

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2015 年 12 月 10 日(10.12.2015)

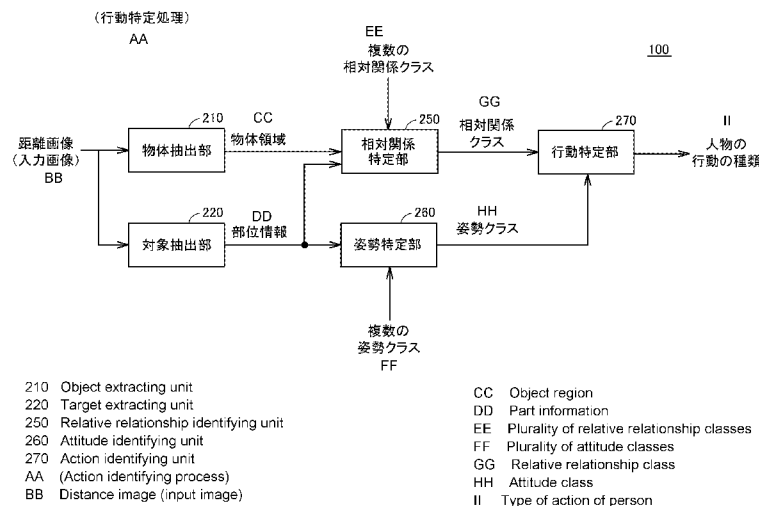


(10) 国際公開番号
WO 2015/186436 A1

- (51) 国際特許分類:
G06T 7/00 (2006.01) **H04N 5/225** (2006.01)
G06T 7/60 (2006.01) **H04N 5/232** (2006.01)
G08B 13/196 (2006.01) **H04N 7/18** (2006.01)
- (21) 国際出願番号: PCT/JP2015/061677
- (22) 国際出願日: 2015 年 4 月 16 日(16.04.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2014-117900 2014 年 6 月 6 日(06.06.2014) JP
- (71) 出願人: コニカミノルタ株式会社(KONICA MINOLTA, INC.) [JP/JP]; 〒1007015 東京都千代田区丸の内二丁目 7 番 2 号 Tokyo (JP).
- (72) 発明者: 指田 岳彦(SASHIDA, Takehiko); 〒1007015 東京都千代田区丸の内二丁目 7 番 2 号 コニカミノルタ株式会社内 Tokyo (JP).
- (74) 代理人: 特許業務法人深見特許事務所(FUKAMI PATENT OFFICE, P.C.); 〒5300005 大阪府大阪市北区中之島二丁目 2 番 7 号 中之島セントラルタワー Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).
- 添付公開書類:
— 国際調査報告 (条約第 21 条(3))

(54) Title: IMAGE PROCESSING DEVICE, IMAGE PROCESSING METHOD, AND IMAGE PROCESSING PROGRAM

(54) 発明の名称: 画像処理装置、画像処理方法、および、画像処理プログラム



(57) Abstract: The invention provides an image processing device, an image processing method and an image processing program capable of identifying to a high degree of accuracy various actions of a monitored target. An image processing device (100) identifies actions of a monitored target, which is a person or an animal. The image processing device (100) is provided with: a target extracting unit (220) which extracts a region in which the monitored target is present, from the three-dimensional positions of subjects contained in the field of view of a camera, and extracts part information indicating the positions of each characteristic part of the monitored target; an attitude identifying unit (260) which identifies the attitude of the monitored target from first relative relationships between the parts of the monitored target, calculated from the part information; a relative relationship identifying unit (250) which identifies second relative relationships between the monitored target and subjects other than the monitored target; and an action identifying unit (270) which identifies an action of the monitored target from a combination of the attitude of the monitored target, identified by the attitude identifying unit (260), and the second relative relationships.

(57) 要約:

[続葉有]

WO 2015/186436 A1



監視対象の様々な行動を高精度に特定することが可能な画像処理装置、画像処理方法、および画像処理プログラムを提供する。画像処理装置（１００）は、人物または動物である監視対象の行動を特定する。画像処理装置（１００）は、カメラの視野に含まれる被写体の３次元位置から、監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するための対象抽出部（２２０）と、部位情報から算出した監視対象の各部位間の第１の相対関係から監視対象の姿勢を特定するための姿勢特定部（２６０）と、監視対象と、当該監視対象以外の被写体との間の第２の相対関係を特定するための相対関係特定部（２５０）と、姿勢特定部（２６０）によって特定された監視対象の姿勢と、第２の相対関係との組み合わせから、監視対象の行動を特定するための行動特定部（２７０）とを備える。

明 細 書

発明の名称：

画像処理装置、画像処理方法、および、画像処理プログラム

技術分野

[0001] 本開示は画像処理に関し、より特定的には、人物または動物である監視対象の行動を特定するための画像処理に関する。

背景技術

[0002] 近年、画像処理の分野においては、画像に写っている監視対象（たとえば、人物や動物）の姿勢や行動を特定するための様々な技術が開発されている。たとえば、特許文献1は、高速かつ高精度に人の姿勢を推定する姿勢推定装置を開示している。非特許文献1は、カメラからの画像から人物の行動を認識する方法を開示している。

先行技術文献

特許文献

[0003] 特許文献1：特開2013-109696号公報

非特許文献

[0004] 非特許文献1：青木茂樹，岩井嘉男，大西正輝，小島篤博，福永邦雄、”人物の位置・姿勢に注目した行動パターンの学習・認識と非日常状態検出への応用”、電子情報通信学会論文誌 D-II，情報・システム，II-パターン処理 J87-D-II(5)，p.1083-1093

発明の概要

発明が解決しようとする課題

[0005] ところで、近年では、監視カメラに映る人物の不正行為（たとえば、万引きなど）を検出したり、病院における患者の行動を見守ったり、様々な種類の行動を検出することが望まれている。

[0006] 特許文献1に開示される姿勢推定装置は、人物の姿勢を推定するものであ

り、人物の行動を特定するものではない。このため、当該姿勢推定装置は、様々な種類の行動を検出するという世の中のニーズを満たすことができない。

[0007] 非特許文献 1 に開示される行動認識方法は、画像に写る人物の手や足などの部位間の位置関係で人物の行動を特定する。このように、部位間の位置関係が用いられるだけでは、特定できる人物の行動の種類が限られる。

[0008] この開示は上述のような問題点を解決するためになされたものであって、ある局面における目的は、監視対象の様々な行動を高精度に特定することが可能な画像処理装置、画像処理方法、および画像処理プログラムを提供することである。

課題を解決するための手段

[0009] 一実施の形態に従うと、人物または動物である監視対象の行動を特定するための画像処理装置が提供される。当該画像処理装置は、カメラの視野に含まれる被写体の 3 次元位置から、監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するための対象抽出部と、部位情報から算出した監視対象の各部位間の第 1 の相対関係から監視対象の姿勢を特定するための姿勢特定部と、監視対象と、当該監視対象以外の被写体との間の第 2 の相対関係を特定するための相対関係特定部と、姿勢特定部によって特定された監視対象の姿勢と、第 2 の相対関係との組み合わせから、監視対象の行動を特定するための行動特定部とを備える。

[0010] 好ましくは、姿勢特定部は、第 1 の相対関係が、監視対象の姿勢の種類を示す複数の姿勢クラスのうちいずれに属するかを特定する。相対関係特定部は、第 2 の相対関係が、監視対象と当該監視対象以外の被写体との間の相対関係の種類を示す複数の相対関係クラスのうちいずれに属するかを特定する。行動特定部は、姿勢特定部によって特定された姿勢クラスと、相対関係特定部によって特定された相対関係クラスとの組み合わせから、監視対象の行動の種類を特定する。

- [0011] 好ましくは、対象抽出部は、異なる時間に得られた３次元位置から、監視対象が存在する領域を順次抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を順次抽出する。画像処理装置は、複数の部位情報を、類似する部位情報ごとにクラスタリングして複数の姿勢クラスを生成するための姿勢クラスタリング部と、対象抽出部により順次抽出された監視対象と、当該監視対象以外の被写体との間の第２の相対関係が互いに類似している集合ごとにクラスタリングして複数の相対関係クラスを生成するための相対関係クラスタリング部とをさらに備える。
- [0012] 好ましくは、姿勢クラスタリング部は、対象抽出部によって順次抽出された複数の部位情報の各々から監視対象の各部位間の第１の相対関係を算出し、第１の相対関係が互いに類似する集合ごとに姿勢クラスを生成する。
- [0013] 好ましくは、画像処理装置は、３次元位置から、監視対象以外の被写体である物体が存在する領域を抽出するための物体抽出部をさらに備える。第２の相対関係は、監視対象と物体との間の相対的な位置関係を含む。
- [0014] 好ましくは、相対関係クラスタリング部は、対象抽出部によって順次抽出された複数の部位情報の各々から、監視対象の特徴的な各部位について、当該部位と物体との間の距離が互いに類似している集合ごとに相対関係クラスを生成する。
- [0015] 好ましくは、第２の相対関係は、監視対象の各関節と物体との間の各距離、および、当該監視対象の各端部と物体との間の各距離の少なくとも一方を含む。
- [0016] 好ましくは、部位情報は、監視対象の各関節の位置、および、当該監視対象の各端部の位置の少なくとも一方を含む。
- [0017] 好ましくは、相対関係特定部は、監視対象の特徴的な各部位と、当該監視対象以外の被写体との間の距離の各々が、距離の長さに応じて予め分けられた複数の相対関係クラスのうちのいずれに属するかを特定する。
- [0018] 好ましくは、画像処理装置は、類似する位置情報の集合ごとに３次元位置を分割して、被写体を複数の領域に分割するための分割部をさらに備える。

第2の相対関係は、監視対象の特徴的な各部位と複数の領域の各々との間の相対関係を含む。

[0019] 好ましくは、複数の姿勢クラスと複数の相対関係クラスとの組み合わせのそれぞれには、監視対象の行動の種類を示す行動カテゴリがそれぞれ対応付けられている。行動特定部は、姿勢特定部によって特定された姿勢クラスと相対関係特定部によって特定された相対関係クラスとの組み合わせに対応する行動カテゴリを特定する。

[0020] 他の実施の形態に従うと、人物または動物である監視対象の行動を特定するための画像処理方法が提供される。当該画像処理方法は、カメラの視野に含まれる被写体の3次元位置から、監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するステップと、部位情報から算出した監視対象の各部位間の第1の相対関係から監視対象の姿勢を特定するステップと、監視対象と、当該監視対象以外の被写体との間の第2の相対関係を特定するステップと、姿勢と第2の相対関係との組み合わせから、監視対象の行動を特定するステップとを備える。

[0021] さらに他の実施の形態に従うと、人物または動物である監視対象の行動を特定するための画像処理プログラムが提供される。当該画像処理プログラムは、コンピュータに、カメラの視野に含まれる被写体の3次元位置から、監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するステップと、部位情報から算出した監視対象の各部位間の第1の相対関係から監視対象の姿勢を特定するステップと、監視対象と、当該監視対象以外の被写体との間の第2の相対関係を特定するステップと、姿勢と第2の相対関係との組み合わせから、監視対象の行動を特定するステップとを実行させる。

発明の効果

[0022] ある局面において、監視対象の様々な行動を高精度に特定することができる。

本発明の上記および他の目的、特徴、局面および利点は、添付の図面と関

連して理解される本発明に関する次の詳細な説明から明らかとなるであろう。

図面の簡単な説明

[0023] [図1]同じ姿勢をとりながら異なる行動をとっている人物の様子を示す図である。

[図2]実施の形態に従う画像処理装置に接続されたカメラが被写体を撮像している様子を表している図である。

[図3]実施の形態に従う画像処理装置が姿勢クラスを生成する過程を概略的に示した概念図である。

[図4]実施の形態に従う画像処理装置が相対関係クラスを生成する過程を概略的に示した概念図である。

[図5]人物の行動を特定するための処理を概略的に示した概念図である。

[図6]実施の形態に従う画像処理装置のハードウェア構成の一例を示すブロック図である。

[図7]実施の形態に従う画像処理装置のハードウェア構成の一例を示すブロック図である。

[図8]実施の形態に従う実施の形態に従う画像処理装置の機能構成の一例を示すブロック図である。

[図9]カメラが椅子を撮像している様子を示している図である。

[図10]人物の特徴的な部位を模式的に表した図である。

[図11]部位情報から算出された各部位間の相対関係のデータ構造を表した図である。

[図12]人物の特徴的な各部位と物体との相対的な位置関係を示した図である。

[図13]物体相対関係の分布を示した図である。

[図14]相対関係クラスのデータ構造を示した図である。

[図15]細分化した姿勢クラスを視覚的に表した図である。

[図16]部位間関係の分布を示した図である。

[図17]姿勢クラスのデータ構造を示した図である。

[図18]物体相対関係の分布を示した図である。

[図19]部位間関係の分布を示した図である。

[図20]行動カテゴリテーブルのデータ構造を示した図である。

[図21]一実施の形態に従う画像処理装置が実行する学習処理の一部を表わすフローチャートである。

[図22]一実施の形態に従う画像処理装置が実行する行動特定処理の一部を表わすフローチャートである。

発明を実施するための形態

[0024] 以下、図面を参照しつつ、本実施の形態について説明する。以下の説明では、同一の部品および構成要素には同一の符号を付してある。それらの名称および機能も同じである。したがって、これらについての詳細な説明は繰り返さない。なお、以下で説明される各実施の形態、および／または、各変形例は、選択的に組み合わせられてもよい。

[0025] [技術原理]

まず、以下の実施の形態に従う画像処理装置100の理解を容易にするために、図1を参照して、画像処理装置100が利用する技術原理について説明する。図1は、同じ姿勢をとりながら異なる行動をとっている人物の様子を示す図である。

[0026] 図1(A)および図1(B)のそれぞれには、同じように指を前に突き出す姿勢をとっている人物25が示される。図1(A)には、人物25が人物27を指さしている様子が示される。図1(B)には、人物25がボタン29を押下している様子が示される。すなわち、人物25は、同じ姿勢をとりながら、異なる行動をとっている。このように、人物の行動は、人の姿勢情報だけでは区別できない場合がある。本実施の形態において、姿勢情報とは、人物の姿勢を特定するための情報のことをいう。

[0027] そこで、本実施の形態に従う画像処理装置100は、監視対象の被写体である人物の姿勢情報だけでなく、人物の周囲にある監視対象以外の被写体の

物体の位置情報をさらに用いて人物の行動を特定する。たとえば、図 1（A）に示される人物 27 を指す行動と、図 1（B）に示されるボタン 29 を押下する行動とは、人物 25 の指の前に物体があるか否かで見分けることができる。このように、人物の周囲にある物体の位置関係は、人物の行動を特定する上で重要な情報になり得る。この点に着目して、画像処理装置 100 は、人物の姿勢情報だけでなく、人物の周囲にある物体の位置情報を用いて、人物の行動を特定する。

[0028] より具体的には、人の姿勢は、手や足などの位置関係や、腕や足などの関節の角度で決まる。このため、画像処理装置 100 は、まず、行動を特定する対象となる人物の手や足などの特徴的な各部位間の相対関係を特定することで人物の姿勢を特定する。本実施の形態において、特徴的な部位とは、人物の姿勢を特定するのに適した部位のことをいう。また、画像処理装置 100 は、人物と、当該人物以外の被写体との相対関係（以下、「物体相対関係」とも称する。）を特定する。画像処理装置 100 は、特定された人物の姿勢と物体相対関係との組み合わせから人物の行動を特定する。このように、物体相対関係が用いられることにより、画像処理装置 100 は、人物の姿勢情報だけでは区別することができない行動を特定することが可能になる。この結果、画像処理装置 100 は、人物の様々な行動を高精度に特定できる。

[0029] 画像処理装置 100 は、人物の行動を特定することにより、監視カメラに映る人物の不正行為（たとえば、万引きなど）を検出したり、病院における患者の行動を見守ったりすることができる。

[0030] なお、行動を特定する対象となる監視対象は、人物だけでなく動物も含み得る。画像処理装置 100 は、動物の行動を特定することにより、監視カメラに映るペットの行動を監視したり、進入が禁止されているエリアに突入した動物を検出したりすることができる。

[0031] [概要]

図 2 ～図 5 を参照して、本実施の形態に従う画像処理装置 100 の概要について説明する。

[0032] (3次元位置)

まず、図2を参照して、画像処理装置100が人物の行動検出に用いる情報である3次元位置について説明する。図2は、画像処理装置100に接続されたカメラ110が被写体を撮像している様子を表している図である。

[0033] 図2に示されるように、画像処理装置100には、カメラ110が接続されている。カメラ110の視野には、人物31が含まれている。画像処理装置100は、カメラ110の視野に含まれる被写体の3次元位置を取得する。3次元位置は、たとえば、カメラ110から被写体の各点までの距離を示した距離画像から取得される。取得された3次元位置は、人物の各部位の位置や、物体の位置などを特定するために用いられる。

[0034] なお、以下では、「3次元位置」とは「距離画像」という概念も含み、「距離画像」とは「3次元位置」という概念も含む。

[0035] (姿勢クラス)

図3を参照して、人物の姿勢の種類を示す姿勢クラスの生成方法の概要について説明する。図3は、画像処理装置100が姿勢クラスを生成する過程を概略的に表した概念図である。人物の姿勢は、手や足などの位置関係や、各関節の角度などで決まる。このため、画像処理装置100は、これらの情報を類似するものごとに自動的にグループ化して、人物の姿勢を示す姿勢クラスを生成する。これにより、画像処理装置100は、人物の姿勢をカテゴライズすることができる。

[0036] より具体的には、カメラ110は、人物の様々な姿勢を撮影するために、人物31を撮影し続ける。これにより、図3に示されるように、画像処理装置100は、人物の様々な姿勢を写した距離画像30を収集することができる。画像処理装置100は、距離画像30から人物を検出するとともに、距離画像30内の人物領域を順次抽出する。画像処理装置100は、抽出した人物領域から、人物の特徴的な各部位の位置からなる部位情報を順次抽出する。部位情報は、カメラ110の視野に含まれる人物の各関節の位置、および、当該人物の各端部の位置の少なくとも一方を含む。

[0037] 画像処理装置 100 は、抽出した部位情報を基に、人物の特徴的な各部位間の相対関係（以下、「部位間関係」ともいう。）を算出する。画像処理装置 100 は、部位間関係が互いに類似している部位情報の集合ごとに姿勢クラスを生成する。これにより、画像処理装置 100 は、撮影中に人物がとった姿勢を、類似する姿勢ごとに自動的にグループ化することができる。

[0038] たとえば、撮影された人物が座る姿勢をとった場合には、座る姿勢を示す姿勢クラス 41 が生成される。撮影された人物が指をさす姿勢をとった場合には、指をさす姿勢を示す姿勢クラス 42 が生成される。撮影された人物が寝転ぶ姿勢をとった場合には、寝転ぶ姿勢を示す姿勢クラス 43 が生成される。撮影された人物が物を保持する姿勢をとった場合には、保持姿勢を表わす姿勢クラス 43 が生成される。

[0039] （相対関係クラス）

図 4 を参照して、人物と物体との相対的な位置関係を示す相対関係クラスの生成方法の概要について説明する。図 4 は、画像処理装置 100 が相対関係クラスを生成する過程を概略的に示した概念図である。上述したように、人物と周囲の物体との位置関係は、人物の行動を特定する上で重要な情報となり得る。このため、画像処理装置 100 は、人物と周囲の物体との位置関係（すなわち、物体相対関係）をカテゴライズし、これにより、人物の行動を特定する。

[0040] より具体的には、カメラ 110 は、人物 31 が周囲の物体に対してとった様々な行動を撮影するために、人物 31 と人物 31 の周辺とを撮影し続ける。これにより、図 4 に示されるように、画像処理装置 100 は、人物と物体との関係を写した様々な距離画像 30 を収集することができる。画像処理装置 100 は、距離画像 30 から人物と物体とを検出する。画像処理装置 100 は、距離画像 30 内の人物の位置と物体の位置との相対関係を算出する。

[0041] 画像処理装置 100 は、物体相対関係が互いに類似する集合ごとに相対関係クラスを生成する。これにより、画像処理装置 100 は、撮影中に物体に対して人物がとった姿勢を、類似するものごとに自動的にグループ化するこ

とができる。

[0042] たとえば、椅子に座る行動をとる人物が撮影された場合には、人物の位置と椅子の位置との相対的な位置関係を示す相対関係クラス 5 1 が生成される。箱を運ぶ行動をとる人物が撮影された場合には、人物の位置と箱の位置との相対的な位置関係を示す相対関係クラス 5 2 が生成される。ボタンを押下する行動をとる人物が撮影された場合には、人物の位置とボタンの位置との相対的な位置関係を示す相対関係クラス 5 3 が生成される。他人を指さす行動をとる人物が撮影された場合には、人物の周囲には物体が無いことを示す相対関係クラス 5 4 が生成される。

[0043] (行動の特定方法)

図 5 を参照して、人物の行動の特定方法について説明する。図 5 は、人物の行動を特定するための処理を概略的に示した概念図である。画像処理装置 100 は、人物の姿勢と、人物および物体間の相対関係（すなわち、物体相対関係）との組み合わせから人物の行動を特定する。

[0044] より具体的には、画像処理装置 100 は、入力画像として、カメラ 110 から距離画像を取得する。画像処理装置 100 は、入力画像から人物を検出し、人物が存在する入力画像内の領域（以下、「人物領域」とも称する。）を抽出する。また、画像処理装置 100 は、入力画像から、人物が有する特徴的な各部位の位置（すなわち、部位情報）を抽出する。画像処理装置 100 は、人物の姿勢を特定するために、抽出した部位情報から各部位間の相対関係を算出する。また、画像処理装置 100 は、人物が物体に対してとった行動を特定するために、人物と当該人物以外の被写体との相対関係（すなわち、物体相対関係）を特定する。画像処理装置 100 は、特定した人物の姿勢と物体相対関係との組み合わせから、当該人物の行動を特定する。

[0045] 以下、図 5 を参照して、画像処理装置 100 が人物の行動を特定する具体例について説明する。図 5 に示されるように、人物 31 がボタン 62 を押下している様子がカメラ 110 によって撮影されたとする。これにより、画像処理装置 100 は、入力画像として、人物 31 がボタン 62 を押下している

様子を写した入力画像 60 を得る。画像処理装置 100 は、入力画像 60 から人物 61 を含む領域を抽出するとともに、抽出した領域から人物 61 の部位情報を抽出し、さらに、抽出した部位情報から部位間関係を算出する。

[0046] 画像処理装置 100 は、人物 61 の部位情報が、予め登録されている複数の姿勢クラスのいずれに属するかを特定する。図 5 の例では、図 3 の例で生成された姿勢クラス 41 ～ 44 が登録されている。この場合、画像処理装置 100 は、入力画像 60 に写る人物 61 の姿勢に類似する姿勢クラス 42 が姿勢情報として特定される。

[0047] また、画像処理装置 100 は、入力画像 60 からボタン 62 の領域を抽出するとともに、人物 61 とボタン 62 との間の位置関係（すなわち、物体相対関係）を特定する。画像処理装置 100 は、特定した物体相対関係が、予め登録されている相対関係クラスのうちいずれに属するかを特定する。図 5 の例では、図 4 の例で生成された相対関係クラス 51 ～ 54 が登録されている。この場合、画像処理装置 100 は、入力画像 60 に写る人物 61 とボタン 62 との物体相対関係が属する相対関係クラスとして、相対関係クラス 54 を特定する。

[0048] 画像処理装置 100 は、特定した姿勢クラス 42 と相対関係クラス 53 との組み合わせから、入力画像に含まれる人物 61 の行動を特定する。人物の行動の種類を示す行動カテゴリは、姿勢クラスと相対関係との組み合わせごとに予め定められているとする。画像処理装置 100 は、姿勢クラス 42 と相対関係クラス 53 との組み合わせに対応する行動カテゴリ「ボタン押下」を人物 61 のとった行動として特定する。

[0049] [ハードウェア構成]

図 6 は、画像処理装置 100 のハードウェア構成の一例を示すブロック図である。図 6 を参照して、画像処理装置 100 のハードウェア構成について説明する。

[0050] 図 6 を参照して、画像処理装置 100 は、主として、汎用的なアーキテクチャを有するコンピュータ上に実装される。画像処理装置 100 は、主た

るコンポーネントとして、ROM (Read Only Memory) 1と、CPU (Central Processing Unit) 2と、RAM (Random Access Memory) 3と、カメラインターフェイス (I/F) 4と、メモリーカードインターフェイス (I/F) 5と、ネットワークインターフェイス (I/F) 6と、記憶装置 20とを含む。

[0051] ROM 1は、画像処理装置 100において起動時に実行される初期プログラム（ブートプログラム）などを格納する。CPU 2は、ROM 1や記憶装置 20などに格納された、オペレーティングシステム（OS : Operating System）や画像処理プログラム 24などの各種プログラムを実行することで、画像処理装置 100の全体を制御する。RAM 3は、CPU 2でプログラムを実行するためのワーキングメモリとして機能し、プログラムの実行に必要な各種データを一次的に格納する。

[0052] カメラ I/F 4は、CPU 2とカメラ 110とのデータ通信を仲介する。たとえば、カメラ I/F 4は、画像バッファを含み、カメラ 110から伝送される距離画像を一旦蓄積する。カメラ I/F 4は、少なくとも1枚分の入力画像のデータが蓄積されると、その蓄積されたデータを記憶装置 20またはROM 1へ転送する。また、カメラ I/F 4は、CPU 2が発生した内部コマンドに従って、カメラ 110に対して撮像指令を与える。カメラ 110は、たとえば、ステレオカメラ、被写体の3次元上の位置情報を検出する距離画像センサ、被写体の3次元上の位置を取得することができるその他のカメラを含む。なお、カメラ 110は、画像処理装置 100に組み込まれてもよく、この場合には、カメラ I/F 4は、必須の構成ではない。

[0053] メモリーカード I/F 5は、SD (Secure Digital) カードやCF (Compact Flash (登録商標)) カードなどの各種メモリーカード（不揮発性記憶媒体）5Aとの間で、データの読み書きを行なう。典型的には、メモリーカード I/F 5には、他の装置で取得した距離画像を格納したメモリーカード 5Aが装着され、そのメモリーカード 5Aから読み出された距離画像が記憶装置 20へ格納される。

- [0054] ネットワーク 1 / F 6 は、各種の通信媒体を介して、他の装置（サーバ装置など）とデータをやり取りする。より具体的には、ネットワーク 1 / F 6 は、イーサネット（登録商標）などの有線回線（L A N（Local Area Network）やW A N（Wide Area Network）など）、および／または、無線L A Nなどの無線回線を介してデータ通信を行なう。
- [0055] 記憶装置 2 0 は、典型的には、ハードディスクなどの大容量磁気記憶媒体などを含む。記憶装置 2 0 は、行動カテゴリテーブル 2 2 と、本実施の形態に従う各種の画像処理を実現するための画像処理プログラム 2 4 を格納する。行動カテゴリテーブル 2 2 についての詳細は後述する。さらに、記憶装置 2 0 には、オペレーティングシステムなどのプログラムが格納されてもよい。
- [0056] カメラ 1 1 0 は、外付けではなく、画像処理装置 1 0 0 の内部に組み込まれて、画像処理装置 1 0 0 本体が被写体を撮像する機能を有していてもよい。また、画像処理装置 1 0 0 は、カメラに類似した機構を用いて距離画像を取得し、取得した距離画像を任意の方法で画像処理装置 1 0 0 へ入力するように構成されてもよい。この場合、距離画像は、メモリーカード 1 / F 5 またはネットワーク 1 / F 6 を介して画像処理装置 1 0 0 に入力される。
- [0057] 記憶装置 2 0 に格納される画像処理プログラム 2 4 は、C D - R O M（Compact Disk-Read Only Memory）などの記憶媒体に格納されて流通し、あるいは、ネットワークを介してサーバ装置などから配信される。画像処理プログラム 2 4 は、画像処理装置 1 0 0 で実行されるオペレーティングシステムの一部として提供されるプログラムモジュールのうち必要なモジュールを、所定のタイミングおよび順序で呼出して処理を実現するようにしてもよい。この場合、画像処理プログラム 2 4 自体には、オペレーティングシステムによって提供されるモジュールは含まれず、オペレーティングシステムと協働して画像処理が実現される。
- [0058] また、画像処理プログラム 2 4 は、単体のプログラムではなく、任意のプログラムの一部に組み込まれて提供されてもよい。このような場合にも、画

画像処理プログラム 24 自体には、任意のプログラムにおいて共通に利用されるようなモジュールは含まれず、任意のプログラムと協働して画像処理が実現される。このような一部のモジュールを含まない画像処理プログラム 24 であっても、本実施の形態に従う画像処理装置 100 の趣旨を逸脱するものではない。さらに、画像処理プログラム 24 によって提供される機能の一部または全部は、専用のハードウェアによって実現されてもよい。

[0059] さらに、画像処理装置 100 は、必ずしも、リアルタイムに処理を行わなくてもよい。たとえば、少なくとも 1 つのサーバ装置が本実施の形態に従う処理を実現する、いわゆるクラウドサービスのような形態で画像処理装置 100 が構成されてもよい。この場合、距離画像がサーバ装置（クラウド側）へ送信され、サーバ装置は、受信した距離画像に対して、本実施の形態に従う画像処理を行なう。さらに、サーバ装置側が全ての機能（処理）を行なう必要はなく、ユーザ側の端末とサーバ装置とが協働して、本実施の形態に従う画像処理を実現するようにしてもよい。

[0060] [機能構成]

図 7 および図 8 を参照して、画像処理装置 100 の機能構成について説明する。図 7 および図 8 は、実施の形態に従う画像処理装置 100 の機能構成の一例を示すブロック図である。画像処理装置 100 は、動作モードとして、上述の姿勢クラスおよび相対関係クラスを生成する学習モードと、入力画像に写る人物の行動を特定する行動特定モードとを有する。

[0061] (学習処理)

まず、図 7 を参照して、学習モード時に実行される学習処理を実現するための画像処理装置 100 の機能構成について説明する。画像処理装置 100 は、動作モードが学習モードである場合には、学習画像である複数の距離画像から姿勢クラスと相対関係クラスとを生成する。学習処理には、たとえば、DecisionTree、DeepLearningなどの画像処理技術が用いられる。画像処理装置 100 は、物体抽出部 210 と、対象抽出部 220 と、相対関係クラスタリング部 230 と、姿勢クラスタリング部 240 とを含む。

[0062] 物体抽出部 210 は、距離画像（学習画像）から監視対象以外の被写体である物体が存在する領域を抽出する。より具体的には、学習画像において物体領域を抽出し、抽出した物体領域から物体の 3 次元上の位置を検出する。物体抽出部 210 の詳細については後述する。物体抽出部 210 は、学習画像内の物体領域を相対関係クラスタリング部 230 に出力する。

[0063] 対象抽出部 220 は、異なる時間に得られた距離画像（学習画像）から監視対象の被写体である人物を検出し、当該人物の部位情報を順次抽出する。対象抽出部 220 の詳細については後述する。対象抽出部 220 は、抽出した部位情報を、相対関係クラスタリング部 230 および姿勢クラスタリング部 240 に出力する。

[0064] 相対関係クラスタリング部 230 は、対象抽出部 220 により順次抽出された人物領域と、当該人物領域の周辺の被写体との間の相対関係（すなわち、物体相対関係）が互いに類似している集合ごとにクラスタリングして複数の相対関係クラスを生成する。より具体的には、相対関係クラスタリング部 230 は、対象抽出部 220 によって順次抽出された複数の部位情報の各々から、人物の特徴的な各部位について、当該部位と物体との間の距離が互いに類似している集合ごとに相対関係クラスを生成する。相対関係クラスタリング部 230 の詳細については後述する。

[0065] 姿勢クラスタリング部 240 は、対象抽出部 220 によって順次抽出された複数の部位情報の各々から監視対象の各部位間の相対関係を算出し、すなわち、姿勢クラスタリング部 240 は、類似する姿勢ごとにグルーピングする。姿勢クラスタリング部 240 の詳細については後述する。

[0066] （行動特定処理）

次に、図 8 を参照して、行動特定モード時に実行される行動特定処理を実現するための画像処理装置 100 の機能構成について説明する。画像処理装置 100 は、物体抽出部 210 と、対象抽出部 220 と、相対関係特定部 250 と、姿勢特定部 260 と、行動特定部 270 とを含む。画像処理装置 100 は、動作モードが行動特定モードである場合には、入力画像を取得して

、人物の行動を特定する。

[0067] 物体抽出部 210 は、距離画像（入力画像）から監視対象以外の被写体である物体が存在する領域を抽出する。より具体的には、入力画像において物体領域を抽出し、抽出した物体領域から物体の 3 次元上の位置を検出する。物体抽出部 210 は、入力画像内の物体領域を相対関係特定部 250 に出力する。

[0068] 対象抽出部 220 は、入力画像から人物を検出し、入力画像内の人物領域を抽出する。対象抽出部 220 は、入力画像内の人物領域から部位情報を抽出する。対象抽出部 220 の詳細については後述する。対象抽出部 220 は、入力画像内の部位情報を相対関係クラスタリング部 230 および姿勢クラスタリング部 240 に出力する。

[0069] 相対関係特定部 250 は、入力画像内の人物領域と、当該人物の周辺領域との間の物体相対関係を特定する。より具体的には、相対関係特定部 250 は、特定した物体相対関係が、人物と当該人物の周辺との間の相対的な関係の種類を示す複数の相対関係クラスのいずれに属するかを特定する。相対関係特定部 250 の詳細については後述する。相対関係特定部 250 は、特定した相対関係クラスを行動特定部 270 に出力する。

[0070] 姿勢特定部 260 は、部位情報から算出した人物の各部位間の相対関係（すなわち、部位間関係）から人物の姿勢を特定する。より具体的には、姿勢特定部 260 は、算出した部位間関係が、人物の姿勢の種類を示す複数の姿勢クラスのいずれに属するかを特定する。姿勢特定部 260 の詳細については後述する。姿勢特定部 260 は、特定した姿勢クラスを行動特定部 270 に出力する。

[0071] 行動特定部 270 は、姿勢特定部 260 によって特定された人物の姿勢と、相対関係特定部 250 によって特定された当該人物の物体相対関係との組み合わせから、入力画像内の人物の行動を特定する。より具体的には、行動特定部 270 は、相対関係特定部 250 によって特定された相対関係クラスと、姿勢特定部 260 によって特定された姿勢クラスとの組み合わせから、

入力画像に写る人物がとる行動の種類を示す行動カテゴリを特定する。行動特定部 270 の詳細については後述する。

[0072] [物体抽出部 210 の詳細]

図 9 を参照して、物体抽出部 210 の詳細について説明する。図 9 は、カメラ 110 が椅子 32 を撮像している様子を示している図である。

[0073] 物体抽出部 210 は、距離画像センサなどのカメラ 110 から得られた 3 次元位置から、監視対象以外の被写体である物体の位置を抽出する。図 9 に示される例においては、物体抽出部 210 は、3 次元位置から地面の推定を行い、地面よりも上にある領域を物体として抽出する。すなわち、物体抽出部 210 は、図 9 に示されるように、領域 211 に対応する位置を物体の位置として特定する。なお、図 9 の例においては、説明を簡単にするために、物体の「位置」を一次元（ x 軸）で表しているが、実際には、「位置」は、2 次元（ x 軸， y 軸）であるため、物体の位置は、2 次元（ x 軸， y 軸）で表される。また、物体の位置は、3 次元（ x 軸， y 軸， z 軸）で表されてもよい。

[0074] 他にも、物体抽出部 210 は、人物がいないときに撮影して得られた背景画像を、取得した距離画像から差分する所謂背景差分により、物体の位置を特定してもよい。この場合、物体抽出部 210 は、背景画像の撮影時（初期状態）から変化した領域を物体の位置として特定する。なお、人物の位置は、対象抽出部 220 により特定されるので、物体抽出部 210 は、特定した物体から人物の位置を除外することができる。また、物体抽出部 210 は、地面自体が一つの物体として特定してもよい。

[0075] 他にも、物体抽出部 210 は、距離画像センサなどで得た 3 次元情報を基に、人物の周辺の物体の種類を特定してもよい。物体の種類を特定する方法としては、一般物体認識技術や特定物体認識技術が用いられる。たとえば、Bag-of-Keypointga が特徴抽出処理で用いられる。ニューラルネットワーク（Neural network）、サポートベクターマシン（SVM : Support Vector Machine）が学習処理で用いられる。

[0076] [対象抽出部 220 の詳細]

図 10 および図 11 を参照して、対象抽出部 220 の詳細について説明する。図 10 は、人物の特徴的な部位を模式的に表した図である。図 11 は、部位情報から算出された各部位間の相対関係（すなわち、部位間関係）のデータ構造を表した図である。

[0077] 対象抽出部 220 では、距離画像センサなどのカメラ 110 から得られた 3 次元情報を基に、人物の検出と、当該人物の姿勢の認識とを行なう。ここでいう姿勢は、人の関節や、頭部などの端部のような特徴的な部位がどのように配置されているかという、身体部位の配置情報で定義される。

[0078] 対象抽出部 220 は、たとえば、マイクロソフト（登録商標）の Kinect（登録商標）センサをカメラ 110 として用いて、人物の各部位の 3 次元位置を示すスケルトン情報を取得する。Kinect センサにより、関節のような特徴的な部位（以下では、「節点」ともいう。）と、各節点の配置とを検出することできる。Kinect センサにより、たとえば、図 10 に示されるような、人体の節点（1）～（15）の 3 次元位置が取得できる。

[0079] 対象抽出部 220 は、Kinect センサにより得られた、節点（1）～（15）の 3 次元位置から、各節点間の相対的な関係（すなわち、部位間関係）を特定する。部位間関係は、各節点のうちの任意の 2 つの接点間の距離、各節点のうちの任意の 3 つの接点で定義される角度などを含む。一例として、図 10 に示されるように、部位間関係は、腕の角度 α 、腕の角度 β 、足の角度 γ 、および足の角度 δ などを含む。

[0080] 図 11 には、部位間関係 23 のデータ構造が示される。部位間関係 23 は、上述した、各節点のうちの任意の 2 つの接点間の距離や、各節点のうちの任意の 3 つの接点で定義される角度などを含む。また、部位間関係 23 には、各節点の 3 次元上の位置や、各節点のうちの任意の 4 つ以上の接点の相対関係などが定義されてもよい。相対関係を示すための節点数を多くすることで姿勢をより正確に表わすことができる。

[0081] なお、上記では、Kinect センサにより人物の検出と、当該人物の姿勢の認

識とを行なう例について説明したが、人物を検出するための手段および人物の姿勢を認識するための手段は、これらに限定されるわけではない。たとえば、対象抽出部220は、予め登録された人物のテンプレートや、人物の各部位のテンプレートを用いて、距離画像に対してテンプレートマッチングを行なうことにより人物の位置や、各部位の位置を抽出してもよい。その他にも、人物の位置および各部位の位置を特定できる方法であればいずれの方法であってもよい。

[0082] また、画像処理装置100は、Kinectセンサの代わりに、TOF (Time of Flight) カメラや、ステレオカメラを用いてもよい。

[0083] [相対関係クラスタリング部230の詳細]

図12～図14を参照して、相対関係クラスタリング部230の詳細について説明する。図12は、人物の特徴的な各部位と物体との相対的な位置関係を示した図である。図13は、物体相対関係の分布を示した図である。図14は、相対関係クラス55のデータ構造を示した図である。

[0084] 相対関係クラスタリング部230は、人物と当該人物の周辺環境情報との間の位置関係を測定することで物体相対関係を特定する。物体相対関係は、人物と物体との間の相対的な位置関係で定義される。より具体的には、物体相対関係は、人物の各関節と物体との間の各距離、および、当該人物の各端部と物体との間の各距離の少なくとも一方を含む。たとえば、物体相対関係には、図12に示されるように、人の関節や、頭部、手先のような人の端部など、人物の特徴的な各部位と物体との間の相対距離が用いられる。

[0085] 当該相対距離には、各端点と周辺環境の各3次元位置とのユークリッド距離が用いられる。他にも、当該相対距離には、物体認識技術や、物体の領域分割技術によって、周辺環境情報を分割した場合、各特徴点と分割領域ごとの最短距離や、各分割領域の重心までの距離などが用いられてもよい。

[0086] 当該相対距離は、算出した距離をそのまま用いてもよいが、各特徴点に関して、接触または非接触というような2値の値に分類されてもよいし、遠い、近い、および接触というような複数のグループに量子化されてもよい。ま

た、以下で説明するように、当該相対距離は、類似するものごとに自動的にクラスタリングされてもよい。量子化や、クラスタリングを行なうことで、物体相対関係の微小な違いを除外することができる。

[0087] 以下では、図 13 を参照して、相対関係クラスタリング部 230 による相対関係クラスの生成方法の詳細について説明する。図 13 には、物体相対関係の一例である、右手および物体 A の間の距離と、右足および物体 A との間の距離とを指標とした、人物と物体との間の相対距離の分布が示されている。当該分布に示される各点は、異なる時間の人物と物体との相対的な位置関係を示す。なお、図 13 では、説明を簡単にするために、物体相対関係の分布が 2 次元のグラフで表されているが、実際には、当該分布は、物体相対関係に定義される人物の各部位と物体との間の相対関係における指標の分の次元数を有する。たとえば、図 13 に示されるように、相対関係クラスタリング部 230 は、右手および物体 A の間の距離と、右足および物体 A との間の距離とが近い点の集合をグループ化して、相対関係クラス A および相対関係クラス B を生成する。

[0088] また、クラスタリングの手法は、特に限定されず、類似する物体相対関係をクラスタリングできる手法であればいずれでもよい。たとえば、クラスタリング処理は、階層的な手法と、非階層的な手法とを含む。階層的なクラスタリングの手法としては、物体相対関係の指標で示される空間上で距離の近い 2 点を順次結合し、当該距離が予め定められた値を越えるまで結合を行なう手法が挙げられる。相対関係クラスタリング部 230 は、当該距離が予め定められた距離を越えた場合には、各点の結合を止め、結合された各点の集合を 1 つの姿勢クラスとする。また、非階層的なクラスタリングの手法としては、k-means 法などが挙げられる。

[0089] 相対関係クラスタリング部 230 が類似する物体相対関係ごとにクラスタリングを行なうことにより、物体に対してとった類似する姿勢同士が同一の相対関係クラスとしてグループ化することができる。

[0090] 相対関係クラスタリング部 230 は、生成した相対関係クラス 55 を記憶

装置 20 に格納する。図 14 に示されるように、相対関係クラス 55 は、人物の各部位と物体と間の相対距離ごとに所定の範囲で定義される。また、相対関係クラスは、図 13 に示される分布上における相対関係クラスの中心座標と、中心から閾範囲 326, 328 までの距離で定義されてもよい。

[0091] なお、上記では、相対関係クラスタリング部 230 が、人物に対する物体相対関係をクラスタリングする例について説明を行なったが、他の方法で人物と当該人物の周辺との関係をクラスタリングしてもよい。たとえば、画像処理装置 100 は、距離画像センサなどで得られた 3 次元情報をいくつかの領域に分割する。すなわち、画像処理装置 100 は、類似する位置情報の集合ごとに 3 次元位置を分割して、被写体を複数の領域に分割する。一例として、画像処理装置 100 は、k-means 法などの画像処理技術により、同一のテクスチャごとに被写体を分割してもよいし、距離が近いもの同士をまとめて被写体を分割してもよい。この場合、相対関係クラスタリング部 230 は、人物の各部位と分割した複数の領域の各々との間の相対関係をクラスタリングして、相対関係クラスを生成する。

[0092] [姿勢クラスタリング部 240 の詳細]

図 15 ～図 17 を参照して、姿勢クラスタリング部 240 の詳細について説明する。図 15 は、細分化した姿勢クラスを視覚的に表した図である。図 16 は、部位間関係の分布を示した図である。図 17 は、姿勢クラス 45 のデータ構造を示した図である。

[0093] 姿勢クラスタリング部 240 は、過去の観測結果を用いて類似する姿勢をクラスタリングする。姿勢をクラスタリングせずに人物の姿勢情報をそのまま用いて人物の姿勢を特定することも考え得るが、その場合には、姿勢が細分化されてしまう可能性がある。たとえば、図 15 に示されるように、「姿勢 A」～「姿勢 C」のような同じ立ち姿勢であっても、足を開いた「姿勢 B」や、手を上げた「姿勢 C」などの少しの違いでそれぞれの姿勢が別の姿勢として扱われてしまう。このため、姿勢クラスタリング部 240 は、類似した姿勢を 1 つの姿勢としてみなすために、類似する部位間関係の集合ごとに

1つの姿勢クラスを生成する。

- [0094] 図16には、部位間関係の一例である腕の角度 α （図10（B）参照）、および足の角度 γ （図10（B）参照）を指標とした分布が示されている。当該分布に示される各点は、繰り返し取得した距離画像から抽出した部位間関係を示す。なお、図16には、説明を簡単にするために、部位間関係の分布が2次元のグラフで表されているが、実際には、当該分布は、部位間関係に定義される各節点間の関係の指標の種類の分の次元数を有する。姿勢クラスタリング部240は、図16に示されるように、たとえば、腕の角度 α と足の角度 γ とが近い点の集合をグループ化して、姿勢クラスAおよび姿勢クラスBを生成する。
- [0095] また、クラスタリングの手法は、特に限定されず、類似する人物の姿勢をクラスタリングできる手法であればいずれでもよい。たとえば、クラスタリング処理は、階層的な手法と、非階層的な手法とを含む。階層的なクラスタリングの手法としては、部位間関係の指標で示される空間上で距離の近い2点を順次結合し、当該距離が予め定められた値を越えるまで結合を行なう手法が挙げられる。姿勢クラスタリング部240は、当該距離が予め定められた距離を越えた場合には、各点の結合を止め、結合された各点の集合を1つの姿勢クラスとする。このことは、各姿勢クラス間で分布位置にある程度差があること、すなわち、各姿勢クラス間で距離がある程度離れるであろうということを前提にしている。当該距離が予め決められた値よりも大きければ処理が打ち切られるため、姿勢クラスが際限なく大きくなることはない。また、非階層的なクラスタリングの手法としては、k-means法などが挙げられる。
- [0096] 姿勢クラスタリング部240が類似する姿勢ごとにクラスタリングを行なうことにより、類似する特徴を有する姿勢同士が同一の姿勢クラスとしてグループ化することができる。また、たとえば、「座る」という姿勢をとる際の手の位置など、姿勢の微小な違いやその姿勢をとる際に重要でない部分を排除することができる。さらに、クラスタリングにより姿勢を適宜学習して

いくことで、事前にユーザが用意した姿勢以外にも、機能を対応させることが可能となる。

[0097] 姿勢クラスタリング部240は、生成した姿勢クラス45を記憶装置20に格納する。たとえば、姿勢クラス45は、図17に示されるように、部位間関係の指標ごとに所定の範囲で定義される。また、姿勢クラスは、図16に示される分布上における姿勢クラスの中心座標と、中心から閾範囲322, 324までの距離で定義されてもよい。

[0098] [相対関係特定部250の詳細]

図18を参照して、相対関係特定部250の詳細について説明する。図18は、各相対関係クラスと物体相対関係との分布を示した図である。なお、図18では、説明を簡単にするために、物体相対関係の分布が2次元のグラフで表されているが、実際には、当該分布は、物体相対関係に定義される人物の各部位と物体との間の相対関係における指標の分の次元数を有する。

[0099] 相対関係特定部250は、入力画像に写る人物と物体との相対的な位置関係が、いずれの相対関係クラスに属するかを特定する。たとえば、相対関係特定部250は、図18に示されるように、入力画像に写る人物の物体相対関係を示す点321と相対関係クラスの中心点との間の距離が所定の範囲内となる相対関係クラスを特定する。当該距離は、ユークリッド距離、およびマハラノビス距離を含む。

[0100] 一例として、点321と相対関係クラスAの中心点との距離 x が、閾範囲326に含まれる場合には、人物と物体との相対関係を相対関係クラスAとして特定する。また、点321と相対関係クラスBの中心点との距離 y が、閾範囲328に含まれる場合に、人物と物体との相対関係を相対関係クラスBとして特定する。

[0101] なお、相対関係特定部250は、人物の各部位と物体との間の相対距離の各々が、距離の長さに応じて予め分けられた複数の相対関係クラスのいずれに属するかを特定するように構成されてもよい。この場合には、相対関係クラスタリング部230によるクラスタリング処理は、実行されなくともよい。

。

[0102] [姿勢特定部 260 の詳細]

図 19 を参照して、姿勢特定部 260 の詳細について説明する。図 19 は、部位間関係の分布を示した図である。なお、図 19 には、説明を簡単にするために、部位間関係の分布が 2 次元のグラフで表されているが、実際には、当該分布は、部位間関係に定義される各節点間の関係の指標の種類の分の次元数を有する。

[0103] 姿勢特定部 260 は、入力画像に写る人物の姿勢が、いずれの姿勢クラスに属するかを特定する。たとえば、姿勢特定部 260 は、物体に接したときの部位間関係で示される点 320 と姿勢クラスの中心点との間の距離が所定の範囲内となる姿勢クラスを特定する。当該距離は、ユークリッド距離、およびマハラノビス距離を含む。

[0104] たとえば、点 320 と姿勢クラス A の中心点との距離 x が、当該中心点と閾範囲 322 との距離以内である場合に、人物は姿勢クラス A で物体に接したと判断する。また、点 320 と姿勢クラス B の中心点との距離 y が、当該中心点と閾範囲 324 との距離以内である場合に、人物は姿勢クラス B で物体に接したと判断する。図 19 に示される例では、人物は姿勢クラス A で物体に接したと判断される。

[0105] [行動特定部 270 の詳細]

図 20 を参照して、行動特定部 270 の詳細について説明する。図 20 は、行動カテゴリテーブルのデータ構造を示した図である。

[0106] 行動特定部 270 は、姿勢情報に加えて、人物の周辺環境との位置関係についても、行動の判断要素に加える。なお、ここでいう周辺環境とは、人物の周囲に物体があるという情報だけでなく、人物の周囲に物体がないという情報も含まれる。これにより、たとえば、同じ姿勢であっても、人物の周辺に何も無い場合、足の近くに何かがある場合、手の近くに何かがある場合の各々を別の行動として認識することができる。

[0107] より具体的には、姿勢クラスと相対関係クラスとの組み合わせのそれぞれ

には、人物の行動の種類を示す行動カテゴリがそれぞれ対応付けられている。すなわち、姿勢クラスと、相対関係クラスと、行動カテゴリとは、互に対応付けられている。これらの対応関係は、たとえば、図20に示されるように、行動カテゴリテーブル22として定義される。

[0108] 行動特定部270は、相対関係特定部250によって特定された相対関係クラスと、姿勢特定部260によって特定された姿勢クラスとの組み合わせから、入力画像に写る人物の行動を特定する。すなわち、行動特定部270は、行動カテゴリテーブル22を参照して、特定した相対関係クラスと姿勢クラスと組み合わせに対応する行動カテゴリを特定する。行動特定部270は、特定した行動カテゴリを入力画像に写る人物がとった行動として出力する。

[0109] なお、物体認識技術などで人物の周辺にある物体の名称が既知である場合には、行動特定部270は、姿勢クラスおよび相対関係クラスの他に、物体の名称をさらに行動の判断要素に加えてもよい。これにより、たとえば、行動特定部270は、人物の近くにある物体が、ただの箱なのか自動販売機なのかによって別の行動として認識することができる。

[0110] また、人物の周辺環境の3次元情報が領域に分割されている場合には、行動特定部270は、姿勢クラスおよび相対関係クラスの他にも、分割されている領域のサイズや、形状などをさらに行動の判断要素に加えてもよい。これにより、行動特定部270は、たとえば、近くにある物体が大きいのか、小さいのか、丸いのか、細長いのかなどの違いによって別の行動として認識することができる。

[0111] [フローチャート]

図21および図22を参照して、画像処理装置100の制御構造について説明する。図21は、画像処理装置100が実行する学習処理の一部を表わすフローチャートである。図22は、画像処理装置100が実行する行動特定処理の一部を表わすフローチャートである。図21および図22の処理は、CPU2がプログラムを実行することにより実現される。他の局面におい

て、処理の一部または全部が、回路素子その他のハードウェアによって実行されてもよい。

[0112] まず、図 21 を参照して、画像処理装置 100 の学習処理のフローについて説明する。画像処理装置 100 の動作モードが学習モードである場合には、図 21 に示されるステップ S510～ステップ S520 の処理が実行される。

[0113] ステップ S510 において、CPU 2 は、カメラ 110 から被写体を撮影し続けることにより複数の距離画像を学習画像として取得する。ステップ S512 において、CPU 2 は、物体抽出部 210 として、入力画像から物体を検出して、物体の 3 次元上の位置を取得する。なお、物体が動かない場合には、CPU 2 は、学習画像のうちの 1 枚についてのみ物体検出処理を行なってもよい。CPU 2 は、学習画像の全てについて物体検出処理を行なってもよい。

[0114] ステップ S514 において、CPU 2 は、対象抽出部 220 として、異なる時間に得られた学習画像の各々に含まれる人物の 3 次元上の位置を取得する。CPU 2 は、学習画像に含まれる人物の位置から当該人物の特徴的な部位の位置（すなわち、部位情報）を取得する。

[0115] ステップ S516 において、CPU 2 は、相対関係クラスタリング部 230 として、順次抽出された人物の各部位と物体との間の相対距離が互いに類似している集合ごとに相対関係クラスを生成する。

[0116] ステップ S518 において、CPU 2 は、複数の部位情報の各々から算出した人物の各部位間の相対関係が互いに類似する部位情報の集合ごとに姿勢クラスを生成する。

[0117] ステップ S520 において、CPU 2 は、本実施の形態に従う学習処理を終了するか否かを判断する。CPU 2 は、たとえば、ユーザが当該学習処理を終了する操作を行なった場合に、当該学習処理を終了する。CPU 2 は、本実施の形態に従う学習処理を終了すると判断した場合には（ステップ S520 において YES）、当該学習処理を終了する。そうでない場合には（ス

テップS 5 2 0においてNO)、CPU 2は、ステップS 5 1 0～ステップS 5 1 8の処理を再び実行する。

[0118] 次に、図22を参照して、画像処理装置100の行動特定処理のフローについて説明する。画像処理装置100の動作モードが行動特定モードである場合には、図22に示されるステップS 5 3 0～ステップS 5 4 2の処理が実行される。

[0119] ステップS 5 3 0において、CPU 2は、カメラ110から被写体を撮影することにより距離画像を入力画像として取得する。ステップS 5 3 2において、CPU 2は、物体抽出部210として、入力画像から物体を検出して、物体の3次元上の位置を取得する。なお、物体が動かない場合には、CPU 2は、最初に得られた入力画像に対してのみ物体検出処理を行なってもよい。また、CPU 2は、入力画像が得られるたびに物体検出処理を行なってもよい。

[0120] ステップS 5 3 4において、CPU 2は、対象抽出部220として、入力画像に含まれる人物の3次元上の位置を取得する。CPU 2は、入力画像に含まれる人物の位置から当該人物の特徴的な部位の位置（すなわち、部位情報）を取得する。

[0121] ステップS 5 3 6において、CPU 2は、相対関係特定部250として、入力画像に含まれる人物と物体との相対的な位置関係（すなわち、物体相対関係）が、複数の相対関係クラスのいずれに属するかを特定する。

[0122] ステップS 5 3 8において、CPU 2は、姿勢特定部260として、入力画像に含まれる人物の特徴的な部位間の相対的な位置関係が、複数の姿勢クラスのいずれに属するかを特定する。

[0123] ステップS 5 4 0において、CPU 2は、行動特定部270として、特定された物体相対クラスと姿勢クラスとの組み合わせから入力画像に含まれる人物がとっている行動を特定する。より具体的には、CPU 2は、行動カテゴリテーブルを参照して、特定された物体相対クラスと姿勢クラスとの組み合わせに対応する行動カテゴリを特定する。

[0124] ステップS542において、CPU2は、本実施の形態に従う行動特定処理を終了するか否かを判断する。CPU2は、たとえば、ユーザが行動特定処理を終了する操作を行なった場合に、当該行動特定処理を終了する。CPU2は、本実施の形態に従う画像処理を終了すると判断した場合には（ステップS542においてYES）、当該画像処理を終了する。そうでない場合には（ステップS542においてNO）、CPU2は、ステップS530～ステップS540の処理を再び実行する。

[0125] [利点]

以上のようにして、本実施の形態に従う画像処理装置100は、人物の姿勢情報だけでなく、人物の周囲にある物体との位置関係を用いることにより、人物の姿勢だけでは区別することができない行動を特定することができる。この結果、画像処理装置100は、人物の様々な行動を高精度に特定することができる。

[0126] 今回開示された実施の形態は全ての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は、上記した説明ではなく、請求の範囲によって示され、請求の範囲と均等の意味および範囲内での全ての変更が含まれることが意図される。

符号の説明

[0127] 1 ROM、2 CPU、3 RAM、4 カメラI/F、5 メモリーカードI/F、5A メモリーカード、6 ネットワークI/F、7 表示装置、20 記憶装置、21 部位情報、22 行動カテゴリテーブル、23 部位間関係、24 画像処理プログラム、25, 27, 31, 61 人物、29, 62 ボタン、30 距離画像、32 椅子、41～45 姿勢クラス、51～54 相対関係クラス、60 入力画像、100 画像処理装置、110 カメラ、210 物体抽出部、211 領域、220 対象抽出部、230 相対関係クラスタリング部、240 姿勢クラスタリング部、250 相対関係特定部、260 姿勢特定部、270 行動特定部、320, 321 点、322, 324, 326, 328 閾範囲。

請求の範囲

[請求項1] 人物または動物である監視対象の行動を特定するための画像処理装置であって、

カメラの視野に含まれる被写体の3次元位置から、前記監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するための対象抽出部と、

前記部位情報から算出した前記監視対象の各部位間の第1の相対関係から前記監視対象の姿勢を特定するための姿勢特定部と、

前記監視対象と、当該監視対象以外の被写体との間の第2の相対関係を特定するための相対関係特定部と、

前記姿勢特定部によって特定された前記監視対象の姿勢と、前記第2の相対関係との組み合わせから、前記監視対象の行動を特定するための行動特定部とを備える、画像処理装置。

[請求項2] 前記姿勢特定部は、前記第1の相対関係が、監視対象の姿勢の種類を示す複数の姿勢クラスのいずれに属するかを特定し、

前記相対関係特定部は、前記第2の相対関係が、監視対象と当該監視対象以外の被写体との間の相対関係の種類を示す複数の相対関係クラスのいずれに属するかを特定し、

前記行動特定部は、前記姿勢特定部によって特定された姿勢クラスと、前記相対関係特定部によって特定された相対関係クラスとの組み合わせから、前記監視対象の行動の種類を特定する請求項1に記載の画像処理装置。

[請求項3] 前記対象抽出部は、異なる時間に得られた前記3次元位置から、監視対象が存在する領域を順次抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を順次抽出し、

前記画像処理装置は、

複数の前記部位情報を、類似する部位情報ごとにクラスタリングして前記複数の姿勢クラスを生成するための姿勢クラスタリング部と

、

前記対象抽出部により順次抽出された監視対象と、当該監視対象以外の被写体との間の第2の相対関係が互いに類似している集合ごとにクラスタリングして前記複数の相対関係クラスを生成するための相対関係クラスタリング部とをさらに備える、請求項2に記載の画像処理装置。

[請求項4] 前記姿勢クラスタリング部は、前記対象抽出部によって順次抽出された複数の前記部位情報の各々から監視対象の各部位間の第1の相対関係を算出し、第1の相対関係が互いに類似する集合ごとに前記姿勢クラスを生成する、請求項3に記載の画像処理装置。

[請求項5] 前記画像処理装置は、前記3次元位置から、監視対象以外の被写体である物体が存在する領域を抽出するための物体抽出部をさらに備え、

前記第2の相対関係は、前記監視対象と前記物体との間の相対的な位置関係を含む、請求項3または4に記載の画像処理装置。

[請求項6] 前記相対関係クラスタリング部は、前記対象抽出部によって順次抽出された複数の前記部位情報の各々から、監視対象の特徴的な各部位について、当該部位と前記物体との間の距離が互いに類似している集合ごとに前記相対関係クラスを生成する、請求項5に記載の画像処理装置。

[請求項7] 前記第2の相対関係は、前記監視対象の各関節と前記物体との間の各距離、および、当該監視対象の各端部と前記物体との間の各距離の少なくとも一方を含む、請求項5または6に記載の画像処理装置。

[請求項8] 前記部位情報は、前記監視対象の各関節の位置、および、当該監視対象の各端部の位置の少なくとも一方を含む、請求項1～7のいずれか1項に記載の画像処理装置。

[請求項9] 前記相対関係特定部は、前記監視対象の特徴的な各部位と、当該監視対象以外の被写体との間の距離の各々が、距離の長さに応じて予め

分けられた前記複数の相対関係クラスの中のいずれに属するかを特定する、請求項2または3に記載の画像処理装置。

[請求項10] 前記画像処理装置は、類似する位置情報の集合ごとに前記3次元位置を分割して、前記被写体を複数の領域に分割するための分割部をさらに備え、

前記第2の相対関係は、前記監視対象の特徴的な各部位と前記複数の領域の各々との間の相対関係を含む、請求項1～9のいずれか1項に記載の画像処理装置。

[請求項11] 前記複数の姿勢クラスと前記複数の相対関係クラスとの組み合わせのそれぞれには、前記監視対象の行動の種類を示す行動カテゴリがそれぞれ対応付けられており、

前記行動特定部は、前記姿勢特定部によって特定された姿勢クラスと前記相対関係特定部によって特定された相対関係クラスとの組み合わせに対応する行動カテゴリを特定する、請求項2～6のいずれか1項に記載の画像処理装置。

[請求項12] 人物または動物である監視対象の行動を特定するための画像処理方法であって、

カメラの視野に含まれる被写体の3次元位置から、前記監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するステップと、

前記部位情報から算出した前記監視対象の各部位間の第1の相対関係から前記監視対象の姿勢を特定するステップと、

前記監視対象と、当該監視対象以外の被写体との間の第2の相対関係を特定するステップと、

前記姿勢と前記第2の相対関係との組み合わせから、前記監視対象の行動を特定するステップとを備える、画像処理方法。

[請求項13] 人物または動物である監視対象の行動を特定するための画像処理プログラムであって、

前記画像処理プログラムは、コンピュータに、

カメラの視野に含まれる被写体の3次元位置から、前記監視対象が存在する領域を抽出するとともに、当該監視対象が有する特徴的な各部位の位置を示す部位情報を抽出するステップと、

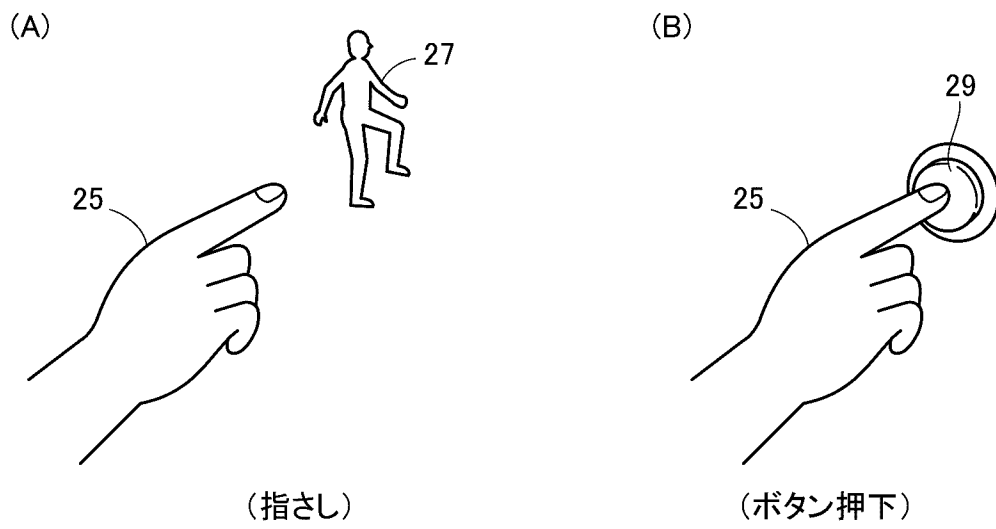
前記部位情報から算出した前記監視対象の各部位間の第1の相対関係から前記監視対象の姿勢を特定するステップと、

前記監視対象と、当該監視対象以外の被写体との間の第2の相対関係を特定するステップと、

前記姿勢と前記第2の相対関係との組み合わせから、前記監視対象の行動を特定するステップとを実行させる、画像処理プログラム。

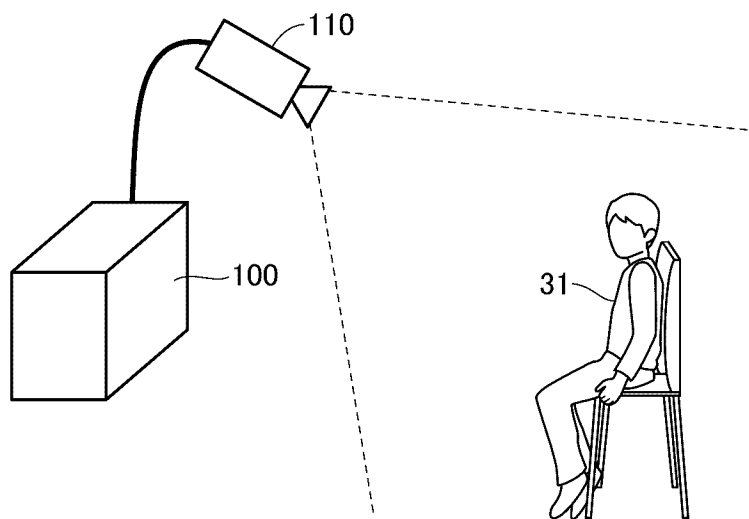
[図1]

図1



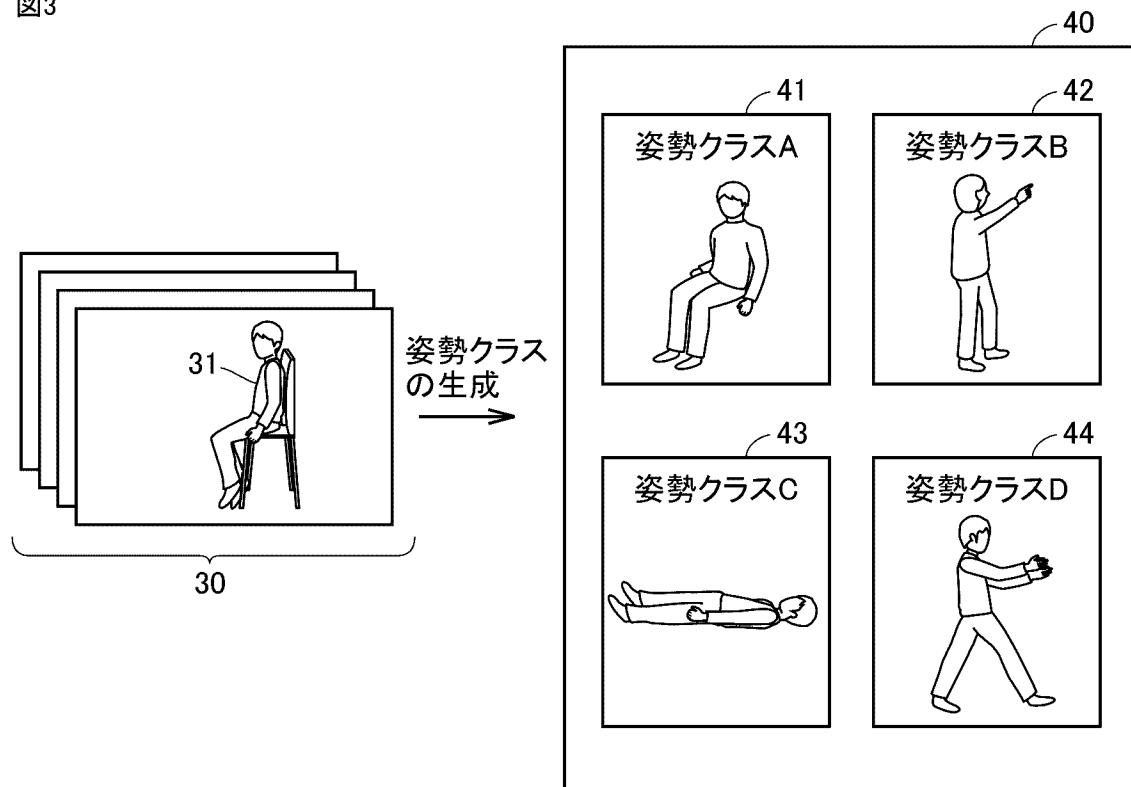
[図2]

図2



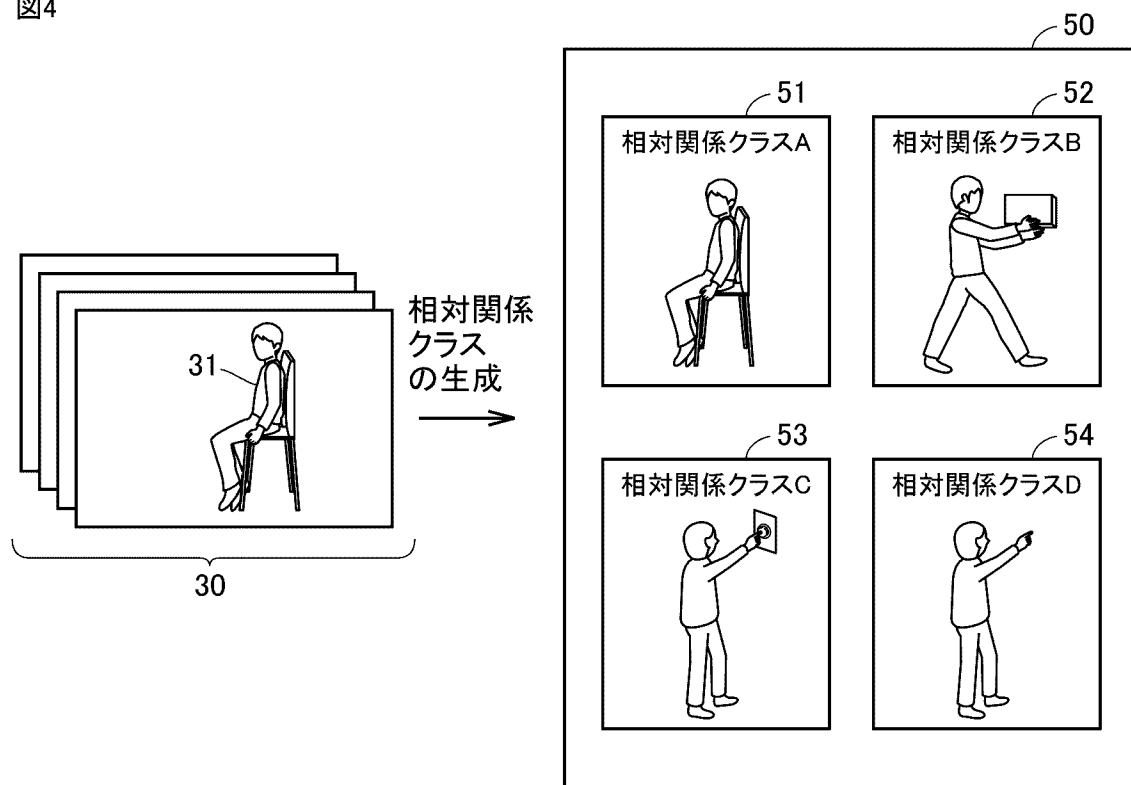
[図3]

図3



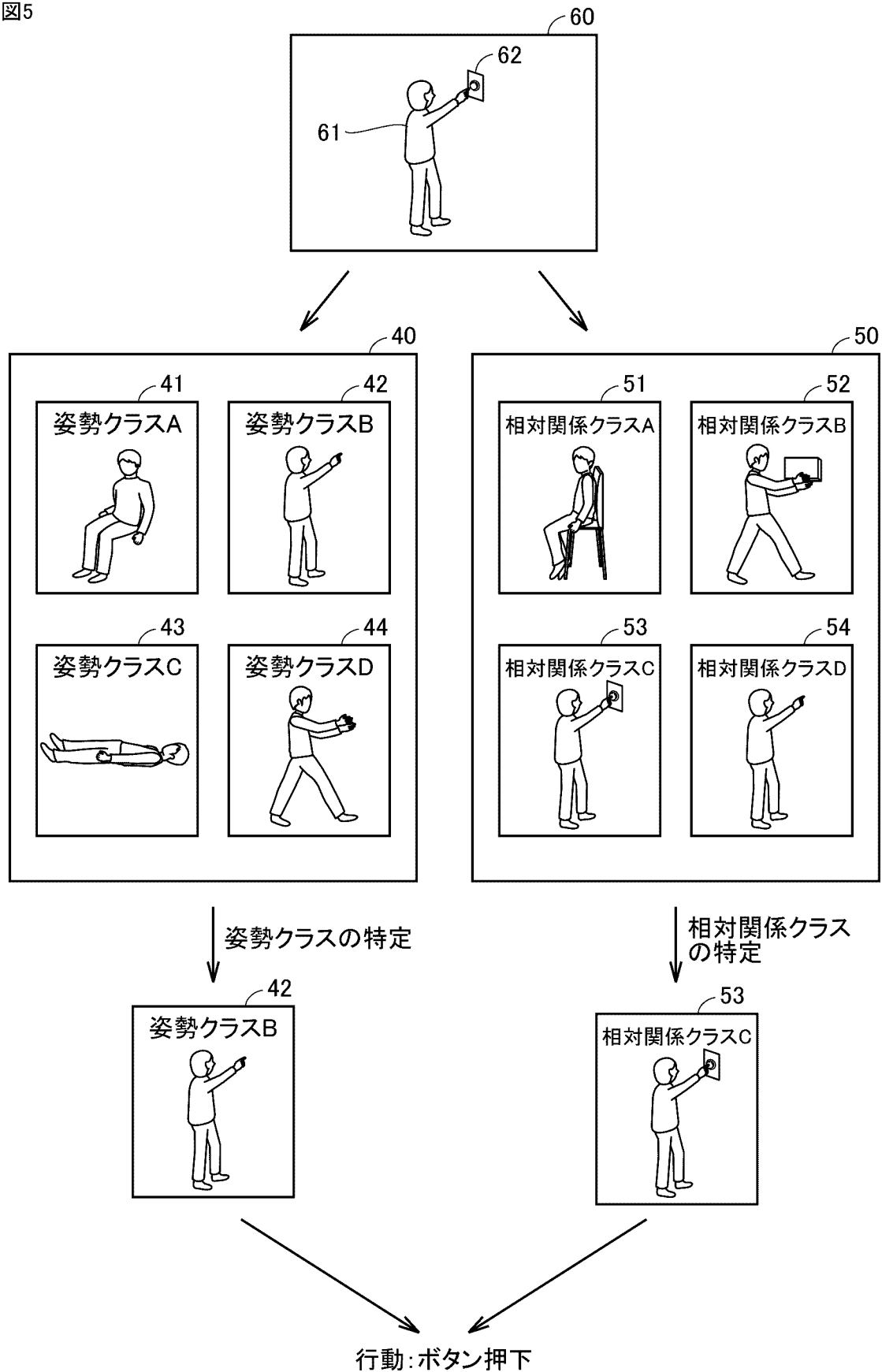
[図4]

図4



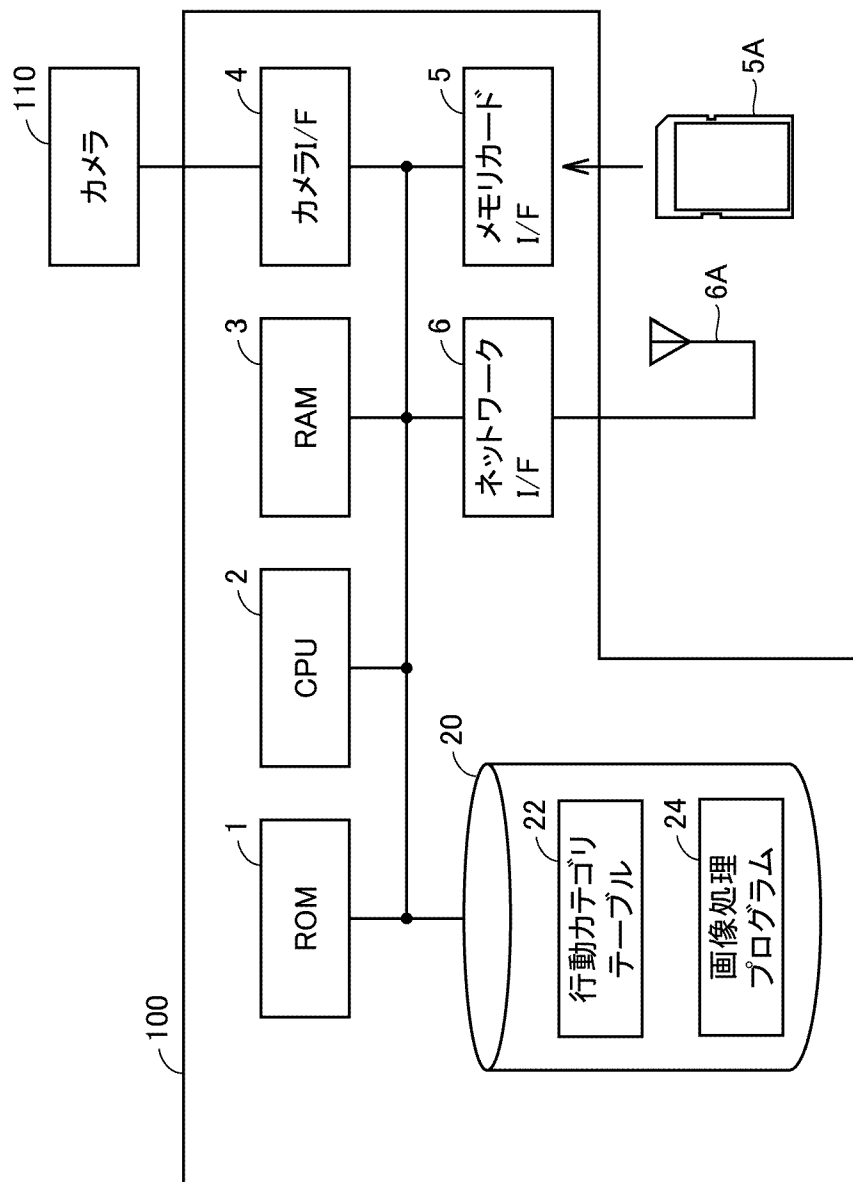
[図5]

図5



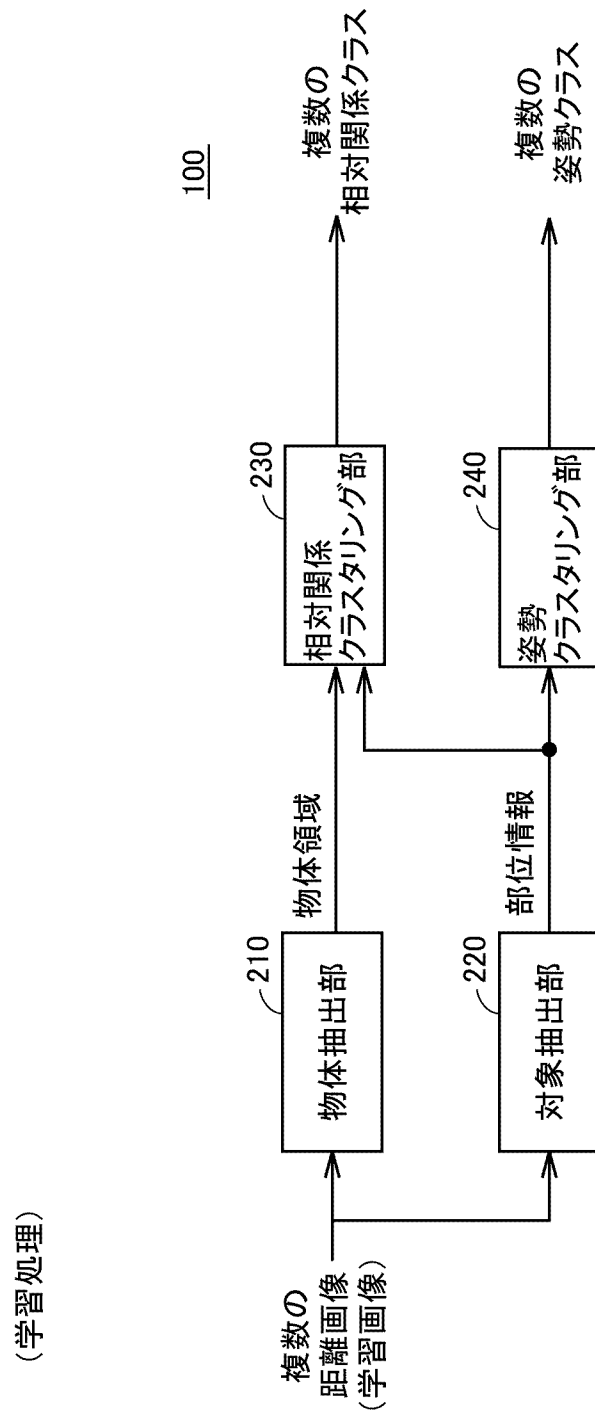
[図6]

図6



[図7]

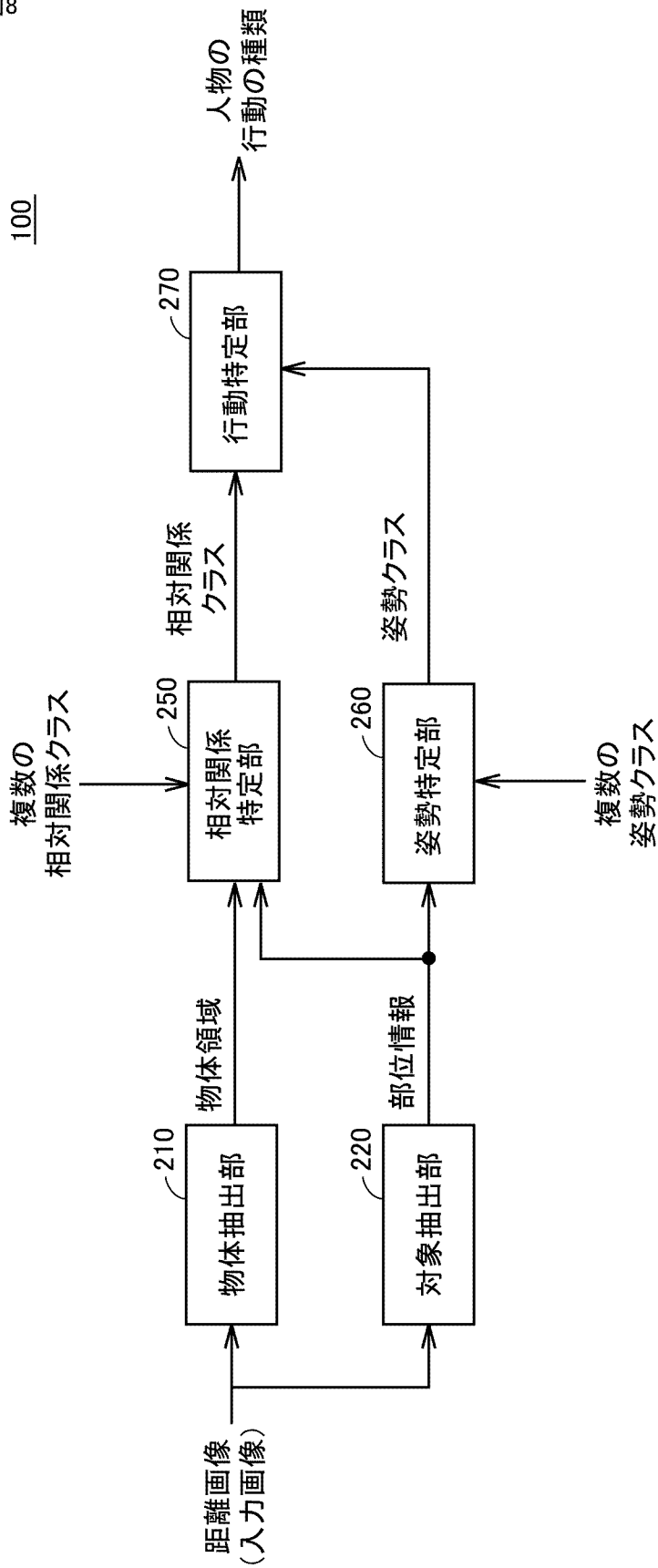
図7



[図8]

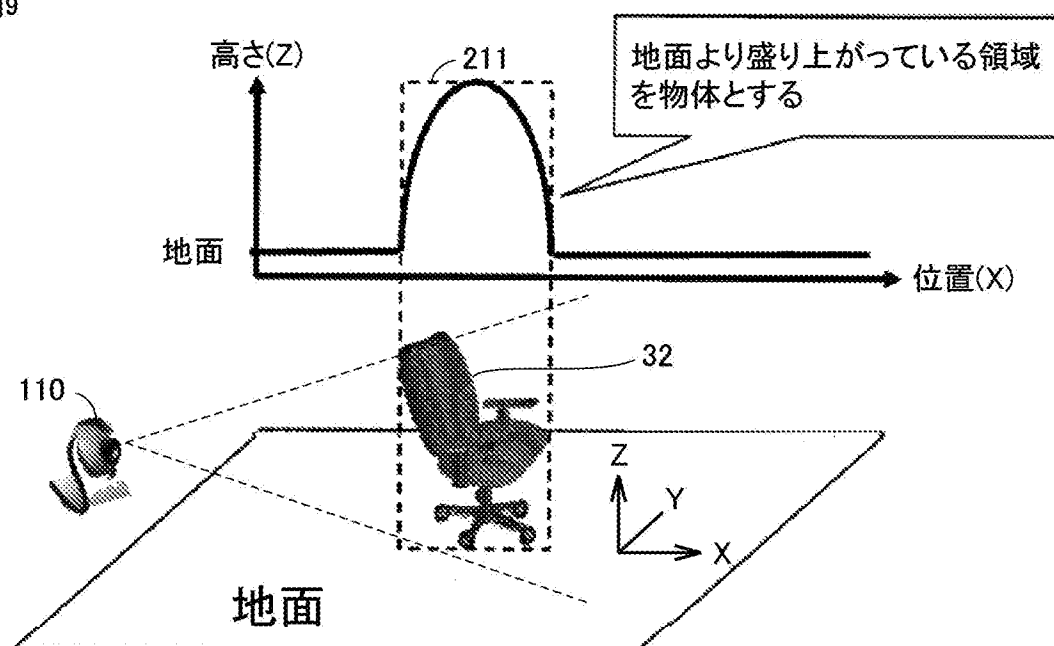
図8

(行動特定処理)



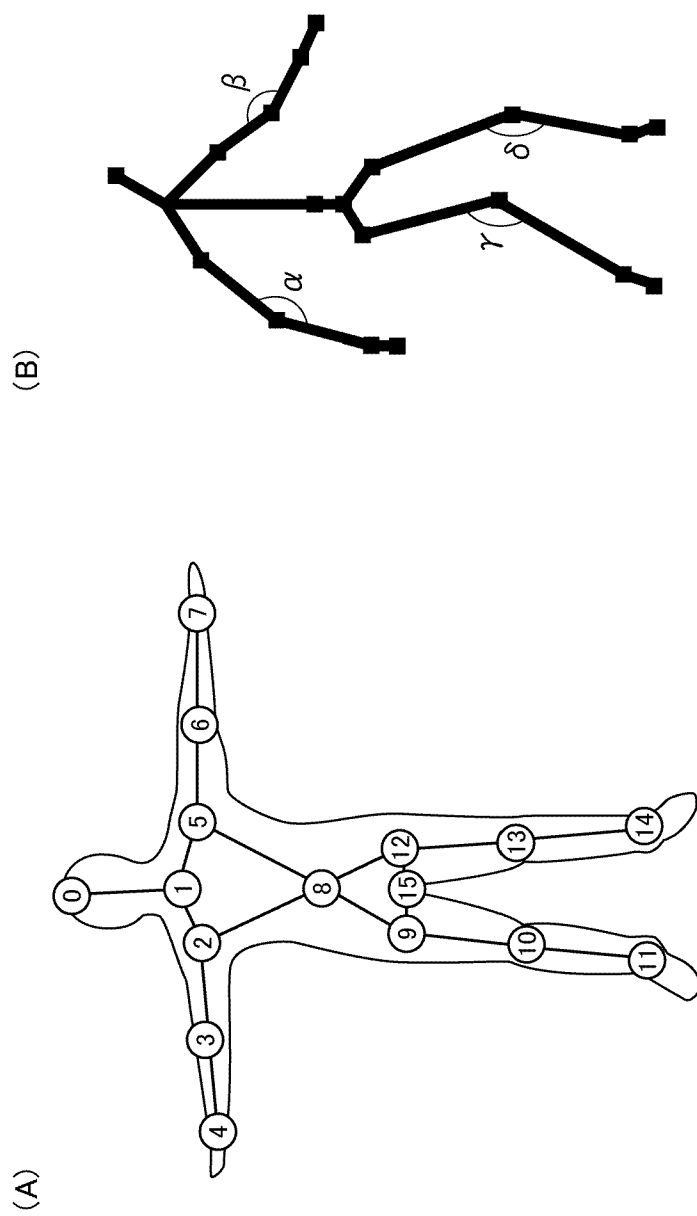
[図9]

図9



[図10]

図10



[図11]

図11

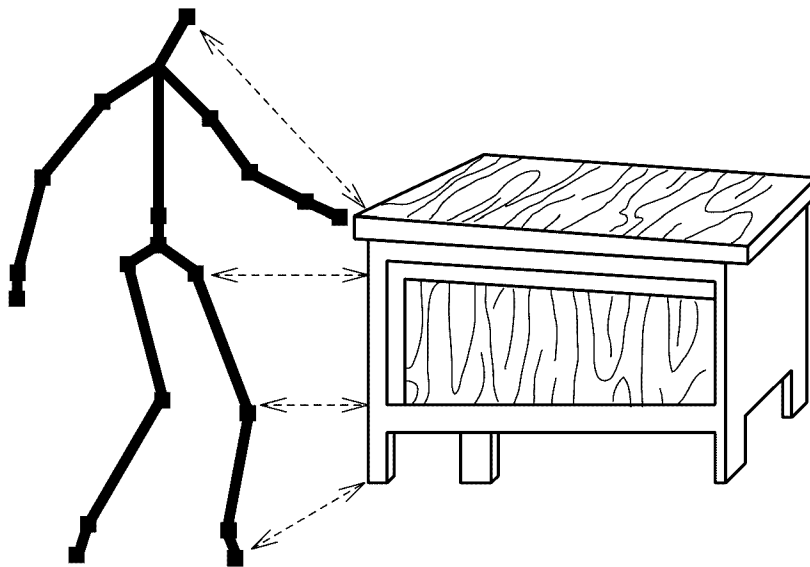
(部位間情報)

23

角度	角度 α	10°
	角度 β	12°
	角度 γ	5°
	⋮	⋮
相対距離	① − ②	x_{1-2}
	① − ③	x_{1-3}
	⋮	⋮
	⑭ − ⑮	x_{14-15}

[図12]

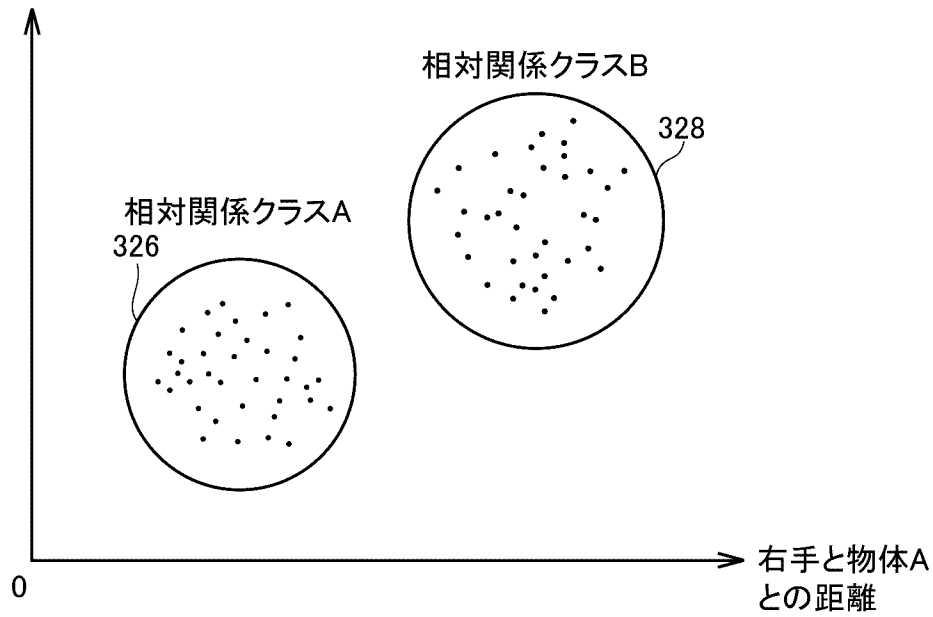
図12



[図13]

図13

右足と物体Aとの距離



[図14]

図14

(相対関係クラスA)

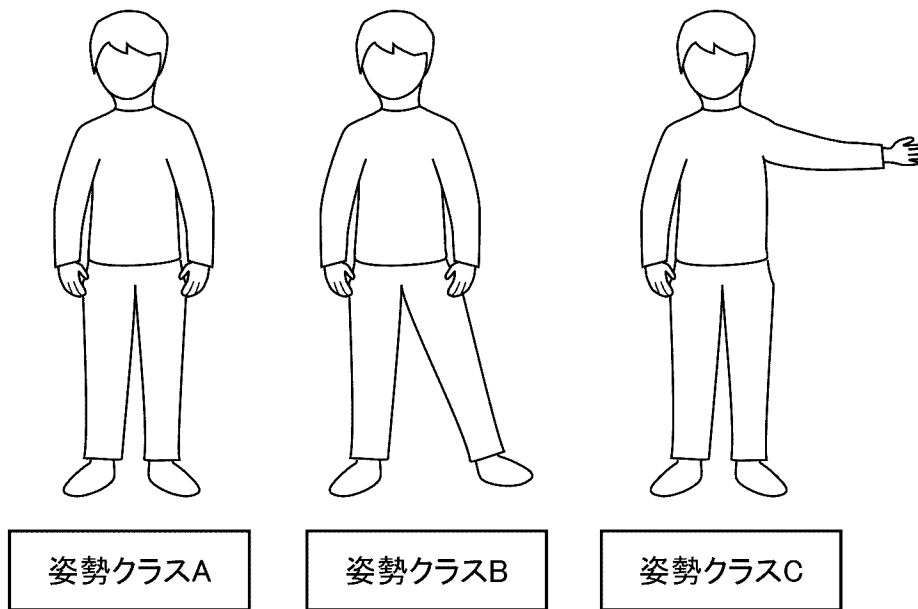
20

55

物体	部位	相対距離
物体A	右手	10～15
	左手	50～53
	⋮	⋮
物体B	右手	30～43
	左手	60～70
	⋮	⋮

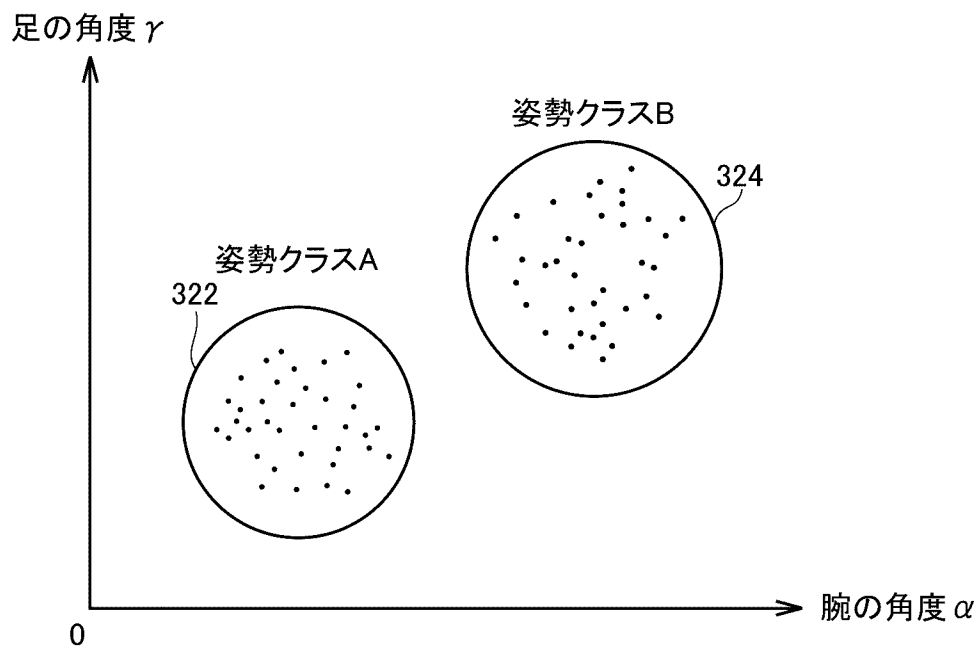
[図15]

図15



[図16]

図16



[図17]

図17

(姿勢クラスA)

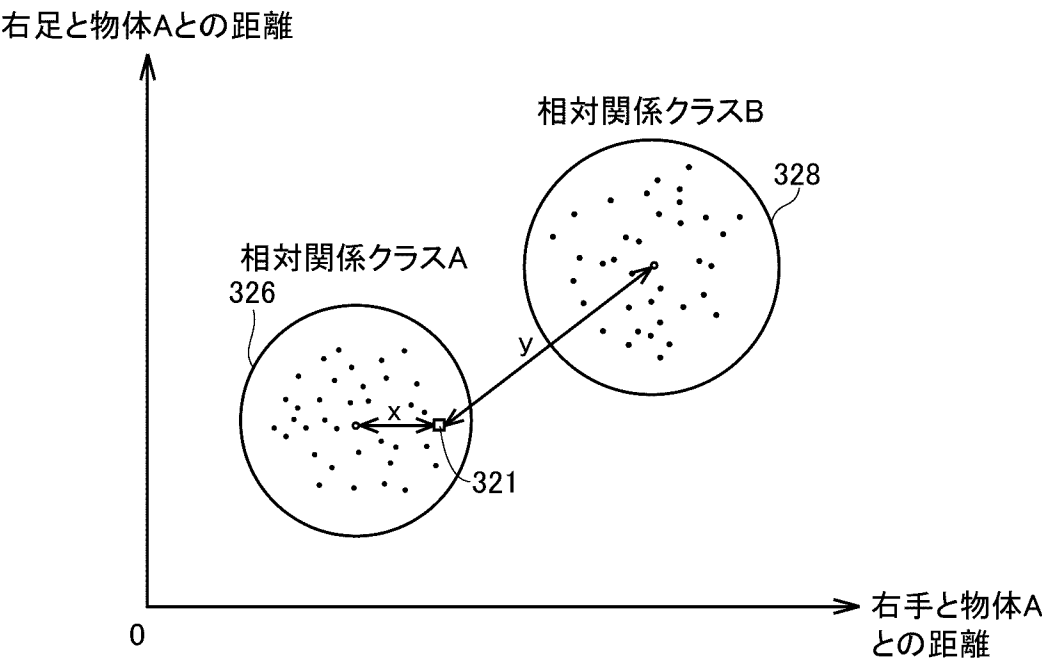
20

45

角度	角度 α	$10^{\circ} \sim 20^{\circ}$
	角度 β	$15^{\circ} \sim 25^{\circ}$
	角度 γ	$0^{\circ} \sim 10^{\circ}$
	$\vdots$	$\vdots$
相対距離	① - ②	$a_{1-2} \sim b_{1-2}$
	① - ③	$a_{1-3} \sim b_{1-3}$
	$\vdots$	$\vdots$
	⑭ - ⑮	$a_{14-15} \sim b_{14-15}$

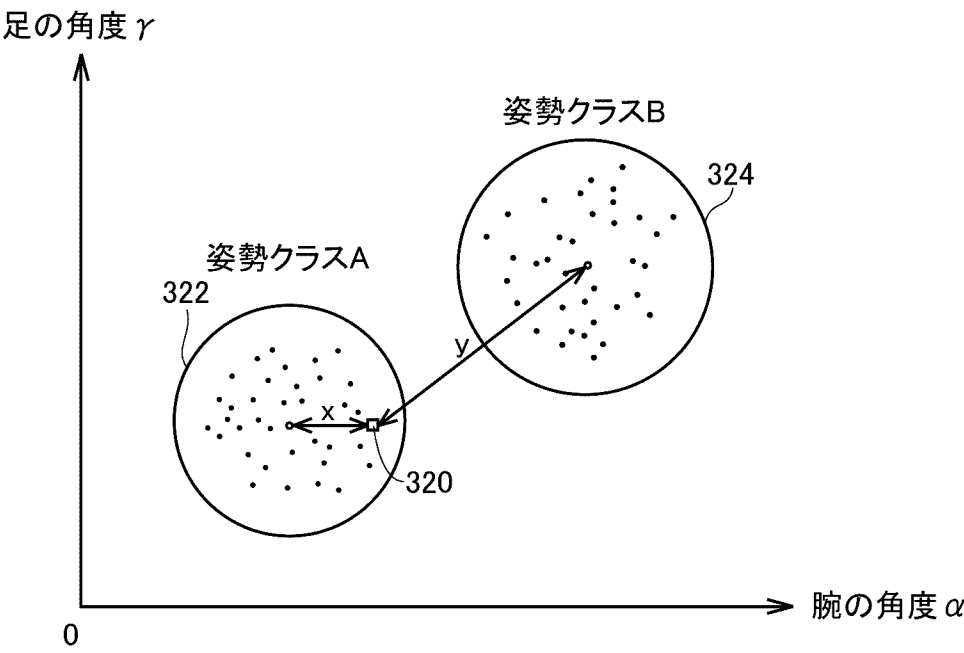
[図18]

図18



[図19]

図19



[図20]

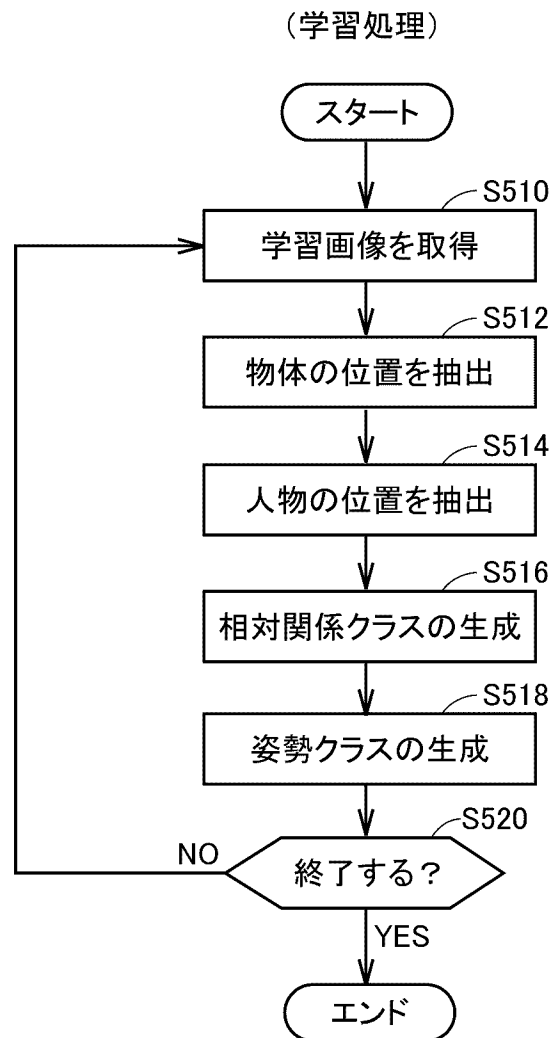
図20

22

姿勢クラス	相対関係クラス	行動カテゴリ
姿勢クラスA	相対関係クラスA	行動カテゴリA
姿勢クラスA	相対関係クラスB	行動カテゴリB
姿勢クラスB	相対関係クラスA	行動カテゴリC
姿勢クラスB	相対関係クラスB	行動カテゴリD

[図21]

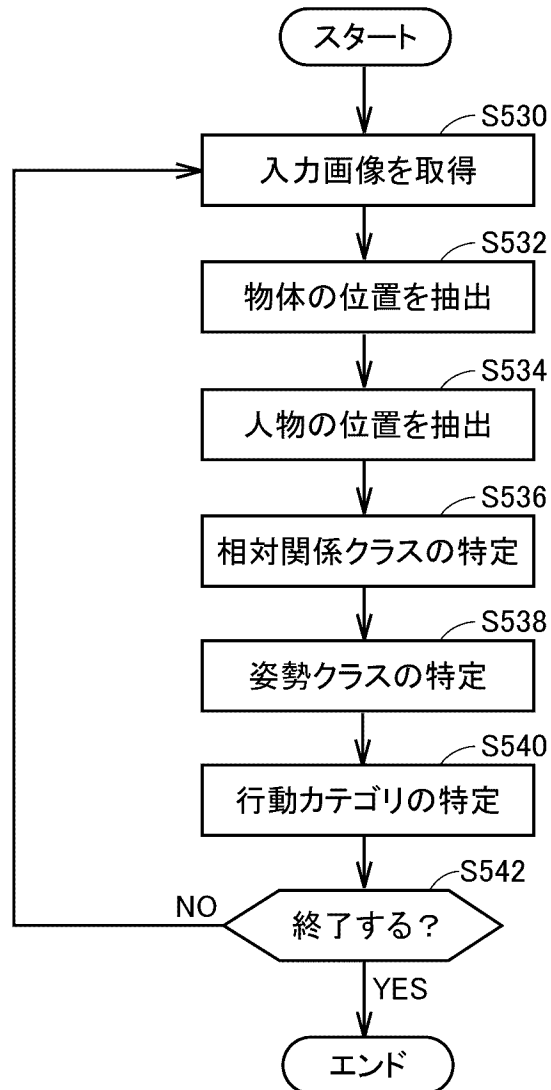
図21



[図22]

図22

(行動特定処理)



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/061677

A. CLASSIFICATION OF SUBJECT MATTER

G06T7/00(2006.01)i, G06T7/60(2006.01)i, G08B13/196(2006.01)i, H04N5/225(2006.01)i, H04N5/232(2006.01)i, H04N7/18(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T7/00, G06T7/60, G08B13/196, H04N5/225, H04N5/232, H04N7/18, G06F3/01

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	JP 2007-128307 A (Advanced Telecommunications Research Institute International), 24 May 2007 (24.05.2007), paragraphs [0032] to [0043] (Family: none)	1, 8, 10, 12, 13 2-7, 9, 11
Y A	JP 2012-133665 A (Sogo Keibi Hosho Co., Ltd.), 12 July 2012 (12.07.2012), paragraphs [0096] to [0115]; fig. 10 (Family: none)	1, 8, 10, 12, 13 2-7, 9, 11
Y	JP 2010-123019 A (Fujitsu Ltd.), 03 June 2010 (03.06.2010), paragraphs [0031], [0046] to [0066] (Family: none)	10

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
21 May 2015 (21.05.15)

Date of mailing of the international search report
02 June 2015 (02.06.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/061677

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Katsunori ASAMURA, "Text Expression of Human Behavior Based on State Transition about Parts of Body", IEICE Technical Report, 18 November 1999 (18.11.1999), vol.99, no.448, pages 17 to 24	1-13

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06T7/00(2006.01)i, G06T7/60(2006.01)i, G08B13/196(2006.01)i, H04N5/225(2006.01)i,
H04N5/232(2006.01)i, H04N7/18(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06T7/00, G06T7/60, G08B13/196, H04N5/225, H04N5/232, H04N7/18, G06F3/01

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y A	JP 2007-128307 A（株式会社国際電気通信基礎技術研究所） 2007.05.24, 段落 [0032] - [0043]（ファミリーなし）	1, 8, 10, 12, 13 2-7, 9, 11
Y A	JP 2012-133665 A（総合警備保障株式会社）2012.07.12, 段落 [0096] - [0115]、図10（ファミリーなし）	1, 8, 10, 12, 13 2-7, 9, 11

☒ C 欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

2 1 . 0 5 . 2 0 1 5

国際調査報告の発送日

0 2 . 0 6 . 2 0 1 5

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

板垣 有紀

5 H

4 4 5 2

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 3 1

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2010-123019 A (富士通株式会社) 2010.06.03, 段落 [0031]、[0046] – [0066] (ファミリーなし)	10
A	浅沼 克紀, 身体各部の状態遷移に基づく人物行動のテキスト表現, 電子情報通信学会技術研究報告, 1999.11.18, 第99巻 第448号, p.17-24	1-13