

(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  
2019年4月4日(04.04.2019)



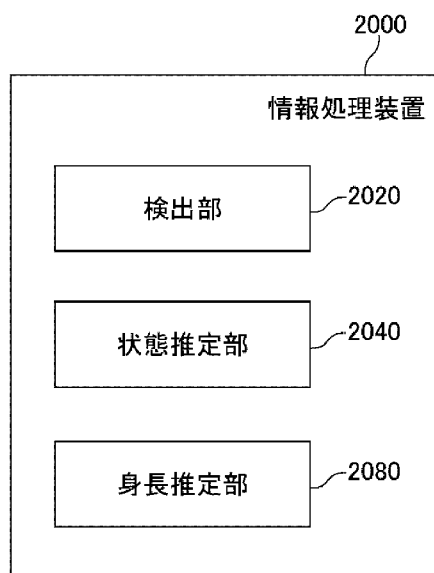
(10) 国際公開番号

**WO 2019/064375 A1**

- (51) 国際特許分類:  
*G06T 7/00* (2017.01)      *G06T 7/20* (2017.01)
- (21) 国際出願番号: PCT/JP2017/034940
- (22) 国際出願日: 2017年9月27日(27.09.2017)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者: 大網 亮磨 (OAMI Ryoma); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).
- (74) 代理人: 速水 進治 (HAYAMI Shinji); 〒1410031 東京都品川区西五反田7丁目9番2号 KDX五反田ビル9階 Tokyo (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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.

(54) **Title:** INFORMATION PROCESSING DEVICE, CONTROL METHOD, AND PROGRAM

(54) 発明の名称: 情報処理装置、制御方法、及びプログラム



2000 Information processing device  
2020 Detection unit  
2040 State estimation unit  
2080 Height estimation unit

(57) **Abstract:** An information processing device (2000) includes a detection unit (2020), a state estimation unit (2040), and a height estimation unit (2080). The detection unit (2020) detects a target person from a video frame. The state estimation unit (2040) estimates the state of the detected target person. In the case where the estimated state satisfies a predetermined condition, the height estimation unit (2080) estimates the height of the target person on the basis of the height of the person in the abovementioned video frame.

[続葉有]

(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:

一 国際調査報告 (条約第21条(3))

---

(57) 要約: 情報処理装置 (2000) は、検出部 (2020)、状態推定部 (2040)、及び身長推定部 (2080) を有する。検出部 (2020) は、動画フレームから対象人物を検出する。状態推定部 (2040) は、検出された対象人物の状態を推定する。身長推定部 (2080) は、推定された状態が所定条件を満たす場合、上記動画フレームにおける対象人物の高さに基づいて、その人物の身長を推定する。

## 明 細 書

発明の名称： 情報処理装置、制御方法、及びプログラム

### 技術分野

[0001] 本発明は画像処理に関する。

### 背景技術

[0002] カメラによって生成される映像に含まれる人物を追跡する技術が開発されている。このような技術を開示する文献として、特許文献1と2がある。特許文献1には、人物の状態を「歩行」、「直立」、「座る」、及び「寝る」に分類し、実空間上での足元位置を推定する方式が記載されている。この方式では、画像から人物を検出した後、人物領域の主軸の傾きによる姿勢判定と、カメラから頭部までの水平距離及びカメラから足元までの水平距離を用いた姿勢判定を統合し、「立つ」、「座る」、及び「寝る」の3つの状態を判別している。このとき、カメラから頭部までの水平距離は、仮定された頭部の高さを用いて、頭部の画像上での座標を実空間上での座標に変換し、求まった位置とカメラの位置との差分をとることで算出される。この際、「立つ」状態については、頭部の高さが身長値であると仮定して算出する。また、「座る」及び「寝る」状態については、あらかじめ定められた値が用いられている。

[0003] また、特許文献1では、一定時間間隔空いたフレーム間での位置を比較することにより、静止か歩行状態かを判定し、上記の姿勢状態と合わせて最終的に「歩行」、「直立」、「座る」、及び「寝る」のいずれの状態であるかが判定される。そして、求まった状態に応じて人物の実空間上での座標をカメラパラメータにより算出している。また、状態間では、人物の軌跡がなめらかになるように内挿し、人物の動線情報を算出している。

[0004] 特許文献2は、カメラから足元がみえるエリアを身長同定領域として指定し、人物の足元と頭部の画像上の位置を実空間上の座標に変換することによって身長を求める方式を開示している。身長同定領域は、足元から頭部まで

がカメラによって撮像できる位置の中で規定されており、この領域に入った時に、足元位置と頭の位置を画像から算出し、身長を求めている。

## 先行技術文献

### 特許文献

[0005] 特許文献1：特開2002-197463号公報

特許文献2：特開2001-056853号公報

## 発明の概要

### 発明が解決しようとする課題

[0006] 特許文献1は、画像に含まれる人物の身長を算出する方法や、「座り」又は「寝る」の状態における頭部の高さを算出する方法について言及していない。また、特許文献2では、身長同定領域における人物の姿勢を考慮していない。そのため、例えば身長同定領域で人物が屈んでいる場合などには、人物の身長を正しく求めることができない。また、人物が身長同定領域に含まれるときにしかその人物の身長を推定できないため、画像上においてこの領域で人物同士が重なり合ってしまう、人物の足元が画像から検出できない場合などでは、身長を推定できなくなる。

[0007] 本発明は、以上の課題に鑑みてなされたものである。本発明の目的の一つは、カメラによって生成される画像に含まれる人物の身長を高い精度で推定する技術を提供することである。

### 課題を解決するための手段

[0008] 本発明の情報処理装置は、1) 動画フレームから人物を検出する検出手段と、2) 検出の結果を用いて、対象人物の状態を推定する状態推定手段と、3) 対象人物の状態が所定条件を満たす場合、動画フレームにおける対象人物の高さに基づいて、対象人物の身長を推定する身長推定手段と、を有する。

[0009] 本発明の制御方法は、コンピュータによって実行される。当該制御方法は、1) 動画フレームから人物を検出する検出ステップと、2) 検出の結果を

用いて、対象人物の状態を推定する状態推定ステップと、3) 対象人物の状態が所定条件を満たす場合、動画フレームにおける対象人物の高さに基づいて、対象人物の身長を推定する身長推定ステップと、を有する。

[0010] 本発明のプログラムは、本発明の制御方法が有する各ステップをコンピュータに実行させる。

### 発明の効果

[0011] 本発明によれば、カメラによって生成される画像に含まれる人物の身長を高い精度で推定する技術が提供される。

### 図面の簡単な説明

[0012] 上述した目的、およびその他の目的、特徴および利点は、以下に述べる好適な実施の形態、およびそれに付随する以下の図面によってさらに明らかになる。

[0013] [図1]実施形態1の情報処理装置の動作の概要を説明するための図である。

[図2]対象人物が含まれる動画フレームを例示する図である。

[図3]実施形態1の情報処理装置の構成を例示する図である。

[図4]情報処理装置を実現するための計算機を例示する図である。

[図5]実施形態1の情報処理装置によって実行される処理の流れを例示する第1のフローチャートである。

[図6]実施形態1の情報処理装置によって実行される処理の流れを例示する第2のフローチャートである。

[図7]追跡情報を例示する図である。

[図8]実施形態2の情報処理装置の機能構成を例示する図である。

[図9]実施形態2の情報処理装置によって実行される処理の流れを例示する第1のフローチャートである。

[図10]実施形態2の情報処理装置によって実行される処理の流れを例示する第2のフローチャートである。

[図11]実施形態3の情報処理装置の機能構成を例示するブロック図である。

[図12]実施形態3の情報処理装置によって実行される処理の流れの概略を示

すフローチャートである。

[図13]第1時点の動画フレームから検出された人物と、追跡対象人物との対応付けを例示する図である。

### 発明を実施するための形態

[0014] 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。また、特に説明する場合を除き、各ブロック図において、各ブロックは、ハードウェア単位の構成ではなく、機能単位の構成を表している。

[0015] [実施形態1]

#### <概要>

図1は、実施形態1の情報処理装置（後述する図3における後述する情報処理装置2000）の動作の概要を説明するための図である。以下で説明する情報処理装置2000の動作は、情報処理装置2000の理解を容易にするための例示であり、情報処理装置2000の動作は以下の例に限定されるわけではない。情報処理装置2000の動作の詳細やバリエーションについては後述する。

[0016] 情報処理装置2000は、カメラ10によって生成された動画データ12から人物20を検出する。図1では、動画フレーム14から4人の人物が検出されている。ここで、動画フレーム14は、動画データ12を構成する時系列の撮像画像の内の1つである。すなわち、動画データ12は、時系列の複数の動画フレーム14によって構成される。

[0017] 情報処理装置2000は、人物20が含まれる複数の動画フレーム14を画像解析することで、人物20の身長を推定する。身長を推定する対象とする人物を、対象人物と表記する。ここで、身長は、一定の観測期間中（例えば、店舗内での動線解析の場合には、ある顧客の入店から退店まで）変化しない、各人物に対して固有の不変量である。一方で、画像から観測される人物の高さは、姿勢変動等によって変化する。そこで、不変量である「身長」に対し、或る時点に生成された画像から観測される人物の高さを「一時的

さ」と呼ぶことにし、身長とは区別して用いる。以下、一部の動画フレーム 14 を画像解析する意義について説明する。

[0018] 図 2 は、対象人物が含まれる動画フレーム 14 を例示する図である。動画フレーム 14-1 では、人物 20 が屈んでいる。一方、動画フレーム 14-2 では、人物 20 が直立している。

[0019] 人物 20 の身長は、直立している状態における人物 20 の高さである。そのため、動画フレーム 14-1 のように、屈んでいる人物 20 が含まれる動画フレーム 14 を画像解析しても、人物 20 の身長を精度よく推定することは難しい。一方で、動画フレーム 14-2 のように、直立している人物 20 が含まれる動画フレーム 14 を画像解析すれば、人物 20 の身長を精度よく推定することができる。

[0020] そこで情報処理装置 2000 は、対象人物が含まれる複数の動画フレーム 14 それぞれについて、動画フレーム 14 に含まれる対象人物の状態を推定する。さらに情報処理装置 2000 は、対象人物が含まれる複数の動画フレーム 14 のうち、推定された対象人物の状態が所定条件を満たす動画フレーム 14 における対象人物の高さに基づいて、対象人物の身長を推定する。ここで、所定条件は、対象人物の状態が直立姿勢又は直立に近い姿勢である場合に満たされる条件であるとする。

[0021] 例えば図 2 のケースでは、動画フレーム 14-1 における対象人物の高さ  $d1$  ではなく、動画フレーム 14-2 における対象人物の高さ  $d2$  を用いて、対象人物の身長を推定する。ここで、推定される身長を「推定身長」とも表記する。そして、身長を推定することを、「推定身長を算出する」とも表記する。

[0022] <作用・効果>

図 2 に示したように、対象人物が含まれる動画フレーム 14 の中には、身長の推定に適さない状態の対象人物が含まれるものもある。このような動画フレーム 14 を利用して対象人物の身長を推定すると、対象人物の身長を精度良く推定することができない。

[0023] そこで本実施形態の情報処理装置 2000 は、動画フレーム 14 に含まれる対象人物の状態を推定し、身長に推定に適した状態の対象人物が含まれる動画フレーム 14 における対象人物の高さに基づいて、対象人物の身長を推定する。こうすることで、対象人物の身長を精度良く推定することができる。

[0024] 以下、本実施形態の情報処理装置 2000 についてさらに詳細に説明する。

[0025] <情報処理装置 2000 の機能構成の例>

図 3 は、実施形態 1 の情報処理装置 2000 の構成を例示する図である。情報処理装置 2000 は、例えば、検出部 2020、状態推定部 2040、及び身長推定部 2080 を有する。検出部 2020 は、動画フレーム 14 から人物 20 を検出する。状態推定部 2040 は、検出された人物 20 の状態を推定する。身長推定部 2080 は、推定された状態が所定条件を満たす場合、上記動画フレーム 14 における人物 20 の高さに基づいて、その人物 20 の身長を推定する。

[0026] <情報処理装置 2000 のハードウェア構成>

情報処理装置 2000 の各機能構成部は、各機能構成部を実現するハードウェア（例：ハードワイヤードされた電子回路など）で実現されてもよいし、ハードウェアとソフトウェアとの組み合わせ（例：電子回路とそれを制御するプログラムの組み合わせなど）で実現されてもよい。以下、情報処理装置 2000 の各機能構成部がハードウェアとソフトウェアとの組み合わせで実現される場合について、さらに説明する。

[0027] 図 4 は、情報処理装置 2000 を実現するための計算機 1000 を例示する図である。計算機 1000 は任意の計算機である。例えば計算機 1000 は、Personal Computer (PC)、サーバマシン、タブレット端末、又はスマートフォンなどである。その他にも例えば、計算機 1000 はカメラ 10 であってもよい。計算機 1000 は、情報処理装置 2000 を実現するために設計された専用の計算機であってもよいし、汎用の計算機であってもよい。

- [0028] 計算機１０００は、バス１０２０、プロセッサ１０４０、メモリ１０６０、ストレージデバイス１０８０、入出力インタフェース１１００、及びネットワークインタフェース１１２０を有する。バス１０２０は、プロセッサ１０４０、メモリ１０６０、ストレージデバイス１０８０、入出力インタフェース１１００、及びネットワークインタフェース１１２０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。ただし、プロセッサ１０４０などを互いに接続する方法は、バス接続に限定されない。プロセッサ１０４０は、CPU (Central Processing Unit) や GPU (Graphics Processing Unit) などの種々のプロセッサである。メモリ１０６０は、RAM (Random Access Memory) などを用いて実現される主記憶装置である。ストレージデバイス１０８０は、ハードディスク、SSD (Solid State Drive)、メモリカード、又は ROM (Read Only Memory) などを用いて実現される補助記憶装置である。ただし、ストレージデバイス１０８０は、RAM など、主記憶装置を構成するハードウェアと同様のハードウェアで構成されてもよい。
- [0029] 入出力インタフェース１１００は、計算機１０００と入出力デバイスとを接続するためのインタフェースである。ネットワークインタフェース１１２０は、計算機１０００を通信網に接続するためのインタフェースである。この通信網は、例えば LAN (Local Area Network) や WAN (Wide Area Network) である。ネットワークインタフェース１１２０が通信網に接続する方法は、無線接続であってもよいし、有線接続であってもよい。
- [0030] 例えば計算機１０００は、ネットワークを介してカメラ１０と通信可能に接続されている。ただし、計算機１０００をカメラ１０と通信可能に接続する方法は、ネットワークを介した接続に限定されない。また、計算機１０００は、カメラ１０と通信可能に接続されていなくてもよい。
- [0031] ストレージデバイス１０８０は、情報処理装置２０００の各機能構成部（検出部２０２０、状態推定部２０４０、身長推定部２０８０）を実現するプログラムモジュールを記憶している。プロセッサ１０４０は、これら各プログラムモジュールをメモリ１０６０に読み出して実行することで、各プログ

ラムモジュールに対応する機能を実現する。

[0032] なお、計算機 1000 は、複数の計算機を利用して実現されてもよい。例えば検出部 2020、状態推定部 2040、及び身長推定部 2080 は、それぞれ異なる計算機で実現することができる。この場合、各計算機のストレージデバイスに記憶されるプログラムモジュールは、その計算機で実現される機能構成部に対応するプログラムモジュールだけでもよい。

[0033] <カメラ 10 について>

カメラ 10 は、繰り返し撮像を行って時系列の動画フレーム 14 を生成することにより、動画データ 12 を生成する任意のカメラである。例えばカメラ 10 は、特定の施設や道路などを監視するために設けられている監視カメラである。

[0034] 前述したように、情報処理装置 2000 を実現する計算機 1000 は、カメラ 10 であってもよい。この場合、カメラ 10 は、自身で生成した動画フレーム 14 を解析することで、人物 20 の検出、対象人物の状態の推定、及び対象人物の身長の推定を行う。このような機能を持たせるカメラ 10 としては、例えば、インテリジェントカメラ、ネットワークカメラ、又は IP (Internet Protocol) カメラなどと呼ばれるカメラを用いることができる。

[0035] なお、情報処理装置 2000 の全ての機能をカメラ 10 で実現するのではなく、情報処理装置 2000 の一部の機能のみがカメラ 10 で実現されてもよい。例えば、動画フレーム 14 から人物 20 を検出する機能のみをカメラ 10 で実現し、それ以外の情報処理装置 2000 の機能をサーバ装置で実現する。この場合、サーバ装置は、カメラ 10 から、検出された人物 20 の位置や画像特徴などの種々の情報を取得する。また、サーバ装置は、カメラ 10 によって生成された動画フレーム 14 のうち、人物 20 及びその周辺が含まれる一部の画像領域のみを取得してもよい。

[0036] <処理の流れ>

図 5 は、実施形態 1 の情報処理装置 2000 によって実行される処理の流れを例示する第 1 のフローチャートである。ループ処理 A は、所定の終了条

件が満たされるまで繰り返し実行されるループ処理である。検出部2020は、最新の動画フレーム14を取得する(S104)。検出部2020は、取得した動画フレーム14から人物20を検出する(S106)。ループ処理Bは、検出された各人物20について実行される処理である。ここで、ループ処理Bにおける処理対象の人物20を、人物iと呼ぶ。人物iが、前述した対象人物となる。

[0037] また、図6は、実施形態1の情報処理装置2000によって実行される処理の流れを例示する第2のフローチャートである。この場合、検出部2020は、さらに、検出した人物を、それまでの追跡結果と対応付け、追跡結果を得る(S120)。ループ処理Bは、追跡処理によって得られた追跡対象に含まれる各人物20について実行される処理である。ここでも、ループ処理Bにおける処理対象の人物20を、人物iと呼ぶ。

[0038] 状態推定部2040は、人物iの状態を推定する(S110、S130)。身長推定部2080は、人物iの状態が所定条件を満たすか否かを判定する(S112、S132)。人物iの状態が所定条件を満たす場合(S112、S132: YES)、S102で取得された動画フレーム14から得られる人物iの高さに基づいて、人物iの推定身長を算出する(S114、S134)。人物iの状態が所定条件を満たさない場合(S112、S132: NO)、人物iの推定身長は算出されない。

[0039] なお、ループ処理Aを終了する条件は任意である。例えば、ユーザから所定の入力操作を受け付けた場合に、ループ処理Aを終了する。

[0040] <動画フレーム14の取得方法: S104>

情報処理装置2000は、処理対象とする1つ以上の動画フレーム14を取得する。情報処理装置2000が動画フレーム14を取得する方法は様々である。例えば情報処理装置2000は、カメラ10から送信される動画フレーム14を受信する。また例えば、情報処理装置2000は、カメラ10にアクセスし、カメラ10に記憶されている動画フレーム14を取得する。

[0041] なお、カメラ10は、カメラ10の外部に設けられている記憶装置に動画

フレーム 14 を記憶させてもよい。この場合、情報処理装置 2000 は、この記憶装置にアクセスして動画フレーム 14 を取得する。そのため、この場合、情報処理装置 2000 とカメラ 10 は通信可能に接続されていなくてもよい。

[0042] 情報処理装置 2000 の一部又は全部の機能がカメラ 10 で実現される場合、情報処理装置 2000 は、情報処理装置 2000 自身によって生成された動画フレーム 14 を取得する。この場合、動画フレーム 14 は、例えば情報処理装置 2000 の内部にある記憶装置（例えばストレージデバイス 1080）に記憶されている。そこで情報処理装置 2000 は、これらの記憶装置から動画フレーム 14 を取得する。

[0043] 情報処理装置 2000 が動画フレーム 14 を取得するタイミングは任意である。例えば情報処理装置 2000 は、カメラ 10 によって動画データ 12 を構成する新たな動画フレーム 14 が生成される度に、その新たに生成された動画フレーム 14 を取得する。その他にも例えば、情報処理装置 2000 は、定期的に未取得の動画フレーム 14 を取得してもよい。例えば情報処理装置 2000 が 1 秒間に 1 回動画フレーム 14 を取得する場合、情報処理装置 2000 は、カメラ 10 によって 1 秒間に生成される複数の動画フレーム 14（例えば動画データ 12 のフレームレートが 30fps（frames/second）であれば、30 枚の動画フレーム 14）をまとめて取得する。

[0044] 情報処理装置 2000 は、動画データ 12 を構成する全ての動画フレーム 14 を取得してもよいし、一部の動画フレーム 14 のみを取得してもよい。後者の場合、例えば情報処理装置 2000 は、カメラ 10 によって生成される動画フレーム 14 を、所定数に 1 つの割合で取得する。

[0045] <人物 20 の検出：S106>

検出部 2020 は、動画フレーム 14 から人物 20 を検出する（S102）。動画フレーム 14 から人物 20 を検出する技術には、既知の種々の技術を利用することができる。例えば検出部 2020 は、人物の画像特徴を学習させた検出器を含む。検出器は、動画フレーム 14 から、学習済みの画像特

徴にマッチする画像領域を、人物20を表す領域（以下、人物領域）として検出する。検出器には、例えば、HOG (Histograms of Oriented Gradients) 特徴に基づいて検出を行うものや、CNN (Convolutional Neural Network) を用いるものが利用できる。なお、検出器は、人物20全体の領域を検出するように学習させたものであってもよいし、人物20の一部の領域を検出するように学習させたものであってもよい。例えば頭部や足元を学習させた検出器を用い、頭部位置と足元位置が検出できれば、人物領域を特定できる。その他にも例えば、背景差分によって求まるシルエット情報（背景モデルと差分がある領域の情報）と頭部検出情報を組み合わせることにより、人物領域を求めるように構成されていてもよい。

[0046] 検出器は、検出した人物20に関する情報（以下、検出情報）を出力する。検出情報は、例えば、人物20の位置及び大きさを示す。ここで、検出情報における人物20の位置は、動画フレーム14上の位置（例えば動画フレーム14の左上端を原点とする座標）で表されていてもよいし、実世界座標で表されてもよい。ここで、カメラによって生成された画像に含まれる物体の実世界座標を算出する技術には、既存の技術を利用することができる。例えば、人物20の実世界座標は、カメラパラメータを用いて、動画フレーム14上の位置から算出することができる。

[0047] 人物20の大きさは、例えば、人物の外接矩形（以下、人物矩形とも表記する）や、頭部や足元などの人物の一部の領域の外接矩形の大きさ（例えば、縦、横の辺の長さ、あるいはそれらの平均値など）で表される。この大きさは、動画フレーム14上の大きさであってもよいし、実世界上の大きさであってもよい。

[0048] <人物20の追跡：S120>

なお、図6に示すフローの場合には、検出部2020は、さらに人物の追跡処理も行う。追跡処理は、現時点で入力された動画フレームから検知された人物と、過去のフレームで検知されて追跡中である人物とを対応付ける処理である。複数の動画フレーム14にわたって同一の人物を追跡する技術に

は、例えば、後述する実施形態3において説明する技術を利用することができる。また、その他の既存の技術を利用することもできる。

[0049] 情報処理装置2000は、人物20を追跡することで、人物20の位置や大きさの履歴を表す追跡情報を生成する。ここで、追跡情報に登録されている人物、すなわち既に動画フレーム14から検出されており、追跡対象となっている人物を、追跡対象人物と呼ぶ。

[0050] 図7は追跡情報を例示する図である。図7の追跡情報では、追跡対象人物ごとにテーブル200が生成されている。テーブル200は、対応づけられている追跡対象人物の位置等の履歴を示す。テーブル200は、フレームID202、位置204、状態206、観測値208、動き210、特徴量212、及び領域214を示す。フレームID202は、動画フレーム14の識別子である。例えばフレームID202がnを示すレコードは、フレームIDがnである動画フレーム14における追跡対象人物の位置等を示す。なお、動き210は、移動状態にある追跡対象人物について、その移動の予測に利用される動きモデルのパラメータを示す。動きモデルについては、後述の実施形態3で説明する。また、領域214におけるTLとBRはそれぞれ、左上端(top left)の座標と右下端(bottom right)の座標を表している。観測値208については後述する。

[0051] 図6に示すフローの場合には、検出部2020は、S106の検出時に追跡に必要な情報も抽出してもよい。すなわち、検出情報は、人物20の外見の特徴を表す特徴量を含んでいてもよい。特徴量としては、人物20の大きさ、表面の色や模様(服の色や模様)などを記述する特徴量を用いることができる。より具体的には、色ヒストグラムやカラーレイアウト、エッジヒストグラムやガボール特徴などの特徴量を用いることができる。また、検出情報は、オブジェクトの形状を表す特徴量を含んでもよい。オブジェクトの形状を表す特徴量としては、例えば、MPEG-7で標準化されたシェイプデスク립タを用いることができる。その他にも例えば、人物20の特徴点を抽出し、特徴点ごとにSIFT(Scale Invariant Feature Transform)やSURF(Spee

ded Up Robust Features)などの局所特徴量を抽出するようにしてもよい。その他に例えば、深層学習により学習されたネットワークによる特徴抽出を用いてもよい。

[0052] <対象人物について：S108>

情報処理装置2000は、動画フレーム14から検出される人物20の内の少なくとも1つを、身長を推定する処理の対象（すなわち対象人物）とする。ここで、動画フレーム14から複数の人物20が検出される場合に、どの人物20を対象人物として扱うかを決める方法は任意である。例えば図5のフローチャートに示す処理では、動画フレーム14から検出された全ての人物20を対象人物として扱っている。

[0053] その他にも例えば、情報処理装置2000は、動画フレーム14内の所定の画像領域に含まれる人物20のみを、対象人物として扱ってもよい。所定の画像領域は、予め定まっている領域であってもよいし、動的に定められる領域であってもよい。後者の場合、例えば情報処理装置2000は、動画フレーム14の中から所定の物体を検出し、その物体を基準とした所定の大きさの画像領域から検出される各人物20を、対象人物として扱う。所定の物体は、常時置かれている物体（椅子、棚、ドア、又は銅像など）であってもよいし、放置物体などのように動的に現れる物体であってもよい。なお、常時置かれている物体や動的に現れる物体を画像から検出する技術には、既存の技術を利用することができる。

[0054] <状態の推定：S110>

状態推定部2040は対象人物の状態を推定する（S110）。状態推定部2040によって推定される状態には、少なくとも、対象人物の姿勢に関する情報（例えば、直立と非直立のどちらであるかを示す情報）が含まれる。対象人物の状態を推定する方法には、様々な方法を採用することができる。以下、その方法の具体例をいくつか説明する。

[0055] <<方法1>>

まず、動画フレーム14の中の1枚の画像のみから身長を求める場合（図

5の処理フローの場合)について述べる。この場合、例えば状態推定部2040は、学習によって状態を分別する識別器(以下、状態識別器)を有し、この識別器を利用して状態判別を行う。例えば、「直立」と「非直立」それぞれの状態(人物姿勢)に該当する画像の学習データを準備し、学習データの画像と人物検出結果を入力としたときに、正解の状態が得られるように、状態識別器を学習させる。この学習には、SVM(Support Vector Machine)やニューラルネットワーク等、様々な識別器を用いることができる。

[0056] より具体的には、しゃがみ・屈みなどの非直立状態を、直立状態と区別するように、状態識別器を学習させておく。

[0057] 例えば状態推定部2040は、状態識別器に対し、動画フレーム14、及び入力された動画フレーム14に対する人物の検出結果を入力する。状態識別器は、その動画フレーム14で検出された各人物20について、その状態を表す情報(以下、状態情報)を出力する。その他にも例えば、状態推定部2040は、動画フレーム14から検出された対象人物の画像領域のみを状態識別器に入力したり、対象人物の検出情報を状態識別器に入力したりしてもよい。

[0058] 状態情報は、対象人物の姿勢がどの状態であることを特定する情報であってもよいし、対象人物が各状態であることの尤度を示す情報であってもよい。後者の場合、状態識別器は、対象人物がとりうる各状態について、対象人物がその状態であることの尤度を算出する。そして状態識別器は、状態情報に各状態の尤度を含める。

[0059] なお、状態推定部2040は、状態として直立状態をさらに細かく分類する姿勢情報を判定するように構成されていてもよい。これらの情報を考慮することで、後述するように、身長推定の精度を高めることが可能となる。

[0060] 例えば、直立して歩いている状態の場合には、同じ直立状態でも、股を開いて歩いている姿勢と足が揃っている姿勢とが考えられる。そのため、これらを区別するように状態識別器を学習させることで、状態識別器が両者を判別するようにしてもよい。さらに、股が開いている状態については、その開

き具合に応じて数段階に場合分けして判定するようになっていてもよい。このような状態識別器からは、対象人物の股が閉じているのか開いているのかを表す情報や、股の開き具合を表す情報なども出力される。

[0061] また、状態情報には人物の向きがさらに含まれるようにしてもよい。向きの判定も、状態識別器を予め学習させておくことで可能となる。

[0062] <<方法2>>

次に、複数の動画フレーム14を用いた追跡処理を含む場合（図6のフローの場合）について述べる。なお、方法2は、状態推定部2040において、画像による姿勢判定を行わずに追跡情報に含まれる時系列の人物矩形情報を用いて状態を判定する方法である。方法1のように画像による姿勢判定を行う場合については、方法3として後述する。

[0063] 状態推定部2040は、或る動画フレーム14における対象人物の状態を、その動画フレーム14及び過去複数の動画フレーム14にわたるその対象人物の追跡結果を利用して推定する。なお、複数の動画フレーム14にわたって同一の人物を追跡する技術には、既存の技術を利用することができる。また、情報処理装置2000は、身長推定部2080によって推定される人物20の推定身長を、その人物20の追跡に利用してもよい。その具体的な方法については、後述の実施形態で説明する。

[0064] 追跡の結果が利用可能な場合、「直立状態」、「非直立状態」以外に、「移動状態」と「静止状態」を区別することが可能となる。この場合、例えば、「直立移動状態」、「直立静止状態」、及び「非直立静止状態」という3つの状態のいずれであるのかが判別される。なお、非直立移動状態という状態も選択肢に含めてもよいが、通常は想定されない状態であるため、以下の説明では選択肢に含めないこととする。以下では、まず、直立か非直立かの区別について述べた後、移動か静止かの区別について述べる。

[0065] 情報処理装置2000は、人物20を追跡することで、人物20の位置や大きさの履歴を表す追跡情報を生成する。ここで、追跡情報に登録されている人物、すなわち既に動画フレーム14から検出されており、追跡対象とな

っている人物を、追跡対象人物と呼ぶ。

[0066] 対象人物の追跡情報の履歴が一定数格納されるまでの間は（対象人物のテーブル200のレコード数が一定数に達するまでの間は）、状態推定部2040は、各動画フレーム14から対象人物の実世界上の高さを算出する。対象人物の実世界上の高さは、例えば、動画フレーム14上における対象人物の高さを、カメラパラメータを利用して実世界上の値に変換することで得られる。なお、動画フレーム14上における対象人物の高さは、例えば、動画フレーム14上における対象人物の足元と頭頂部との間の距離として算出することができる。以下、動画フレーム14から算出される、対象人物の実世界上の高さを、「観測値」とも表記する。この観測値が、前述した観測値206に示される値である。すなわち、動画フレーム14から算出した対象人物の観測値が、追跡情報の観測値206に格納される。

[0067] 対象人物について一定数の観測値が蓄積されたら、状態推定部2040は、蓄積された観測値を用いて、対象人物が直立しているか否かを識別するための閾値を算出する。この閾値は、観測値の分布において、人物20が直立していると考えられる領域と、人物20が直立していないと考えられる領域との境界線を意味する。

[0068] この閾値を算出した後から、状態推定部2040は、上記閾値を用いて、動画フレーム14から得られる対象人物の姿勢を推定する。例えば状態推定部2040は、対象人物の観測値が閾値と比較して有意に大きい場合に（例えば、観測値が閾値以上である場合に）、対象人物の状態が直立状態であると推定する。一方、観測値が閾値より有意に大きいとはいえなければ（例えば、観測値が閾値未満であれば）、状態推定部2040は、対象人物の状態が非直立状態であると推定する。

[0069] その他にも例えば、状態推定部2040は、観測値が閾値から乖離している程度を表す指標値を算出し、この指標値を、対象人物の状態が直立状態である尤度としてもよい。

[0070] ここで、閾値の決め方について説明する。履歴が一定数格納されたときに

、実世界上の高さに有意な変化がある場合には、一番高い値が身長に近いと考えられるため、この高さを基準に一定の値だけ減じた値（例えば 5cm を減じた値）を閾値として設定する。なお、この減じる値は、観測値の誤差を考慮して決めることが好適である。例えば、誤差の標準偏差の値に基づいて決定することができる。一方、履歴が一定数格納されたときに、実世界上の高さに有意な変化がない場合（観測値の分布が誤差の範囲と考えられる場合）には、その代表値（平均値、中央値、モード値など）を基準に一定の値だけ減じた値を閾値として設定する。途中で、実世界の高さが定めた閾値よりも有意な差をもって大きくなる場合には、もともと直立ではなかったと考えられるため、新たに得られた観測値を基準にして、閾値を設定し直す。そして、それまでの状態を直立ではなく、非直立と修正するようにする。

[0071] <<移動・静止の考慮>>

前述したように、対象人物がとりうる状態は、直立移動状態、非直立移動状態、及び直立静止状態であってもよい。このためには、移動の有無を判定する必要がある。以下では、移動の有無を判定し、直立移動状態、非直立移動状態、及び直立静止状態のいずれに該当するかを判定する方法について述べる。

[0072] 状態推定部 2040 は、前述した追跡情報を利用して、対象人物が移動しているか否かを判別する。例えば状態推定部 2040 は、動画フレーム 14 から検出された対象人物の位置と、所定時間前（例えば直前のフレーム）における対象人物の位置を比較し、位置の変化が小さければ（例えば所定値以下であれば）、対象人物が静止していると判定する。一方、位置の変化が大きければ、状態推定部 2040 は、対象人物が移動していると判定する。対象人物の位置としては、頭部の位置、人物矩形の所定位置（例えば中心や左上端など）、又は足元位置などを利用することができる。なお、所定時間前の動画フレーム 14 における対象人物の位置だけでなく、過去複数フレームにおける対象人物の位置との比較を行ってもよい。

[0073] 状態推定部 2040 は、対象人物の位置が変化した方向をさらに考慮して

、対象人物が移動しているか否かを判定してもよい。例えば対象人物の頭部が鉛直方向に変化することは、対象人物が立ち上がったたり座ったりすること、すなわち同じ場所で姿勢を変えていることを表すと言える。そこで例えば、状態推定部2040は、対象人物の位置の変化の方向が鉛直方向に近い方向である場合、状態推定部2040は移動していないと判定する。一方、状態推定部2040は、対象人物の頭部の位置の変化の方向が鉛直方向に近い方向でない場合、状態推定部2040は移動していると判定する。

[0074] 対象人物の頭部の位置の変化の方向が鉛直方向に近いかな否かは、例えば、対象人物の位置の変化の方向と鉛直方向のベクトルとのなす角を求めることによって判定することができる。なお、状態推定部2040は、位置の変化の方向と鉛直方向とのなす角が大きい場合に、対象人物が移動していると判定してもよい。

[0075] また、画像上での頭部の動きが鉛直方向に近い場合、状態推定部2040は、足元位置の変動も考慮して、姿勢変動と移動とを判別してもよい。具体的には、足元位置の変動が小さく、静止とみなせる場合には、姿勢変動が起きているだけで人物は移動していないと考えられる。よって、状態推定部2040の状態は、非直立静止状態となる。一方、足元位置が頭部の動きと連動して動いている場合、直立移動状態となる。

[0076] なお、足元が隠れていて見えない場合もある。この場合も、足元が頭部と連動していないと見なせる場合（例えば、頭部がカメラ10の画像の鉛直方向に移動しているにもかかわらず、足元は障害物などに隠れたままである場合など）は、足元位置は静止していると仮定して、非直立静止状態と判定してもよい。また、椅子があり、座ることが想定されている場所の場合には、足元位置がみえなくても、着席する蓋然性が高いと考えられるため、非直立静止状態と判定してもよい。また、静止していると仮定すると矛盾が生じる場合（例えば、人物矩形の高さが想定される高さよりも大幅にずれるなど）には、直立移動状態であると判定してもよい。もし、どちらとも判定が難しい場合には、両方の状態の尤度を同程度にし、将来のどこかのフレームでど

ちらの状態かがはっきりした時点で状態推定をする（状態の尤度に大きな差をつける）ようにしてもよい。

### <<方法3>>

次に、複数の動画フレーム14を用いた追跡処理を含む場合（図6のフローの場合）で、かつ、状態推定部2040において、画像による姿勢判定を行う場合について述べる。この場合、人物の姿勢を画像によって判定可能なため、直立、非直立を直接的に分類することができる。これと、方法2で述べた移動情報と組み合わせることで、「直立移動」、「直立静止」、「非直立静止」の3状態を区別できるようになる。

[0077] さらに、方法1で述べたように、詳細な状態を出力するようになっていてもよい。すなわち、股を開いて歩いている状態か、さらに、その開き具合はどのぐらいか、あるいは、人物の向きはどちらかといった情報も合わせて出力するようになっていてもよい。これらの識別は、方法1のところで述べたように、画像による識別器を生成することで実現できる。これらの情報を考慮することで、後述するように、身長推定の精度を高めることが可能となる。

### [0078] <身長の推定：S114>

図5のフローの場合、身長推定部2080は、動画フレーム14を用いて推定された対象人物の状態が所定条件を満たすときに、その動画フレーム14における対象人物の高さに基づいて、対象人物の身長を推定する（S114）。所定条件は、「対象人物の状態が、直立状態又はこれに近い状態であること」を表す条件である。

[0079] 例えば状態情報が、対象人物の状態を特定する情報であるとする。この場合、所定条件は、「対象人物の状態が直立状態である」という条件となる。

[0080] その他にも例えば、状態推定部2040が対象人物に関し、複数の状態それぞれについて尤度を算出するとする。この場合、所定条件は、「対象人物の状態が直立状態である尤度が閾値以上である」という条件である。

[0081] 身長推定部2080は、動画フレーム14における対象人物の状態が所定

条件を満たしている場合に、その動画フレーム 14 における対象人物の観測値を算出する。前述したように、動画フレーム 14 における対象人物の観測値は、対象人物の動画フレーム 14 上のける高さを、カメラパラメータなどを用いて、実世界上の高さに変換した値である。ここで、対象人物の状態にかかわらず、検出部 2020 が対象人物の観測値を算出して検出情報に含めるように構成されていてもよい。この場合、身長推定部 2080 は、検出情報から観測値を取得する。

[0082] 例えば身長推定部 2080 は、動画フレーム 14 における対象人物の状態が所定条件を満たしている場合、その動画フレーム 14 における対象人物の観測値を、対象人物の推定身長とする。

[0083] その他にも例えば、状態推定部 2040 がより細かい状態情報（例えば、股を開いて歩いている状態か、股の開き具合、人物の向き）も一緒に出力する場合には、所定条件にこれらの条件も含めてもよい。例えば、同じ直立状態であっても、股を開いていない（あるいは開き具合が小さい）場合のみ、身長を推定するようになっていてもよい。あるいは、人物の向きがカメラの光軸方向と大きく異なる場合のみ、身長を推定するようになっていてもよい。

[0084] <身長の推定：S 134>

図 6 のフローの場合も、身長推定部 2080 は、動画フレーム 14 を用いて推定された対象人物の状態が所定条件を満たす場合、その動画フレーム 14 における対象人物の高さに基づいて、対象人物の身長を推定する（S 134）。所定条件は、「対象人物の状態が、直立状態又はこれに近い状態であること」を表す条件である。すなわち、状態が直立静止状態か、あるいは直立移動状態のときとなる。その他にも例えば、状態推定部 2040 が対象人物に関し、複数の状態それぞれについて尤度を算出する場合、所定条件は、「対象人物の状態が直立状態である尤度が閾値以上である」という条件である。

[0085] 身長推定部 2080 は、動画フレーム 14 における対象人物の状態が所定

条件を満たしている場合に、その動画フレーム 14 における対象人物の観測値(実世界上の高さ)を算出する。ここで、対象人物の状態にかかわらず、検出部 2020 が対象人物の観測値を算出して検出情報に含めるように構成されていてもよい。この場合、身長推定部 2080 は、検出情報から観測値を取得する。

[0086] 例えば、身長推定部 2080 は、最新の動画フレーム 14 で得られた観測値に加え、対象人物の状態が所定条件を満たしている過去の動画フレーム 14 (1 つでも複数でもよい) から得られた観測値も用いて、対象人物の推定身長を算出する。具体的には、最新の動画フレーム 14 から得られた観測値と、過去の動画フレーム 14 から得られた観測値の統計値を算出し、その統計値を対象人物の推定身長とする。

[0087] 例えば身長推定部 2080 は、最新の動画フレーム 14 において対象人物の状態が所定条件を満たしていたら、その動画フレーム 14 から得られた観測値、及び対象人物の状態が所定条件を満たす過去所定数の動画フレーム 14 それぞれから得られた観測値を用いて、対象人物の推定身長を算出する。対象人物の状態が所定条件を満たす動画フレーム 14 が得られる度にこのように対象人物の推定身長を算出していくことで、対象人物の推定身長を段々と精度の高い値に更新していくことができる。

[0088] 上記統計値には、様々なものを利用することができる。例えば統計処理は、平均値である。ただし、誤検出の場合など、推定身長が正しく求まらない場合もある。このため、ロバスト推定等の手法により、外れ値を除外する統計処理で統計値を算出してもよい。例えば、RANSAC (Random Sampling Consensus) などの手法を利用することができる。

[0089] その他にも例えば、対象人物の状態の尤度に応じて観測値に重みづけをして加算するようにしてもよい。例えば、対象人物がカメラの近くに位置するために俯角が大きくなると、対象人物の頭頂部分の推定が難しくなり、対象人物の観測値が誤差を含みやすくなる。一方、対象人物がカメラから離れると対象人物の解像度(動画フレーム 14 上での大きさ)が下がり、検出位置

のわずかなずれが身長に推定に影響を与えるようになる。よって、カメラからの俯角や、カメラからの距離に応じて、観測値の誤差の生じやすさを考慮して重みづけして加算し、平均を算出してもよい。誤差の生じやすさについては、あらかじめ身長が既知の人物をカメラで撮影することによって、カメラの俯角や距離に応じた誤差の傾向を把握できるため、この結果に基づいて予め決めておくことができる。すなわち、誤差が大きくなるにつれ、重みが小さくなるように重みを決定し、重みづけ平均で用いるようにすればよい。

[0090] また、人物の状態の種類を考慮した重みをつけるようにしてもよい。上述の状態分類によれば、移動しているときの姿勢は直立のみであるのに対し、静止しているときの姿勢には直立と非直立の2通りがある。よって、移動状態であれば、対象人物の観測値が身長を表わす蓋然性が高い。一方で、移動しているときは多少高さが上下するため、対象人物の観測値は一定の誤差を含む蓋然性が高い。また、静止しているときは観測値が誤差を含む蓋然性が低い。このことから、直立静止状態のとき、もっとも高い確度で対象人物の身長が推定できる。

[0091] そこで、対象人物の状態が直立静止状態である動画フレーム14から得られる対象人物の観測値に大きい重みを付与する。例えば、移動状態から静止状態に移した直後で、計測される観測値の状態が安定しており、かつ、移動状態の時に算出された推定身長に近い場合は、直立の姿勢で静止している確度が高いと見做せる。よって、このような状況で計測された観測値の重みを高くして推定するようにしてもよい。また、同じ移動状態でも、移動速度によって求まりやすさも変わる。例えば、早く移動している場合には、歩幅が大きくなる分、人物の高さの変動も大きくなる。よって、速度が大きい場合には、重みを下げるように制御してもよい。

[0092] さらに、対象人物の向きや股の開き具合なども考慮して重み付けをしてもよい。対象人物が直立移動状態の場合、移動の向きやその時の人物の姿勢によって、対象人物の観測値が身長を表す程度が異なる。具体的には、股が開いている状態では、足が揃っている場合に比べ、高さが低くなる。よって、

対象人物の股が開いている場合、その時の対象人物の観測値は、実際の身長よりも小さい値になりやすい。そこで、足が揃っている姿勢の対象人物の観測値に与える重みを、股が開いている姿勢の対象人物の観測値に与える重みよりも大きくする。

[0093] また、移動の方向がカメラの光軸方向に近い場合には、人物領域として求まる矩形の足元位置が定まりにくくなり、対象人物の観測値に誤差が生じやすい。このことは、股が開いている状態ではより顕著になる。よって、カメラの光軸方向に近い方向に移動している対象人物の観測値の重みを小さくして統計処理を行うようにする。

[0094] また、直立移動状態のときに、身長の概算値を求めておき、静止状態に移行した時点で、概算値から詳細な値を推定するようになっていてもよい。すなわち、人物が直立静止状態に移行した後（あるいは移行する前）、直立移動状態で求まった概算値に近い値で観測値が安定している場合には、正しい身長が求められる蓋然性が高い。よって、このような状態の観測値に高い重みをつけて身長を求めるようになっていてもよい。上述した身長の概算値は、例えば、直立移動状態の対象人物について求まる一時的な高さである。

[0095] なお、推定身長を精度良く求めるためには、動画フレーム14において対象人物の足元と頭部の双方が見えていることが好適であるといえる。しかしながら、障害物や他の人物との重なりによって、足元が隠れる場合がある。身長の推定には、足元が隠れている場合を除くことが望ましい。そこで、上記所定条件に、「対象人物の足元が見えている」という条件を加えてもよい。なお、対象人物の足元が見えているか否かを判定する方法には、様々な方法を利用できる。具体的には、直接的に足元の検知を行ってもよいし、障害物に隠れているかどうか、あるいは、人物と重なっているかどうかの判定を行ってもよい。棚などの障害物は、通常予め定まった位置にあるため、カメラから見たときに足元位置が隠れる床面上の領域は、あらかじめ定めることができる。よって、対象人物の位置がこの領域内にあるか否かを判定することで、対象人物の足元が見えるか否かを判定することができる。また、人と

の重なりについては、検出部 2020 によって検出された各人物 20 の位置に基づいて判定可能である。例えば、対象人物の外接矩形の下端に他の人物 20 の外接矩形が重なっている場合には、対象人物の足元が見えないと判定する。この際、各人物 20 の 3 次元位置を考慮し、カメラから見た対象人物と他の人物 20 との前後関係も用いて、対象人物の足元が見えるか否かを判定してもよい。すなわち、対象人物と他の人物 20 の外接矩形が重なっており、なおかつ他の人物 20 が対象人物よりもカメラの近くにいる場合に、対象人物の足元が見えないと判定する。

[0096] [実施形態 2]

図 8 は、実施形態 2 の情報処理装置 2000 の機能構成を例示する図である。下記で説明する事項を除き、実施形態 2 の情報処理装置 2000 は、実施形態 1 の情報処理装置 2000 と同様の機能を有する。

[0097] 実施形態 2 の情報処理装置 2000 は、一時的高さ推定部 2100 を有する。一時的高さ推定部 2100 は、或る時点における対象人物の一時的な高さを算出する。一時的な高さとは、対象人物が直立しているか否かにかかわらず求まる高さである。例えば対象人物がしゃがんでいる場合、その状態における対象人物の足元から頭頂部までの高さが、その時における対象人物の一時的な高さである。直立状態の時には、一時的な高さは、観測誤差範囲内で身長と一致する。

[0098] <処理の流れ>

図 9 は、実施形態 2 の情報処理装置 2000 によって実行される処理の流れを例示する第 1 のフローチャートである。図 9 のフローチャートは、図 5 のフローチャートに比べ、身長推定 (S114) の後に、対象人物の一時的な高さを算出する処理 (S202) が実行される点が異なる。それ以外は、図 5 のフローチャートと同様である。

[0099] 図 10 は、実施形態 2 の情報処理装置 2000 によって実行される処理の流れを例示する第 2 のフローチャートである。図 10 のフローチャートは、図 6 のフローチャートに比べ、身長推定 (S134) の後に、対象人物の一

時的高さを算出する処理（S 2 1 2）が実行される点が異なる。それ以外の部分については、図 6 のフローチャートと同様である。

[0100] <一時的高さの推定：S 2 0 2>

図 9 のフローの場合には、例えば一時的高さ推定部 2 1 0 0 は、対象人物の状態が前述した所定条件を満たさない動画フレーム 1 4 も含むすべてのフレームにおける対象人物の観測値を、その動画フレーム 1 4 が生成された時点における対象人物の一時的な高さとする。

[0101] <一時的高さの推定：S 2 1 2>

追跡処理が含まれる場合（図 1 0 のフローの場合には）には、その他にも例えば、一時的高さ推定部 2 1 0 0 は、過去複数の動画フレーム 1 4 から得られる情報も利用して、対象人物の一時的な高さを算出してもよい。対象人物の一時的な高さは、時間的に連続的に変化する特性、及び推定身長以下になるという特性を有する。よって、障害物等によって対象人物の足元が見えない場合など、1 つの動画フレーム 1 4 から算出される観測値の信頼度が低いと考えられる場合には、連続性を考慮して時間方向に補完（内挿や外挿）することにより、対象人物の一時的な高さを算出することが好適である。なお、補完の方法は、既存の任意の方式を用いることができる。

[0102] 例えば、時点  $t$  で生成された動画フレーム 1 4 が、推定身長を算出するための所定条件を満たさず、なおかつその動画フレーム 1 4 において対象人物の足元が見えないとする。この場合、例えば一時的高さ推定部 2 1 0 0 は、時点  $t$  よりも過去の複数の動画フレーム 1 4 それぞれから算出された対象人物の一時的な高さの時間変化に対して補完を行うことで、時点  $t$  における対象人物の一時的な高さを算出する。

[0103] なお、時間的に近傍の複数の動画フレーム 1 4 を用いて、対象人物の一時的な高さを 1 つ算出してもよい。例えば、各動画フレーム 1 4 について算出された対象人物の観測値を、ある一定の幅の時間窓内で平均する。こうすることで、各時間窓内について、対象人物の一時的な高さが得られる。この際、各動画フレーム 1 4 から得られる観測値に、観測値の確度に応じた重みづ

けをしてもよい。

[0104] また、姿勢変動が少なく、対象人物の観測値が安定しているときは時間窓の幅を長めにとり、座り込みや立ち上がりなどで、高さが急激に一方向に変化しているときは、時間窓の幅を短くするようにしてもよい。

[0105] ただし、特定の人物状態においては、以下に述べる制約を以って推定することが望ましい。しゃがみや屈み、座りなどの非直立静止状態においては、人物の位置は変化しないと考えられる。特に椅子などがあり、座ることが想定されるエリアにおいては、特にこの傾向が強い。よって、非直立静止状態においては、足元が見えなくても、足元位置は動かさず、頭頂位置のみを変化させて、一時的な高さを算出するようにすればよい。すなわち、複数の動画フレーム14間で、足元位置を共通の位置に設定する。

[0106] <ハードウェア構成の例>

実施形態2の情報処理装置2000を実現する計算機のハードウェア構成は、実施形態1と同様に、例えば図4によって表される。ただし、本実施形態の情報処理装置2000を実現する計算機1000のストレージデバイス1080には、本実施形態の情報処理装置2000の機能を実現するプログラムモジュールがさらに記憶される。

[0107] <作用・効果>

本実施形態の情報処理装置2000によれば、人物20の状態に基づいて、人物20の推定身長と、人物20の一時的な高さ（人物20が直立していないケースにおける高さ）とが、区別して算出される。よって、人物20の推定身長を正確に算出しつつ、その時々における人物20の一時的な高さも精度良く求めることができる。

[0108] [実施形態3]

図11は、実施形態3の情報処理装置2000の機能構成を例示するブロック図である。情報処理装置2000は、位置推定部2120及び更新部2140を有する。下記で説明する事項を除き、実施形態3の情報処理装置2000は、実施形態2の情報処理装置2000と同様の機能を有する。

[0109] 実施形態3の情報処理装置2000は、動画データ12に含まれる人物を追跡する機能を有する。具体的には、前述した追跡情報の生成・更新を行うことで、人物の追跡を行う。

[0110] まず、フローチャートを用いて、人物追跡処理の全体的な流れを説明する。図12は、実施形態3の情報処理装置2000によって実行される処理の流れの概略を示すフローチャートである。情報処理装置2000は、最新の動画フレーム14から人物20を検出する(S302)。情報処理装置2000は、検出した人物20を追跡対象人物として、追跡情報を生成する(S304)。追跡対象人物は、情報処理装置2000によって既に検出された人物20を意味する。

[0111] ループ処理Cは、所定の終了条件が満たされるまで繰り返し実行される処理である。S306において、所定の終了条件が満たされている場合、情報処理装置2000は、図12の処理を終了する。一方、所定の終了条件が満たされていない場合、図12の処理はS308に進む。

[0112] S308において、検出部2020は、最新の動画フレーム14から人物20を検出する。ここで、取得される動画フレーム14の生成時点を第1時点と表記する。位置推定部2120は、追跡情報を用いて、第1時点における各追跡対象人物の位置を推定する(S310)。更新部2140は、動画フレーム14から検出された人物20と、追跡対象人物との対応付けを行う(S312)。更新部2140は、対応付けの結果に基づいて追跡情報を更新する(S314)。

[0113] なお、実施形態3の情報処理装置2000が対象人物の推定身長や一時的な高さを算出するタイミングは任意である。例えば情報処理装置2000は、ループ処理Cの中で、対象人物の推定身長又は一時的な高さの算出を行う。

[0114] <推定位置の算出：S310>

位置推定部2120は、追跡情報を用いて、第1時点における各追跡対象人物の位置を推定する(S310)。追跡情報に示されている追跡対象人物

の位置は、過去の位置（例えば1つ前の動画フレーム14における位置）である。そこで位置推定部2120は、追跡対象人物の過去の位置から、第1時点における追跡対象人物の位置を推定する。

[0115] 上記推定の方法は様々である。例えば位置推定部2120は、追跡情報に示されている追跡対象人物の位置及び人物の動きモデルに基づいて、第1時点における追跡対象人物の位置を予測する。この手法としては、カルマンフィルタを用いる方法や、パーティクルフィルタを用いる方法など、既存の様々な方法を用いることができる。なお、予測する位置情報は画像上の座標であってもよいし、実世界座標であってもよい。画像上の座標の場合には、人物矩形情報として予測結果を生成するようになっていてもよい。

[0116] 人物の動きモデルは追跡対象人物ごとに異なってもよい。例えば位置推定部2120は、追跡対象人物について過去に推定された状態を利用して、その追跡対象人物の動きモデルを決定する。例えば、直立移動状態、直立静止状態、及び非直立静止状態それぞれに対応する予測モデルを予め用意しておく。位置推定部2120は、追跡対象人物の直近の状態（最新の追跡情報に示されている追跡対象人物の状態）に対応した予測モデルを用いて、その追跡対象人物の推定位置を算出する。

[0117] 例えば直立移動状態では、追跡情報に示される追跡対象人物の位置の履歴に基づいて動きを予測するモデルを用いる。具体的な例としては、以下のような等速直線運動モデルを用いることができる。まず、等速直線運動モデルは、過去所定期間の追跡情報を用いて追跡対象人物の速度ベクトルを求め、その速度ベクトル及びフレーム時間間隔に応じて、追跡対象人物の動きベクトルを求める。そして、等速直線運動モデルは、直近の追跡情報に示される追跡対象人物の位置に上記動きベクトルを加算することで、第1時点における追跡対象人物の推定位置を算出する。

[0118] 直立静止状態では、追跡対象人物は動かないと予測できるため、移動量0のモデルを用いる。すなわち、第1時点における追跡対象人物の位置は、直近の追跡情報に示されている位置と同一となる。

[0119] 非直立静止状態では、歩行による移動はないものの、姿勢変動が起こりうる。例えば、足元位置は動かなくても、頭部位置は、鉛直に近い方向に動き得る。そこで、非直立静止状態では、このような姿勢変動を予測するモデルを用いる。

[0120] ここで、位置推定部 2120 は、上述のように追跡対象人物の状態を考慮した位置推定を、最も確度が高い状態（最尤状態）についてのみ行ってもよいし、複数の状態それぞれについて行ってもよい。後者の場合、複数の状態それぞれに対応した推定位置が算出される。

[0121] <人物 20 と追跡対象人物との対応付け：S312>

更新部 2140 は、第 1 時点の動画フレーム 14 から検出された人物 20 と、追跡対象人物との対応付けを行う。この対応付けは、第 1 時点の動画フレーム 14 から検出された各人物 20 が、どの追跡対象人物に相当するのかを特定する処理である。図 13 は、第 1 時点の動画フレーム 14 から検出された人物 20 と、追跡対象人物との対応付けを例示する図である。図 13 において、互いに両矢印で結ばれている人物 20 と追跡対象人物が、互いに対応付けられた人物 20 と追跡対象人物である。

[0122] 上記対応付けを行う方法は様々である。例えば身長推定部 2080 は、予測した追跡対象人物の位置と検出された人物 20 の位置の近さや、両者の外見特徴量の類似性等に基づいて、両者が対応づく可能性を表す尤度（以後対応づけ尤度と呼ぶ）を算出し、対応付けを行うことができる。例えば、尤度をコストに変換し、追跡対象人物と検出された人物との対応付け問題をコスト最小化問題に帰着させ、ハンガリアン法などのアルゴリズムによって解くことができる。

[0123] ここで、対応づけ尤度を求める際、人物の近さ情報は、例えば実空間上の距離で判定する。すなわち、動画フレーム 14 上の位置をカメラパラメータと人物の高さ情報を用いて実空間上の位置座標に変換し、座標間の距離を求める。

[0124] この際、更新部 2140 は、上記変換に利用する人物の高さを、人物の状

態に応じて決定する。前述したように、情報処理装置 2000 は、対象人物の推定身長と、一時的な高さとを区別して算出している。また、各時点における対象人物の一時的な高さは、追跡情報に記憶されている。

[0125] そこで更新部 2140 は、第 1 時点における人物 20 の状態に応じて、人物 20 の推定身長と一時的な高さのいずれかを用いて位置を算出する。具体的には、直立静止状態又は直立移動状態では、立った状態で静止しているか移動しているかであるため、人物 20 の高さは身長と等しいと見なしてよい。よって、更新部 2140 は、動画フレーム 14 から検出された人物 20 の頭頂位置の高さが身長であると見なし、カメラパラメータを用いて、人物 20 の動画フレーム 14 上の座標を実世界座標に変換する。一方、非直立静止状態の場合には、人物 20 の高さは身長とは異なるため、第 1 時点における人物 20 の一時的な高さを用いて算出する。

[0126] なお、前述のように、位置推定部 2120 において非直立静止状態の場合に人物 20 の頭部の動きも予測している場合には、この動きによって人物 20 の一時的な高さが変化することも反映させて、人物 20 の位置を算出してもよい。すなわち、過去の追跡情報に含まれる人物 20 の一時的な高さをそのまま用いるのではなく、頭部の動きによって変化が生じる分を人物 20 の一時的な高さに反映させ、位置情報を推定してもよい。

[0127] また、人物の近さ情報は、画像上の座標で計算するようになっていてもよい。たとえば、人物矩形の特定の位置(例えば足元位置)の画像上での座標を求め、その座標間の距離によって近さを判定するようになっていてもよい。あるいは、人物矩形同士の重なりによって、近さを判定するようになっていてもよい。この場合は、矩形同士の重なりが大きいほど、人物同士が近づいていると判定される。

[0128] なお、追跡対象人物の推定位置が、最尤状態についてのみ求まっているとする。この場合、更新部 2140 は、各追跡対象人物の最尤状態についての位置と、第 1 時点の動画フレーム 14 から検出された人物 20 の位置との間で、対応づけ尤度を求めるようにすればよい。一方、複数の状態について追

跡対象人物の推定位置が求まっている場合には、それぞれの状態に対して対応付け尤度を算出し、最も対応づけ尤度が高くなる状態とそのときの対応付け尤度を選択する。

[0129] また、外見特徴量の類似性も一緒に考慮してもよい。この場合、特徴量間の距離や類似性を表す尺度を用いて判定する。尺度としては、ユークリッド距離やヒストグラムインターセクション等、既存の様々なものを用いることができる。

[0130] <追跡情報の更新>

更新部2140は、上記対応付けの結果に基づいて、追跡情報に示される追跡対象人物の情報を更新する。具体的には、第1時点における各追跡対象人物に関する情報が、追跡情報に追加される。

[0131] 例えば更新部2140は、追跡対象人物の位置情報を更新する。例えば更新部2140は、第1時点における追跡対象人物の位置を、その追跡対象人物と対応づけられた人物20の位置とする。その他にも例えば、更新部2140は、第1時点における追跡対象人物の位置情報を、第1時点におけるその追跡対象人物の推定位置と、その追跡対象人物と対応づけられた人物20の位置とを重みづけ加算した位置としてもよい。

[0132] 追跡対象人物の動き情報については、更新した位置情報と動きの予測情報との差異に基づいて、動きモデルのパラメータを更新する。例えば等速直線運動を仮定できる人物の場合は、推定位置と更新位置の差分を現在の動きに加算することで、更新することが考えられる。あるいは、カルマンフィルタにより追跡を行っている場合には、既知のカルマンフィルタの更新式に基づいて、位置情報とともに動き情報を更新してもよい。

[0133] 領域情報の更新については、領域情報が人物の外接矩形である場合には、更新された位置情報に基づいて、矩形を再計算してもよい。このとき、カメラのキャリブレーション情報を用いて、位置の移動だけでなく、見かけの大きさの変化等も考慮して、矩形情報を更新してもよい。

[0134] ここで、動画フレーム14から検出された人物20の中には、対応する追

跡対象人物が存在しないものもありうる。この人物20については、新たに現れた（カメラ10の撮像範囲に入った）人物かどうかを判定し、新たに現れた人物と判定された場合には、新たな追跡対象人物として追跡情報に追加する。一方で、新たに現れた人物でなければ、追跡情報への追加は行われない。例えば、この人物20の領域が、既存の追跡対象人物の領域と大きく重なっている場合には、誤検知である可能性が高いと判定し、新規追加を行わない。

[0135] 一方で、追跡対象人物の中には、対応する人物20が存在しないものもありうる。この追跡対象人物については、カメラ10の撮像範囲から外れた人物かどうかを判定し、撮像範囲から外れた人物と判定された場合には、追跡情報から該当人物を除く。例えば、前回の更新時にカメラ10の撮像範囲の縁のあたりで、撮像範囲の外に向かって移動していた追跡対象人物や、前回更新時に出口付近に存在していた追跡対象人物が、どの人物20とも対応付かない場合には、カメラ10の撮像範囲から外れた人物であるといえる。ここで、追跡情報から該当人物を除く代わりに、追跡情報に、該当人物がカメラ10の撮像範囲から外れたことを示す情報を付加する（例えば、除外フラグのビットを1にする）ようにしてもよい。

[0136] ＜ハードウェア構成の例＞

実施形態3の情報処理装置2000を実現する計算機のハードウェア構成は、実施形態1と同様に、例えば図4によって表される。ただし、本実施形態の情報処理装置2000を実現する計算機1000のストレージデバイス1080には、本実施形態の情報処理装置2000の機能を実現するプログラムモジュールがさらに記憶される。

[0137] ＜作用・効果＞

本実施形態の情報処理装置2000によれば、追跡対象人物の推定位置の算出や、追跡対象人物と動画フレーム14から検出された人物20との対応付けにおいて、人物の状態が考慮され、人物の推定身長と一時的な高さのうち、いずれか適切な方が用いられる。こうすることで、人物の追跡をより高

精度に行えるようになる。

[0138] 人物の追跡をより高精度に行うことで、その時々における人物の位置を高精度に推定することができる。これにより、従来よりも高精度な動線解析が可能になる。例えば、店舗において、顧客の動線を分析してマーケティングに活用したり、店員の動きを可視化して、作業効率を計測したりする用途で用いることができる。また、倉庫や工場において、作業員がどのように動いているかを可視化し、ワークフローや業務効率の改善の検討に活用できる。

[0139] [実施形態4]

実施形態4の情報処理装置2000は、実施形態3の情報処理装置2000と同様に、例えば図11で表される。以下で説明する事項を除き、実施形態4の情報処理装置2000は、実施形態3の情報処理装置2000と同様の機能を有する。

[0140] 実施形態4の情報処理装置2000では、複数のカメラ10から動画データ12を得ることを前提とする。そのため、以下の特徴を有する。

[0141] <検出情報の生成について>

検出部2020は、検出された各人物20がどのカメラ10によって撮像されたものであるかを特定できるように、検出情報を生成する。具体的には、検出情報に、どのカメラ10によって生成された動画フレーム14から検出された情報であることを示す識別子（カメラ識別子）を設ける。例えば検出部2020は、複数のカメラ10それぞれについて個別の検出情報を生成し、各検出情報にカメラ識別子を対応づける。その他にも例えば、検出部2020は、複数のカメラ10から検出される人物20を全て示した一つの検出情報を生成し、各レコードにどのカメラ10から検出した人物であることを示すようにしてもよい。

[0142] <追跡情報の更新について>

複数のカメラ10それぞれから得た動画フレーム14を得ると、動画フレーム14ごとに、検出される人物20が異なるといえる。ただし、カメラ10の撮像範囲が一部重なっていれば、複数の動画フレーム14から同一の人

物 20 が検出されることもある。

[0143] 情報処理装置 2000 は、このように動画フレーム 14 ごとに検出される人物 20 が異なることを考慮し、或るカメラ 10 から得られる動画フレーム 14 を利用して追跡情報を更新する際に、更新対象の追跡対象人物を、追跡情報に含まれる追跡対象人物の一部に絞り込む。例えば情報処理装置 2000 が、第 1 カメラによって生成された動画フレーム 14 を用いて、追跡情報の更新を行うとする。この場合、位置推定部 2120 は、追跡情報に示される追跡対象人物のうち、第 1 カメラの撮像範囲に存在する蓋然性が高い追跡対象人物のみを対象として、推定位置の算出を行う。例えば、位置推定部 2120 は、予め第 1 カメラから撮像可能な実空間上の領域を求めておき、各追跡対象人物がその領域に含まれるかどうかを位置情報から判定することにより、第 1 カメラの撮像範囲に含まれると推定される追跡対象人物を抽出する。そして位置推定部 2120 は、抽出された追跡対象人物に対してのみ推定位置の算出を行う。

[0144] また、更新部 2140 は、上記抽出された追跡対象人物のみを対象として、動画フレーム 14 から検出される人物 20 との対応付け、及び追跡情報の更新を行う。この際、追跡対象人物の中で、第 1 カメラの撮像範囲に含まれない人物に対しては、第 1 カメラの撮像範囲に含まれないために追跡情報の更新を行っていないことを表す情報を追跡情報に含めるようにし、後段の処理で用いるようにしてもよい。

[0145] <状態の推定について>

カメラ 10 が複数ある場合、カメラによって人物の状態判定のしやすさが異なることがある。実施形態 4 の状態推定部 2040 は、この点を考慮して人物の状態を推定する。例えば、人物がカメラに近い方が、動画フレーム 14 上での人物領域の大きさ（解像度）が大きくなり、人物の動き等が判定しやすくなる。また、人物の動きを判定する際、カメラの光軸方向の動きに比べ、光軸と垂直な方向に対する動きの方が判定しやすい。このように、カメラと人物の位置関係等によって、人物の状態判定のしやすさ、確からしさが

変化する。

[0146] そこで状態推定部2040は、複数のカメラ10それぞれから得られる動画フレーム14を利用して各人物20の状態を推定し、その中で最も信頼度が高い状態を利用する。例えば状態推定部2040は、各動画フレーム14について、その動画フレーム14を生成したカメラ10と検出された人物20との距離（近い方が信頼度が高い）や、人物20の動きとカメラ10の光軸とのなす角の関係（垂直に近い方が信頼度が高い）に基づいて、状態推定の信頼度を算出する。そして、信頼度が最も高い動画フレーム14を利用して推定された人物20の状態を、その人物20の状態として利用する。

[0147] また、各動画フレーム14について、対象人物が他の人物や障害物によって隠されているかどうかを判定し、状態判定を行うカメラの優先度に反映させてもよい。すなわち、対象人物が他の人物や障害物に隠されていない動画フレーム14を優先的に利用して、その対象人物の状態を推定する。

[0148] <推定身長や一時的な高さの補正>

対象人物が複数のカメラ10によって撮像される場合、身長推定部2080は、同一時刻について複数のカメラ10がそれぞれ生成する動画フレーム14を利用して（すなわち、それぞれ異なるカメラ10によって生成される複数の動画フレーム14を利用して）、推定身長の精度を高める。人物の推定身長が実際の身長と異なる場合、その推定身長を用いて実空間上での人物位置を推定しようとする、カメラの奥行き方向にずれた位置に射影される。例えば、実際の身長よりも推定身長が小さい場合には、カメラパラメータを用いて画像上の座標から実空間上の座標に変換すると、実際の位置よりもカメラから遠くにいるように変換されてしまう。逆に、推定身長の方が大きければ、実際の位置よりも近くにいるように変換される。よって、推定身長と実際の身長にずれがあると、同じ時刻に複数のカメラでそれぞれ位置を求めた場合に、同一人物であるにもかかわらず、カメラごとに異なった位置に人物が投影される。逆に言えば、このように位置がずれている場合に、位置が合うように推定身長を補正することで、推定身長を正しい値に近づけるこ

とができる。

[0149] 具体的には、推定身長に一定値を加算したときに、元の推定身長の場合よりも、カメラ間で求まる実空間上での位置の差が縮まる場合には、推定身長を増加させる。逆に推定身長に一定値を減算したときに、カメラ間で求まる実空間位置の差が縮まる場合には、推定身長を減らすようにする。その結果、カメラ間で位置が十分近くなれば、そのときの推定身長は真値に近いと考えられるため、求めた値を推定身長とする。このように各時点における推定身長の精度が高まれば、各時点における推定身長の統計値として最終的に算出される推定身長の精度も高まる。

[0150] なお、人物の一時的な高さについても、同様の方法で補正することができる。

[0151] ここで、人物検出の結果に誤検出や未検出が含まれる場合もあるため、一度に推定身長や一時的な高さを大きく変更させずに、時間の経過とともに徐々に変更させて合わせていく方法も考えられる。この場合、徐々に推定身長や一時的な高さが変化するため、なめらかな動線情報を得ることが可能になる。また、突発的に誤った検出結果の影響を大きく受けることなく、高精度な位置推定が可能となる。

[0152] なお、複数のカメラから捉えることができる人物であっても、あるカメラから見た時には、他の人物と重なって、人物矩形を高い信頼度で推定できない場合には、そのカメラの情報は、上述の比較に用いないようにする。追跡対象人物同士に重なりが生じるかどうかは、カメラパラメータを用いて追跡対象人物の位置情報を画像上の位置に変換し、各追跡対象人物の画像上での存在領域を推定し、その領域が他の人物の領域と重なるかどうかによって判定できる。

[0153] <ハードウェア構成の例>

実施形態4の情報処理装置2000を実現する計算機のハードウェア構成は、実施形態1と同様に、例えば図4によって表される。ただし、本実施形態の情報処理装置2000を実現する計算機1000のストレージデバイス

1080には、本実施形態の情報処理装置2000の機能を実現するプログラムモジュールがさらに記憶される。

[0154] <作用・効果>

本実施形態の情報処理装置2000によれば、複数のカメラ10から得られる動画フレーム14を利用して、追跡情報の更新や推定身長の実算などが行われる。よって、人物の追跡や身長の実定などを、より高精度に行うことができる。

[0155] 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、上記各実施形態を組み合わせた構成や、上記以外の様々な構成を採用することもできる。

## 請求の範囲

- [請求項1]            動画フレームから人物を検出する検出手段と、  
                     前記検出の結果を用いて、対象人物の状態を推定する状態推定手段と、  
                     前記対象人物の状態が所定条件を満たす場合、前記動画フレームにおける前記対象人物の高さに基づいて、前記対象人物の身長を推定する身長推定手段と、  
                     を有する、情報処理装置。
- [請求項2]            前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、  
                     前記対象人物の状態が前記直立かつ移動の状態又は前記直立かつ静止の状態である場合に前記所定条件が満たされる、請求項1に記載の情報処理装置。
- [請求項3]            前記身長推定手段は、前記推定された状態が所定条件を満たす複数の動画フレームそれぞれから、前記対象人物の実世界上の高さを算出し、算出された複数の高さを統計処理することで、前記対象人物の推定身長を算出する、請求項1又は2に記載の情報処理装置。
- [請求項4]            前記身長推定手段は、複数の動画フレームそれぞれから算出した前記対象人物の実世界上の高さに対し、前記対象人物と前記動画フレームを生成するカメラの向きによって定まる俯角、及び前記対象人物の前記動画フレーム上の解像度に基づく重みを付与して、前記統計処理を行う、請求項3に記載の情報処理装置。
- [請求項5]            前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、  
                     前記状態推定手段は、前記対象人物の股の開き具合か、又は前記対象人物の向きか少なくとも一方をさらに推定し、  
                     前記統計処理において、前記対象人物の状態が前記直立かつ移動の状態である場合、前記対象人物の股の開き具合が小さいほど、前記対

象人物の実世界上の高さに対して大きい重みが付与されるか、又は前記対象人物の向きが前記動画フレームを撮像するカメラの向きと異なる度合いが大きいほど、前記対象人物の実世界上の高さに対して大きい重みが付与される、請求項3に記載の情報処理装置。

[請求項6] 前記状態推定手段は、

前記対象人物が含まれる複数の動画フレームそれぞれから、前記対象人物の実世界上における高さを算出し、前記算出された高さに基づいて、人物が直立しているか否かを判定する閾値を決定し、

その後に生成される前記動画フレームについては、前記対象人物について実世界上の高さを算出し、その高さと前記閾値とを比較することで、その動画フレームにおいて前記対象人物が直立しているか否かを判定し、

前記対象人物が直立している場合に前記所定条件が満たされる、請求項3乃至5いずれか一項に記載の情報処理装置。

[請求項7] 前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、

前記身長推定手段は、前記対象人物の状態を考慮して重みづけを行って身長を推定し、直立移動状態の時を基準にして身長の概算値を算出し、直立移動状態の前後で身長の概算値に近い値で安定している観測値の重みを高くして身長を推定する、請求項3に記載の情報処理装置。

[請求項8] 前記動画フレームにおける前記対象人物の高さに基づいて、前記対象人物の一時的な高さを推定する一時的高さ推定手段を有する、請求項1乃至7いずれか一項に記載の情報処理装置。

[請求項9] 前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、

前記一時的高さ推定手段は、人物の足元位置が見えない場合でも、直立静止状態、非直立静止状態においては、画像上の足元位置は動い

ていないことを仮定して頭部位置のみからでも一時的な高さの観測値を算出する、請求項 8 に記載の情報処理装置。

[請求項10] 前記検出手段は、2 つ以上の、カメラによって生成されたそれぞれの画像に対して人物検出処理を行って、カメラごとに人物検出結果を生成し、

前記一時的高さ推定手段及び前記身長推定手段は、複数のカメラで観測可能な領域に存在する追跡対象人物に対しては、それぞれのカメラから取得した、生成時刻の差が所定時間以内である画像に対する検出結果に基づく追跡結果に含まれる位置座標を比較し、その座標間の距離が小さくなるように、身長や一時的高さの観測値を修正する、請求項 8 又は 9 に記載の情報処理装置。

[請求項11] 前記一時的高さ推定手段及び前記身長推定手段は、人物が直立移動状態のときは、人物が同一位置に存在すると見做せる程度に前記所定時間を短く設定し、人物が直立静止状態か非直立静止状態の場合には、人物が同一位置に存在し、姿勢変動が生じていないと見做せる範囲で前記所定時間を長く設定する、請求項 8 乃至 10 いずれか一項に記載の情報処理装置。

[請求項12] 前記検出手段は、第 1 時点に生成された第 1 動画フレームから人物を検出し、

前記第 1 時点よりも前に生成された前記動画フレームから検出された追跡対象人物に関する情報の履歴を示す追跡情報を利用して、前記第 1 時点における各追跡対象人物の位置を推定する推定位置算出手段と、

前記第 1 動画フレームから検出された各人物の実世界上又は画像上の位置と、前記第 1 時点における各追跡対象人物の実世界上又はまたは画像上の位置とを比較することで、前記第 1 動画フレームから検出された人物と前記追跡対象人物との対応付けを行い、その結果を用いて、前記追跡情報を更新する更新手段と、を有し、

前記更新手段は、前記第 1 動画フレームから検出された人物が直立している場合、その人物の推定身長を用いてその人物の実世界上の位置を推定し、前記第 1 動画フレームから検出された人物が直立していない場合、その人物の一時的な高さを用いてその人物の実世界上の位置を推定する、請求項 8 乃至 11 いずれか一項に記載の情報処理装置。

[請求項13] コンピュータによって実行される制御方法であって、  
動画フレームから人物を検出する検出ステップと、  
前記検出の結果を用いて、対象人物の状態を推定する状態推定ステップと、  
前記対象人物の状態が所定条件を満たす場合、前記動画フレームにおける前記対象人物の高さに基づいて、前記対象人物の身長を推定する身長推定ステップと、  
を有する、制御方法。

[請求項14] 前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、  
前記対象人物の状態が前記直立かつ移動の状態又は前記直立かつ静止の状態である場合に前記所定条件が満たされる、請求項 13 に記載の制御方法。

[請求項15] 前記身長推定ステップにおいて、前記推定された状態が所定条件を満たす複数の動画フレームそれぞれから、前記対象人物の実世界上の高さを算出し、算出された複数の高さを統計処理することで、前記対象人物の推定身長を算出する、請求項 13 又は 14 に記載の制御方法。

[請求項16] 前記身長推定ステップにおいて、複数の動画フレームそれぞれから算出した前記対象人物の実世界上の高さに対し、前記対象人物と前記動画フレームを生成するカメラの向きによって定まる俯角、及び前記対象人物の前記動画フレーム上の解像度に基づく重みを付与して、前

記統計処理を行う、請求項 15 に記載の制御方法。

[請求項17] 前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、

前記状態推定ステップにおいて、前記対象人物の股の開き具合か、又は前記対象人物の向きか少なくとも一方をさらに推定し、

前記統計処理において、前記対象人物の状態が前記直立かつ移動の状態である場合、前記対象人物の股の開き具合が小さいほど、前記対象人物の実世界上の高さに対して大きい重みが付与されるか、又は前記対象人物の向きが前記動画フレームを撮像するカメラの向きと異なる度合いが大きいほど、前記対象人物の実世界上の高さに対して大きい重みが付与される、請求項 15 に記載の制御方法。

[請求項18] 前記状態推定ステップにおいて、

前記対象人物が含まれる複数の動画フレームそれぞれから、前記対象人物の実世界上における高さを算出し、前記算出された高さに基づいて、人物が直立しているか否かを判定する閾値を決定し、

その後に生成される前記動画フレームについては、前記対象人物について実世界上の高さを算出し、その高さと前記閾値とを比較することで、その動画フレームにおいて前記対象人物が直立しているか否かを判定し、

前記対象人物が直立している場合に前記所定条件が満たされる、請求項 15 乃至 17 いずれか一つに記載の制御方法。

[請求項19] 前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、

前記身長推定ステップにおいて、前記対象人物の状態を考慮して重みづけを行って身長を推定し、直立移動状態の時を基準にして身長の概算値を算出し、直立移動状態の前後で身長の概算値に近い値で安定している観測値の重みを高くして身長を推定する、請求項 15 に記載の制御方法。

- [請求項20] 前記動画フレームにおける前記対象人物の高さに基づいて、前記対象人物の一時的な高さを推定する一時的高さ推定ステップを有する、請求項13乃至19いずれか一つに記載の制御方法。
- [請求項21] 前記対象人物がとりうる状態は、直立かつ移動の状態、直立かつ静止の状態、及び非直立かつ静止の状態を含み、  
前記一時的高さ推定ステップにおいて、人物の足元位置が見えない場合でも、直立静止状態、非直立静止状態においては、画像上の足元位置は動いていないことを仮定して頭部位置のみからでも一時的な高さの観測値を算出する、請求項20に記載の制御方法。
- [請求項22] 前記検出ステップにおいて、2つ以上の、カメラによって生成されたそれぞれの画像に対して人物検出処理を行って、カメラごとに人物検出結果を生成し、  
前記一時的高さ推定ステップ及び前記身長推定ステップにおいて、複数のカメラで観測可能な領域に存在する追跡対象人物に対しては、それぞれのカメラから取得した、生成時刻の差が所定時間以内である画像に対する検出結果に基づく追跡結果に含まれる位置座標を比較し、その座標間の距離が小さくなるように、身長や一時的高さの観測値を修正する、請求項20又は21に記載の制御方法。
- [請求項23] 前記一時的高さ推定ステップ及び前記身長推定ステップにおいて、人物が直立移動状態のときは、人物が同一位置に存在すると見做せる程度に前記所定時間を短く設定し、人物が直立静止状態か非直立静止状態の場合には、人物が同一位置に存在し、姿勢変動が生じていないと見做せる範囲で前記所定時間を長く設定する、請求項20乃至22いずれか一項に記載の制御方法。
- [請求項24] 前記検出ステップにおいて、第1時点に生成された第1動画フレームから人物を検出し、  
前記第1時点よりも前に生成された前記動画フレームから検出された追跡対象人物に関する情報の履歴を示す追跡情報を利用して、前記

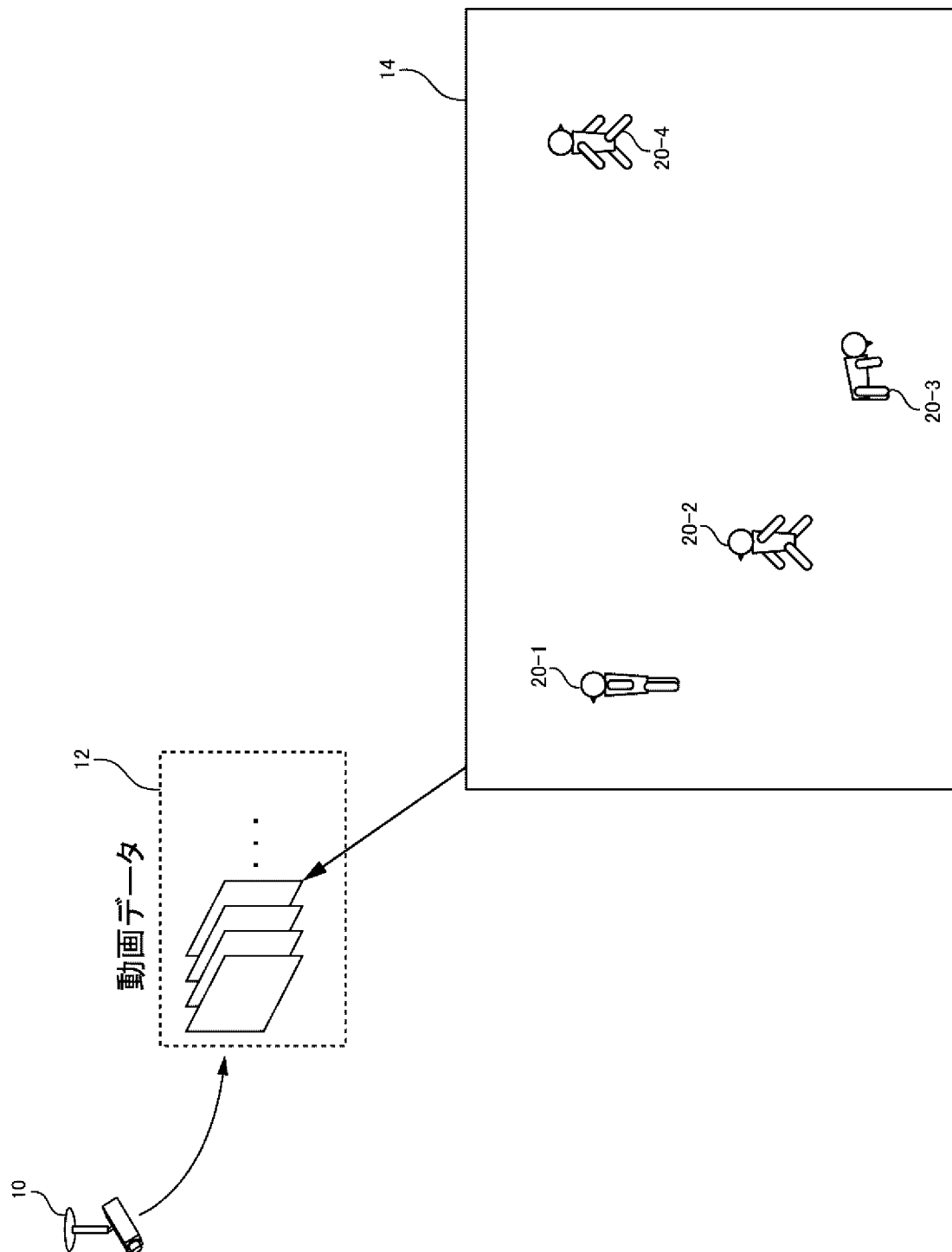
第1時点における各追跡対象人物の位置を推定する推定位置算出ステップと、

前記第1動画フレームから検出された各人物の実世界上又は画像上の位置と、前記第1時点における各追跡対象人物の実世界上又は画像上の位置とを比較することで、前記第1動画フレームから検出された人物と前記追跡対象人物との対応付けを行い、その結果を用いて、前記追跡情報を更新する更新ステップと、を有し、

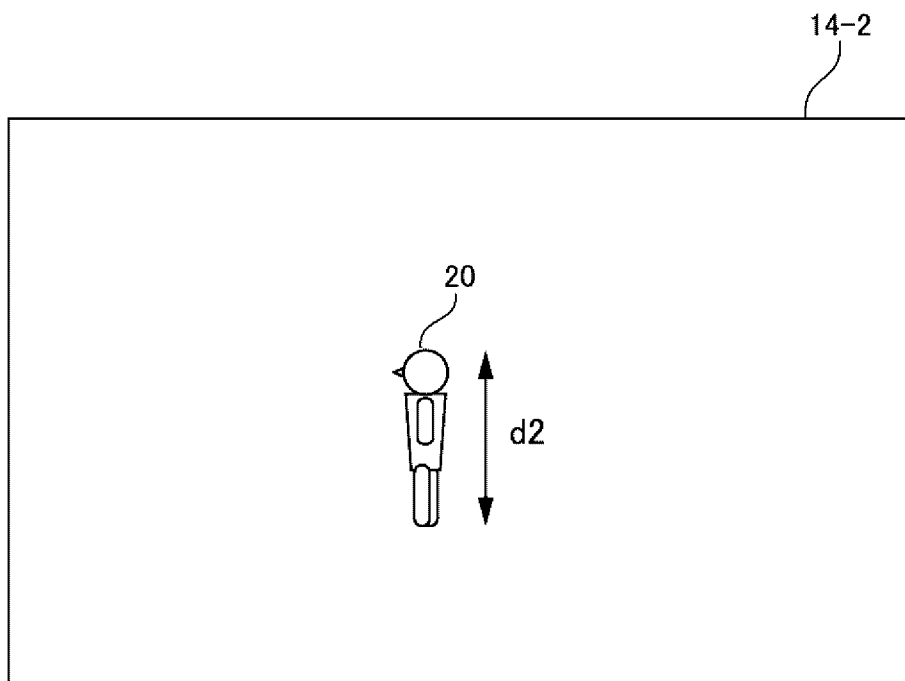
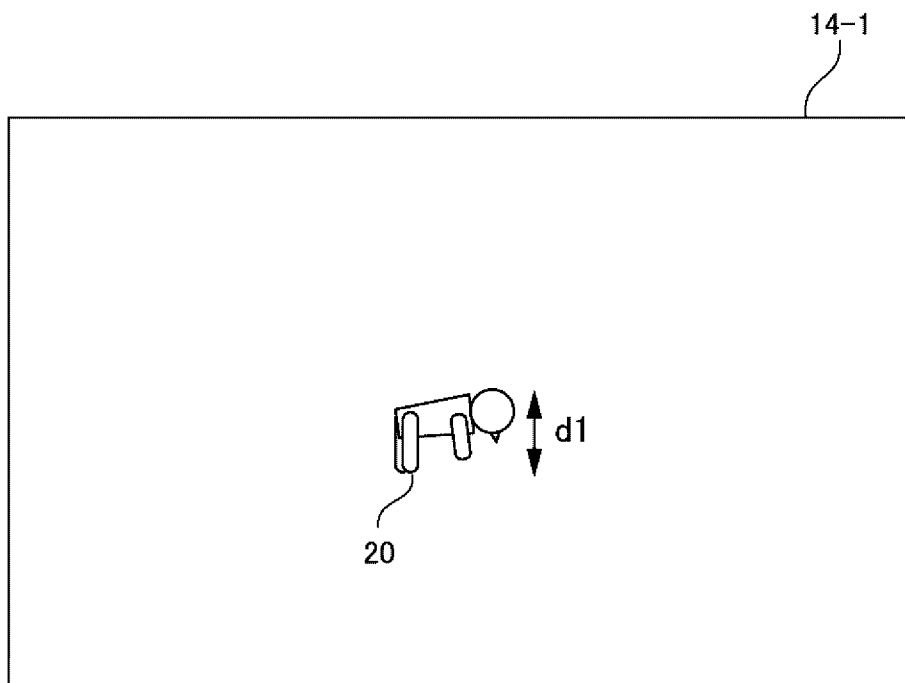
前記更新ステップにおいて、前記第1動画フレームから検出された人物が直立している場合、その人物の推定身長を用いてその人物の実世界上の位置を推定し、前記第1動画フレームから検出された人物が直立していない場合、その人物の一時的な高さを用いてその人物の実世界上の位置を推定する、請求項20乃至23いずれか一項に記載の制御方法。

[請求項25]      請求項13乃至24いずれか一項に記載の制御方法の各ステップをコンピュータに実行させるプログラム。

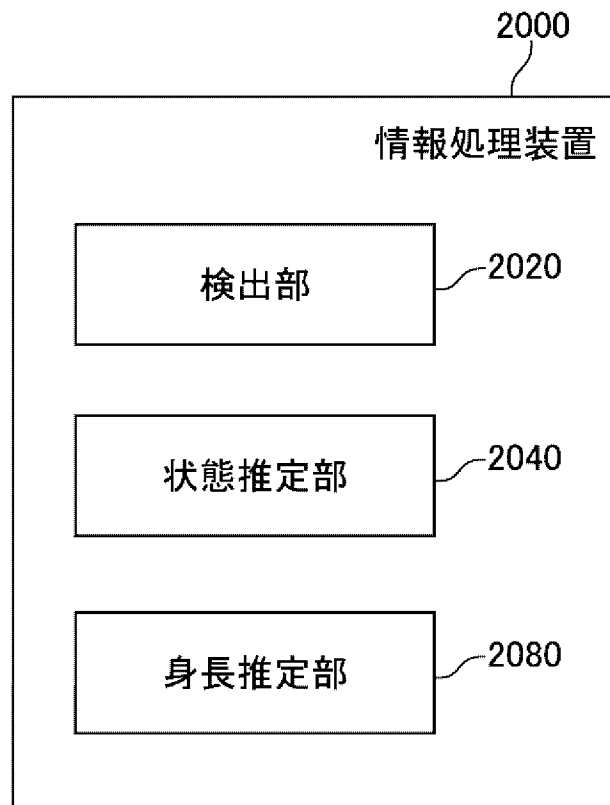
[図1]



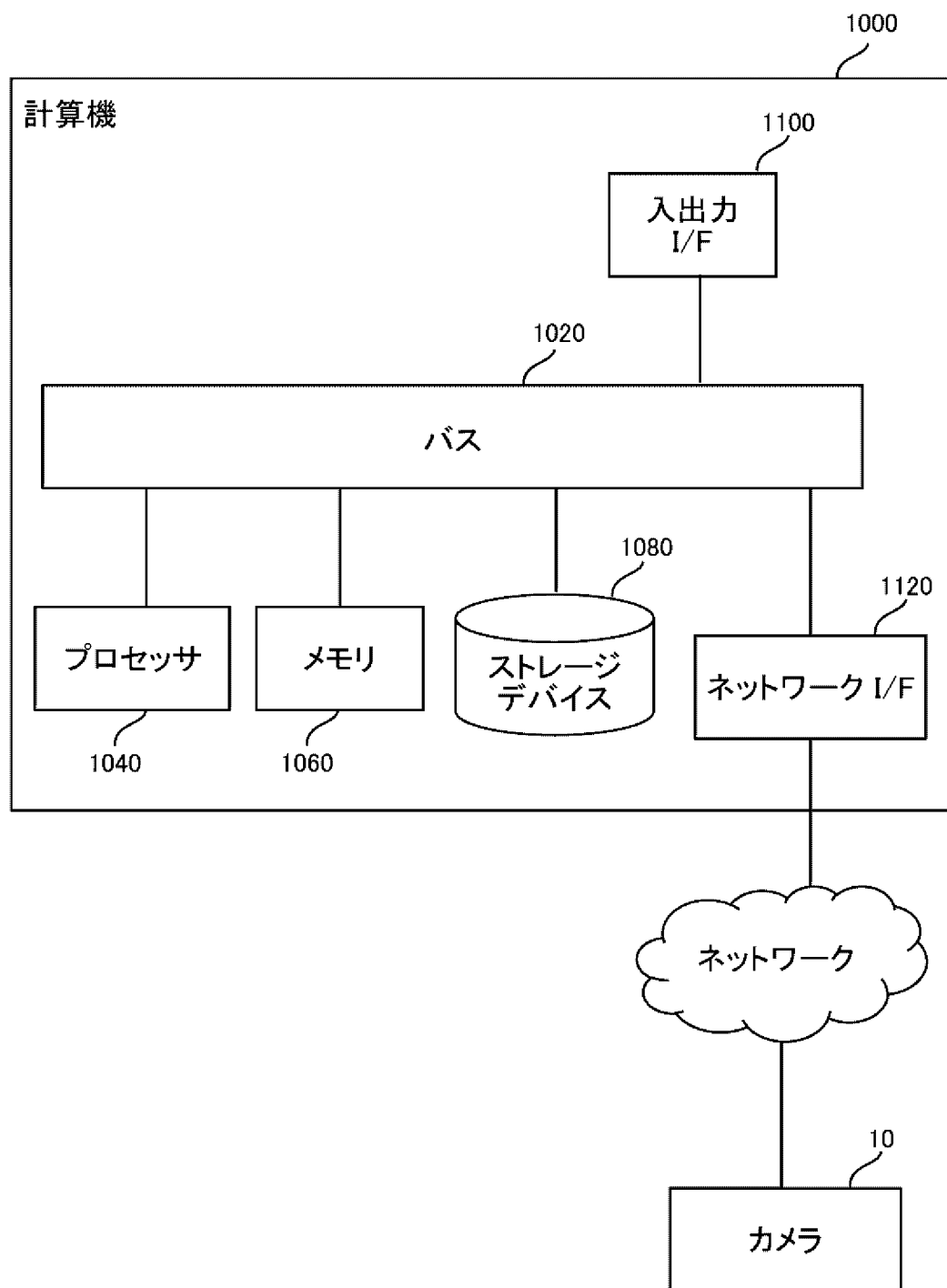
[図2]



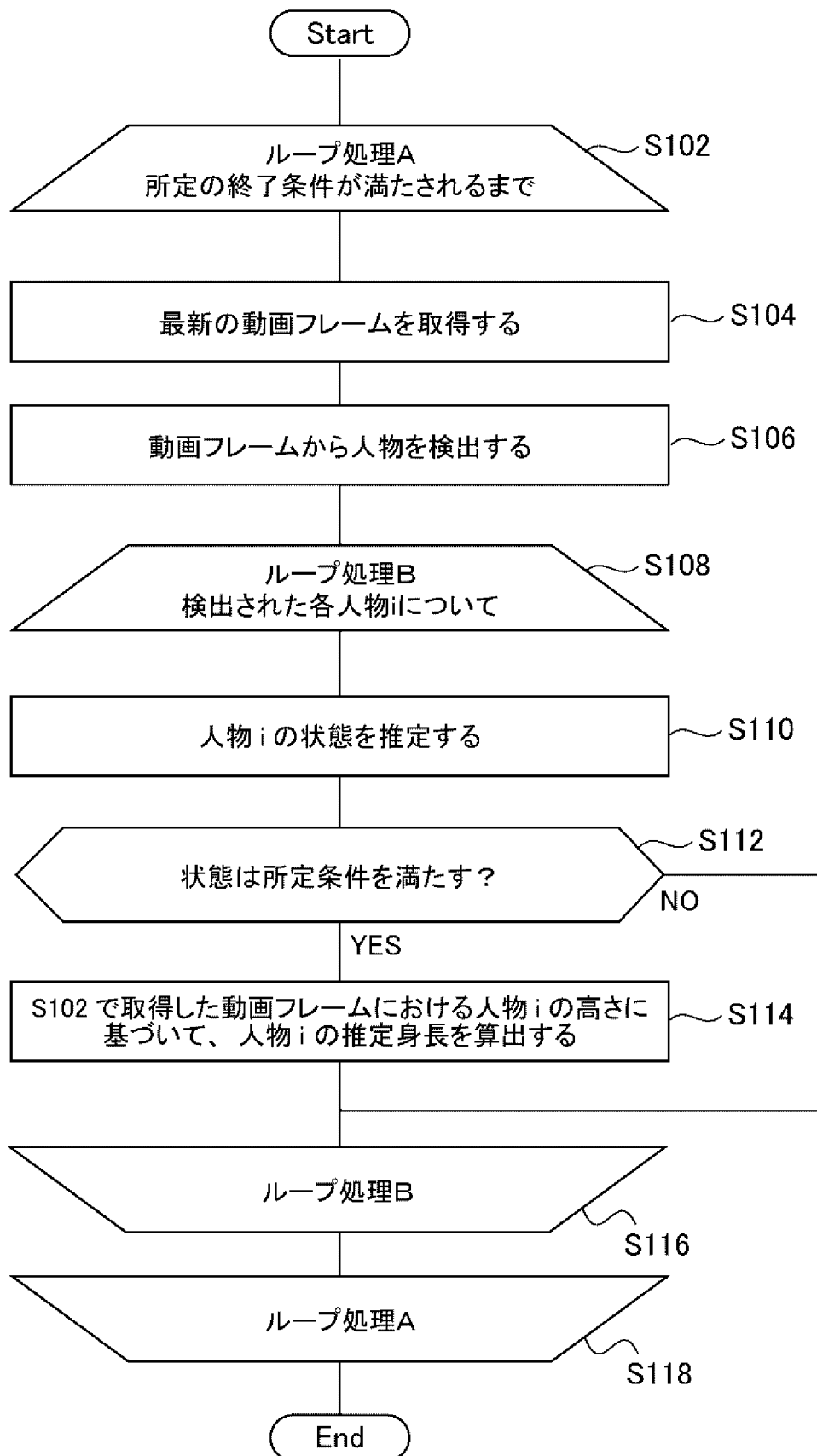
[図3]



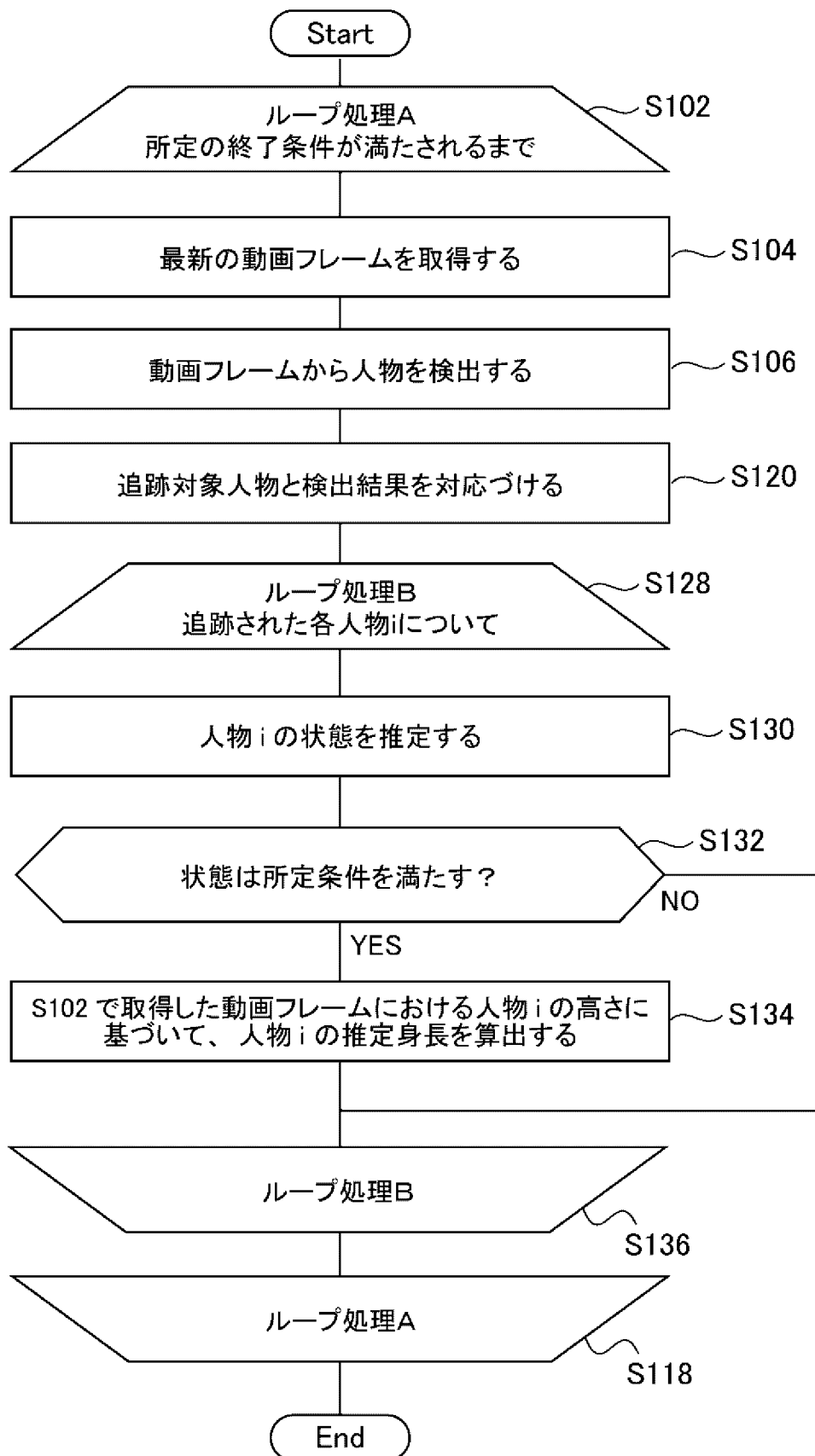
[図4]



[図5]



[図6]

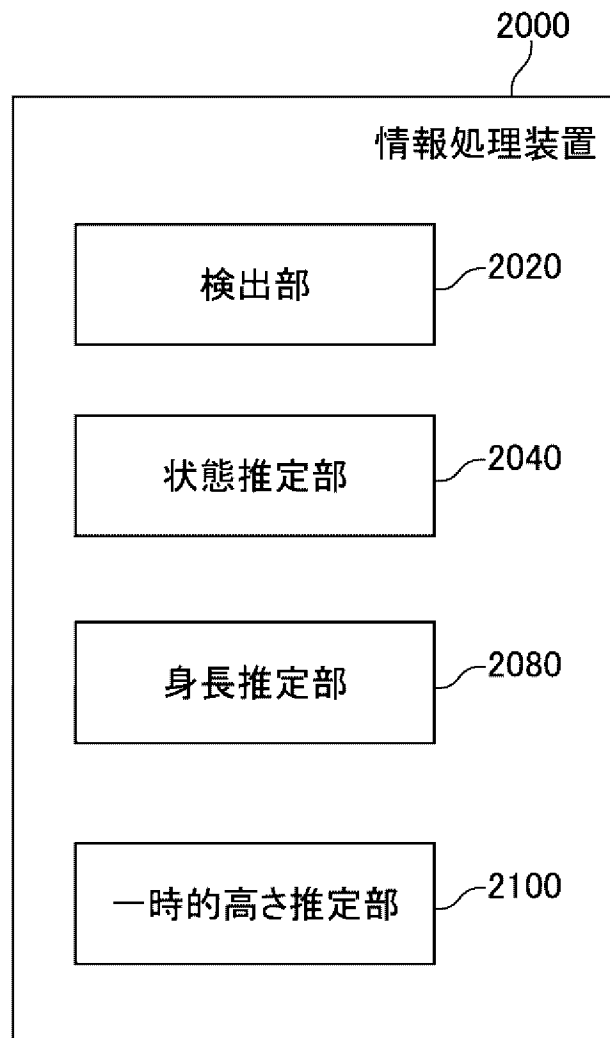


[図7]

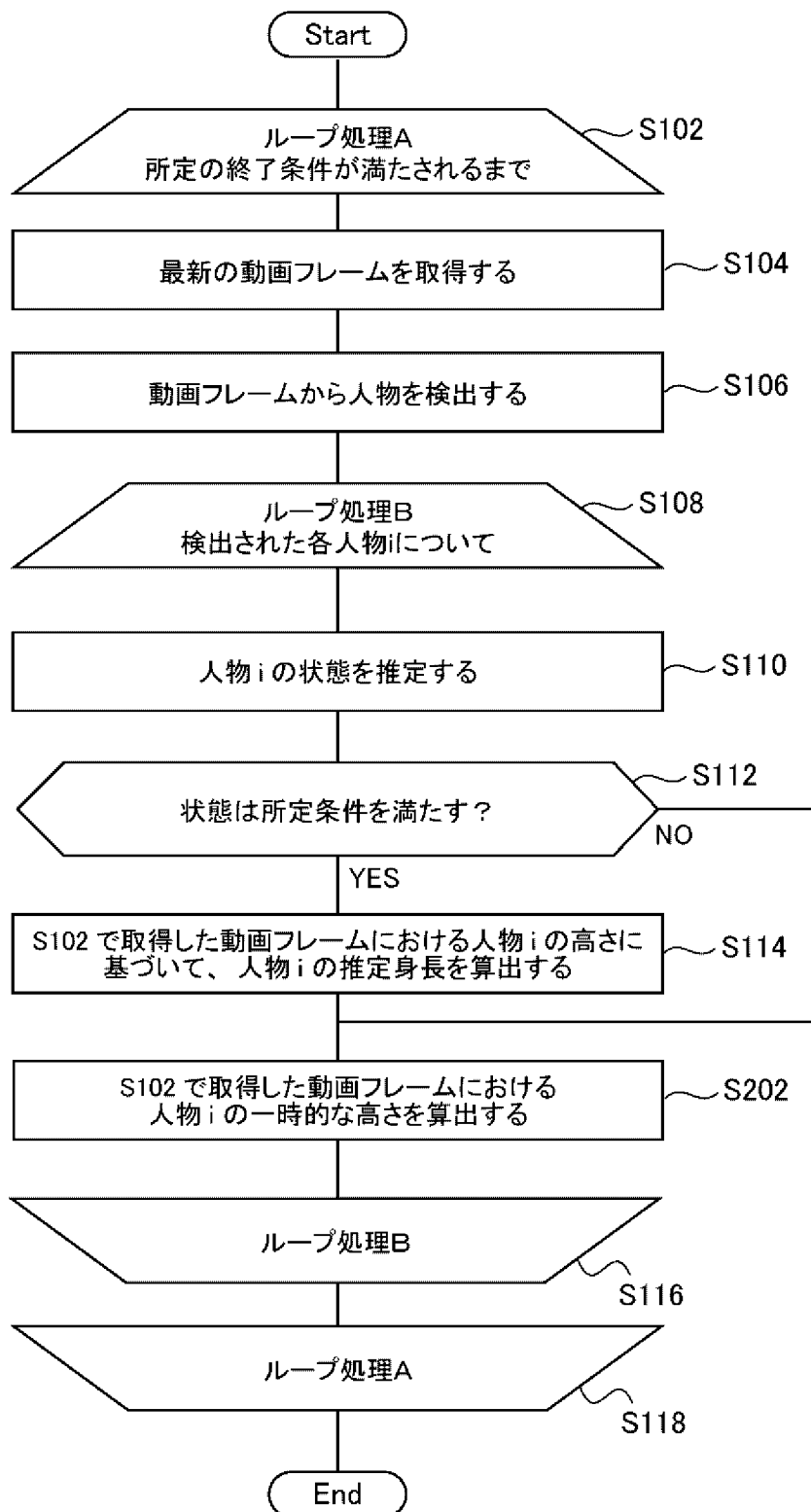
200

|                                                                                         |        |  |            |  |      |  |     |  |          |  |          |  |                                |  |
|-----------------------------------------------------------------------------------------|--------|--|------------|--|------|--|-----|--|----------|--|----------|--|--------------------------------|--|
| <div>追跡ID</div> <div>id0001</div> <div>id0002</div> <div>id0003</div> <div>id0004</div> | 202    |  | 204        |  | 206  |  | 208 |  | 210      |  | 212      |  | 214                            |  |
|                                                                                         | フレームID |  | 位置         |  | 状態   |  | 観測値 |  | 動き       |  | 特徴量      |  | 領域                             |  |
|                                                                                         | 1      |  | (x11, y11) |  | 直立静止 |  | d1  |  | -        |  | f11, f12 |  | TL=(x12, y12)<br>BR=(x13, y13) |  |
|                                                                                         | 2      |  | (x21, y21) |  | 直立静止 |  | d2  |  | -        |  | f21, f22 |  | TL=(x22, y22)<br>BR=(x23, y23) |  |
|                                                                                         | ...    |  | ...        |  | ...  |  | ... |  | ...      |  | ...      |  | ...                            |  |
|                                                                                         | n      |  | (x31, y31) |  | 直立移動 |  | d3  |  | p31, p32 |  | f31, f32 |  | TL=(x32, y32)<br>BR=(x33, y33) |  |
|                                                                                         | n+1    |  | (x41, y41) |  | 直立移動 |  | d4  |  | p41, p42 |  | f41, f42 |  | TL=(x42, y42)<br>BR=(x43, y43) |  |
|                                                                                         | ...    |  | ...        |  | ...  |  |     |  | ...      |  | ...      |  | ...                            |  |

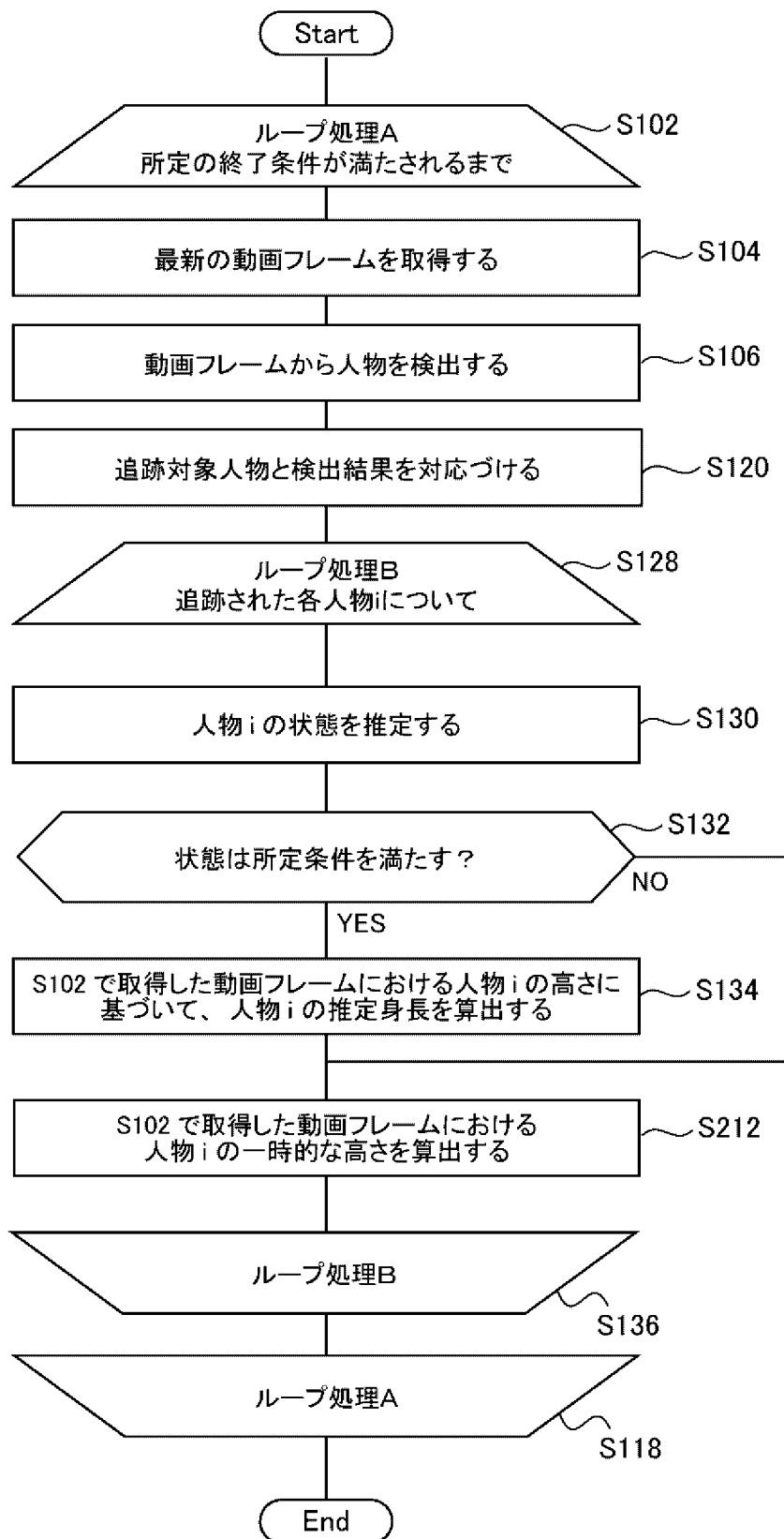
[図8]



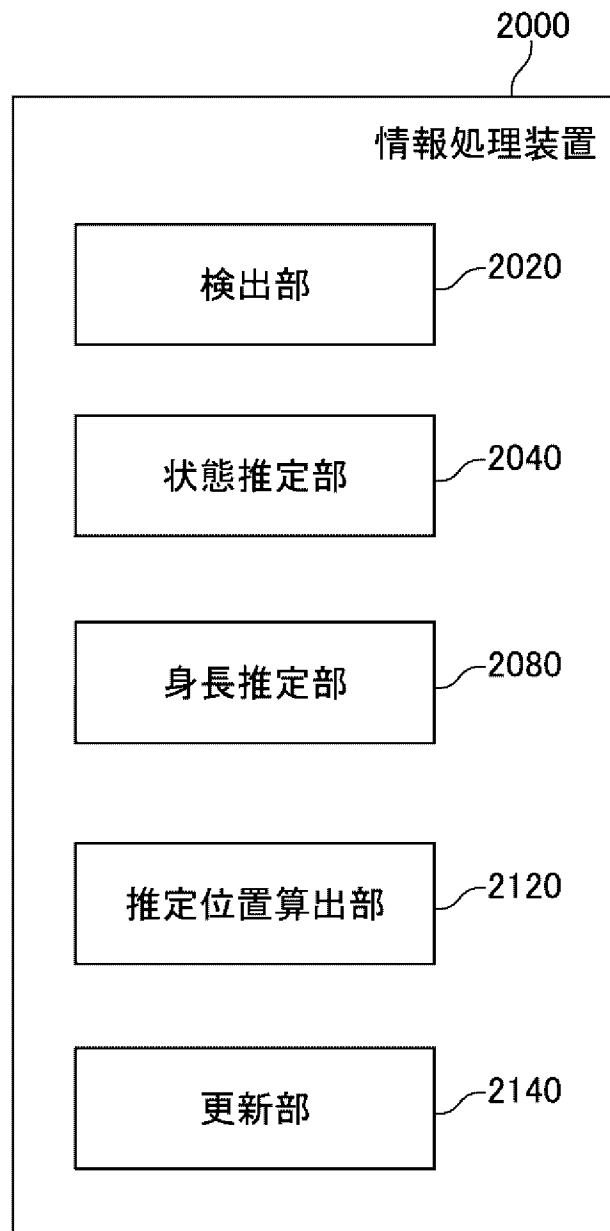
[図9]



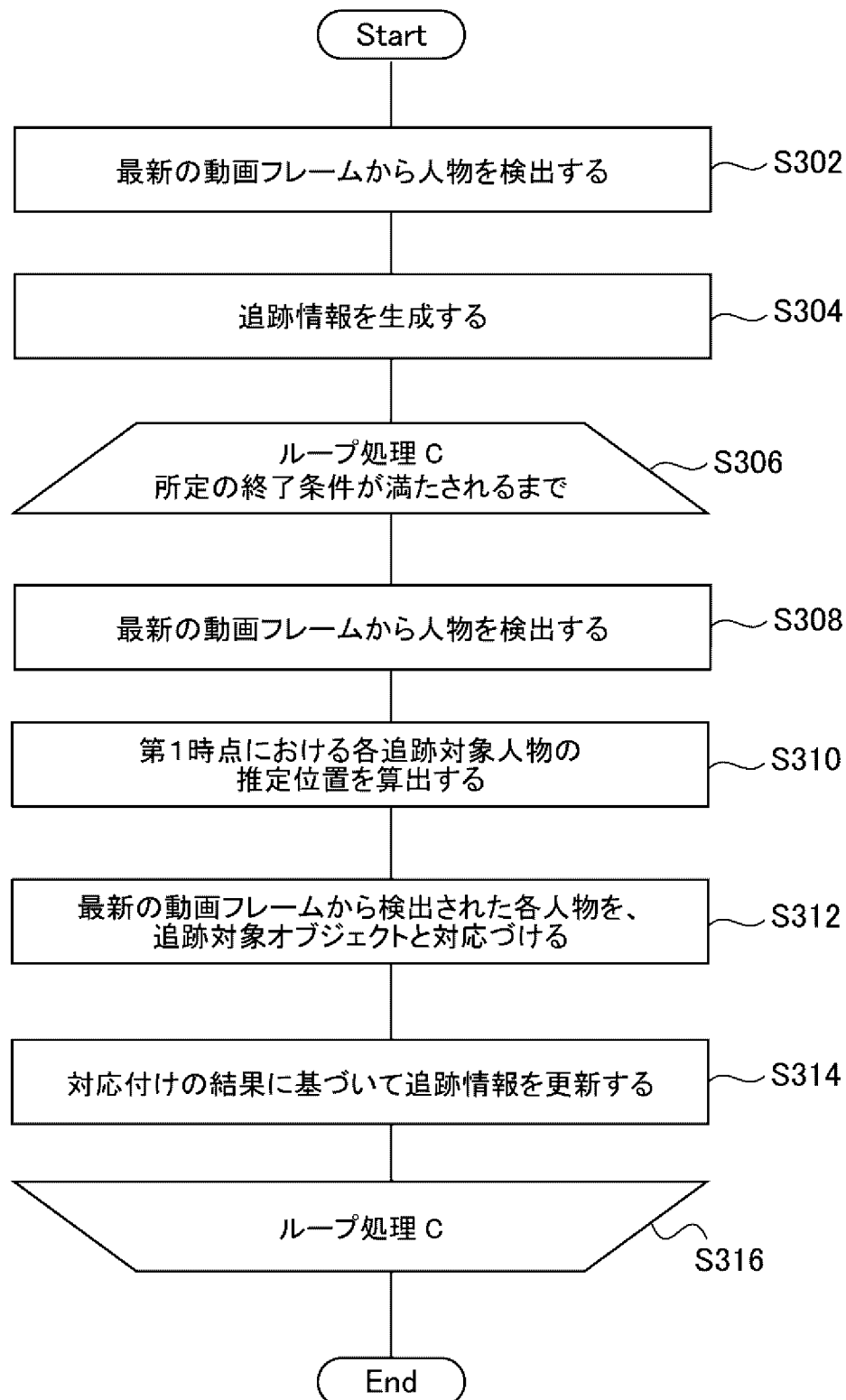
[図10]



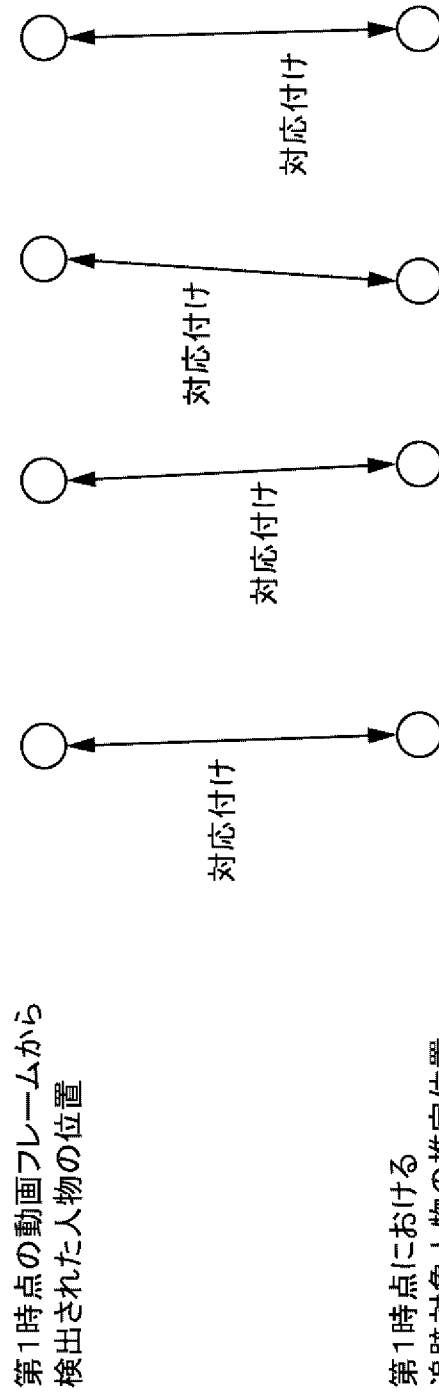
[図11]



[図12]



[図13]



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/034940

## A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06T7/00 (2017.01) i, G06T7/20 (2017.01) 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)

Int. Cl. G06T7/00, G06T7/20

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

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2017

Registered utility model specifications of Japan 1996-2017

Published registered utility model applications of Japan 1994-2017

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.      |
|-----------|--------------------------------------------------------------------------------------------------|----------------------------|
| X         | JP 2007-78354 A (TOA CORP.) 29 March 2007,<br>paragraphs [0039]-[0047], fig. 5, 6 (Family: none) | 1-3, 8-9, 13-15, 20-21, 25 |
| A         |                                                                                                  | 4-7, 10-12, 16-19, 22-24   |
| A         | JP 2013-37406 A (SOGO KEIBI HOSHO CO., LTD.) 21<br>February 2013, claim 1 (Family: none)         | 1-25                       |

☐ 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  
20.11.2017

Date of mailing of the international search report  
28.11.2017

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

Authorized officer

Telephone No.

## A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06T7/00(2017.01)i, G06T7/20(2017.01)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（IPC））

Int.Cl. G06T7/00, G06T7/20

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   | 1922-1996年 |
| 日本国公開実用新案公報 | 1971-2017年 |
| 日本国実用新案登録公報 | 1996-2017年 |
| 日本国登録実用新案公報 | 1994-2017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<br>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<br>請求項の番号             |
|-----------------|----------------------------------------------------------------------------------|----------------------------|
| X               | JP 2007-78354 A（ティーオーエー株式会社）2007.03.29, 段落 [0039] - [0047], [図5] - [図6]（ファミリーなし） | 1-3, 8-9, 13-15, 20-21, 25 |
| A               |                                                                                  | 4-7, 10-12, 16-19, 22-24   |

☒ 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

「&amp;」同一パテントファミリー文献

国際調査を完了した日

20.11.2017

国際調査報告の発送日

28.11.2017

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

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）

佐藤 実

5H

3247

電話番号 03-3581-1101 内線 3531

| C (続き) . 関連すると認められる文献 |                                                                 |                |
|-----------------------|-----------------------------------------------------------------|----------------|
| 引用文献の<br>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<br>請求項の番号 |
| A                     | JP 2013-37406 A (総合警備保障株式会社) 2013. 02. 21, [請求項 1]<br>(ファミリーなし) | 1 - 2 5        |