矩形によって人物検索し、歩行者9の外見情報を用いて計算した同一人物である確率と、歩行者9の歩行状態を用いて計算した同一人物である確率の線形和によって、歩行者9が同一人物である確率を計算する。係数 $\alpha$ は、 $0 < \alpha < 1$ の実数であり、例えば、 $0.3 < \alpha < 0.7$ 程度の範囲で実験により好適な値を設定する。

歩行追跡装置1は、外見情報と歩行状態を併せて判定した同一人物である確率が所定の閾値以上である場合に、歩行者9を同一人物であると同定する。閾値は、実験により求める。

このように、歩行追跡装置1は、被写体の外見情報を取得する外見情報取得手段を備え、歩行状態と、当該外見情報と、を用いて被写体を同定する。

【0028】

フレーム画像から人物を矩形で画像認識する技術や外見情報による同一人物程度の判定技術は、一般に用いられている技術を用いる。

外見情報を用いて同定する技術には、色分布によるカラーヒストグラム特徴を用いるもの、輝度勾配分布によるHOG特徴量を用いるもの、及び、深層学習で構築した距離空間に基づくもの（同一人物同士を近づけ、異なる人物同士を遠ざける距離空間を学習したもの）などがある。

これらは、何れも外見 (Appearance) に基づくものであり、例えば、服の色や種類が類似している場合や、照明が変化する場合、あるいは、カメラ設置位置による見え方が違う場合に誤同定することがある。

【0029】

一方、歩行状態による同一人物程度は、歩行者9の姿勢 (Pose) に着目して人物同定を行う技術であって、本願発明者が今回新たに開発した技術である。

近接する時刻であれば、カメラ位置によらず同一人物は同じ姿勢をとるため、姿勢に基づいて人物同定を行うことが可能となる。

【0030】

外見情報と歩行状態は、異なる観点から得られる情報であるため、相補効果が顕著である。

そのため、式(1)に示したように、外見情報による同定と、歩行状態による同定と、を相補的に組み合わせることによって、外見情報で生じる誤同定の原因に対しても頑健となり、人物同定判断の精度を大幅に高めることができる。

【0031】

例えば、人々が集団で歩行していて、歩行者9の近くを歩行している他の歩行者の外見が類似している場合は、歩行状態による区別が可能であり、逆に、歩行状態が類似している場合は、外見情報による区別が可能である。

以下では、今回新たに導入した歩行状態の検出について説明する。

【0032】

図2は、歩行状態の検出手順の全体像を説明するための図である。

まず、歩行追跡装置1は、フレーム画像から歩行者9を矩形により人物検出する。そして、検出した歩行者9の姿勢情報を骨格情報15によって抽出する。

骨格情報15の抽出には、例えば、OpenPoseやPoseNetといったソフトウェアで一般的に利用されている技術を用いることができる。

このように、歩行追跡装置1は、被写体の骨格情報を取得する骨格情報取得手段を備えている。

【0033】

次に、歩行追跡装置1は、抽出した骨格情報15を、予め骨格情報に対応する歩行状態を学習しておいたMLP16に入力して歩行状態sを推定する。

このように、歩行追跡装置1は、各種の歩行状態を骨格情報によって学習したニューラルネットワークを備えている。

【0034】

後述するように、本実施形態では、歩行者9が歩行する際の周期的な歩行動作に基づいて歩行状態sを定義し、歩行方向の2次元と、歩行段階の2次元で張られた4次元の歩行状態空間17における4次元ベクトルとして歩行状態sを表す。

なお、ベクトルのsは、図では太字で表しているが、文字コードの誤変換を防ぐために、明細書中では通常の英文字sで表す。他のベクトル量についても同様とする。

このように、歩行追跡装置1は、ニューラルネットワークを用いて被写体の周期的な歩行動作の連続性に基づく歩行状態を取得する。

【0035】

そして、歩行追跡装置1は、歩行状態空間17で、1つ前の動画フレーム画像で検出した歩行状態(s2とする)と、今回の動画フレームで検出した歩行状態(s1とする)の差をユークリッド距離Uによって計測し、これによって人物同定を行う。

【0036】

平面上の2点間の距離は、ピタゴラスの定理で与えられるが、これを一般化したのがユークリッド距離であって、ユークリッド距離Uは、s1とs2の各成分の差分の二乗の和の平方根で求めることができる。

歩行状態が近いほどユークリッド距離Uは小さくなるため、これによって2つの歩行状態の類似度を定量的に比較することができる。そのため、ユークリッド距離Uが近いほど同一人物である確率が高くなるように同定判断を行うことができる。なお、これは一例であって、他の計量を用いてもよい。

【0037】

このように、歩行追跡装置1は、骨格情報をニューラルネットワークに入力して、当該ネットワークが出力する歩行状態を用いて、動画を構成するフレーム画像間での被写体を同定する同定手段を備えている。

そして、歩行追跡装置1は、フレーム画像での同定処理を逐次行うことにより、歩行者9を動画内で追跡することができる。

このように、歩行追跡装置1は、動画で同定した被写体を追跡する追跡手段を備えてい

10

20

30

40

50

る。

#### 【0038】

図3は、骨格情報を説明するための図である。

本実施形態では、骨格情報をジョイント番号0～24の合計25個のジョイントで定義した。以下、ジョイント番号0のジョイントをJ0などと略記する。ジョイントは、骨格情報において関節として機能する部位であって、ジョイントを中心としてこれらを結ぶ線分を回転させることにより姿勢を変化させることができる。本実施形態で用いたジョイントと人体部位との対応関係は次の通りである。

#### 【0039】

J0：鼻、J1：首、J2：右肩、J3：右肘、J4：右手首、J8：腰部中心、J9：腰部右、J10：右膝、J11：右足首、J15：右目、J17：右耳、J22：右足親指、J23：右足小指、J24：右かかと、以下、対応する左半身も同様である。

#### 【0040】

歩行追跡装置1は、矩形で検出した人物の画像において、各ジョイントの画像上での2次元座標値を取得することにより、姿勢に応じた骨格情報を2次元データとして抽出する。

座標を取得する元となる画像は、フレーム画像から人物検出で切り出した矩形の画像における座標でもよいし、フレーム画像上での座標でもよい。

#### 【0041】

ジョイントは、更に多く定義したり、あるいは少なく定義したり、自由に設定することができる。例えば、背骨に該当する領域に更にジョイントを追加してより細かい姿勢を検出することもできるが、通常歩行は背筋を伸ばした状態で行うため必要ないと考えられる。このように、検出したい姿勢の属性に応じて適切なジョイントを定義することによって不要な計算を避けることができる。実験の結果、上記25点のジョイントで良好に歩行状態の推定を行うことができた。

#### 【0042】

図4は、歩行方向を説明するための図である。

図4(a)は、撮影エリアを上から見たところを示しており、歩行者9の歩行方向dを、歩行面に平行な平面上の単位ベクトルで表す。

図の例では、撮影方向のD1軸と、これに垂直なD2軸で張られるD平面を設定し、歩行者9を中心として回転する単位ベクトルによって歩行方向dを定義した。

#### 【0043】

計測する角度の基点は、どこでもよいが、図の例のようにD1を基点とする反時計回りの角度 $\theta$  ( $0^\circ \leq \theta < 360^\circ$ )で計測すると、歩行方向dの成分は、図4(b)に示したように、 $d = (d_1, d_2) = (\cos \theta, \sin \theta)$ となる。

#### 【0044】

図5は、歩行段階を説明するための図である。

図5(a)に示したように、歩行動作(足や手を運動させて体の重心を移動させる全身の連続動作)を歩行方向に垂直な方向から観察すると、歩行者9は、同じ周期的な動作を連続的に繰り返している。

#### 【0045】

1歩行周期は、例えば、右足を着地する(歩行者9a)→右足に重心を乗せる(歩行者9b)→左足を振り出す(歩行者9c)→左足を着地する(歩行者9d)→右足を振り出す(歩行者9e、図は省略)、の各歩行段階で構成されている。

このように歩行動作は周期的な動作であるため、歩行段階は1歩行周期の位相で表すことができる。

#### 【0046】

そこで、図5(b)で示したように直交するP1、P2軸で張られるP平面を想定し、歩行の位相の位相角 $\phi$  ( $0^\circ \leq \phi < 360^\circ$ )に対応してP平面上を回転する単位ベクトル( $p_1, p_2$ )によって歩行段階pを定義した。

位相角の起点は、どこでもよいが、図の例では、歩行者9aの段階を位相角 $\phi$ の起点と

10

20

30

40

50

し、歩行者 9 b ~ 9 e の段階に進むに従って反時計回りに増大するように  $\phi$  を定義した。

なお、図 5 (b) では、歩行者 9 e を省略し、P 1、P 2 軸による単位円上に歩行者 9 a ~ 9 d を順に表示した概念図であって、P 1、P 2 軸と、歩行者 9 a ~ 9 d の正確な対応関係を示すものではない。

【0047】

このように歩行段階  $p$  を 2 次元ベクトルで表すと、その成分は、図 5 (c) に示したように、 $p = (p_1, p_2) = (\cos \phi, \sin \phi)$  となる。

【0048】

図 6 は、歩行状態を説明するための図である。

図 6 (a) に示したように、歩行方向  $d$  と歩行段階  $p$  を合わせた 4 次元ベクトルで歩行状態  $s = (d_1, d_2, p_1, p_2)$  を定義する。

図示すると、歩行状態  $s$  は、図 6 (b) で示したような D 1、D 2、P 1、P 2 軸で張られる 4 次元の歩行状態空間 17 内のベクトル  $s$  で表される。

4 次元空間は図示することはできないため、図では、模式的に D 1 ~ P 2 軸を示してある。

【0049】

このように歩行方向  $d$  と歩行段階  $p$  を組み合わせて歩行状態  $s$  を定義すると、歩行方向  $d$  や歩行段階  $p$  のばらつきによる歩行状態  $s$  のずれが小さくなるように、これらが相補的に機能するため、歩行状態検出の頑健性を高めることができる

【0050】

歩行状態をこのように 4 次元ベクトルで表現したのは、後述する機械学習や人物同定の計算が容易になるからであり、他の手法で歩行状態の表現をすることもできる。

歩行状態は、同一時刻（多少前後してもよい）の歩行方向  $\theta$  と歩行段階の位相角  $\phi$  の組み合わせで規定されるため、直交する  $\theta$  軸と  $\phi$  軸からなる平面での点  $(\theta, \phi)$  ( $0 \leq \theta, \phi < 2\pi$ ) で表したり、単位球からなる球座標  $(\theta, \phi)$  で表すこともできる。これらのうち、理論計算や実装など、それぞれの場面で便利な座標系を用いればよい。

【0051】

なお、本実施形態では、 $\theta$  と  $\phi$  の組み合わせによって歩行状態を定義したが、 $\phi$  だけ、即ち歩行段階  $p$  だけで歩行状態を定義することも可能である。

また、更に他の要素を加えて更に高次元のベクトルとすることもできる。

歩行追跡装置 1 が備える同定手段は、歩行方向と歩行段階から歩行状態を生成することにより、歩行動作に加えて、被写体の歩行する方向を歩行状態に含めて同定を行うことができる。

【0052】

ところで、歩行者が歩行する様子、即ち歩容によって歩行者を識別する技術は、従来各種のものが提案されているが、これらは何れも、歩行者がある時間歩いたシーケンスにおいて歩容を取得している。これに対し、本実施形態の歩行追跡装置 1 は、歩行者 9 が歩いている一場面を切り取ったワンショットの画像から歩行状態を推定できるところが大きな特徴の 1 つである。

【0053】

図 7 は、MLP を説明するための図である。

歩行追跡装置 1 は、MLP（多層パーセプトロン：ニューラルネットワークの一種）による深層学習によって歩行状態  $s$  を計算した。

MLP 16 は、入力層、中間層、及び出力層を備えている。中間層の層数は 3、各中間層のノード数は 64 とした。

【0054】

入力層は、フレーム画像上での J 0 の座標値 ( $x_0, y_0$ )、J 1 の座標値 ( $x_1, y_1$ )、・・・、J 24 の座標値 ( $x_{24}, y_{24}$ ) の、それぞれを入力する 50 個のノードを備えている。

出力層は、歩行状態  $s = (d_1, d_2, p_1, p_2)$  の各成分を出力する 4 個のノード

10

20

30

40

50

を備えている。

#### 【0055】

活性化関数はReLUを用い、エポック数を100、バッチサイズを256とした。ReLUは、MLPで一般的に使われる非線形の関数である。活性化関数の非線形性により、対象（ここでは歩行状態）を非線形分離することができる。

エポック数は、学習回数であり、バッチサイズは、学習データをバッチに分けて学習させる際の各バッチに含まれる学習データ数である。

#### 【0056】

学習については、様々な歩行状態の画像を用意し、これらのジョイント座標値  $j = [x_0, y_0, \dots, x_{24}, y_{24}]$  と歩行状態  $s = [d_1, d_2, p_1, p_2]$  からなる学習データ  $(j, s)$  を用いて教師あり学習を行った。

つまり、学習データ  $(j, s)$  の  $j$  を入力すると、MLP16は  $s$  を出力するが、これを正解であるところの学習データ  $(j, s)$  の  $s$  と比較し、その誤差が小さくなるように中間層などのパラメータを調節した。

このようにパラメータが調整されたMLP16を用いると、あるフレーム画像から抽出した骨格情報15のジョイント座標値を入力するだけで、歩行状態  $s$ 、即ち、歩行方向  $d$  と歩行段階  $p$  を同時に算出することができる。

#### 【0057】

図8は、歩行状態を用いた人物同定方法を説明するための図である。

人物同定は、先に述べたように歩行状態空間17でのユーグリッド距離  $U$  を用いて行う。

例えば、歩行状態  $s_1$ 、 $s_2$  を、それぞれ、最新のフレーム画像の歩行者9から抽出した歩行状態と、これより1つ前のフレーム画像の歩行者9から抽出した歩行状態とする。

なお、第4実施形態で、複数台のカメラ8a、8bからの動画で歩行者9を同定する場合は、カメラ8aで撮影したフレーム画像aから抽出した歩行状態と、カメラ8bで同一時刻に撮影したフレーム画像bから抽出した歩行状態とする。

#### 【0058】

カメラ8による動画撮影のフレームレートは例えば30fps程度であり、この場合時間的に隣接するフレーム画像の撮影時間差は、0.03秒程度である。これを人間の歩行動作の速さと比較すると非常に速く、隣接するフレーム画像内での姿勢変化はないと考えられる。

このように同一時刻の概念は、追跡対象の動く早さと、フレーム画像の撮影間隔の対比により規定され、撮影間隔での歩行状態  $s$  の変化が同定処理に実質的に影響を与えない範囲内である時間差は、同一時刻と考えられる。

#### 【0059】

これにより、最新のフレーム画像と1つ前のフレーム画像は、ほぼ同一時刻に撮影した画像と考えられ、最新のフレーム画像から計算した歩行状態  $s_1$  と、1つ前のフレーム画像から計算した歩行状態  $s_2$  が、同一人物によるものであった場合、これらの差は非常に小さいことが期待できる。

#### 【0060】

そこで、本実施形態では、歩行状態  $s_1$ 、 $s_2$  の差を式(2)で示したユーグリッド距離  $U$  で表し、両歩行状態の人物が同一人物である確率  $prob$  を式(3)で定義した。

式(2)で示したユーグリッド距離  $U$  が小さいほど歩行状態  $s_1$ 、 $s_2$  の差が小さく、両フレーム画像での歩行者9の姿勢が近くなり、同一人物である可能性が高くなる。

そして、式(3)によれば、ユーグリッド距離  $U$  が小さいほど確率が1に近づいて同一人物である確率が高まり、その近づき方を定数  $\gamma$  によって調節することができる。

両人物が同一であると同定判断するための確率の閾値と定数  $\gamma$  は、実験により適当な値を設定する。

#### 【0061】

変形例として、歩行状態  $s_1$ 、 $s_2$  の内積を用いて人物同定することも可能である。

例えば、歩行状態  $s_1$ 、 $s_2$  の大きさが、それぞれ1となるように規格化して内積を計

10

20

30

40

50

算すると、歩行状態  $s_1$ 、 $s_2$  の距離を内積によって  $-1$  から  $1$  までの値で測量することができる。内積計算は、計算機で容易に計算できるため、これを用いることも考えられる。

これらは、一例であって、他の計算式によって人物同定を行ってもよい。

#### 【0062】

以上の例では、時間的に隣接する2つのフレーム画像での歩行状態について説明したが、実質的に同一時刻と考えられる範囲内で時間的に前後したフレーム画像であれば、間欠的に取得したフレーム画像で解析可能である。

このように、歩行追跡装置1は、実質的に同一時刻での歩行状態  $s$  を抽出して比較するため、歩行者9が方向を変えたり歩調を変化させたりしても、これに追隨して追跡することができる。

#### 【0063】

図9は、実験結果を説明するための図である。

図9(a)の左図は、歩行方向  $d$  の実験結果を表したグラフである。

横軸は実線31で示したグラントゥールスを表しており、縦軸は、実際に動画から計算した歩行者の歩行方向  $d$  の予測値 (MLP16が出力した値) の分布 (縦線) を表している。

グラントゥールスとは、機械学習の精度を評価するための比較用のデータである。

図に示したように、予測値は、グラントゥールス上に分布しており、高い精度で歩行方向  $d$  を予測することができている。

#### 【0064】

図9(a)の右図は、歩行段階  $p$  の実験結果を表したグラフである。

同様に、実線32でグラントゥールスを表し、縦線で予測値の分布を表している。歩行方向  $d$  に比べると、予測値のばらつきが若干広がるものの、予測値は、概ねグラントゥールス上に分布しており、高い精度で歩行段階  $p$  を予測することができている。

#### 【0065】

図9(b)は、これら実験による歩行方向  $d$ 、歩行段階  $p$ 、及び歩行状態  $s$  のRMSEを示した表である。

RMSE (Root Mean Square Error: 二乗平均平方根誤差) は、式(4)で表され、グラントゥールス ( $y_{obs, i}$ ) と予測値 ( $y_{pred, i}$ ) とのずれの二乗の総和の平均の平方根である。この値が小さいほどMLP16の性能が良く、予測精度が高いことを意味している。

#### 【0066】

表に示したように、歩行方向  $d$ 、歩行段階  $p$ 、及び歩行状態  $s$  のRMSEは、それぞれ、0.110、0.229、0.179となり、これらは、十分に実用に耐える値である。

なお、表では、歩行方向  $d$  と歩行段階  $p$  について、平均誤差をベクトルの角度に換算した値も併記してあり、それぞれ、 $6.3^\circ$ 、 $13.1^\circ$  である。

#### 【0067】

図10は、歩行追跡装置1が行う歩行追跡処理を説明するためのフローチャートである。

以下の処理は、記憶部10に記憶した歩行追跡プログラムに従ってCPU2が行うものである。

まず、CPU2は、カメラ8が撮影した動画のフレーム画像を取得してRAM4に記憶する(ステップ5)。

#### 【0068】

次に、CPU2は、ステップ5でRAM4に記憶したフレーム画像において、矩形によって人物検出を行い、これによって追跡対象となる歩行者9を設定する(ステップ10)。

人物検出で複数の歩行者9、9、...が検出された場合は、これら全員を追跡対象としてもよいし、これから選択した単数、又は複数の歩行者9、9、...を追跡対象としてもよい。複数の歩行者9、9、...を追跡対象とする場合は、個々の歩行者9の追跡を並列処理にて同時に行う。

また、追跡対象とする歩行者9を選択する場合は、何らかのアルゴリズムに従って自動

10

20

30

40

50

的に選択してもよいし、手動で選択してもよい。

【0069】

次に、CPU 2は、追跡対象として設定した歩行者9の外見情報を取得してRAM 4に記憶する（ステップ15）。

複数の歩行者9、9、・・・を追跡対象として設定した場合は、それぞれの歩行者9について外見情報を取得する。以下同様に、複数の歩行者9、9、・・・を追跡する場合は、歩行者ごとに処理する。

【0070】

次に、CPU 2は、歩行者9の骨格情報15を抽出してMLP 16に入力することにより歩行状態を取得し、これをRAM 4に記憶する（ステップ20）。

以上の外見情報と歩行状態が、追跡に用いる初期値となり、以降の動作で外見情報と歩行状態を連続的に追跡していく際の起点となる。

【0071】

次に、CPU 2は、カメラ8の撮影した動画から次のフレーム画像を取得してRAM 4に記憶する（ステップ25）。

次に、CPU 2は、外見情報計算処理を行って、式（1）右辺第1項の外見情報による確率を計算してRAM 4に記憶する（ステップ30）。

【0072】

CPU 2は、この外見情報による確率の計算を、例えば、次のようにして行う。

まず、CPU 2は、ステップ25でRAM 4に記憶したフレーム画像から歩行者9を人物検出し、その外見情報を取得してRAM 4に記憶する。

そして、CPU 2は、ステップ15でRAM 4に記憶した1つ前のフレーム画像から取得した外見情報を読み取り、それを今回記憶した外見情報と比較して、その類似度によって同一人物である確率を計算する。

この計算は、一例であって、外見情報を用いて一般的に行われている他の方法を用いてもよい。

【0073】

次に、CPU 2は、歩行状態計算処理を行って、ステップ25でRAM 4に記憶したフレーム画像による歩行状態を計算してRAM 4に記憶すると共に、式（1）右辺第2項の歩行状態による確率を計算してRAM 4に記憶する（ステップ35）。

次に、CPU 2は、ステップ30で記憶した外見情報による確率とステップ35で記憶した歩行状態による確率をRAM 4から読み出し、これらを式（1）に代入して、歩行者9が同一人物である確率を計算してRAM 4に記憶する。

【0074】

そして、CPU 2は、同一人物である確率が所定の閾値以上である場合は、今回のフレーム画像から検出した人物が、追跡対象の歩行者9であると同定する（ステップ40）。

同定により、ステップ25でRAM 4に記憶したフレーム画像での追跡が成功したことになる。

【0075】

確率が閾値未満であった場合、CPU 2は、今回の判定はスキップして、次のフレーム画像で再度判定を試みるとか、検索した人物が複数いる場合は、歩行者9に近接する人物が歩行者9である可能性があるため、これについて判定するとか、あるいは、アラームを発して処理を終了するなど、何らかの処理を行う。

【0076】

追跡を続行する場合（ステップ45；Y）、CPU 2は、ステップ25に戻って次のフレーム画像について判定を行う。この場合は、前回にステップ30、35でRAM 4に記憶した外見情報と歩行状態を用いて判定を行う。

一方、追跡を続行しない場合（ステップ45；N）、CPU 2は、処理を終了する。

【0077】

図11は、ステップ35（図10）の歩行状態計算処理を説明するためのフローチャー

10

20

30

40

50

トである。

まず、CPU 2は、矩形領域によって人物検出した歩行者9から骨格情報15を抽出してRAM 4に記憶する（ステップ60）。

CPU 2は、例えば、矩形領域によって切り出した画像におけるJ0～J24の座標値によって骨格情報15を構成する。

【0078】

次に、CPU 2は、ステップ60でRAM 4に記憶した骨格情報15をMLP 16に入力し（ステップ65）、次いで、MLP 16によって計算した歩行状態s1を取得してRAM 4に記憶する（ステップ70）。

次に、CPU 2は、前のフレーム画像で取得した歩行状態s2をRAM 4から読み出して取得する（ステップ75）。

【0079】

そして、CPU 2は、ステップ70で記憶した歩行状態s1（今回のフレーム画像での歩行状態）をRAM 4から読み出して取得し、これとステップ75で読み出した歩行状態s2（1つ前のフレーム画像での歩行状態）を図8の式（3）に代入して歩行状態による確率を計算し、これをRAM 4に記憶する（ステップ80）。

【0080】

[第2実施形態]

第1実施形態では、例えば、OpenPoseやPoseNetなどで用いられている骨格検出可能な手法を利用して歩行者9の姿勢を算出した。

しかし、障害物の存在や人とのすれ違いなど、環境によっては歩行者9の一部が隠れによって観測できず、骨格座標が部分的にしか検出できない場合があり、このようなときは歩行状態の推定が困難となる場合がある。

【0081】

そこで、第2実施形態では、隠れによって骨格情報15の一部が観測できない学習データを人為的に作成し、これをMLP 16に入力して隠れを含めた学習を行った。

これにより、隠れによって未検出のジョイントがあっても頑健な歩行状態の推定を実施することができる。

【0082】

図12は、隠れを有する学習データを説明するための図である。

本実施形態では、骨格情報15の一部の領域に含まれるジョイントを未検出に設定することにより、隠れによって未検出となったジョイントを含む学習データ（遮蔽学習データ）を用意した。

例えば、図に示した領域35を隠すと右頭部が未検出となった骨格情報15が得られ、領域36を隠すと左手が未検出となった骨格情報15が得られ、領域37を隠すと右足が未検出となった骨格情報15が得られる。

【0083】

このように骨格情報15を部分的に隠すことにより、骨格の先端領域から中心領域に向けて骨格情報が欠如するように、また、隠れ領域の面積が全面積の0～80%の間となるように隠れ領域を設定した学習データ（遮蔽学習データ）を多数用意した。

骨格の先端から骨格の中心に向けて領域が隠れるようにしたのは、歩行中には手足などの四肢や片半身が隠れる場合が多いためである。

【0084】

このようにMLP 16の学習を行う学習データとしては、完全学習データ（j、s）と、遮蔽学習データ（cj、s）を使用する。

完全学習データ（j、s）は、第1実施形態で説明した学習データと同じで、全ジョイント座標値jが揃った骨格情報（完全骨格情報j）と歩行状態s（教師信号）からなる学習データである。

遮蔽学習データ（cj、s）は、完全学習データの完全骨格情報jから、末端側のジョイントから優先的に位置情報を削除（欠如）して作成した遮蔽骨格情報cjと、完全骨格

10

20

30

40

50

情報  $j$  に対する歩行状態  $s$  (教師信号) からなる学習データである。

遮蔽学習データの遮蔽骨格情報  $c_j$  は、完全骨格情報  $j$  から、図 12 に示すように、末端側のジョイントを削除することで作成する。

【0085】

すなわち、顔の中心や胴などの身体の中心線に沿ったジョイントを主要ジョイントとし、この主要ジョイントは必ず保持する。

一方、遮蔽が発生しやすい手やつま先等の骨格の末端側から優先して一定の割合で削除する。

主要ジョイントに直結しているジョイントを1次ジョイントとした場合、この1次ジョイントについては、隠れ領域の全面積(削除するジョイントの割合)が最大80%以下、好ましくは60%以下となる数の範囲まで削除する。

【0086】

ここで、図3のジョイント  $j_0 \sim j_{24}$  のうち、主要ジョイントは身体の中心線に沿った鼻( $j_0$ )、首( $j_1$ )、腰部中心( $j_8$ )が該当し、この主要ジョイントに直結しているジョイント  $j_2$ 、 $j_5$ 、 $j_9$ 、 $j_{12}$ 、 $j_{15}$ 、 $j_{16}$  が1次ジョイントに該当する。

全ジョイント数が25で、主要ジョイント数が3なので、6つの1次ジョイントのうち、少なくとも2つ以上のジョイントは削除対象から除外することで、隠れ領域の全面積を80%以下としている。

このように、主要ジョイントと、それに直結する1次ジョイントの一部を保持することで、歩行状態  $s$  の推定が可能になる。

【0087】

本実施形態による学習は、完全学習データと遮蔽学習データにより行うことで、より多くの学習データを使用した学習が可能なり、学習精度を向上させることができる。

なお、上述のように、遮蔽学習データは完全学習データから作成するが、次の何れの方法によることも可能である。

すなわち、第1の方法では、予め全学習データから対応する全ての遮蔽学習データを作成しておき、両学習データを使用してMPL16の学習を行う。

一方、第2の方法では、完全学習データにおける個々の学習データを使用してMPL16の学習を行う際に、対応する個々の個別学習データを作成しながらMPL16の学習を行う。

【0088】

図18は、隠れを有する学習データを使用したMPL16の隠れ学習処理を表したフローチャートである。

この隠れ学習処理については、上記第2の方法によるMPL16の学習方法である。

なお、本実施形態としては、MPL16の学習を、歩行追跡装置1のCPU2が行う場合について説明する。但し、取得したフレーム画像から歩行者9の歩行状態  $s$  を推定する場合、歩行追跡装置1は学習済みのMPL16を使用することになるので、歩行追跡装置1以外のコンピュータシステムによりMPL16の学習を行うことも可能である。

【0089】

歩行追跡装置1のCPU2は、予め用意した様々な歩行状態の画像による多数の完全学習データをRAM4に記憶しておく。RAM4に保存する完全学習データは、第1実施形態でも説明したように、完全骨格情報  $j$  のジョイント座標値  $j = [x_0, y_0, \dots, x_{24}, y_{24}]$  と、教師信号である歩行状態  $s = [d_1, d_2, p_1, p_2]$  で構成される。

そして、CPU2は、RAM4から1つの完全学習データ( $j, s$ )を取得する(ステップ200)。

【0090】

次にCPU2は、遮蔽学習データ( $c_j, s$ )を作成する(ステップ210)。

すなわち、CPU2は、取得した完全学習データ( $j, s$ )の完全骨格情報  $j$  のジョイント座標値  $j = [x_0, y_0, \dots, x_{24}, y_{24}]$  から、図12で説明したように

10

20

30

40

50

、骨格の末端側の座標値を一定の割合で削除することで、遮蔽骨格情報  $c_j$  を作成する。

なお、遮蔽骨格情報  $c_j$  において、削除したジョイントの座標値は  $(x_n, y_n)$  は存在しない。このため、MLP16の層入力(図7参照)における50個のノードのうち、削除した1ジョイント当たり対応する2か所のノードの入力はされない。

そしてCPU2は、作成した遮蔽骨格情報  $c_j$  と、完全学習データの歩行情報  $s$  とから写生学習データ  $(c_j, s)$  を作成しRAM4に記憶する。

CPU2は、1の完全学習データ  $(j, s)$  に対して作成する遮蔽学習データ  $(c_j, s)$  は1つではなく、異なる割合いでジョイントを削除した複数の遮蔽データ  $(c_{j1}, s)$ 、 $(c_{j2}, s)$ 、…、を作成する。

【0091】

削除するジョイントの一例として、骨格情報15のうち、右手、左手、右足、左足の少なくとも1つの先端側から身体の中心線に沿った主要ジョイントに向けて任意の範囲の骨格情報(ジョイントの値)を欠如させた遮蔽学習データ  $(c_j, s)$  を複数作成する。

この遮蔽学習データ  $(c_j, s)$  のパターンとしては、手足の4つの組み合わせのパターン(13通り)×それぞれの欠如範囲を変えたパターン(任意通り)の組み合わせが可能である。

この場合にも、上述したように、主要ジョイントである、鼻  $(j_0)$ 、首  $(j_1)$ 、腰部中心  $(j_8)$  と、この主要ジョイントに直結している1次ジョイント  $j_2$ 、 $j_5$ 、 $j_9$ 、 $j_{12}$ 、 $j_{15}$ 、 $j_{16}$  のうちの少なくとも2つの1次ジョイントとを残すことで、隠れ領域の全面積80%以下の条件を満たす必要がある。

【0092】

次にCPU2は、完全学習データと遮蔽学習データで歩行状態の学習を行う(ステップ220)。

すなわち、CPU2は、取得した完全学習データ  $(j, s)$  と、作成した複数の遮蔽学習データ  $(c_j, s)$  を使用し、MLP16に対する学習を行う。

【0093】

次にCPU2は、未学習である他の完全学習データがRAM4に存在するか否かを判断し(ステップ230)、存在する場合には(ステップ230; Y)、ステップ200に戻って、次の完全学習データに対する処理を行う。

一方、CPU2は、他の完全学習データが存在しなければ(ステップ230; N)、隠れ学習処理を終了する。

【0094】

本実施形態では、全身の骨格情報15(完全学習データ)に加えて、このようなランダムな位置に隠れ(未検出ジョイント)を発生させた骨格情報15の学習データ(遮蔽学習データ)をMLP16に学習させた。

このように骨格の先端領域から中心領域に向けて骨格情報が欠如するように、全身の骨格情報の少なくとも一部の領域から構成された学習データを用いて学習したMLP16を用いて実験したところ、全身の骨格情報15がある場合に加えて、隠れがある場合でも歩行状態を高い精度で予測することができた。

実験結果は、第3実施形態において、第1～第3実施形態での実験結果を比較する形で説明する。

【0095】

[第3実施形態]

本実施形態では、骨格情報15を上半身・下半身などの部分領域に分割して、それぞれの部分領域を個別に学習したMLP16を用意し、これらによる結果を統合して歩行状態を推定する。

【0096】

図13は、部分領域を説明するための図である。

本実施形態では、図の破線で示したように、骨格情報15を、全身(R0)、右半身(R1)、左半身(R2)、上半身(R3)、下半身(R4)の5領域に分類し、領域R0

10

20

30

40

50

～領域 R 4 の各領域について隠れありの学習を行った。以下、領域 R 1 の骨格情報 1 5 を骨格情報 ( R 1 ) などと記すことにする。

なお、この領域分割例は一例であって、更に多数の領域、あるいは少数の領域に分割してもよい。

#### 【 0 0 9 7 】

図 1 4 は、本実施形態に係る M L P 1 6 の全体構成を説明するための図である。

本実施形態では M L P を 6 個用いるが、これらを M L P ( R 0 ) 、 M L P ( R 1 ) 、 M L P ( R 2 ) 、 M L P ( R 3 ) 、 M L P ( R 4 ) 、 M L P ( 重み ) とする。

第 2 実施形態では、全身の骨格情報 1 5 や、ジョイントの一部が欠如し、未検出となった骨格情報 1 5 を多数用いて M L P 1 6 を学習させたが、本実施形態でも同じく、全身や隠れのある骨格情報 1 5 を用いて学習を行った。

#### 【 0 0 9 8 】

M L P ( R 0 ) は、骨格情報 ( R 0 ) を用いて学習しており、骨格情報 1 5 を入力すると、領域 R 0 に含まれるジョイントの座標値を用いて歩行状態を出力する。

同様に、M L P ( R 1 ) ～ M L P ( R 4 ) は、それぞれ骨格情報 ( R 1 ) ～ 骨格情報 ( R 4 ) を用いて学習しており、骨格情報 1 5 を入力すると、それぞれ、領域 R 1 ～ R 4 に含まれるジョイントの座標値を用いて領域別歩行状態を出力する。

このように、本実施形態では、隠れを含めた骨格情報 1 5 で領域ごとに歩行状態推定の学習を行った複数の M L P を並列に配置して使用する。

#### 【 0 0 9 9 】

このように、領域ごとに学習した M L P を用いると、隠れの発生した骨格情報 1 5 に対して、隠れの発生していない領域での推定が得意な M L P が存在するため、隠れに対する頑健性が向上する。

例えば、主に右半身のジョイントが検出された骨格情報 1 5 に関しては、右半身用の M L P ( R 1 ) での推定値の精度が高く、左半身用の M L P ( R 2 ) での推定値は、精度が低いと予想される。また、下半身用の領域 R 4 など、他の領域での推定値も、M L P ( R 1 ) ほどではないものの、M L P ( R 2 ) よりは、ある程度確かであると思われる。

#### 【 0 1 0 0 】

この場合、M L P ( R 1 ) による歩行状態の重み付けを最も大きく、M L P ( R 2 ) による歩行状態の重み付けを最も小さく、他の M L P による歩行状態を、その間で適度に重み付けし、これらを合算すると、最も確からしい歩行状態を求めることができる。

M L P ( 重み ) は、骨格情報 1 5 の隠れのパターンに対して、このような M L P ( R 0 ) ～ M L P ( R 4 ) による領域別歩行状態の適切な重み付けを学習している。

#### 【 0 1 0 1 】

隠れのパターンは、検出フラグによって与えられる。検出フラグは、各ジョイントに対して、例えば、検出したものを 1 で表し、未検出のものを 0 で表したフラグ情報である。

主に右半身のジョイントが検出された骨格情報 1 5 に関しては、その領域あたりのフラグが主に 1 になるなど、ジョイントの検出／未検出のパターンを検出フラグで表すことができる。

#### 【 0 1 0 2 】

M L P ( 重み ) の学習は、骨格情報 1 5 によって M L P ( R 0 ) ～ M L P ( R 4 ) が出力した各領域別歩行状態と、骨格情報 1 5 から取得した検出フラグを M L P ( 重み ) に入力し、これによって M L P ( 重み ) が出力した歩行状態と正解の歩行状態の誤差が小さくなるように M L P ( 重み ) の各パラメータを調節することにより行った。

#### 【 0 1 0 3 】

このように調節された M L P ( 重み ) は、検出フラグの各パターンに従って、M L P ( R 0 ) ～ M L P ( R 4 ) が出力した領域別歩行状態を、検出フラグに基づいて動的に重み付けして合成し (例えば、単純に加算して)、最終的な推定結果となる歩行状態を出力する。

上述したように、ニューラルネットワーク ( M L P ( R 0 ) ～ M L P ( R 4 ) 、 M L P

10

20

30

40

50

(重み))は、骨格の領域ごとに学習した複数の領域ネットワーク(MLP(R0)~MLP(R4))と、当該領域ネットワークのそれぞれの出力を、骨格情報の欠如に応じて重み付けする重み付けネットワーク(MLP(重み))と、によって歩行状態を出力している。

【0104】

図15は、本実施形態における歩行状態計算処理を説明するためのフローチャートである。

図11と同じステップには同じ番号を付して説明を簡略化する。

まず、CPU2は、矩形領域によって検出した歩行者9から骨格情報15を抽出してRAM4に記憶する(ステップ60)。

次に、CPU2は、当該記憶した骨格情報15で、ジョイントの検出／未検出を調べて検出フラグを生成し、これをRAM4に記憶する(ステップ105)。

【0105】

次に、CPU2は、ステップ60で記憶した骨格情報15を分割して骨格情報(R0)~骨格情報(R4)を生成し、これらをRAM4に記憶する(ステップ110)。

次に、CPU2は、ステップ110でRAM4に記憶した領域別の骨格情報を、それぞれ対応する領域別のMLP(R0)~MLP(R4)に入力する(ステップ115)。

次に、CPU2は、MLP(R0)~MLP(R4)が出力した領域別歩行状態を取得してRAM4に記憶する(ステップ120)。

【0106】

次に、CPU2は、ステップ120でRAM4に記憶した領域別歩行状態と、ステップ105でRAM4に記憶した検出フラグをMLP(重み)に入力する(ステップ125)。

そしてCPU2は、MLP(重み)によって計算した歩行状態s1を取得してRAM4に記憶する(ステップ70)。

【0107】

次に、CPU2は、前のフレーム画像で取得した歩行状態s2をRAM4から読み出して取得する(ステップ75)。

そして、CPU2は、ステップ70で記憶した歩行状態s1とステップ75で読み出した歩行状態s2を用いて歩行状態による確率を計算し、これをRAM4に記憶する(ステップ80)。

以上のように、第3実施形態では、各領域の学習モデルにより歩行状態を推定し、各ジョイントの検出／未検出により、動的に各領域の推定結果に重み付けを付与して歩行状態を算出することができる。

【0108】

図16は、第1実施形態、第2実施形態、及び第3実施形態による学習モデルの精度を比較した実験結果を説明するための図である。

第1実施形態、第2実施形態、第3実施形態で行った処理形態を、それぞれ、隠れ未学習モデル、隠れ学習モデル、領域分割モデルと呼ぶことにする。

実験では、テストデータの未検出ジョイントの割合rを変化させることにより、隠れに対する頑健性を評価した。

【0109】

各モデルの評価は、図16(a)のグラフのようになった。

縦軸はRMSEを表し、横軸は未検出ジョイントの割合rを示している。ジョイントの未検出(隠れ)割合は、最大80%とし、20%間隔で測定を行った。RMSEが小さいほど精度が高いことを示している。具体的な数値は、図16(b)の表に示した通りである。

【0110】

グラフ中の実線41、42、43は、それぞれ、隠れ未学習モデル、隠れ学習モデル、領域分割モデルを表している。

グラフから分かるように、隠れがない場合(r=0)は、どのモデルも高い精度を示し

10

20

30

40

50

ており、隠れ未学習モデルが他のモデルよりも若干よい精度を示している。

しかし、隠れ未学習モデルは、隠れが生じた場合 ( $0 < r \leq 0.8$ )、精度が大幅に低下し、 $r = 0.8$  では、RMSE が 1.0 近くまで達している。

【0111】

一方、隠れ学習モデルと領域分割モデルは、隠れが生じた場合でも精度の低下が軽減されており、隠れを含む学習の効果が現れている。

特に、 $r = 0.8$  においては、領域分割モデルは、隠れ学習モデルよりも 10% 近く高い精度を示した。

【0112】

表に示したように、 $0 \leq r \leq 0.8$  で総合的に評価した場合、領域分割モデル (RMSE = 0.316)、隠れ学習モデル (RMSE = 0.353)、隠れ未学習モデル (RMSE = 0.794) の順で評価が高かった。

このように、隠れ学習モデルと領域分割モデルは、隠れに対して頑健な状態推定を行うことができ、領域分割モデルは特に効果が高いことがわかった。

【0113】

[第4実施形態]

本実施形態では、複数のカメラの映像に写った歩行者 9 をカメラ間で対応づけて追跡する場合について説明する。

この例では、例えば、イベント会場などに設置した複数台のカメラ 8、8、・・・の動画データを通信ネットワーク 11 経由で歩行追跡装置 1 に送信する。

このように、歩行追跡装置 1 の有する動画取得手段は、複数のカメラで撮影した被写体の複数の動画を取得する。

そして、これらカメラ 8、8、・・・の映像の重なったエリアにおいて、複数のカメラによる同一時刻のフレーム画像で歩行者 9 を同定する。

これにより、複数のカメラ映像を連携させて歩行者 9 を追跡することができる。

【0114】

あるいは、カメラ 8 をステレオカメラとして自律走行ロボットに搭載することもできる。

この場合、ロボットは、右フレーム画像と左フレーム画像で歩行者 9 を対応させて歩行者 9 を立体視しながら、これによって得られる 3 次元情報を用いて歩行者 9 を追跡する。

【0115】

このように複数台のカメラ 8、8、・・・を用いる場合、カメラの取付位置などの影響によって歩行者 9 の見え方が各映像で異なる。外見情報は、これによる影響を受けやすいため、歩行状態を付加情報として用いることは、誤った対応付けを抑止するために特に有効である。

【0116】

また、一般の複数カメラによる追跡では、歩行者の座標値を歩行者の特定に用いるため、各カメラの相対位置や姿勢を正確に設定するキャリブレーションが必要であるが、歩行追跡装置 1 は、歩行者 9 に固有の歩行状態を手がかりとして同定を行うため、正確なキャリブレーションは必ずしも必要ない。

そのため、固定カメラの設置が容易であるほか、複数のドローンに搭載したカメラを配置、移動させながら歩行者 9 を追跡することも可能である。

【0117】

本実施形態では、2 台のカメラ 8 a、8 b で、それぞれ同時刻に撮影したフレーム画像 a での歩行者 9 と、フレーム画像 b での歩行者 9 を対応づける場合について説明する。

ここで、カメラ 8 a とカメラ 8 b は、異なる位置から歩行者 9 を撮影するため、カメラ 8 a、8 b から見た歩行方向の角度  $\theta$  が異なる。そのため、例えば、カメラ 8 b による D 平面の座標軸がカメラ 8 a による座標軸に対して  $\Delta \theta$  の角度を成すとき、カメラ 8 b による歩行方向から  $\Delta \theta$  を減算して歩行状態を補正するなどして、両カメラ 8 の歩行方向を揃えて比較する。

なお、歩行段階の位相角  $\phi$  は、両カメラ 8 に共通である。また、歩行状態を歩行段階か

10

20

30

40

50

ら成る2次元ベクトルで定義し、 $\theta$ によらない歩行状態を定義することもできる。

#### 【0118】

図17は、歩行状態による対応付け手順を説明するためのフローチャートである。

図11と同じステップには、a、bを付して処理対象となるフレーム画像のa、bを区別しながら説明を簡略化する。

まず、CPU2は、フレーム画像aから骨格情報15を抽出して（ステップ60a）、MLP16に入力し（ステップ65a）、MLP16が出力した歩行状態s1を取得してRAM4に記憶する（ステップ70a）。

#### 【0119】

次に、CPU2は、フレーム画像bから骨格情報15を抽出して（ステップ60b）、MLP16に入力し（ステップ65b）、MLP16が出力した歩行状態s2を取得してRAM4に記憶する（ステップ70b）。

#### 【0120】

次に、CPU2は、ステップ70aでRAM4に記憶した歩行状態s1と、ステップ70bでRAM4に記憶した歩行状態s2を用いて式(3)（図8の式(3)）による歩行状態による確率を計算し、これをRAM4に出力する（ステップ80）。

#### 【0121】

また、CPU2は、外見情報についても、フレーム画像a、bについて外見情報による確率を計算し、式(1)に従って、同一人物である確率を計算する。カメラのキャリブレーションが行われている場合は、外見情報と歩行状態に加えて、歩行者9の座標値を用いて同定することもできる。

このように、歩行追跡装置1が有する同定手段は、複数の動画のフレーム画像間で被写体を対応させることにより被写体を同定している。

#### 【0122】

以上、異なるカメラ8の映像間で歩行者9を同定する場合について説明したが、本実施形態に、第1実施形態～第3実施形態の技術を組み合わせることも可能である。

この場合、カメラ8aによる時間的に隣接するフレーム画像a1、a2で歩行者9を同定し、カメラ8bによる時間的に隣接するフレーム画像b1、b2で歩行者9を同定すると共に、更に、同一時刻のフレーム画像a1、b1間（あるいは、フレーム画像a2、b2間、撮影時刻が近ければフレーム画像a1、b2などの組み合わせも可能）で歩行者9を同定して対応づける。

ここで、2台のカメラ8a、8bで歩行者9を同定する場合について説明したが、3台以上のカメラ8で撮影する場合も同様にして歩行者9の同定を行うことができる。

#### 【0123】

以上、各実施形態について説明したが、これらは、歩行を検出するという要素技術であるため、様々な技術的、及びビジネス的な展開が考えられる。

歩行追跡装置1を自律走行ロボットに搭載する場合、例えば、ホテルで顧客を追跡しながら顧客の荷物を運んだり、イベント会場の入り口で入場者を追跡して検温したり、あるいは、工場で資材を積載した状態で担当者の後についていたり、といった用途に用いることができる。

#### 【0124】

監視カメラ映像を用いる場合、例えば、店内のカメラと商店街に設置したカメラを連携させて、店に忘れ物をした顧客を追跡して担当者が忘れ物を届けたり、迷子の子供を追跡して保護したりなど、色々な用途に用いることができる。

また、歩行追跡装置1は、未来に向かって歩行者9を追跡するほか、撮影した動画を逆に辿ることにより（即ち、動画を逆に再生して）、歩行者9が歩行してきた経路を過去に向かって追跡することもできる。

これにより、ある店で商品を購入した顧客の歩行経路を過去に向かって追跡し、どのような動線を経て当該店にやってきたかを調べる、といったマーケティング分析に用いることもできる。

## 【0125】

更に、歩行は歩行者の健康状態を反映するため、医療分野への展開が考えられる。

例えば、リハビリで歩行訓練している患者の歩行を撮影して歩行状態の時間的推移を計測し、歩行方向や歩行周期のばらつきなどを統計学的手法を用いて解析して訓練の進捗状況を把握する、といった歩行診断装置としての利用方法が考えられる。

## 【0126】

この場合、歩行追跡装置1は、記憶部10に記憶した歩行状態の時間的変化を統計的な処理を施すなどして表示部5で提示する。

このように、歩行追跡装置1は、ニューラルネットワークが出力した歩行状態の時間変化を提示する提示手段を備えている。

## 【0127】

また、健康診断の際に高齢者に歩行してもらって歩行状態を保存しておき、その経年変化を観察することにより、老化の進行具合を観察したり、病気の早期発見を行ったりすることも考えられる。

更に、スマートフォンなどを用いて患者に自宅で歩行するところを動画撮影してもらい、これを病院に設置した歩行追跡装置1に送信して、病院で歩行状態を解析するといったような、遠隔診断も可能である。

## 【0128】

また、発明者の実験によると、歩行追跡装置1は、歩行状態を継続的に観察するため、歩行の異常を検知する能力に優れている。

そこで、歩行追跡装置1は、異常歩行検出装置として用いることが可能である。

例えば、ホテルのロビーで異常歩行を検出した場合、顧客が重い荷物を運んでいる場合や気分が悪くなった場合などが考えられ、これを異常歩行検出装置で検出して早期の支援を行うことができる。

あるいは、重い荷物を運んでいる顧客の歩行状態を学習させておき、事務所に設置した異常歩行検出装置でこれを検出した場合は、歩行追跡装置1を搭載した自律走行ロボットが顧客の場所に急行して荷物を受け取り、顧客の歩行を追跡しながら顧客の荷物を搬送する、といった運用も考えられる。

## 【0129】

更に、歩行追跡装置1は、歩行状態を付加情報として用いるため、同じ外見をした集団で歩行者9を追跡する場合に特に効果を発揮する。

これは、例えば、同じ制服を着用した高校生などの集団が団体旅行している場合に適用できる。

また、同じ外見の集団が同一の歩行状態である場合は行進を行っている、とか、人々の歩行状態が一斉に変化して一方向に急激に移動を始めた場合は、何らかの緊急の事象が起こって人々が避難行動を始めた、などと歩行が行われているシーンを推論することも可能である。

## 【0130】

更に、骨格情報によって姿勢を把握できればよいので、歩行追跡装置1を動物に適用することも可能である。特に、群れで暮らす動物は外見が同じため、歩行追跡装置1の歩行状態による追跡が効果的である。

例えば、近年、農業分野や牧畜分野のIT化が急速に進んでおり、牧羊の数理解析モデルの研究に基づく牧羊ロボットの開発が行われている。

## 【0131】

羊は外見が同じなので、外見で羊の個体を識別するのは困難であるが、歩行状態はばらばらであるので歩行追跡装置1で認識・追跡することができる。

これにより、例えば、複数の羊オーナーから集めた羊の群れを牧畜しているところを複数のドローンで撮影し、オーナーごとに個別の羊を追跡して動画配信したり、あるいは、各羊を追跡しながら画像解析することにより食べた牧草の量を羊ごとに推定し、その代金をオーナーに課金する、といった新たなビジネスを生む可能性を秘めている。

10

20

30

40

50

## 【0132】

また、原理的には節足動物に対しても適用可能であり、蟻のような集団で行動する昆虫や蟹のような甲殻類に適用して、学術的な研究に利用することも考えられる。

これらは一例であって、歩行状態を用いた観察対象の同定は要素的な技術であるため、裾野の広い応用分野を有している。

## 【0133】

なお、以上に説明した第1実施形態～第4実施形態により、歩行追跡装置、歩行追跡プログラムを次の様に構成することも可能である。

(1) 構成1では、歩行している被写体を撮影した動画を取得する動画取得手段と、前記被写体の周期的な歩行動作に基づく歩行状態を用いて、前記取得した動画を構成するフレーム画像間での前記被写体を同定する同定手段と、前記取得した動画で前記同定した被写体を追跡する追跡手段と、を具備したことを特徴とする歩行追跡装置を提供する。

10

(2) 構成2では、前記同定手段が、前記歩行動作に加えて、前記被写体の歩行する方向を前記歩行状態に含めて前記同定を行うことを特徴とする構成1に記載の歩行追跡装置を提供する。

(3) 構成3では、前記同定手段が、前記歩行動作の連続性に基づく歩行状態を用いることを特徴とする構成1、又は構成2に記載の歩行追跡装置を提供する。

(4) 構成4では、前記被写体の外見情報を取得する外見情報取得手段を具備し、前記同定手段は、前記歩行状態と、前記取得した外見情報と、を用いて前記被写体を同定することを特徴とする構成1、構成2、又は構成3に記載の歩行追跡装置を提供する。

20

(5) 構成5では、前記被写体の骨格情報を取得する骨格情報取得手段と、各種の歩行状態を骨格情報によって学習したニューラルネットワークと、を具備し、前記同定手段は、前記取得した骨格情報を前記ニューラルネットワークに入力して、当該ニューラルネットワークが出力する歩行状態を取得する、ことを特徴とする構成1から構成4までのうちの何れか1の構成に記載の歩行追跡装置を提供する。

(6) 構成6では、前記ニューラルネットワークが、全身の骨格情報の少なくとも一部の領域から構成された学習データを用いて学習した、ことを特徴とする構成5に記載の歩行追跡装置を提供する。

(7) 構成7では、前記ニューラルネットワークが、骨格の先端領域から中心領域に向けて骨格情報が欠如するように構成された学習データで学習した、ことを特徴とする構成6に記載の歩行追跡装置を提供する。

30

(8) 構成8では、前記ニューラルネットワークが、骨格の領域ごとに学習した複数の領域ネットワークと、当該領域ネットワークのそれぞれの出力を、骨格情報の欠如に応じて重み付けする重み付けネットワークと、によって前記歩行状態を出力する、ことを特徴とする構成6、又は構成7に記載の歩行追跡装置を提供する。

(9) 構成9では、前記動画取得手段が、複数の前記カメラで撮影した前記被写体の複数の動画を取得し、前記同定手段は、前記取得した複数の動画のフレーム画像間で前記被写体を対応させることにより前記被写体を同定する、ことを特徴とする構成1から構成8までのうちの何れか1の構成に記載の歩行追跡装置を提供する。

(10) 構成10では、前記ニューラルネットワークが出力した歩行状態の時間変化を提示する提示手段を具備したことを特徴とする構成5から構成9までのうちの何れか1の構成に記載の歩行追跡装置を提供する。

40

(11) 構成11では、歩行している被写体を撮影した動画を取得する動画取得機能と、前記被写体の周期的な歩行動作に基づく歩行状態を用いて、前記取得した動画を構成するフレーム画像間での前記被写体を同定する同定機能と、前記取得した動画で前記同定した被写体を追跡する追跡機能と、をコンピュータで実現する歩行追跡プログラムを提供する。

## 【0134】

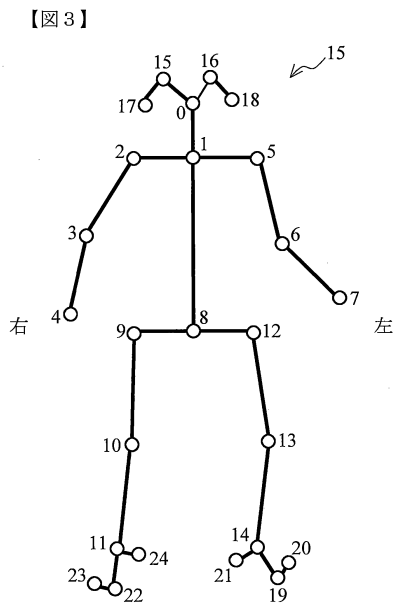
以上の各構成によれば、歩行状態を用いて対象を同定することにより、歩行者追跡の頑健性を高めることができる。

## 【符号の説明】

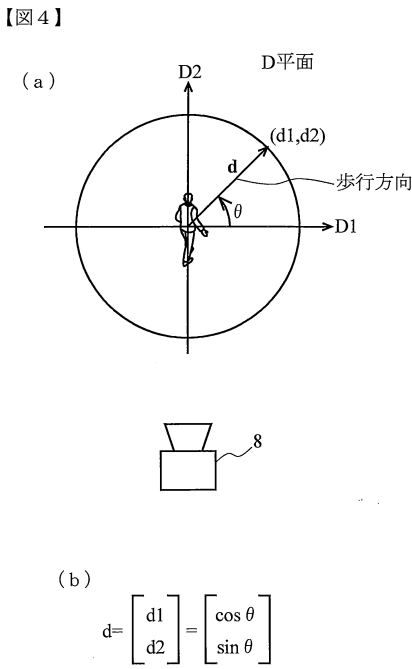
50



【図 3】



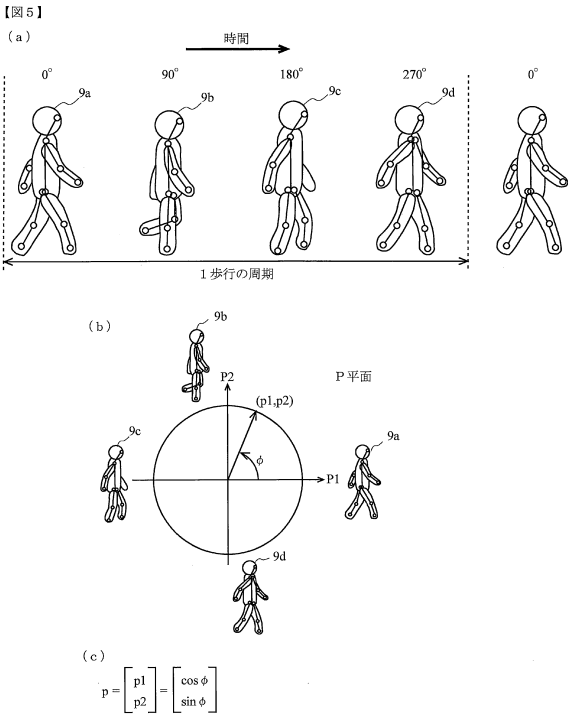
【図 4】



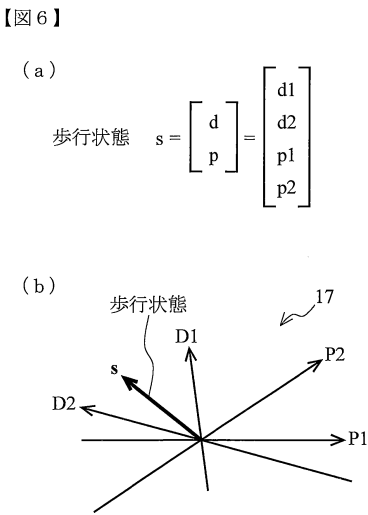
10

20

【図 5】



【図 6】

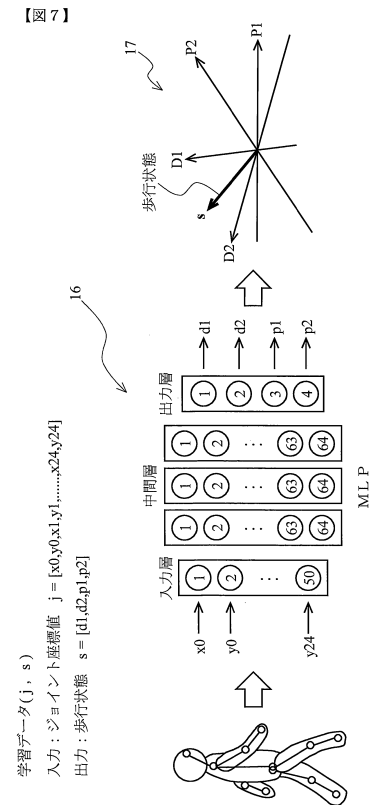


30

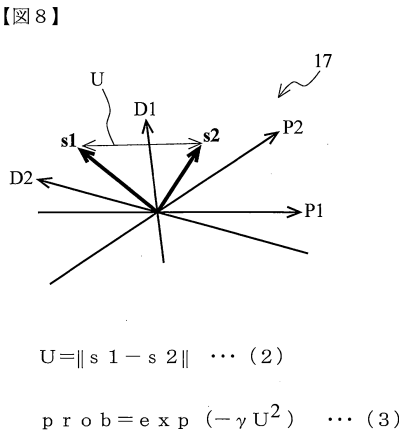
40

50

【図 7】



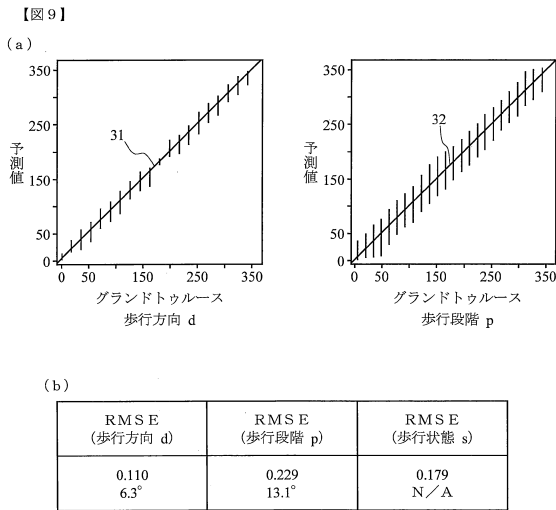
【図 8】



10

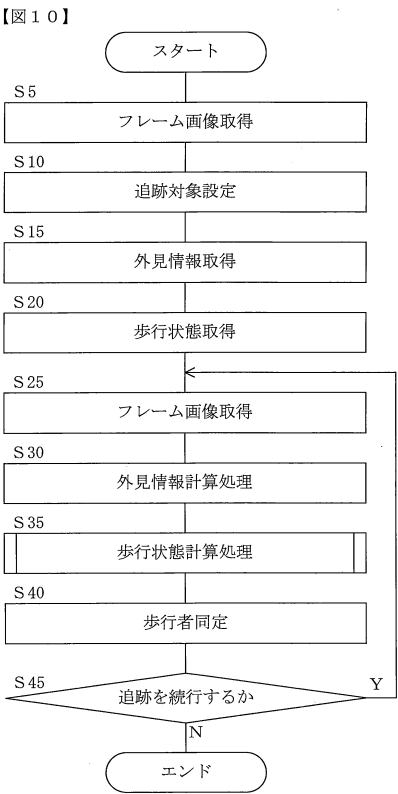
20

【図 9】



$$\text{RMSE} = \sqrt{\sum_i (y_{\text{obs},i} - y_{\text{pred},i})^2 / n} \dots (4)$$

【図 10】



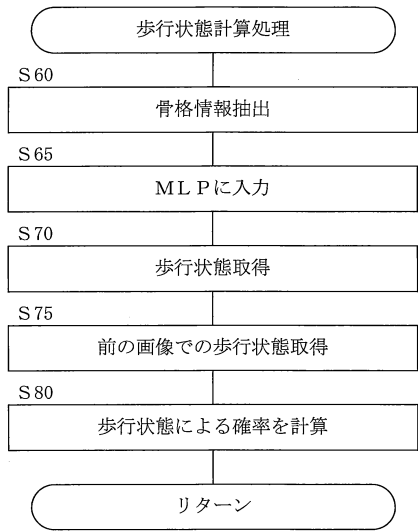
30

40

50

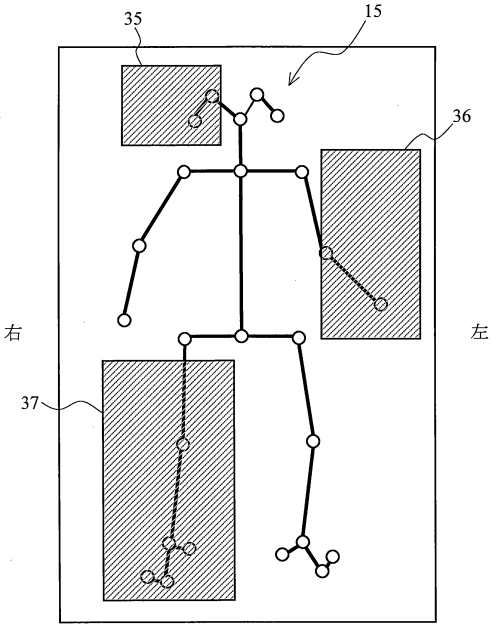
【図 1 1】

【図 1 1】



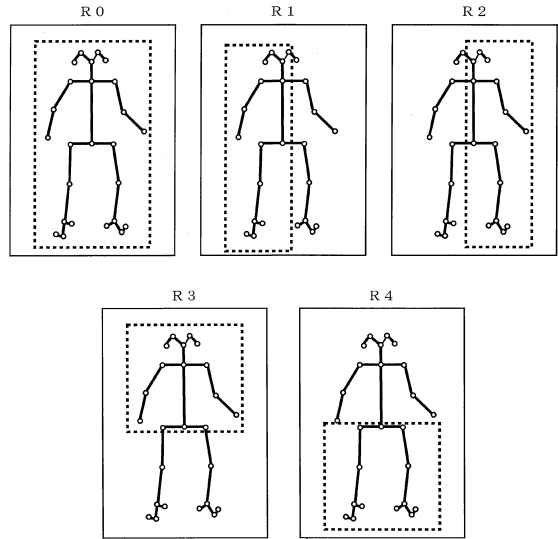
【図 1 2】

【図 1 2】



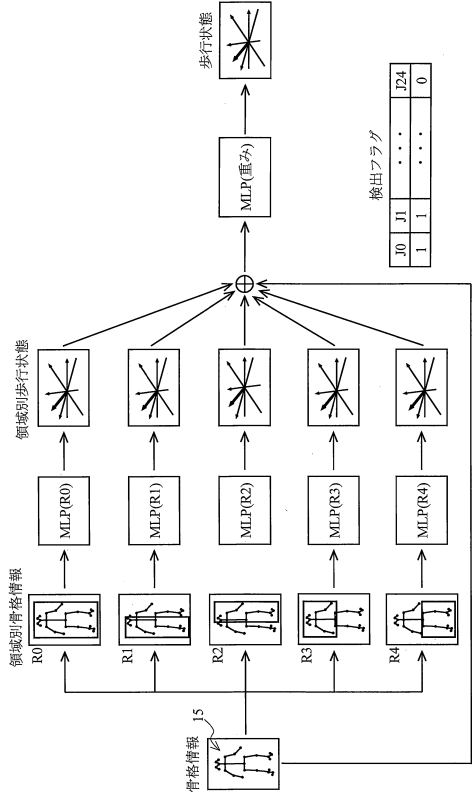
【図 1 3】

【図 1 3】



【図 1 4】

【図 1 4】



10

20

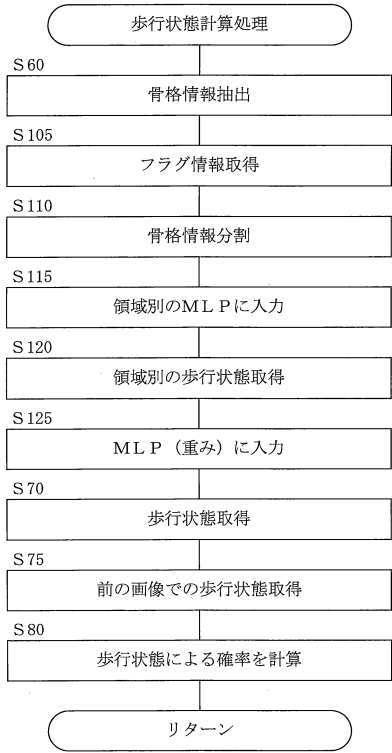
30

40

50

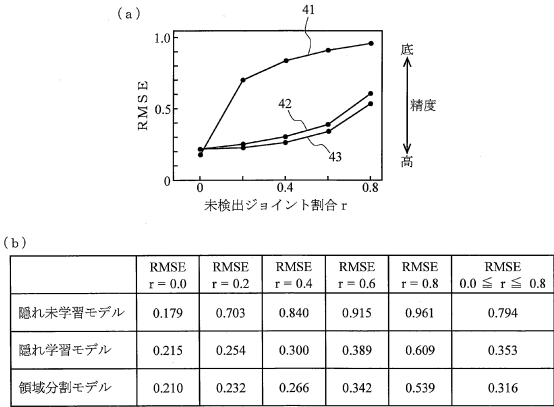
【図 1 5】

【図 1 5】



【図 1 6】

【図 1 6】

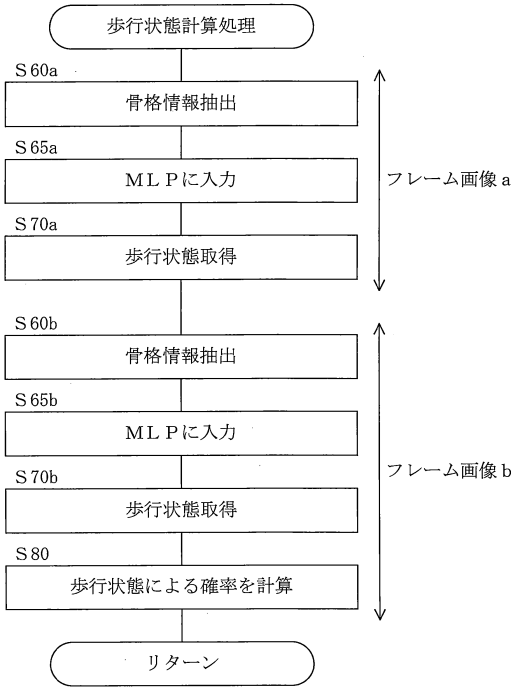


10

20

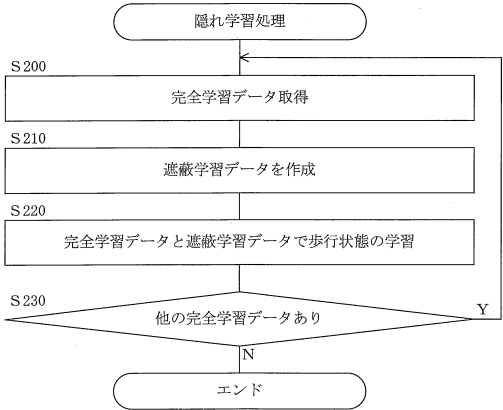
【図 1 7】

【図 1 7】



【図 1 8】

【図 1 8】



30

40

50

## フロントページの続き

ス・リサーチ内

(72)発明者 柴田 雅聡

東京都千代田区外神田1丁目18番13号 秋葉原ダイビル7階 株式会社エクス・リサーチ内

(72)発明者 榎田 修一

福岡県飯塚市川津680-4 国立大学法人九州工業大学 情報工学研究院 知能情報工学科内

(72)発明者 今里 元樹

福岡県飯塚市川津680-4 国立大学法人九州工業大学 情報工学研究院 知能情報工学科内

審査官 伊知地 和之

(56)参考文献 特開2019-025134 (JP, A)

特開2019-012304 (JP, A)

特開2020-123105 (JP, A)

中国特許出願公開第109766868 (CN, A)

中国特許出願公開第104463118 (CN, A)

岩下友美 外1名, 人影に着目した個人認証手法の提案と開発, 画像ラボ, 日本工業出版株式会社, 2018年03月10日, 第29巻 第3号, pp.9~13

(58)調査した分野 (Int.Cl., DB名)

A61B 5/06 - 5/22

G06F 18/00 - 18/40

G06N 3/00 - 99/00

G06T 7/00 - 7/90

G06V 10/00 - 20/90

G06V 30/418

G06V 40/16

G06V 40/20

H04N 7/18

CSDB (日本国特許庁)