

明 細 書

発明の名称：画像処理装置、画像処理方法、及びプログラム

技術分野

[0001] 本発明は、画像処理装置、画像処理方法、及びプログラムに関する。

背景技術

[0002] 近年、監視システム等において、監視カメラの画像から人物の姿勢や行動等の状態の検出や検索を行う技術が利用されている。関連する技術として、例えば、特許文献1及び2が知られている。特許文献1には、深さ映像に含まれる人物の頭や手足等のキージョイントに基づいて、類似する人物の姿勢を検索する技術が開示されている。特許文献2には、人物の姿勢と関連しないが、画像に付加された傾き等の姿勢情報を利用して類似画像を検索する技術が開示されている。特許文献3には、画像上で検索したい人物を指定する入力を受け、指定された人物を含む画像を検索する技術が開示されている。特許文献4には、画像から複数の特徴点で構成される検索対象の姿勢情報を取得し、その姿勢情報で特定される姿勢と類似した姿勢を含む画像を検索する技術が開示されている。なお、その他に、人物の骨格推定に関連する技術として、非特許文献1が知られている。

先行技術文献

特許文献

[0003] 特許文献1：特表2014-522035号公報

特許文献2：特開2006-260405号公報

特許文献3：特開2020-13290号公報

特許文献4：特開2019-91138号公報

非特許文献

[0004] 非特許文献1：Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 201

7, P. 7291-7299

発明の概要

発明が解決しようとする課題

[0005] 特許文献 1 及び 4 は、所定の状態の人物を含む画像を検索する技術であるが、検索クエリの入力方法に改善の余地があった。特許文献 2 及び 3 は、所定の状態の人物を含む画像を検索する技術でない。

[0006] 本発明の目的は、所定の状態の人物を含む画像を検索するシステムにおいて、ユーザフレンドリーな検索クエリの入力手段を実現することである。

課題を解決するための手段

[0007] 本発明によれば、
画像を取得する画像取得手段と、
ユーザ入力に基づき、前記画像に含まれる少なくとも 1 人の人物を選択する選択手段と、
前記画像に含まれる人物の 2 次元骨格構造を検出する骨格構造検出手段と、
、
前記検出された 2 次元骨格構造の特徴量を算出する特徴量算出手段と、
前記 2 次元骨格構造の特徴量の類似度に基づき、解析対象画像の中から、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する検索手段と、
を備える画像処理装置が提供される。

[0008] また、本発明によれば、
コンピュータが、
画像を取得し、
ユーザ入力に基づき、前記画像に含まれる少なくとも 1 人の人物を選択し、
前記画像に含まれる人物の 2 次元骨格構造を検出し、
前記検出された 2 次元骨格構造の特徴量を算出し、
前記 2 次元骨格構造の特徴量の類似度に基づき、解析対象画像の中から

、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する画像処理方法が提供される。

- [0009] また、本発明によれば、
- コンピュータを、
- 画像を取得する画像取得手段、
- ユーザ入力に基づき、前記画像に含まれる少なくとも 1 人の人物を選択する選択手段、
- 前記画像に含まれる人物の 2 次元骨格構造を検出する骨格構造検出手段、
- 、
- 前記検出された 2 次元骨格構造の特徴量を算出する特徴量算出手段、
- 前記 2 次元骨格構造の特徴量の類似度に基づき、解析対象画像の中から、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する検索手段、
- として機能させるプログラムが提供される。

発明の効果

- [0010] 本発明によれば、所定の状態の人物を含む画像を検索するシステムにおいて、ユーザフレンドリーな検索クエリの入力手段が実現される。

図面の簡単な説明

- [0011] [図1]実施の形態に係る画像処理装置の概要を示す構成図である。
- [図2]実施の形態 1 に係る画像処理装置の構成を示す構成図である。
- [図3]実施の形態 1 に係る画像処理方法を示すフローチャートである。
- [図4]実施の形態 1 に係る分類方法を示すフローチャートである。
- [図5]実施の形態 1 に係る検索方法を示すフローチャートである。
- [図6]実施の形態 1 に係る骨格構造の検出例を示す図である。
- [図7]実施の形態 1 に係る人体モデルを示す図である。
- [図8]実施の形態 1 に係る骨格構造の検出例を示す図である。
- [図9]実施の形態 1 に係る骨格構造の検出例を示す図である。
- [図10]実施の形態 1 に係る骨格構造の検出例を示す図である。

- [図11]実施の形態1に係る分類方法の具体例を示すグラフである。
- [図12]実施の形態1に係る分類結果の表示例を示す図である。
- [図13]実施の形態1に係る検索方法を説明するための図である。
- [図14]実施の形態1に係る検索方法を説明するための図である。
- [図15]実施の形態1に係る検索方法を説明するための図である。
- [図16]実施の形態1に係る検索方法を説明するための図である。
- [図17]実施の形態1に係る検索結果の表示例を示す図である。
- [図18]実施の形態2に係る画像処理装置の構成を示す構成図である。
- [図19]実施の形態2に係る画像処理方法を示すフローチャートである。
- [図20]実施の形態2に係る身長画素数算出方法の具体例1を示すフローチャートである。
- [図21]実施の形態2に係る身長画素数算出方法の具体例2を示すフローチャートである。
- [図22]実施の形態2に係る身長画素数算出方法の具体例2を示すフローチャートである。
- [図23]実施の形態2に係る正規化方法を示すフローチャートである。
- [図24]実施の形態2に係る人体モデルを示す図である。
- [図25]実施の形態2に係る骨格構造の検出例を示す図である。
- [図26]実施の形態2に係る骨格構造の検出例を示す図である。
- [図27]実施の形態2に係る骨格構造の検出例を示す図である。
- [図28]実施の形態2に係る人体モデルを示す図である。
- [図29]実施の形態2に係る骨格構造の検出例を示す図である。
- [図30]実施の形態2に係る身長画素数算出方法を説明するためのヒストグラムである。
- [図31]実施の形態2に係る骨格構造の検出例を示す図である。
- [図32]実施の形態2に係る3次元人体モデルを示す図である。
- [図33]実施の形態2に係る身長画素数算出方法を説明するための図である。
- [図34]実施の形態2に係る身長画素数算出方法を説明するための図である。

[図35]実施の形態2に係る身長画素数算出方法を説明するための図である。

[図36]実施の形態2に係る正規化方法を説明するための図である。

[図37]実施の形態2に係る正規化方法を説明するための図である。

[図38]実施の形態2に係る正規化方法を説明するための図である。

[図39]画像処理装置が処理する情報の一例を模式的に示す図である。

[図40]画像処理装置が処理する情報の一例を模式的に示す図である。

[図41]画像処理装置のハードウェア構成例を示す図である。

[図42]画像処理装置が出力する画面の一例を模式的に示す図である。

[図43]画像処理装置が出力する画面の一例を模式的に示す図である。

[図44]画像処理装置が出力する画面の一例を模式的に示す図である。

[図45]画像処理装置が出力する画面の一例を模式的に示す図である。

[図46]画像処理装置が出力する画面の一例を模式的に示す図である。

[図47]画像処理装置が出力する画面の一例を模式的に示す図である。

[図48]画像処理装置が出力する画面の一例を模式的に示す図である。

[図49]画像処理装置の処理の流れの一例を示すフローチャートである。

[図50]画像処理装置の処理の流れの一例を示すフローチャートである。

[図51]画像処理装置が出力する画面の一例を模式的に示す図である。

[図52]画像処理装置が出力する画面の一例を模式的に示す図である。

発明を実施するための形態

[0012] 以下、本発明の実施の形態について、図面を用いて説明する。尚、すべての図面において、同様な構成要素には同様の符号を付し、適宜説明を省略する。

[0013] (実施の形態に至る検討)

近年、ディープラーニング等の機械学習を活用した画像認識技術が様々なシステムに応用されている。例えば、監視カメラの画像により監視を行う監視システムへの適用が進められている。監視システムに機械学習を活用することで、画像から人物の姿勢や行動等の状態をある程度把握することが可能とされつつある。

[0014] しかしながら、このような関連する技術では、必ずしもオンデマンドにユーザが望む人物の状態を把握できない場合がある。例えば、ユーザが検索し把握したい人物の状態を事前に特定できている場合もあれば、未知の状態のように具体的に特定できていない場合もある。そうすると、場合によっては、ユーザが検索したい人物の状態を詳細に指定することができない。また、人物の体の一部が隠れているような場合には検索等を行うことができない。関連する技術では、特定の検索条件のみからしか人物の状態を検索できないため、所望の人物の状態を柔軟に検索や分類することが困難である。

[0015] 発明者らは、オンデマンドに画像からユーザ所望の人物の状態を認識するため、非特許文献1などの骨格推定技術を利用する方法を検討した。非特許文献1に開示されたOpenPose等のように、関連する骨格推定技術では、様々なパターンの正解付けされた画像データを学習することで、人物の骨格を推定する。以下の実施の形態では、このような骨格推定技術を活用することで、人物の状態を柔軟に認識することを可能とする。

[0016] なお、OpenPose等の骨格推定技術により推定される骨格構造は、関節等の特徴的な点である「キーポイント」と、キーポイント間のリンクを示す「ボーン（ボーンリンク）」とから構成される。このため、以下の実施の形態では、骨格構造について「キーポイント」と「ボーン」という用語を用いて説明するが、特に限定されない限り、「キーポイント」は人物の「関節」に対応し、「ボーン」は人物の「骨」に対応している。

[0017] （実施の形態の概要）

図1は、実施の形態に係る画像処理装置10の概要を示している。図1に示すように、画像処理装置10は、骨格検出部11、特徴量算出部12、及び認識部13を備えている。骨格検出部11は、カメラ等から取得される2次元画像に基づいて、複数の人物の2次元骨格構造（以下、単に「骨格構造」という場合がある）を検出する。特徴量算出部12は、骨格検出部11により検出された複数の2次元骨格構造の特徴量を算出する。認識部13は、特徴量算出部12により算出された複数の特徴量の類似度に基づいて、複数

の人物の状態の認識処理を行う。認識処理は、人物の状態の分類処理や検索処理等である。

[0018] このように、実施の形態では、２次元画像から人物の２次元骨格構造を検出し、この２次元骨格構造から算出される特徴量に基づいて人物の状態の分類や検索等の認識処理を行う。

[0019] （実施の形態１）

以下、図面を参照して実施の形態１について説明する。図２は、本実施の形態に係る画像処理装置１００の構成を示している。画像処理装置１００は、カメラ２００及び記憶手段（データベース（ＤＢ）１１０）、とともに画像処理システム１を構成する。画像処理装置１００を含む画像処理システム１は、画像から推定される人物の骨格構造に基づき、人物の姿勢や行動等の状態を分類及び検索するシステムである。

[0020] カメラ２００は、２次元の画像を生成する監視カメラ等の撮像部である。カメラ２００は、所定の箇所に設置されて、その設置箇所から撮像領域における人物等を撮像する。カメラ２００は、撮像した画像（映像）を画像処理装置１００へ出力可能に有線又は無線で直接接続、もしくは任意の通信ネットワーク等を介して接続されている。なお、カメラ２００を画像処理装置１００の内部に設けてもよい。

[0021] データベース１１０は、画像処理装置１００の処理に必要な情報（データ）や処理結果等を格納するデータベースである。データベース１１０は、画像取得部１０１が取得した画像や、骨格構造検出部１０２の検出結果、機械学習用のデータ、特徴量算出部１０３が算出した特徴量、分類部１０４の分類結果、検索部１０５の検索結果等を記憶する。データベース１１０は、画像処理装置１００と必要に応じてデータを入出力可能に有線又は無線で直接接続、もしくは任意の通信ネットワーク等を介して接続されている。なお、データベース１１０をフラッシュメモリなどの不揮発性メモリやハードディスク装置等として、画像処理装置１００の内部に設けてもよい。

[0022] 図２に示すように、画像処理装置１００は、画像取得部１０１、骨格構造

検出部１０２、特徴量算出部１０３、分類部１０４、検索部１０５、入力部１０６、表示部１０７、及び選択部１０９を備えている。なお、各部（ブロック）の構成は一例であり、後述の方法（動作）が可能であれば、その他の各部で構成されてもよい。また、画像処理装置１００は、例えば、プログラムを実行するパーソナルコンピュータやサーバ等のコンピュータ装置で実現されるが、１つの装置で実現してもよいし、ネットワーク上の複数の装置で実現してもよい。例えば、入力部１０６や表示部１０７等を外部の装置としてもよい。また、分類部１０４及び検索部１０５の両方を備えていてもよいし、いずれか一方のみを備えていてもよい。分類部１０４及び検索部１０５の両方、もしくは一方は、人物の状態の認識処理を行う認識部１３である。また、検索部１０５及び選択部１０９は、検索処理を実行する機能部であり、図１の認識部１３に該当する。

[0023] 画像処理装置１００は、データ蓄積処理と、分類処理と、検索処理とをこの順に実行する。なお、以下で説明するように、画像処理装置１００は分類処理を実行しなくてもよい。

[0024] ＜データ蓄積処理＞

データ蓄積処理は、解析対象の画像（以下、「解析対象画像」）を取得し、複数の解析対象画像各々から人物の２次元の骨格構造を検出し、検出された２次元の骨格構造の特徴量を算出し、算出した特徴量を各解析対象画像に紐付けてデータベース１１０に格納する処理である。以下、データ蓄積処理に関わる機能部の構成を説明する。

[0025] 画像取得部１０１は、解析対象画像を取得する。本明細書において、「取得」とは、ユーザ入力に基づき、又は、プログラムの指示に基づき、「自装置が他の装置や記憶媒体に格納されているデータを取りに行くこと（能動的な取得）」、たとえば、他の装置にリクエストまたは問い合わせして受信すること、他の装置や記憶媒体にアクセスして読み出すこと等、および、ユーザ入力に基づき、又は、プログラムの指示に基づき、「自装置に他の装置から出力されるデータを入力すること（受動的な取得）」、たとえば、配信（

または、送信、プッシュ通知等）されるデータを受信すること、また、受信したデータまたは情報の中から選択して取得すること、及び、「データを編集（テキスト化、データの並び替え、一部データの抽出、ファイル形式の変更等）などして新たなデータを生成し、当該新たなデータを取得すること」の少なくともいずれか一方を含む。

[0026] 例えば、画像取得部101は、所定の監視期間にカメラ200が撮像した人物を含む2次元の画像を解析対象画像として取得する。その他、画像取得部101は、データベース110等の記憶手段に記憶されている人物を含む2次元の画像を解析対象画像として取得してもよい。

[0027] 骨格構造検出部102は、取得された解析対象画像各々から人物の2次元の骨格構造を検出する。骨格構造検出部102は、解析対象画像の中で認識される全ての人物について、骨格構造を検出することができる。骨格構造検出部102は、機械学習を用いた骨格推定技術を用いて、認識される人物の関節等の特徴に基づき人物の骨格構造を検出する。骨格構造検出部102は、例えば、非特許文献1のOpenPose等の骨格推定技術を用い、関節等の特徴的な点であるキーポイントを抽出したりする。

[0028] 特徴量算出部103は、検出された2次元の骨格構造の特徴量を算出し、算出した特徴量を、その2次元の骨格構造が検出された解析対象画像に紐付けてデータベース110に格納する。骨格構造の特徴量は、人物の骨格の特徴を示しており、人物の骨格に基づいて人物の状態を分類や検索するための要素となる。通常、この特徴量は、複数のパラメータ（例えば後述する分類要素）を含んでいる。特徴量は、骨格構造の全体の特徴量でもよいし、骨格構造の一部の特徴量でもよく、骨格構造の各部のように複数の特徴量を含んでもよい。特徴量の算出方法は、機械学習や正規化等の任意の方法でよく、正規化として最小値や最大値を求めてもよい。一例として、特徴量は、骨格構造を機械学習することで得られた特徴量や、骨格構造の頭部から足部までの画像上の大きさ等である。骨格構造の大きさは、画像上の骨格構造を含む骨格領域の上下方向の高さや面積等である。上下方向（高さ方向または縦方

向)は、画像における上下の方向(Y軸方向)であり、例えば、地面(基準面)に対し垂直な方向である。また、左右方向(横方向)は、画像における左右の方向(X軸方向)であり、例えば、地面に対し平行な方向である。

[0029] なお、ユーザが望む分類や検索を行うためには、分類や検索処理に対しロバスト性を有する特徴量を用いることが好ましい。例えば、ユーザが、人物の向きや体型に依存しない分類や検索を望む場合、人物の向きや体型にロバストな特徴量を使用してもよい。同じ姿勢で様々な方向に向いている人物の骨格や同じ姿勢で様々な体型の人物の骨格を学習することや、骨格の上下方向のみの特徴を抽出することで、人物の向きや体型に依存しない特徴量を得ることができる。

[0030] <分類処理>

分類処理は、データ蓄積処理でデータベース110に格納されたデータ(解析対象画像と、各解析対象画像から検出された2次元の骨格構造の特徴量とを紐付けたデータ)に基づき、解析対象画像から検出された複数の2次元の骨格構造をその特徴量が類似するもの同士でまとめて分類(グループ分け)する処理である。なお、解析対象画像と、各解析対象画像から検出された2次元の骨格構造とは互いに紐付けられている。このため、分類処理による複数の2次元の骨格構造の分類は、複数の解析対象画像の分類にもなる。分類処理により、複数の解析対象画像は、類似する2次元の骨格構造を含むもの同士でまとめられる。以下、分類処理に関わる機能部の構成を説明する。

[0031] 分類部104は、データベース110に格納された複数の骨格構造を、骨格構造の特徴量の類似度に基づいて分類する(クラスタリングする)。分類部104は、人物の状態の認識処理として、骨格構造の特徴量に基づいて複数の人物の状態进行分类しているとも言える。類似度は、骨格構造の特徴量間の距離である。分類部104は、骨格構造の全体の特徴量の類似度により分類してもよいし、骨格構造の一部の特徴量の類似度により分類してもよく、骨格構造の第1の部分(例えば両手)及び第2の部分(例えば両足)の特徴量の類似度により分類してもよい。なお、各画像における人物の骨格構造の

特徴量に基づいて人物の姿勢を分類してもよいし、時系列に連続する複数の画像における人物の骨格構造の特徴量の変化に基づいて人物の行動を分類してもよい。すなわち、分類部104は、骨格構造の特徴量に基づいて人物の姿勢や行動を含む人物の状態を分類できる。例えば、分類部104は、所定の監視期間に撮像された複数の画像における複数の骨格構造を分類対象とする。分類部104は、分類対象の特徴量間の類似度を求め、類似度の高い骨格構造が同じクラス（似た姿勢のグループ）となるように分類する。なお、検索と同様に、分類条件をユーザが指定できるようにしてもよい。分類部104は、骨格構造の分類結果をデータベース110に格納するとともに、表示部107に表示することができる。

[0032] <検索処理>

検索処理は、データ蓄積処理でデータベース110に格納されたデータ（解析対象画像と、各解析対象画像から検出された2次元の骨格構造の特徴量とを紐付けたデータ）に基づき、解析対象画像から検出された複数の2次元の骨格構造の中から所定の骨格構造を検索する処理である。なお、解析対象画像と、各解析対象画像から検出された2次元の骨格構造とは互いに紐付けられている。このため、上記「解析対象画像から検出された複数の2次元の骨格構造の中から所定の骨格構造を検索する処理」により、所定の骨格構造を含む解析対象画像を検索することができる。以下、検索処理に関わる機能部の構成を説明する。

[0033] 画像取得部101は、クエリ画像を取得する。画像取得部101は、例えば以下の取得例のいずれかにより、クエリ画像を取得することができる。

[0034] 「取得例1」

当該例では、画像取得部101は、解析対象画像の中のいずれかをクエリ画像として取得する。例えば、画像取得部101は、上記分類処理で生成した複数のグループ各々から所定のルールで所定数の解析対象画像を選択し、選択した解析対象画像をクエリ画像としてもよい。

[0035] 「取得例2」

当該例では、画像取得部 101 は、ユーザが用意し、画像処理装置 100 に入力したクエリ画像を取得する。

[0036] 「取得例 3」

当該例では、画像取得部 101 は、ユーザが指定したキーワードで検索した画像をクエリ画像として取得する。図 42 に、当該検索で利用される UI (user interface) 画面の一例を示す。キーワードの入力欄、及び、検索結果の表示欄が示されている。当該 UI 画面の場合、検索結果として表示された画像の中からユーザが選択した画像が、クエリ画像となる。キーワードは、「座る」、「立つ」等のように、人物の状態（姿勢、行動等）に関する内容が想定される。キーワードの入力は、例えばテキストボックス、ドロップダウン、チェックボックス等の周知の GUI を利用して実現できる。

[0037] 例えば、予め、図 39 に示すように、クエリ画像として利用されるために用意された画像（以下、「クエリ用の画像」）と、キーワード（各画像に含まれる人物の状態を示すワード）とを紐付けた情報が、データベース 110 に登録されてもよい。そして、画像取得部 101 は、入力されたキーワードが紐付けられたクエリ用の画像を当該情報の中から検索し、検索結果に含まれるクエリ用の画像の一部又は全部をクエリ画像として取得してもよい。

[0038] その他、図 40 に示すように、解析対象画像の一部と、キーワード（各画像に含まれる人物の状態を示すワード）とを紐付けた情報が、データベース 110 に登録されてもよい。そして、画像取得部 101 は、入力されたキーワードが紐付けられた解析対象画像を当該情報の中から検索し、検索結果に含まれる解析対象画像の一部又は全部をクエリ画像として取得してもよい。

[0039] その他、画像取得部 101 は、キーワードに関連する画像を検索する検索エンジンに、入力されたキーワードを送信し、当該検索エンジンから検索結果を取得してもよい。そして、画像取得部 101 は、検索結果に含まれる画像の一部又は全部をクエリ画像として取得してもよい。

[0040] 選択部 109 は、ユーザ入力に基づき、クエリ画像に含まれる少なくとも 1 人の人物を選択する。クエリ画像は、1 人の人物を含む場合もあれば、複

数の人物を含む場合もある。クエリ画像が複数の人物を含む場合、選択部 109 は、ユーザ入力に基づき、クエリ画像に含まれる複数の人物の中から少なくとも 1 人の人物を選択することとなる。選択部 109 は、例えば以下の 2 つの手法のいずれかで、当該選択を実現することができる。

[0041] 「クエリ画像に含まれる少なくとも 1 人の人物を選択する処理例 1」

当該例では、選択部 109 は、自由に画像内の一部領域を指定するユーザ入力を受付ける。そして、選択部 109 は、指定された一部領域内で検出された人物を選択する。「指定された一部領域内で検出された人物」は、指定された一部領域内に身体が完全に包含される人物、または、指定された一部領域内に身体の少なくとも一部が含まれる人物である。以下、詳細に説明する。

[0042] 例えば、画像取得部 101 が、図 43 に示すようなクエリ画像 P を取得したとする。選択部 109 は、当該クエリ画像 P を表示部 107 に表示させ、当該クエリ画像 P の中で少なくとも一部領域を指定するユーザ入力を受付ける。例えば、選択部 109 は、図 44 に示すように枠 W で一部領域を囲むユーザ入力により、一部領域を指定するユーザ入力を受付けてもよい。枠 W の画像内の位置及び大きさはユーザが自由に指定可能となっている。

[0043] クエリ画像 P の中で少なくとも一部領域を指定するユーザ入力を受付けた後、選択部 109 は、指定された一部領域内で検出された人物を選択する。以下、当該処理の一例を説明する。

[0044] ー指定された一部領域内で検出された人物を選択する処理例 1ー

選択部 109 は、指定された一部領域内の画像、又は、指定された一部領域及びその周辺を含む画像に対し、人物を検出する処理を実行する。そして、選択部 109 は、指定された一部領域内に身体が完全に包含される人物、または、指定された一部領域内に身体の少なくとも一部が含まれる人物を選択する。画像内から人物を検出する処理は、周知のあらゆる技術を採用できる。

[0045] なお、図 45 に示すように指定された一部領域（枠 W 内の領域）内に複数

の人物が存在し、当該領域内で複数の人物が検出された場合、選択部109は、図46に示すように、検出された複数の人物各々を識別可能に表示し、その中から少なくとも一部（一部又は全部）を指定するユーザ入力をさらに受付けてもよい。そして、選択部109は、当該ユーザ入力で指定された人物を選択してもよい。図46では、枠Yにより、指定された一部領域内で検出された人物各々が識別可能になっている。

[0046] ー指定された一部領域内で検出された人物を選択する処理例2ー

選択部109は、クエリ画像全体（画像内の全領域）に対し人物を検出する処理を実行し、人物が存在する領域である人領域を検出する。人領域は、矩形領域であってもよいし、人物の輪郭に沿った領域であってもよい。画像内から人物を検出する処理は、周知のあらゆる技術を採用できる。人物の輪郭に沿った領域を検出する技術は、セマンティックセグメンテーション、インスタンスセグメンテーション、予め用意された背景画像との差分算出、輪郭抽出処理等が挙げられるが、これらに限定されない。そして、選択部109は、検出した人領域の中から、指定された一部領域内に完全に包含される人領域、または、指定された一部領域に一部が含まれる人領域を特定し、特定した人領域内に存在する人物を選択する。

[0047] なお、図45に示すように指定された一部領域（枠W内の領域）内に複数の人物が存在し、当該領域内で複数の人領域が検出された場合、選択部109は、図46に示すように、検出された複数の人領域各々を識別可能に表示し、その中から少なくとも一部（一部又は全部）を指定するユーザ入力を受付けてもよい。そして、選択部109は、当該ユーザ入力で指定された人領域内に存在する人物を選択してもよい。図46では、枠Yにより、指定された一部領域内で検出された複数の人領域各々が識別可能になっている。

[0048] ー指定された一部領域内で検出された人物を選択する処理例3ー

骨格構造検出部102は、指定された一部領域内の画像、又は、指定された一部領域及びその周辺を含む画像に対し、人物の2次元骨格構造を検出する処理を実行する。そして、選択部109は、指定された一部領域内に検出

された２次元骨格構造が完全に包含される人物、または、指定された一部領域内に検出された２次元骨格構造の少なくとも一部が含まれる人物を選択する。

[0049] なお、図４５に示すように指定された一部領域（枠Ｗ内の領域）内に複数の人物が存在し、当該領域内で複数の２次元骨格構造が検出された場合、選択部１０９は、図４６に示すように、検出された複数の２次元骨格構造各々を識別可能に表示し、その中から少なくとも一部（一部又は全部）を指定するユーザ入力を受付けてもよい。そして、選択部１０９は、当該ユーザ入力で指定された２次元骨格構造を有する人物を選択してもよい。図４６では、枠Ｙにより、指定された一部領域内で検出された２次元骨格構造各々が識別可能になっている。

[0050] ー指定された一部領域内で検出された人物を選択する処理例４ー

骨格構造検出部１０２は、クエリ画像全体（画像内の全領域）に対し人物の２次元骨格構造を検出する処理を実行する。そして、選択部１０９は、指定された一部領域内に検出された２次元骨格構造が完全に包含される人物、または、指定された一部領域内に検出された２次元骨格構造の少なくとも一部が含まれる人物を選択する。

[0051] なお、図４５に示すように指定された一部領域（枠Ｗ内の領域）内に複数の人物が存在し、当該領域内で複数の２次元骨格構造が検出された場合、選択部１０９は、図４６に示すように、検出された複数の２次元骨格構造各々を識別可能に表示し、その中から少なくとも一部（一部又は全部）を指定するユーザ入力を受付けてもよい。そして、選択部１０９は、当該ユーザ入力で指定された２次元骨格構造を有する人物を選択してもよい。図４６では、枠Ｙにより、指定された一部領域内で検出された２次元骨格構造各々が識別可能になっている。

[0052] 「クエリ画像に含まれる少なくとも１人の人物を選択する処理例２」

当該例では、人物を選択するユーザ入力を受付ける前に、クエリ画像全体（画像内の全領域）に対して人物を検出する処理が実行される。そして、選

択部 109 は、図 47 に示すように、クエリ画像上に、当該クエリ画像内で検出された人物を選択可能な人物として識別可能に表示した加工画像を表示する。そして、選択部 109 は、図 48 に示すように選択可能な人物の中から少なくとも 1 人を選択するユーザ入力を受付ける。ここで、人物を検出する処理、及び、選択可能な人物として表示する人物を決定する処理の一例を説明する。

[0053] 一人物を検出する処理、及び、選択可能な人物として表示する人物を決定する処理例 1 -

選択部 109 は、周知の人物を検出する処理に基づき、クエリ画像全体（画像内の全領域）に対して人物を検出する処理を実行する。そして、選択部 109 は、検出したすべての人物を、選択可能な人物とする。その他、選択部 109 は、検出した人物の中の、画像内で占める領域（画像内の人物の大きさを意味する）が基準値以上の人物を、選択可能な人物としてもよい。画像内で占める領域が所定レベルより小さい人物は、2次元骨格構造の検出や特徴量の算出が困難になる。このような人物の姿勢を検索のためのクエリとすると、所望の検索結果が得られにくくなる。そこで、このような人物は選択可能な人物から除外してもよい。

[0054] 一人物を検出する処理、及び、選択可能な人物として表示する人物を決定する処理例 2 -

骨格構造検出部 102 は、クエリ画像全体（画像内の全領域）に対し人物の 2次元骨格構造を検出する処理を実行する。そして、選択部 109 は、骨格構造検出部 102 による 2次元骨格構造の検出結果が所定条件を満たす人物を、選択可能な人物とする。

[0055] 所定条件を満たす人物は、「抽出されたキーポイントの数が基準値以上の人物」、又は、「抽出されたキーポイントの数、及び抽出されたキーポイント各々の信頼度の少なくとも一方に基づき算出された評価値が基準値以上の人物」である。

[0056] 評価値算出の詳細なアルゴリズムは特段制限されないが、以下の内容を満

たすように設計される。

- ・ 抽出されたキーポイントの数が多いほど、評価値が高い。
- ・ 抽出されたキーポイントの信頼度が高いほど、評価値が高い。
- ・ 画像内における人物が大きいほど評価値が高い。

[0057] 骨格構造検出部 102 は、クエリ画像に含まれる人物の 2 次元骨格構造を検出する。骨格構造検出部 102 の当該検出処理は、データ蓄積処理で説明した通りである。骨格構造検出部 102 は、クエリ画像全体（画像内の全領域）に対し人物の 2 次元骨格構造を検出する処理を実行してもよいし、クエリ画像の一部（ユーザ入力で指定された領域、選択部 109 が選択した人物が含まれる一部領域等）に対し人物の 2 次元骨格構造を検出する処理を実行してもよい。

[0058] 特徴量算出部 103 は、選択部 109 が選択した人物の 2 次元骨格構造の特徴量を算出する。特徴量算出部 103 の当該抽出処理は、データ蓄積処理で説明した通りである。

[0059] 検索部 105 は、データ蓄積処理でデータベース 110 に格納された複数の骨格構造の中から、検索クエリ（クエリ状態）の特徴量（選択部 109 が選択した人物の 2 次元骨格構造の特徴量）と類似度の高い骨格構造を検索する。

[0060] 例えば、検索部 105 は、検索クエリの特徴量と、複数の解析対象画像各々から検出された骨格構造の特徴量とを照合することで、検索クエリの特徴量と類似度の高い骨格構造を検索してもよい。この構成の場合、上述した分類処理は不要となる。しかし、照合対象が複数の解析対象画像の全てとなるので、照合におけるコンピュータの処理負担が大きくなる。

[0061] そこで、検索部 105 は、分類処理で得られたグループ毎に 2 次元の骨格構造の特徴量の代表を任意の手段で決定し、代表と上記検索クエリの特徴量との照合により、検索クエリの特徴量と類似度の高い骨格構造を検索してもよい。この構成の場合、照合対象の数が少なくなるので、照合におけるコンピュータの処理負担が小さくなる。

[0062] なお、解析対象画像と、各解析対象画像から検出された２次元の骨格構造とは互いに紐付けられている。このため、上記「解析対象画像から検出された複数の２次元の骨格構造の中から所定の骨格構造を検索する処理」により、所定の骨格構造（検索クエリの特徴量と類似度の高い骨格構造）を含む解析対象画像を検索することができる。すなわち、解析対象画像の中から、クエリ画像に含まれる人物の状態と類似する状態の人物を含む解析対象画像を検索することができる。

[0063] 類似度は、骨格構造の特徴量間の距離である。検索部１０５は、骨格構造の全体の特徴量の類似度により検索してもよいし、骨格構造の一部の特徴量の類似度により検索してもよく、骨格構造の第１の部分（例えば両手）及び第２の部分（例えば両足）の特徴量の類似度により検索してもよい。なお、各画像における人物の骨格構造の特徴量に基づいて人物の姿勢を検索してもよいし、時系列に連続する複数の画像における人物の骨格構造の特徴量の変化に基づいて人物の行動を検索してもよい。すなわち、検索部１０５は、骨格構造の特徴量に基づいて人物の姿勢や行動を含む人物の状態を検索できる。例えば、検索部１０５は、所定の監視期間に撮像された複数の解析対象画像における複数の骨格構造の特徴量を検索対象とする。

[0064] 入力部１０６は、画像処理装置１００を操作するユーザから入力された情報を取得する入力インターフェイスである。例えば、ユーザは、監視カメラの画像から不審な状態の人物を監視する監視者である。入力部１０６は、例えば、ＧＵＩ（Graphical User Interface）であり、キーボード、マウス、タッチパネル、マイク、物理ボタン等の入力装置から、ユーザの操作に応じた情報が入力される。

[0065] 表示部１０７は、画像処理装置１００の動作（処理）の結果等を表示する表示部であり、例えば、液晶ディスプレイや有機ＥＬ（Electro Luminescence）ディスプレイ等のディスプレイ装置である。表示部１０７は、分類部１０４の分類結果、検索部１０５の検索結果、上述したクエリ画像の候補の評価値等を表示する。

[0066] 次に、画像処理装置１００のハードウェア構成の一例を説明する。画像処理装置１００の各機能部は、任意のコンピュータのＣＰＵ（Central Processing Unit）、メモリ、メモリにロードされるプログラム、そのプログラムを格納するハードディスク等の記憶ユニット（あらかじめ装置を出荷する段階から格納されているプログラムのほか、ＣＤ（Compact Disc）等の記憶媒体やインターネット上のサーバ等からダウンロードされたプログラムをも格納できる）、ネットワーク接続用インターフェイスを中心にハードウェアとソフトウェアの任意の組合せによって実現される。そして、その実現方法、装置にはいろいろな変形例があることは、当業者には理解されるところである。

[0067] 図４１は、画像処理装置１００のハードウェア構成を例示するブロック図である。図４１に示すように、画像処理装置１００は、プロセッサ１Ａ、メモリ２Ａ、入出力インターフェイス３Ａ、周辺回路４Ａ、バス５Ａを有する。周辺回路４Ａには、様々なモジュールが含まれる。画像処理装置１００は周辺回路４Ａを有さなくてもよい。なお、画像処理装置１００は物理的及び／又は論理的に分かれた複数の装置で構成されてもよいし、物理的及び／又は論理的に一体となった１つの装置で構成されてもよい。画像処理装置１００が物理的及び／又は論理的に分かれた複数の装置で構成される場合、複数の装置各々が上記ハードウェア構成を備えることができる。

[0068] バス５Ａは、プロセッサ１Ａ、メモリ２Ａ、周辺回路４Ａ及び入出力インターフェイス３Ａが相互にデータを送受信するためのデータ伝送路である。プロセッサ１Ａは、例えばＣＰＵ、ＧＰＵ（Graphics Processing Unit）などの演算処理装置である。メモリ２Ａは、例えばＲＡＭ（Random Access Memory）やＲＯＭ（Read Only Memory）などのメモリである。入出力インターフェイス３Ａは、入力装置、外部装置、外部サーバ、外部センサー、カメラ等から情報を取得するためのインターフェイスや、出力装置、外部装置、外部サーバ等に情報を出力するためのインターフェイスなどを含む。入力装置は、例えばキーボード、マウス、マイク、物理ボタン、タッチパネル等である。

。出力装置は、例えばディスプレイ、スピーカ、プリンター、メーラ等である。プロセッサ1 Aは、各モジュールに指令を出し、それらの演算結果をもとに演算を行うことができる。

[0069] 図3～図5、図49、図50は、本実施の形態に係る画像処理装置100の処理の流れを示している。図3は、画像処理装置100における画像取得から検索処理までの流れを示し、図4は、図3の分類処理(S104)の流れを示し、図5、図49、図50は、図3の検索処理(S105)の流れを示している。

[0070] 図3に示すように、画像取得部101は、複数の解析対象画像を取得する(S101)。続いて、骨格構造検出部102は、取得された複数の解析対象画像各々から人物の2次元骨格構造を検出する(S102)。図6は、骨格構造の検出例を示している。図6に示すように、解析対象画像には複数の人物が含まれている場合がある。この場合、骨格構造検出部102は、解析対象画像に含まれる各人物について骨格構造を検出する。

[0071] 図7は、このとき検出する人体モデル300の骨格構造を示しており、図8～図10は、骨格構造の検出例を示している。骨格構造検出部102は、OpenPose等の骨格推定技術を用いて、2次元の画像から図7のような人体モデル(2次元骨格モデル)300の骨格構造を検出する。人体モデル300は、人物の関節等のキーポイントと、各キーポイントを結ぶボーンから構成された2次元モデルである。

[0072] 骨格構造検出部102は、例えば、画像の中からキーポイントとなり得る特徴点を抽出し、キーポイントの画像を機械学習した情報を参照して、人物の各キーポイントを検出する。図7の例では、人物のキーポイントとして、頭A1、首A2、右肩A31、左肩A32、右肘A41、左肘A42、右手A51、左手A52、右腰A61、左腰A62、右膝A71、左膝A72、右足A81、左足A82を検出する。さらに、これらのキーポイントを連結した人物の骨として、頭A1と首A2を結ぶボーンB1、首A2と右肩A31及び左肩A32をそれぞれ結ぶボーンB21及びボーンB22、右肩A3

1 及び左肩 A 3 2 と右肘 A 4 1 及び左肘 A 4 2 をそれぞれ結ぶボーン B 3 1 及びボーン B 3 2、右肘 A 4 1 及び左肘 A 4 2 と右手 A 5 1 及び左手 A 5 2 をそれぞれ結ぶボーン B 4 1 及びボーン B 4 2、首 A 2 と右腰 A 6 1 及び左腰 A 6 2 をそれぞれ結ぶボーン B 5 1 及びボーン B 5 2、右腰 A 6 1 及び左腰 A 6 2 と右膝 A 7 1 及び左膝 A 7 2 をそれぞれ結ぶボーン B 6 1 及びボーン B 6 2、右膝 A 7 1 及び左膝 A 7 2 と右足 A 8 1 及び左足 A 8 2 をそれぞれ結ぶボーン B 7 1 及びボーン B 7 2 を検出する。骨格構造検出部 1 0 2 は、検出した人物の骨格構造をデータベース 1 1 0 に格納する。

[0073] 図 8 は、起立した状態の人物を検出する例である。図 8 では、起立した人物が正面から撮像されており、正面から見たボーン B 1、ボーン B 5 1 及びボーン B 5 2、ボーン B 6 1 及びボーン B 6 2、ボーン B 7 1 及びボーン B 7 2 がそれぞれ重ならず検出され、右足のボーン B 6 1 及びボーン B 7 1 は左足のボーン B 6 2 及びボーン B 7 2 よりも多少折れ曲がっている。

[0074] 図 9 は、しゃがみ込んでいる状態の人物を検出する例である。図 9 では、しゃがみ込んでいる人物が右側から撮像されており、右側から見たボーン B 1、ボーン B 5 1 及びボーン B 5 2、ボーン B 6 1 及びボーン B 6 2、ボーン B 7 1 及びボーン B 7 2 がそれぞれ検出され、右足のボーン B 6 1 及びボーン B 7 1 と左足のボーン B 6 2 及びボーン B 7 2 は大きく折れ曲がり、かつ、重なっている。

[0075] 図 1 0 は、寝ている状態の人物を検出する例である。図 1 0 では、寝ている人物が左斜め前から撮像されており、左斜め前から見たボーン B 1、ボーン B 5 1 及びボーン B 5 2、ボーン B 6 1 及びボーン B 6 2、ボーン B 7 1 及びボーン B 7 2 がそれぞれ検出され、右足のボーン B 6 1 及びボーン B 7 1 と左足のボーン B 6 2 及びボーン B 7 2 は折れ曲がり、かつ、重なっている。

[0076] 続いて、図 3 に示すように、特徴量算出部 1 0 3 は、検出された骨格構造の特徴量を算出する (S 1 0 3)。例えば、骨格領域の高さや面積を特徴量とする場合、特徴量算出部 1 0 3 は、骨格構造を含む領域を抽出し、その領

域の高さ（画素数）や面積（画素面積）を求める。骨格領域の高さや面積は、抽出される骨格領域の端部の座標や端部のキーポイントの座標から求められる。特徴量算出部103は、求めた骨格構造の特徴量をデータベース110に格納する。なお、この骨格構造の特徴量は、人物の状態を示す情報としても用いられる。

[0077] 図8の例では、起立した人物の骨格構造から全てのボーンを含む骨格領域を抽出する。この場合、骨格領域の上端は頭部のキーポイントA1、骨格領域の下端は左足のキーポイントA82、骨格領域の左端は右肘のキーポイントA41、骨格領域の右端は左手のキーポイントA52となる。このため、キーポイントA1とキーポイントA82のY座標の差分から骨格領域の高さを求める。また、キーポイントA41とキーポイントA52のX座標の差分から骨格領域の幅を求め、骨格領域の高さと幅から面積を求める。

[0078] 図9の例では、しゃがみ込んだ人物の骨格構造から全てのボーンを含む骨格領域を抽出する。この場合、骨格領域の上端は頭部のキーポイントA1、骨格領域の下端は右足のキーポイントA81、骨格領域の左端は右腰のキーポイントA61、骨格領域の右端は右手のキーポイントA51となる。このため、キーポイントA1とキーポイントA81のY座標の差分から骨格領域の高さを求める。また、キーポイントA61とキーポイントA51のX座標の差分から骨格領域の幅を求め、骨格領域の高さと幅から面積を求める。

[0079] 図10の例では、寝ている人物の骨格構造から全てのボーンを含む骨格領域を抽出する。この場合、骨格領域の上端は左肩のキーポイントA32、骨格領域の下端は左手のキーポイントA52、骨格領域の左端は右手のキーポイントA51、骨格領域の右端は左足のキーポイントA82となる。このため、キーポイントA32とキーポイントA52のY座標の差分から骨格領域の高さを求める。また、キーポイントA51とキーポイントA82のX座標の差分から骨格領域の幅を求め、骨格領域の高さと幅から面積を求める。

[0080] 続いて、図3に示すように、分類部104は、分類処理を行う（S104）。分類処理では、図4に示すように、分類部104は、算出された骨格構

造の特徴量の類似度を算出し（S 1 1 1）、算出された特徴量に基づいて骨格構造を分類する（S 1 1 2）。分類部 1 0 4 は、分類対象であるデータベース 1 1 0 に格納されている全ての骨格構造間の特徴量の類似度を求め、最も類似度が高い骨格構造（姿勢）を同じクラスに分類する（クラスタリングする）。さらに、分類したクラス間の類似度を求めて分類し、所定の数のクラスとなるまで分類を繰り返す。図 1 1 は、骨格構造の特徴量の分類結果のイメージを示している。図 1 1 は、2 次元の分類要素によるクラスタ分析のイメージであり、2 つ分類要素は、例えば、骨格領域の高さと骨格領域の面積等である。図 1 1 では、分類の結果、複数の骨格構造の特徴量が 3 つのクラス C 1 ～C 3 に分類されている。クラス C 1 ～C 3 は、例えば、立っている姿勢、座っている姿勢、寝ている姿勢のように各姿勢に対応し、似ている姿勢ごとに骨格構造（人物）が分類される。

[0081] 本実施の形態では、人物の骨格構造の特徴量に基づいて分類することにより、多様な分類方法を用いることができる。なお、分類方法は、予め設定されていてもよいし、ユーザが任意に設定できるようにしてもよい。また、後述する検索方法と同じ方法により分類を行ってもよい。つまり、検索条件と同様の分類条件により分類してもよい。例えば、分類部 1 0 4 は、次の分類方法により分類を行う。いずれかの分類方法を用いてもよいし、任意に選択された分類方法を組み合わせてもよい。

[0082] <分類方法 1>

複数の階層による分類全身の骨格構造による分類や、上半身や下半身の骨格構造による分類、腕や脚の骨格構造による分類等を階層的に組み合わせて分類する。すなわち、骨格構造の第 1 の部分や第 2 の部分の特徴量に基づいて分類し、さらに、第 1 の部分や第 2 の部分の特徴量に重みづけを行って分類してもよい。

[0083] <分類方法 2>

時系列に沿った複数枚の画像による分類時系列に連続する複数の画像における骨格構造の特徴量に基づいて分類する。例えば、時系列方向に特徴量を

積み重ねて、累積値に基づいて分類してもよい。さらに、連続する複数の画像における骨格構造の特徴量の変化（変化量）に基づいて分類してもよい。

[0084] <分類方法3>

骨格構造の左右を無視した分類人物の右側と左側が反対の骨格構造を同じ骨格構造として分類する。

[0085] さらに、分類部104は、骨格構造の分類結果を表示する（S113）。分類部104は、データベース110から必要な骨格構造や人物の画像を取得し、分類結果として似ている姿勢（クラス）ごとに骨格構造及び人物を表示部107に表示する。図12は、姿勢を3つに分類した場合の表示例を示している。例えば、図12に示すように、表示ウィンドウW1に、姿勢ごとの姿勢領域WA1～WA3を表示し、姿勢領域WA1～WA3にそれぞれ該当する姿勢の骨格構造及び人物（イメージ）を表示する。姿勢領域WA1は、例えば立っている姿勢の表示領域であり、クラスC1に分類された、立っている姿勢に似た骨格構造及び人物を表示する。姿勢領域WA2は、例えば座っている姿勢の表示領域であり、クラスC2に分類された、座っている姿勢に似た骨格構造及び人物を表示する。姿勢領域WA3は、例えば寝ている姿勢の表示領域であり、クラスC2に分類された、寝ている姿勢に似た骨格構造及び人物を表示する。

[0086] 続いて、図3に示すように、画像処理装置100は、検索処理を行う（S105）。検索処理では、図5に示すように、画像処理装置100は、検索条件の入力を受け付ける（S121）。

[0087] S121の処理の一例を、図49を用いて説明する。画像取得部101は上述した取得例1乃至3のいずれかに基づき、クエリ画像を取得する（S301）。次いで、選択部109は上述したクエリ画像に含まれる少なくとも1人の人物を選択する処理例1又は2に基づき、クエリ画像に含まれる少なくとも1人の人物を選択する（S302）。次いで、特徴量算出部103は、S302で選択された人物の2次元骨格構造の特徴量を算出する（S303）。当該2次元骨格構造の特徴量が、検索クエリとなる。

- [0088] S 1 2 1 の処理の他の一例を、図 5 0 を用いて説明する。画像取得部 1 0 1 は上述した取得例 1 乃至 3 のいずれかに基づき、クエリ画像を取得する (S 3 1 1)。次いで、画像処理装置 1 0 0 は、クエリ画像に対して人物を検出する処理を行う (S 3 1 2)。例えば、選択部 1 0 9 は、周知の人物を検出する処理に基づき、クエリ画像全体 (画像内の全領域) に対して人物を検出する処理を実行してもよい。その他、骨格構造検出部 1 0 2 は、クエリ画像全体 (画像内の全領域) に対して人物の 2 次元骨格構造を検出する処理を実行してもよい。
- [0089] そして、選択部 1 0 9 は、S 3 1 2 の検出処理の結果に基づき特定された画像の中に含まれる人物の数が複数の場合 (S 3 1 3 の Y e s)、画像に含まれる複数の人物の中から少なくとも 1 人の人物を選択する (S 3 1 4)。その後、特徴量算出部 1 0 3 は、S 3 1 4 で選択された少なくとも 1 人の人物の 2 次元骨格構造の特徴量を算出する (S 3 1 6)。当該 2 次元骨格構造の特徴量が、検索クエリとなる。この S 3 1 4 及び S 3 1 6 の処理は、図 4 9 の S 3 0 2 及び S 3 0 3 の処理と同様である。
- [0090] 一方、S 3 1 2 の検出処理の結果に基づき特定された画像の中に含まれる人物の数が 1 人の場合 (S 3 1 3 の N o)、選択部 1 0 9 は、その検出された 1 人の人物を選択する (S 3 1 5)。そして、特徴量算出部 1 0 3 は、S 3 1 5 で選択された 1 人の人物の 2 次元骨格構造の特徴量を算出する (S 3 1 6)。当該 2 次元骨格構造の特徴量が、検索クエリとなる。
- [0091] なお、「S 3 1 2 の検出処理の結果に基づき特定された画像の中に含まれる人物の数」は、S 3 1 2 で検出された人物の数であってもよいし、上述した「選択可能な人物」の数であってもよい。「選択可能な人物」の概念は、「人物を検出する処理、及び、選択可能な人物として表示する人物を決定する処理例 1 及び 2」で説明した通りである。
- [0092] 図 5 に戻り、その後、検索部 1 0 5 は、検索条件に基づいて骨格構造を検索する (S 1 2 2)。検索部 1 0 5 は、ユーザにより指定された骨格構造を検索クエリとして、検索対象であるデータベース 1 1 0 に格納されている骨

格構造の中から特徴量の類似度が高い骨格構造を検索する。検索部 105 は、検索クエリの骨格構造の特徴量と検索対象の骨格構造の特徴量（解析対象画像から検出された骨格構造の特徴量）との類似度を算出し、算出した類似度が所定の閾値よりも高い骨格構造を抽出する。検索クエリの骨格構造の特徴量は、予め算出された特徴量を使用してもよいし、検索時に求めた特徴量を使用してもよい。なお、検索クエリは、ユーザの操作に応じて骨格構造の各部を動かすことで入力してもよいし、ユーザがカメラの前で実演した姿勢を検索クエリとしてもよい。

[0093] 本実施の形態では、分類方法と同様に、人物の骨格構造の特徴量に基づいて検索することにより、多様な検索方法を用いることができる。なお、検索方法は、予め設定されていてもよいし、ユーザが任意に設定できるようにしてもよい。例えば、検索部 105 は、次の検索方法により検索を行う。いずれかの検索方法を用いてもよいし、任意に選択された検索方法を組み合わせてもよい。複数の検索方法（検索条件）を論理式（例えばAND（論理積）、OR（論理和）、NOT（否定））により組み合わせ検索してもよい。例えば、検索条件を「（右手を挙げている姿勢）AND（左足を挙げている姿勢）」として検索してもよい。

[0094] <検索方法 1>

高さ方向の特徴量のみによる検索人物の高さ方向の特徴量のみを用いて検索することで、人物の横方向の変化の影響を抑えることができ、人物の向きや人物の体型の変化に対しロバスト性が向上する。例えば、図 13 の骨格構造 501～503 のように、人物の向きや体型が異なる場合でも、高さ方向の特徴量は大きく変化しない。このため、骨格構造 501～503 では、検索時（分類時）に同じ姿勢であると判断することができる。

[0095] <検索方法 2>

部分検索画像において人物の体の一部が隠れている場合、認識可能な部分の情報のみを用いて検索する。例えば、図 14 の骨格構造 511 及び 512 のように、左足が隠れていることにより、左足のキーポイントが検出できな

い場合でも、検出されている他のキーポイントの特徴量を使用して検索できる。このため、骨格構造 5 1 1 及び 5 1 2 では、検索時（分類時）に同じ姿勢であると判断することができる。つまり、全てのキーポイントではなく、一部のキーポイントの特徴量を用いて、分類や検索を行うことができる。図 1 5 の骨格構造 5 2 1 及び 5 2 2 の例では、両足の向きが異なっているものの、上半身のキーポイント（A 1、A 2、A 3 1、A 3 2、A 4 1、A 4 2、A 5 1、A 5 2）の特徴量を検索クエリとすることで、同じ姿勢であると判断することができる。また、検索したい部分（特徴点）に対して、重みを付けて検索してもよいし、類似度判定の閾値を変化させてもよい。体の一部が隠れている場合、隠れた部分を見捨てて検索してもよいし、隠れた部分を加味して検索してもよい。隠れた部分も含めて検索することで、同じ部位が隠れているような姿勢を検索することができる。

[0096] <検索方法 3>

骨格構造の左右を見捨てた検索人物の右側と左側が反対の骨格構造を同じ骨格構造として検索する。例えば、図 1 6 の骨格構造 5 3 1 及び 5 3 2 のように、右手を挙げている姿勢と、左手を挙げている姿勢を同じ姿勢として検索（分類）できる。図 1 6 の例では、骨格構造 5 3 1 と骨格構造 5 3 2 は、右手のキーポイント A 5 1、右肘のキーポイント A 4 1、左手のキーポイント A 5 2、左肘のキーポイント A 4 2 の位置が異なるものの、その他のキーポイントの位置は同じである。骨格構造 5 3 1 の右手のキーポイント A 5 1 及び右肘のキーポイント A 4 1 と骨格構造 5 3 2 の左手のキーポイント A 5 2 及び左肘のキーポイント A 4 2 のうち、一方の骨格構造のキーポイントを左右反転させると、他方の骨格構造のキーポイントと同じ位置となり、また、骨格構造 5 3 1 の左手のキーポイント A 5 2 及び左肘のキーポイント A 4 2 と骨格構造 5 3 2 の右手のキーポイント A 5 1 及び右肘のキーポイント A 4 1 のうち、一方の骨格構造のキーポイントを左右反転させると、他方の骨格構造のキーポイントと同じ位置となるため、同じ姿勢と判断する。

[0097] <検索方法 4>

縦方向と横方向の特徴量による検索人物の縦方向（Ｙ軸方向）の特徴量のみで検索を行った後、得られた結果をさらに人物の横方向（Ｘ軸方向）の特徴量を用いて検索する。

[0098] ＜検索方法５＞

時系列に沿った複数枚の画像による検索時系列に連続する複数の画像における骨格構造の特徴量に基づいて検索する。例えば、時系列方向に特徴量を積み重ねて、累積値に基づいて検索してもよい。さらに、連続する複数の画像における骨格構造の特徴量の変化（変化量）に基づいて検索してもよい。

[0099] さらに、検索部１０５は、骨格構造の検索結果を表示する（Ｓ１２３）。検索部１０５は、データベース１１０から必要な骨格構造や人物の画像を取得し、検索結果として得られた骨格構造及び人物を表示部１０７に表示する。例えば、検索クエリ（検索条件）が複数指定されている場合、検索クエリごとに検索結果を表示する。図１７は、３つの検索クエリ（姿勢）により検索した場合の表示例を示している。例えば、図１７に示すように、表示ウィンドウＷ２において、左端部に指定された検索クエリＱ１０、Ｑ２０、Ｑ３０の骨格構造及び人物を表示し、検索クエリＱ１０、Ｑ２０、Ｑ３０の右側に各検索クエリの検索結果Ｑ１１、Ｑ２１、Ｑ３１の骨格構造及び人物を並べて表示する。

[0100] 検索結果を検索クエリの隣から並べて表示する順番は、該当する骨格構造が見つかった順でもよいし、類似度が高い順でもよい。部分検索の部分（特徴点）に重みを付けて検索した場合に、重み付けて計算した類似度順に表示してもよい。ユーザが選択した部分（特徴点）のみから計算した類似度順に表示してもよい。また、検索結果の画像（フレーム）を中心に、時系列の前後の画像（フレーム）を一定時間分切り出して表示してもよい。

[0101] ここで、上記検索処理の変形例を説明する。選択部１０９は、クエリ画像の中で少なくとも一部領域を指定するユーザ入力に基づき、クエリ画像に含まれる少なくとも１人の人物を選択するとともに、クエリ画像の中で少なくとも一部領域を指定するユーザ入力に基づき、選択した人物の身体の一部を

選択する。そして、検索部105は、選択された身体の一部の重みを他の部分よりも大きくして、選択された人物の状態と類似する状態の人物を含む解析対象画像を検索する。

[0102] 例えば、選択部109は、図51に示すように、枠Xでクエリ画像の一部領域を囲むユーザ入力により、身体の一部を指定するユーザ入力を受付けることができる。枠Xの画像内の位置及び大きさはユーザが自由に指定可能となっている。また、例えば、検索部105は、選択された身体の一部に含まれるキーポイント（特徴点）の重みを他のキーポイント（特徴点）よりも大きくして、選択された人物の状態と類似する状態の人物を含む解析対象画像を検索することができる。

[0103] ここで、上記検索処理の他の変形例を説明する。画像処理装置100は、予め用意されたクエリ用の画像の中から、選択部109が選択した人物と2次元骨格構造の特徴量が類似する人物を含む画像を検索する。そして、画像処理装置100は、図52に示すように、検索結果Zを表示し、その中からクエリ画像を指定するユーザ入力を受付ける。この場合、検索部105は、図52に示すような検索結果の中から指定されたクエリ画像に含まれる人物の2次元骨格構造の特徴量を検索クエリとして、所定の解析対象画像を検索する。

[0104] 以上のように、本実施の形態では、2次元画像から人物の骨格構造を検出し、検出した骨格構造の特徴量に基づいて分類や検索を行うことを可能とした。これにより、類似度が高い似た姿勢ごとに分類することができ、また、検索クエリ（検索キー）と類似度が高い似た姿勢を検索することができる。画像から似ている姿勢を分類し表示することで、ユーザが姿勢等を指定することなく、画像中の人物の姿勢を把握することができる。分類結果の中からユーザが検索クエリの姿勢を指定できるため、予めユーザが検索したい姿勢を詳細に把握していない場合でも、所望の姿勢を検索することができる。例えば、人物の骨格構造の全体や一部等を条件として分類や検索を行うことができるため、柔軟な分類や検索が可能となる。

- [0105] また、本実施形態の場合、検索したい状態の人物を含むクエリ画像を画像処理装置１００に与える必要がある。検索したい状態の人物のみを含む画像が見つければ特に問題ないが、検索したい状態の人物に加えて、無関係な状態の人物を含む画像しか見つからない場合もある。このようなクエリ画像を画像処理装置１００に与えた場合に、クエリ画像に含まれる複数の人物すべての状態を検索クエリとして検索されると、所望の状態の人物を含む画像のみならず、無関係な状態の人物を含む画像も検索結果に含まれてしまう。この問題に対し、画像処理装置１００は、検索処理において、ユーザ入力に基づきクエリ画像に含まれる少なくとも１人の人物を選択し、解析対象画像の中から、選択された人物の状態と類似する状態の人物を含む解析対象画像を検索する。このような画像処理装置１００によれば、検索したい状態の人物に加えて、無関係な状態の人物を含む画像をクエリ画像とした場合であっても、所望の人物の状態のみを検索クエリとした検索を実行させることができる。結果、検索結果の中に検索対象と異なる画像が含まれる不都合を抑制できる。
- [0106] また、本実施形態の画像処理装置１００は、クエリ画像の中で少なくとも一部領域を指定するユーザ入力を受け付け、指定された一部領域内で検出された人物を選択することができる。このような画像処理装置１００によれば、ユーザは、自分の入力によりどの人物が選択されるかを容易に把握できる。
- [0107] また、本実施形態の画像処理装置１００は、クエリ画像に対して人物を検出する処理を実行した後、クエリ画像上にその画像内で検出された人物を選択可能な人物として識別可能に表示した加工画像を表示し、当該選択可能な人物の中から少なくとも１人を選択するユーザ入力を受け付けることができる。このような画像処理装置１００によれば、ユーザは、自分の入力によりどの人物が選択されるかを容易に把握できる。また、ユーザは、複数の選択肢の中から所望の選択肢を選択するだけでよいので、ユーザ入力により容易になる。
- [0108] また、本実施形態の画像処理装置１００は、検出した人物の中の、画像内

で占める領域（画像内の人物の大きさを意味する）が基準値以上の人物や、骨格構造検出部 102 による 2 次元骨格構造の検出結果が所定条件を満たす人物を、選択可能な人物として上記加工画像に表示することができる。結果、2 次元骨格構造の検出や特徴量の算出が十分な精度で行うことができる人物のみを選択可能な人物とすることができ、検索クエリの質の向上、及び、検索結果の精度の向上が実現される。

[0109] また、本実施形態の画像処理装置 100 は、クエリ画像上で人物の身体の一部を指定するユーザ入力に基づき人物の身体の一部を選択し、選択された身体の一部の重みを他の部分よりも大きくして、選択された人物の状態と類似する状態の人物を含む解析対象画像を検索することができる。このような画像処理装置 100 によれば、ユーザは、自分の入力によりどの身体の一部が選択されるかを容易に把握できる。

[0110] また、本実施形態の画像処理装置 100 は、クエリ画像の中に含まれる人物の数を特定することができる。そして、画像処理装置 100 は、特定した人物の数が複数である場合、ユーザ入力に基づき画像に含まれる複数の人物の中から少なくとも 1 人を選択し、特定した人物の数が 1 人である場合、その 1 人の人物を選択することができる。このような画像処理装置 100 によれば、ユーザ入力に基づきクエリ画像の中から少なくとも 1 人を選択する処理の実行を、クエリ画像の中に含まれる人物が複数の場合に限定することができる。結果、コンピュータの処理負担の軽減や、ユーザ入力の軽減によるユーザ負担の軽減が実現される。

[0111] （実施の形態 2）

以下、図面を参照して実施の形態 2 について説明する。本実施の形態では、実施の形態 1 における特徴量算出の具体例について説明する。本実施の形態では、人物の身長を用いて正規化することで特徴量を求める。その他については、実施の形態 1 と同様である。

[0112] 図 18 は、本実施の形態に係る画像処理装置 100 の構成を示している。図 18 に示すように、画像処理装置 100 は、実施の形態 1 の構成に加えて

、さらに身長算出部108を備える。なお、特徴量算出部103と身長算出部108を一つの処理部としてもよい。

[0113] 身長算出部（身長推定部）108は、骨格構造検出部102により検出された2次元の骨格構造に基づき、2次元の画像内の人物の起立時の高さ（身長画素数という）を算出（推定）する。身長画素数は、2次元の画像における人物の身長（2次元画像空間上の人物の全身の長さ）であるとも言える。身長算出部108は、検出された骨格構造の各ボーンの長さ（2次元画像空間上の長さ）から身長画素数（ピクセル数）を求める。

[0114] 以下の例では、身長画素数を求める方法として具体例1～3を用いる。なお、具体例1～3のいずれかの方法を用いてもよいし、任意に選択される複数の方法を組み合わせて用いてもよい。具体例1では、骨格構造の各ボーンのうち、頭部から足部までのボーンの長さを合計することで、身長画素数を求める。骨格構造検出部102（骨格推定技術）が頭頂と足元を出力しない場合は、必要に応じて定数を乗じて補正することもできる。具体例2では、各ボーンの長さと全身の長さ（2次元画像空間上の身長）との関係を示す人体モデルを用いて、身長画素数を算出する。具体例3では、3次元人体モデルを2次元骨格構造にフィッティング（あてはめる）することで、身長画素数を算出する。

[0115] 本実施の形態の特徴量算出部103は、算出された人物の身長画素数に基づいて、人物の骨格構造（骨格情報）を正規化する正規化部である。特徴量算出部103は、正規化した骨格構造の特徴量（正規化値）をデータベース110に格納する。特徴量算出部103は、骨格構造に含まれる各キーポイント（特徴点）の画像上での高さを、身長画素数で正規化する。本実施の形態では、例えば、高さ方向は、画像の2次元座標（X-Y座標）空間における上下の方向（Y軸方向）である。この場合、キーポイントの高さは、キーポイントのY座標の値（画素数）から求めることができる。あるいは、高さ方向は、実世界の3次元座標空間における地面（基準面）に対し垂直な鉛直軸の方向を、2次元座標空間に投影した鉛直投影軸の方向（鉛直投影方向）

でもよい。この場合、キーポイントの高さは、実世界における地面に対し垂直な軸を、カメラパラメータに基づいて2次元座標空間に投影した鉛直投影軸を求め、この鉛直投影軸に沿った値（画素数）から求めることができる。なお、カメラパラメータは、画像の撮像パラメータであり、例えば、カメラパラメータは、カメラ200の姿勢、位置、撮像角度、焦点距離等である。カメラ200により、予め長さや位置が分かっている物体を撮像し、その画像からカメラパラメータを求めることができる。撮像された画像の両端ではひずみが発生し、実世界の鉛直方向と画像の上下方向が合わない場合がある。これに対し、画像を撮影したカメラのパラメータを使用することで、実世界の鉛直方向が画像中でどの程度傾いているのかが分かる。このため、カメラパラメータに基づいて画像中に投影した鉛直投影軸に沿ったキーポイントの値を身長で正規化することで、実世界と画像のずれを考慮してキーポイントを特徴量化することができる。なお、左右方向（横方向）は、画像の2次元座標（X-Y座標）空間における左右の方向（X軸方向）であり、または、実世界の3次元座標空間における地面に対し平行な方向を、2次元座標空間に投影した方向である。

[0116] 図19～図23は、本実施の形態に係る画像処理装置100の処理の流れを示している。図19は、画像処理装置100における画像取得から検索処理までの流れを示し、図20～図22は、図19の身長画素数算出処理（S201）の具体例1～3の流れを示し、図23は、図19の正規化処理（S202）の流れを示している。

[0117] 図19に示すように、本実施の形態では、実施の形態1における特徴量算出処理（S103）として、身長画素数算出処理（S201）及び正規化処理（S202）を行う。その他については実施の形態1と同様である。

[0118] 画像処理装置100は、画像取得（S101）及び骨格構造検出（S102）に続いて、検出された骨格構造に基づいて身長画素数算出処理を行う（S201）。この例では、図24に示すように、画像における直立時の人物の骨格構造の高さを身長画素数（h）とし、画像の人物の状態における骨格

構造の各キーポイントの高さをキーポイント高さ (y_i) とする。以下、身長画素数算出処理の具体例 1 ～ 3 について説明する。

[0119] <具体例 1>

具体例 1 では、頭部から足部までのボーンの長さをを用いて身長画素数を求める。具体例 1 では、図 20 に示すように、身長算出部 108 は、各ボーンの長さを取得し (S_{211})、取得した各ボーンの長さを合計する (S_{212})。

[0120] 身長算出部 108 は、人物の頭部から足部の 2 次元の画像上のボーンの長さを取得し、身長画素数を求める。すなわち、骨格構造を検出した画像から、図 24 のボーンのうち、ボーン B 1 (長さ L_1)、ボーン B 5 1 (長さ L_{21})、ボーン B 6 1 (長さ L_{31}) 及びボーン B 7 1 (長さ L_{41})、もしくは、ボーン B 1 (長さ L_1)、ボーン B 5 2 (長さ L_{22})、ボーン B 6 2 (長さ L_{32}) 及びボーン B 7 2 (長さ L_{42}) の各長さ (画素数) を取得する。各ボーンの長さは、2 次元の画像における各キーポイントの座標から求めることができる。これらを合計した、 $L_1 + L_{21} + L_{31} + L_{41}$ 、もしくは、 $L_1 + L_{22} + L_{32} + L_{42}$ に補正定数を乗じた値を身長画素数 (h) として算出する。両方の値を算出できる場合、例えば、長い方の値を身長画素数とする。すなわち、各ボーンは正面から撮像された場合が画像中での長さが最も長くなり、カメラに対して奥行き方向に傾くと短く表示される。従って、長いボーンの方が正面から撮像されている可能性が高く、真実の値に近いと考えられる。このため、長い方の値を選択することが好ましい。

[0121] 図 25 の例では、ボーン B 1、ボーン B 5 1 及びボーン B 5 2、ボーン B 6 1 及びボーン B 6 2、ボーン B 7 1 及びボーン B 7 2 がそれぞれ重ならずに出検されている。これらのボーンの合計である、 $L_1 + L_{21} + L_{31} + L_{41}$ 、及び、 $L_1 + L_{22} + L_{32} + L_{42}$ を求め、例えば、検出されたボーンの長さが長い左足側の $L_1 + L_{22} + L_{32} + L_{42}$ に補正定数を乗じた値を身長画素数とする。

[0122] 図26の例では、ボーンB1、ボーンB51及びボーンB52、ボーンB61及びボーンB62、ボーンB71及びボーンB72がそれぞれ検出され、右足のボーンB61及びボーンB71と左足のボーンB62及びボーンB72が重なっている。これらのボーンの合計である、 $L1 + L21 + L31 + L41$ 、及び、 $L1 + L22 + L32 + L42$ を求め、例えば、検出されたボーンの長さが長い右足側の $L1 + L21 + L31 + L41$ に補正定数を乗じた値を身長画素数とする。

[0123] 図27の例では、ボーンB1、ボーンB51及びボーンB52、ボーンB61及びボーンB62、ボーンB71及びボーンB72がそれぞれ検出され、右足のボーンB61及びボーンB71と左足のボーンB62及びボーンB72が重なっている。これらのボーンの合計である、 $L1 + L21 + L31 + L41$ 、及び、 $L1 + L22 + L32 + L42$ を求め、例えば、検出されたボーンの長さが長い左足側の $L1 + L22 + L32 + L42$ に補正定数を乗じた値を身長画素数とする。

[0124] 具体例1では、頭から足までのボーンの長さを合計することで身長を求めることができるため、簡易な方法で身長画素数を求めることができる。また、機械学習を用いた骨格推定技術により、少なくとも頭から足までの骨格を検出できればよいため、しゃがみ込んでいる状態など、必ずしも人物の全体が画像に写っていない場合でも精度よく身長画素数を推定することができる。

[0125] <具体例2>

具体例2では、2次元骨格構造に含まれる骨の長さと2次元画像空間上の人物の全身の長さとの関係を示す2次元骨格モデルを用いて身長画素数を求める。

[0126] 図28は、具体例2で用いる、2次元画像空間上の各ボーンの長さと2次元画像空間上の全身の長さとの関係を示す人体モデル（2次元骨格モデル）301である。図28に示すように、平均的な人物の各ボーンの長さと全身の長さとの関係（全身の長さに対する各ボーンの長さの割合）を、人体モデ

ル301の各ボーンに対応付ける。例えば、頭のボーンB1の長さは全身の長さ $\times 0.2$ (20%)であり、右手のボーンB41の長さは全身の長さ $\times 0.15$ (15%)であり、右足のボーンB71の長さは全身の長さ $\times 0.25$ (25%)である。このような人体モデル301の情報をデータベース110に記憶しておくことで、各ボーンの長さから平均的な全身の長さを求めることができる。平均的な人物の人体モデルの他に、年代、性別、国籍等の人物の属性ごとに人体モデルを用意してもよい。これにより、人物の属性に応じて適切に全身の長さ(身長)を求めることができる。

[0127] 具体例2では、図21に示すように、身長算出部108は、各ボーンの長さを取得する(S221)。身長算出部108は、検出された骨格構造において、全てのボーンの長さ(2次元画像空間上の長さ)を取得する。図29は、しゃがみ込んでいる状態の人物を右斜め後ろから撮像し、骨格構造を検出した例である。この例では、人物の顔や左側面が写っていないことから、頭のボーンと左腕及び左手のボーンが検出できていない。このため、検出されているボーンB21、B22、B31、B41、B51、B52、B61、B62、B71、B72の各長さを取得する。

[0128] 続いて、身長算出部108は、図21に示すように、人体モデルに基づき、各ボーンの長さから身長画素数を算出する(S222)。身長算出部108は、図28のような、各ボーンと全身の長さとの関係を示す人体モデル301を参照し、各ボーンの長さから身長画素数を求める。例えば、右手のボーンB41の長さが全身の長さ $\times 0.15$ であるため、ボーンB41の長さ $\div 0.15$ によりボーンB41に基づいた身長画素数を求める。また、右足のボーンB71の長さが全身の長さ $\times 0.25$ であるため、ボーンB71の長さ $\div 0.25$ によりボーンB71に基づいた身長画素数を求める。

[0129] このとき参照する人体モデルは、例えば、平均的な人物の人体モデルであるが、年代、性別、国籍等の人物の属性に応じて人体モデルを選択してもよい。例えば、撮像した画像に人物の顔が写っている場合、顔に基づいて人物の属性を識別し、識別した属性に対応する人体モデルを参照する。属性ごと

の顔を機械学習した情報を参照し、画像の顔の特徴から人物の属性を認識することができる。また、画像から人物の属性が識別できない場合に、平均的な人物の人体モデルを用いてもよい。

[0130] また、ボーンの長さから算出した身長画素数をカメラパラメータにより補正してもよい。例えばカメラを高い位置において、人物を見下ろすように撮影した場合、二次元骨格構造において肩幅のボーン等の横の長さはカメラの俯角の影響を受けないが、首―腰のボーン等の縦の長さは、カメラの俯角が大きくなる程小さくなる。そうすると、肩幅のボーン等の横の長さから算出した身長画素数が実際より大きくなる傾向がある。そこで、カメラパラメータを活用すると、人物がどの程度の角度でカメラに見下ろされているかがわかるため、この俯角の情報を使って正面から撮影したような二次元骨格構造に補正することができる。これによって、より正確に身長画素数を算出できる。

[0131] 続いて、身長算出部108は、図21に示すように、身長画素数の最適値を算出する(S223)。身長算出部108は、ボーンごとに求めた身長画素数から身長画素数の最適値を算出する。例えば、図30に示すような、ボーンごとに求めた身長画素数のヒストグラムを生成し、その中で大きい身長画素数を選択する。つまり、複数のボーンに基づいて求められた複数の身長画素数の中で他よりも長い身長画素数を選択する。例えば、上位30%を有効な値とし、図30ではボーンB71、B61、B51による身長画素数を選択する。選択した身長画素数の平均を最適値として求めてもよいし、最も大きい身長画素数を最適値としてもよい。2次元画像のボーンの長さから身長を求めるため、ボーンを正面からできていない場合、すなわち、ボーンがカメラから見て奥行き方向に傾いて撮像された場合、ボーンの長さが正面から撮像した場合よりも短くなる。そうすると、身長画素数が大きい値は、身長画素数が小さい値よりも、正面から撮像された可能性が高く、より尤もらしい値となることから、より大きい値を最適値とする。

[0132] 具体例2では、2次元画像空間上のボーンと全身の長さとの関係を示す人

体モデルを用いて、検出した骨格構造のボーンに基づき身長画素数を求めるため、頭から足までの全ての骨格が得られない場合でも、一部のボーンから身長画素数を求めることができる。特に、複数のボーンから求められた値のうち、より大きい値を採用することで、精度よく身長画素数を推定することができる。

[0133] <具体例3>

具体例3では、2次元骨格構造を3次元人体モデル（3次元骨格モデル）にフィッティングさせて、フィッティングした3次元人体モデルの身長画素数を用いて全身の骨格ベクトルを求める。

[0134] 具体例3では、図22に示すように、身長算出部108は、まず、カメラ200の撮像した画像に基づき、カメラパラメータを算出する（S231）。身長算出部108は、カメラ200が撮像した複数の画像の中から、予め長さが分かっている物体を抽出し、抽出した物体の大きさ（画素数）からカメラパラメータを求める。なお、カメラパラメータを予め求めておき、求めておいたカメラパラメータを必要に応じて取得してもよい。

[0135] 続いて、身長算出部108は、3次元人体モデルの配置及び高さを調整する（S232）。身長算出部108は、検出された2次元骨格構造に対し、身長画素数算出用の3次元人体モデルを用意し、カメラパラメータに基づいて、同じ2次元画像内に配置する。具体的には、カメラパラメータと、2次元骨格構造から、「実世界におけるカメラと人物の相対的な位置関係」を特定する。例えば、仮にカメラの位置を座標（0，0，0）としたときに、人物が立っている（または座っている）位置の座標（ x ， y ， z ）を特定する。そして、特定した人物と同じ位置（ x ， y ， z ）に3次元人体モデルを配置して撮像した場合の画像を想定することで、2次元骨格構造と3次元人体モデルを重ね合わせる。

[0136] 図31は、しゃがみ込んでいる人物を左斜め前から撮像し、2次元骨格構造401を検出した例である。2次元骨格構造401は、2次元の座標情報を有する。なお、全てのボーンを検出していることが好ましいが、一部のボ

ーンが検出されていなくてもよい。この２次元骨格構造４０１に対し、図３２のような、３次元人体モデル４０２を用意する。３次元人体モデル（３次元骨格モデル）４０２は、３次元の座標情報を有し、２次元骨格構造４０１と同じ形状の骨格のモデルである。そして、図３３のように、検出した２次元骨格構造４０１に対し、用意した３次元人体モデル４０２を配置し重ね合わせる。また、重ね合わせるとともに、３次元人体モデル４０２の高さを２次元骨格構造４０１に合うように調整する。

[0137] なお、このとき用意する３次元人体モデル４０２は、図３３のように、２次元骨格構造４０１の姿勢に近い状態のモデルでもよいし、直立した状態のモデルでもよい。例えば、機械学習を用いて２次元画像から３次元空間の姿勢を推定する技術を用いて、推定した姿勢の３次元人体モデル４０２を生成してもよい。２次元画像の関節と３次元空間の関節の情報を学習することで、２次元画像から３次元の姿勢を推定することができる。

[0138] 続いて、身長算出部１０８は、図２２に示すように、３次元人体モデルを２次元骨格構造にフィッティングする（Ｓ２３３）。身長算出部１０８は、図３４のように、３次元人体モデル４０２を２次元骨格構造４０１に重ね合わせた状態で、３次元人体モデル４０２と２次元骨格構造４０１の姿勢が一致するように、３次元人体モデル４０２を変形させる。すなわち、３次元人体モデル４０２の身長、体の向き、関節の角度を調整し、２次元骨格構造４０１との差異がなくなるように最適化する。例えば、３次元人体モデル４０２の関節を、人の可動範囲で回転させていき、また、３次元人体モデル４０２の全体を回転させたり、全体のサイズを調整する。なお、３次元人体モデルと２次元骨格構造のフィッティング（あてはめ）は、２次元空間（２次元座標）上で行う。すなわち、２次元空間に３次元人体モデルを写像し、変形させた３次元人体モデルが２次元空間（画像）でどのように変化するかを考慮して、３次元人体モデルを２次元骨格構造に最適化する。

[0139] 続いて、身長算出部１０８は、図２２に示すように、フィッティングさせた３次元人体モデルの身長画素数を算出する（Ｓ２３４）。身長算出部１０

8は、図35のように、3次元人体モデル402と2次元骨格構造401の差異がなくなり、姿勢が一致すると、その状態の3次元人体モデル402の身長画素数を求める。最適化された3次元人体モデル402を直立させた状態として、カメラパラメータに基づき、2次元空間上の全身の長さを求める。例えば、3次元人体モデル402を直立させた場合の頭から足までのボーンの長さ（画素数）により身長画素数を算出する。具体例1と同様に、3次元人体モデル402の頭部から足部までのボーンの長さを合計してもよい。

[0140] 具体例3では、カメラパラメータに基づいて3次元人体モデルを2次元骨格構造にフィッティングさせて、その3次元人体モデルに基づいて身長画素数を求めることで、全てのボーンが正面に写っていない場合、すなわち、全てのボーンが斜めに映っているため誤差が大きい場合でも、精度よく身長画素数を推定することができる。

[0141] <正規化処理>

図19に示すように、画像処理装置100は、身長画素数算出処理に続いて、正規化処理（S202）を行う。正規化処理では、図23に示すように、特徴量算出部103は、キーポイント高さを算出する（S241）。特徴量算出部103は、検出された骨格構造に含まれる全てのキーポイントのキーポイント高さ（画素数）を算出する。キーポイント高さは、骨格構造の最下端（例えばいずれかの足のキーポイント）からそのキーポイントまでの高さ方向の長さ（画素数）である。ここでは、一例として、キーポイント高さを、画像におけるキーポイントのY座標から求める。なお、上記のように、キーポイント高さは、カメラパラメータに基づいた鉛直投影軸に沿った方向の長さから求めてもよい。例えば、図24の例で、首のキーポイントA2の高さ（ y_i ）は、キーポイントA2のY座標から右足のキーポイントA81または左足のキーポイントA82のY座標を引いた値である。

[0142] 続いて、特徴量算出部103は、正規化のための基準点を特定する（S242）。基準点は、キーポイントの相対的な高さを表すための基準となる点である。基準点は、予め設定されていてもよいし、ユーザが選択できるよう

にしてもよい。基準点は、骨格構造の中心もしくは中心よりも高い（画像の上下方向における上である）ことが好ましく、例えば、首のキーポイントの座標を基準点とする。なお、首に限らず頭やその他のキーポイントの座標を基準点としてもよい。キーポイントに限らず、任意の座標（例えば骨格構造の中心座標等）を基準点としてもよい。

[0143] 続いて、特徴量算出部103は、キーポイント高さ（ y_i ）を身長画素数で正規化する（S243）。特徴量算出部103は、各キーポイントのキーポイント高さ、基準点、身長画素数を用いて、各キーポイントを正規化する。具体的には、特徴量算出部103は、基準点に対するキーポイントの相対的な高さを身長画素数により正規化する。ここでは、高さ方向のみに着目する例として、Y座標のみを抽出し、また、基準点を首のキーポイントとして正規化を行う。具体的には、基準点（首のキーポイント）のY座標を（ y_c ）として、次の式（1）を用いて、特徴量（正規化値）を求める。なお、カメラパラメータに基づいた鉛直投影軸を用いる場合は、（ y_i ）及び（ y_c ）を鉛直投影軸に沿った方向の値に変換する。

[数1]

$$f_i = (y_i - y_c) / h \quad \dots (1)$$

[0144] 例えば、キーポイントが18個の場合、各キーポイントの18点の座標（ x_0, y_0 ）、（ x_1, y_1 ）、 $\dots$ （ x_{17}, y_{17} ）を、上記式（1）を用いて、次のように18次元の特徴量に変換する。

[数2]

$$\begin{aligned} f_0 &= (y_0 - y_c) / h \\ f_1 &= (y_1 - y_c) / h \\ &\vdots \\ f_{17} &= (y_{17} - y_c) / h \end{aligned} \quad \dots (2)$$

[0145] 図36は、特徴量算出部103が求めた各キーポイントの特徴量の例を示している。この例では、首のキーポイントA2を基準点とするため、キーポ

イントA 2の特徴量は0.0となり、首と同じ高さの右肩のキーポイントA 3 1及び左肩のキーポイントA 3 2の特徴量も0.0である。首よりも高い頭の特徴量は-0.2である。首よりも低い右手のキーポイントA 5 1及び左手のキーポイントA 5 2の特徴量は0.4であり、右足のキーポイントA 8 1及び左足のキーポイントA 8 2の特徴量は0.9である。この状態から人物が左手を挙げると、図37のように左手が基準点よりも高くなるため、左手のキーポイントA 5 2の特徴量は-0.4となる。一方で、Y軸の座標のみを用いて正規化を行っているため、図38のように、図36に比べて骨格構造の幅が変わっても特徴量は変わらない。すなわち、本実施の形態の特徴量（正規化値）は、骨格構造（キーポイント）の高さ方向（Y方向）の特徴を示しており、骨格構造の横方向（X方向）の変化に影響を受けない。

[0146] 以上のように、本実施の形態では、2次元画像から人物の骨格構造を検出し、検出した骨格構造から求めた身長画素数（2次元画像空間上の直立時の高さ）を用いて、骨格構造の各キーポイントを正規化する。この正規化された特徴量を用いることで、分類や検索等を行った場合のロバスト性を向上することができる。すなわち、本実施の形態の特徴量は、上記のように人物の横方向の変化に影響を受けないため、人物の向きや人物の体型の変化に対しロバスト性が高い。

[0147] さらに、本実施の形態では、OpenPose等の骨格推定技術を用いて人物の骨格構造を検出することで実現できるため、人物の姿勢等を学習する学習データを用意する必要がない。また、骨格構造のキーポイントを正規化し、データベースに格納しておくことで、人物の姿勢等の分類や検索が可能となるため、未知な姿勢に対しても分類や検索を行うことができる。また、骨格構造のキーポイントを正規化することで、明確でわかりやすい特徴量を得ることができるため、機械学習のようにブラックボックス型のアルゴリズムと異なり、処理結果に対するユーザの納得性が高い。

[0148] 以上、図面を参照して本発明の実施形態について述べたが、これらは本発

明の例示であり、上記以外の様々な構成を採用することもできる。

[0149] また、上述の説明で用いた複数のフローチャートでは、複数の工程（処理）が順番に記載されているが、各実施形態で実行される工程の実行順序は、その記載の順番に制限されない。各実施形態では、図示される工程の順番を内容的に支障のない範囲で変更することができる。また、上述の各実施形態は、内容が相反しない範囲で組み合わせることができる。

[0150] 上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限定されない。

1. 画像を取得する画像取得手段と、

ユーザ入力に基づき、前記画像に含まれる少なくとも1人の人物を選択する選択手段と、

前記画像に含まれる人物の2次元骨格構造を検出する骨格構造検出手段と、

、

前記検出された2次元骨格構造の特徴量を算出する特徴量算出手段と、

前記2次元骨格構造の特徴量の類似度に基づき、解析対象画像の中から、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する検索手段と、

を備える画像処理装置。

2. 前記選択手段は、

前記画像の中で少なくとも一部領域を指定するユーザ入力を受け、

指定された前記一部領域内で検出された人物を選択する1に記載の画像処理装置。

3. 前記選択手段は、

前記画像上に、前記画像内で検出された人物を選択可能な人物として識別可能に表示した加工画像を表示し、

前記選択可能な人物の中から少なくとも1人を選択するユーザ入力を受け、1に記載の画像処理装置。

4. 前記選択手段は、

前記 2 次元骨格構造の検出結果が所定条件を満たす人物を、前記選択可能な人物として識別可能に表示した前記加工画像を表示する 3 に記載の画像処理装置。

5. 前記骨格構造検出手段は、

前記 2 次元骨格構造を検出する処理において、身体複数のキーポイントを抽出し、

前記所定条件を満たす人物は、

抽出された前記キーポイントの数が基準値以上の人物、又は、

抽出された前記キーポイントの数、及び抽出された前記キーポイント各々の信頼度の少なくとも一方に基づき算出された評価値が基準値以上の人物である 4 に記載の画像処理装置。

6. 前記選択手段は、

前記画像の中で少なくとも一部領域を指定するユーザ入力に基づき、選択した前記人物の身体の一部を選択し、

前記検索手段は、

選択された前記身体の一部の重みを他の部分よりも大きくして、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する 1 から 5 のいずれかに記載の画像処理装置。

7. 前記選択手段は、

前記画像取得手段が取得した前記画像の中に含まれる人物の数を特定し、

特定した人物の数が複数である場合、ユーザ入力に基づき、前記画像に含まれる複数の人物の中から少なくとも 1 人を選択し、

特定した人物の数が 1 人である場合、その 1 人の人物を選択する 1 から 6 のいずれかに記載の画像処理装置。

8. コンピュータが、

画像を取得し、

ユーザ入力に基づき、前記画像に含まれる少なくとも 1 人の人物を選択

し、

前記画像に含まれる人物の２次元骨格構造を検出し、

前記検出された２次元骨格構造の特徴量を算出し、

前記２次元骨格構造の特徴量の類似度に基づき、解析対象画像の中から、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する画像処理方法。

９． コンピュータを、

画像を取得する画像取得手段、

ユーザ入力に基づき、前記画像に含まれる少なくとも１人の人物を選択する選択手段、

前記画像に含まれる人物の２次元骨格構造を検出する骨格構造検出手段

、

前記検出された２次元骨格構造の特徴量を算出する特徴量算出手段、

前記２次元骨格構造の特徴量の類似度に基づき、解析対象画像の中から、前記選択された人物の状態と類似する状態の人物を含む前記解析対象画像を検索する検索手段、

として機能させるプログラム。

符号の説明

[0151]	１	画像処理システム
	１０	画像処理装置
	１１	骨格検出部
	１２	特徴量算出部
	１３	認識部
	１００	画像処理装置
	１０１	画像取得部
	１０２	骨格構造検出部
	１０３	特徴量算出部
	１０４	分類部

1 0 5	検索部
1 0 6	入力部
1 0 7	表示部
1 0 8	身長算出部
1 0 9	選択部
1 1 0	データベース
2 0 0	カメラ
3 0 0、3 0 1	人体モデル
4 0 1	2次元骨格構造
4 0 2	3次元人体モデル