(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7228569号
(P7228569)

(45)発行日　令和5年2月24日(2023.2.24)

(24)登録日　令和5年2月15日(2023.2.15)

(51)国際特許分類

		FI		
G06T	*7/20 (2017.01)*	G06T	7/20	300Z
G06T	*7/00 (2017.01)*	G06T	7/00	350C

請求項の数　13　（全69頁）

(21)出願番号　特願2020-507669(P2020-507669)
(86)(22)出願日　平成30年7月26日(2018.7.26)
(65)公表番号　特表2020-530170(P2020-530170A)
(43)公表日　令和2年10月15日(2020.10.15)
(86)国際出願番号　PCT/US2018/043933
(87)国際公開番号　WO2019/032304
(87)国際公開日　平成31年2月14日(2019.2.14)
審査請求日　令和3年7月9日(2021.7.9)
(31)優先権主張番号　62/542,077
(32)優先日　平成29年8月7日(2017.8.7)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　15/847,796
(32)優先日　平成29年12月19日(2017.12.19)
最終頁に続く

(73)特許権者　518394684
スタンダード コグニション コーポレーション
アメリカ合衆国 カリフォルニア州 ９４１０３ サンフランシスコ，７番フロア ミッション ストリート ９６５
(74)代理人　100114476
弁理士　政木 良文
(72)発明者　フィッシャー， ジョーダン
アメリカ合衆国 カリフォルニア州 ９４１０７ サンフランシスコ， アパートメント ７１６ キング ストリート ２５０
(72)発明者　ヴァルドマン， デイヴィッド
アメリカ合衆国 カリフォルニア州 ９４１０３ サンフランシスコ， ナンバー１
最終頁に続く

(54)【発明の名称】　画像認識を用いた被写体識別及び追跡

(57)【特許請求の範囲】

【請求項１】

実空間のエリア内の多関節被写体を追跡するシステムであって、

複数のカメラ内のカメラが前記実空間内の対応する視野のそれぞれの画像シーケンスを生成し、前記複数のカメラにおいて各カメラの前記視野が少なくとも１つの他のカメラの前記視野と重なる、前記複数のカメラと、

前記複数のカメラと結合された処理システムと、を備えてなり、

前記処理システムが、

前記複数のカメラから前記画像シーケンスを受信し、画像を処理し、関節タイプ、特定の画像の時間、及び前記特定の画像内の要素の座標によって、前記特定の画像の要素を分類する関節データ構造の配列を、前記特定の画像毎に、生成する画像認識エンジン、

重なる視野を有するカメラからの画像シーケンス内の画像に対応する前記関節データ構造の配列を受信し、様々な画像シーケンス内の画像に対応する前記関節データ構造の配列内の前記要素の前記座標を、実空間内の座標を有する候補関節に変換するように構成された追跡エンジン、及び、

実空間内の座標を有する候補関節のセットを前記実空間内の多関節被写体として識別するロジック、を備え、

前記画像認識エンジンが、画像を処理して、

前記特定の画像の要素毎に信頼度配列をそれぞれ生成し、且つ、

前記信頼度配列に基づいて、前記特定の画像の特定の要素の前記関節データ構造の関節

タイプを選択し、
前記特定の画像の特定の要素についての信頼度配列が、複数の関節タイプの各関節タイプについての対応する信頼値を含み、関節タイプ毎に、前記特定の要素が前記各関節タイプである信頼度を示すものであることを特徴とするシステム。
【請求項２】
前記処理システムが、前記識別された候補関節のセットを処理して、前記画像シーケンス内の前記多関節被写体の手の画像を含む有界ボックスを指定するように構成された有界ボックス生成器を、更に備える請求項１に記載のシステム。
【請求項３】
多関節被写体として識別された前記候補関節のセットを記憶するロジックを備え、
候補関節のセットを識別する前記ロジックが、特定の時間に撮影された画像において識別された候補関節が先行する画像において多関節被写体として識別された前記候補関節のセットのうちの１つのメンバーに対応するかどうかを判定するロジックを含む請求項１または２に記載のシステム。
【請求項４】
実空間のエリア内の多関節被写体を追跡する方法であって、
各カメラの視野が少なくとも１つの他のカメラの視野と重なり合う複数のカメラを使用して、前記実空間内の対応する視野のそれぞれの画像シーケンスを生成すること、
関節タイプ、特定の画像の時間、及び前記特定の画像内の要素の座標によって、前記特定の画像の要素を分類する関節データ構造の配列を、前記特定の画像毎に、生成するために、前記画像シーケンス内の画像を処理すること、
様々な画像シーケンス内の画像に対応する前記関節データ構造の配列内の前記要素の前記座標を、前記実空間内の座標を有する候補関節に変換すること、及び、
実空間内の座標を有する候補関節のセットを前記実空間内の多関節被写体として識別すること、を備え、
前記画像を処理することが、
前記特定の画像の要素毎に信頼度配列をそれぞれ生成すること、及び、
前記信頼度配列に基づいて、前記特定の画像の特定の要素の前記関節データ構造の関節タイプを選択することを含み、
前記特定の画像の特定の要素についての信頼度配列が、複数の関節タイプの各関節タイプについての対応する信頼値を含み、関節タイプ毎に、前記特定の要素が前記各関節タイプである信頼度を示すものであることを特徴とする方法。
【請求項５】
前記画像を処理することが、畳み込みニューラル・ネットワークを使用することを含む請求項４に記載の方法。
【請求項６】
前記候補関節のセットを識別することが、前記識別された候補関節のセットを処理して、前記画像シーケンス内の前記多関節被写体の手の画像を含む有界ボックスを指定することを含む請求項４または５に記載の方法。
【請求項７】
多関節被写体として識別された前記候補関節のセットを記憶することを含み、
前記候補関節のセットを識別することが、特定の時間に撮影された画像において識別された候補関節が先行する画像において多関節被写体として識別された前記候補関節のセットのうちの１つのメンバーに対応するかどうかを判定することを含む請求項４～６のいずれか１項に記載の方法。
【請求項８】
前記画像シーケンスが同期されている請求項４～７のいずれか１項に記載の方法。
【請求項９】
前記複数のカメラが、前記実空間内のエリアのそれぞれの部分を包含する視野を有し、その上に配置されたカメラを備え、

多関節被写体として識別された候補関節のセットのメンバーの実空間内の座標は、前記多関節被写体の前記エリア内の位置を識別する請求項４～８のいずれか１項に記載の方法。
【請求項１０】
前記実空間のエリア内の複数の多関節被写体の位置を追跡することを含む請求項４～９のいずれか１項に記載の方法。
【請求項１１】
前記複数の多関節被写体内の多関節被写体が前記実空間のエリアを離れるときを特定することを含む請求項１０に記載の方法。
【請求項１２】
特定の多関節被写体として識別された候補関節のセットのメンバーである複数の候補関節の前記実空間のエリア内の座標を追跡することを含む請求項４～１１のいずれか１項に記載の方法。
【請求項１３】
非一時的なコンピュータ可読記憶媒体であって、
請求項４～１２のいずれか１項に係る実空間のエリア内の多関節被写体を追跡する方法のためのコンピュータ命令が格納されていることを特徴とする非一時的なコンピュータ可読記憶媒体。
【発明の詳細な説明】
【著作権通知】
【０００１】
本特許書類の開示の一部分は、著作権保護の対象となる資料を含んでいる。著作権所有者は、特許文献や特許開示を誰でも特許庁の特許ファイルや記録に記載されている通りにファクシミリで複写することに異議はないが、それ以外はあらゆる著作権を保有するものとする。
【技術分野】
【０００２】
本発明は、レジレス・チェックアウトに使用可能なシステム及びその構成要素に関する。
【背景技術】
【０００３】
画像処理における困難な問題は、大きな空間上に配置された複数のカメラからの画像が被写体の行為を識別し追跡するために使用される場合に生じる。
【０００４】
ショッピングストア内の人々のような実空間のエリア内の被写体の行為を追跡することは、多くの技術的課題を提示する。例えば、複数の顧客がショッピングストア内の棚と棚の間の通路及びオープンスペースを移動するショッピングストアに配備される当該画像処理システムを考える。顧客は棚から商品を取り、それらをそれぞれのショッピングカートまたはバスケットに置く。顧客は、商品を望まない場合には商品を棚に置くこともできる。
【０００５】
顧客がこれらの行為を実行している間、顧客の異なる部分、及び、棚の異なる部分、または店舗の在庫を保持する他の陳列構成は、他の顧客、棚、及び製品陳列などの存在のために、異なるカメラからの画像において、塞がれることになる。また、いつでも店舗内に多くの顧客がいる可能性があり、個人及びその行為を経時的に識別し追跡することが困難になる。
【０００６】
より効果的かつ自動的に、大きなスペース内の被写体の取る及び置くという行為を識別及び追跡し、レジレス・チェックアウトなどの機能を含む、被写体とその環境との複雑な交流をサポートする他の処理を実行することができるシステムを提供することが望ましい。
【発明の概要】
【０００７】
システム及びシステムを操作する方法は、画像処理を使用して、実空間のエリア内の人

物などの被写体による変化、及び被写体とその環境との他の複雑な交流を追跡するために提供される。画像処理による変化を追跡するこの機能は、処理されるべき画像データのタイプ、画像データの如何なる処理を実行すべきか、及び、如何にして画像データから高い信頼性で行為を決定するかに関連して、コンピュータ工学の複雑な問題を提示する。本明細書に記載のシステムは、実空間の頭上に配置されたカメラからの画像のみを使用してこれらの機能を実行することができ、その結果、所与の設定での展開のために、店舗の棚及びフロアスペースにセンサなどを改装する必要がない。

【０００８】

在庫陳列構造の上方に配置された複数のカメラを使用して、各カメラの視野が複数のカメラ内の少なくとも１つの他のカメラの視野と重なる、実空間内の対応する視野内に在庫陳列構造の画像のそれぞれのシーケンスを生成することを備えた在庫陳列構造を含む、実空間のエリア内の被写体による在庫商品の置くこと及び取ることを追跡するシステム及び方法が提供される。これらの画像シーケンスを使用して、在庫陳列構造上の在庫商品に関連する画像シーケンスにおける意味的に重要な変化を識別し、意味的に重要な変化を画像シーケンス内に表される被写体に関連付けることによって、在庫商品を置くこと及び取ることを検出するシステム及び方法が説明される。

【０００９】

実空間のエリア内の被写体による在庫商品を置くこと及び取ることを追跡するためのシステム及び方法が提供され、このシステム及び方法は在庫陳列構造の上方に配置された複数のカメラを使用して、実空間内の対応する視野内に在庫陳列構造の画像のそれぞれのシーケンスを生成することを含み、各カメラの視野は、複数のカメラ内の少なくとも１つの他のカメラの視野と重複する。これらの画像シーケンスを使用して、画像シーケンス内の前景データを処理することにより、被写体のジェスチャ及び該ジェスチャに関連する在庫商品を識別することによって、在庫商品を置くこと及び取ることを検出するシステム及び方法が説明される。

【００１０】

また、前景処理と背景処理とを同じ画像シーケンスで組み合わせるシステム及び方法が説明される。この組み合わされたアプローチでは、提供されるシステム及び方法が、画像シーケンス内の前景データを処理することにより、被写体のジェスチャ及び該ジェスチャに関連付けられた在庫商品を識別することによって、在庫商品を置くこと及び取ることを検出するためにこれらの画像のシーケンスを使用することと、画像シーケンス内の背景データを処理することによって、在庫陳列構造上の在庫商品に関連する画像シーケンス内の意味的に重要な変化を識別することによって、在庫商品を置くこと及び取ることを検出するためにこれらの画像シーケンスを使用することと、意味的に重要な変化を画像シーケンス内に表される被写体に関連付けることとを含む。

【００１１】

本明細書で説明される実施形態では、システムが複数のカメラを使用して、実空間内の対応する視野のそれぞれの画像シーケンスを生成する。各カメラの視野は、複数のカメラのうちの少なくとも１つの他のカメラの視野と重なる。このシステムは、複数のカメラから対応する画像シーケンスを受信する、被写体画像認識エンジンを含む第１の画像プロセッサを含む。第１の画像プロセッサは、画像を処理して、対応する画像シーケンス内の画像に表される被写体を識別する。システムは、複数のカメラから対応する画像シーケンスを受信する、背景画像認識エンジンを含む第２の画像プロセッサを更に含む。第２の画像プロセッサは、識別された被写体をマスクしてマスクされた画像を生成し、マスクされた画像を処理して、対応する画像シーケンス内の画像に表される背景変化を識別し且つ分類する。

【００１２】

一実施形態では、背景画像認識エンジンは、畳み込みニューラル・ネットワークを含む。システムは、識別された背景変化を識別された被写体に関連付けるロジックを含む。

【００１３】

一実施形態では、第２の画像プロセッサが、対応する画像シーケンスの背景画像を格納する背景画像格納装置を含む。第２の画像プロセッサは、識別された被写体を表す前景画像データを背景画像データで置き換えるために、画像シーケンス内の画像を処理するマスクロジックを更に含む。背景画像データは、マスクされた画像を提供するために、対応する画像シーケンスの背景画像から収集される。

【００１４】

一実施形態では、マスクロジックが画像シーケンス内のＮ個のマスクされた画像のセットを組み合わせて、各カメラのファクタ化画像のシーケンスを生成する。第２の画像プロセッサは、ファクタ化画像のシーケンスを処理することによって、背景変化を識別し且つ分類する。

【００１５】

一実施形態では、第２の画像プロセッサが、対応する画像シーケンスのための変化データ構造を生成するロジックを含む。変化データ構造は、識別された背景変化のマスクされた画像内の座標、識別された背景変化の在庫商品被写体の識別子、及び識別された背景変化の分類を含む。第２の画像プロセッサは更に、重なり合う視野を有するカメラのセットからの変化データ構造を処理して、実空間内での識別された背景変化の位置を見つけるための調整ロジックを含む。

【００１６】

一実施形態では、変化データ構造における識別された背景変化の分類が識別された在庫商品が背景画像に対して追加されたか除去されたかを示す。

【００１７】

別の実施形態では、変化データ構造における識別された背景変化の分類が識別された在庫商品が背景画像に対して追加されたか除去されたかを示す。システムは、背景変化を識別された被写体に関連付けるためのロジックを更に含む。最後に、システムは、識別された被写体による在庫商品を取ること、及び識別された被写体による在庫陳列構造上に在庫商品を置くことの検出を行うロジックを含む。

【００１８】

別の実施形態では、システムは、背景変化を識別された被写体に関連付けるロジックを含む。システムは、識別された被写体による在庫商品を取ること、及び識別された被写体による在庫陳列構造上に在庫商品を置くことの検出を行うロジックを更に含む。

【００１９】

システムは、複数のカメラから対応する画像シーケンスを受信する前景画像認識エンジンを含む、本明細書で説明する第３の画像プロセッサを含むことができる。第３の画像プロセッサは画像を処理して、対応する画像シーケンス内の画像に表される前景変化を識別し且つ分類する。

【００２０】

システム及びシステムを操作する方法が、実空間において、人などの多関節被写体を追跡するために提供される。このシステムは、複数のカメラを使用して、実空間における対応する視野のそれぞれの画像シーケンスを生成する。各カメラの視野は、複数のカメラのうちの少なくとも１つの他のカメラの視野と重なる。このシステムは、画像シーケンス内の画像を処理して、各画像に対応する関節データ構造の配列を生成する。特定の画像に対応する関節データ構造の配列は、関節タイプ、特定の画像の時間、及び特定の画像内の要素の座標によって、特定の画像の要素を分類する。次に、システムは、様々な画像シーケンスに対応する関節データ構造の配列内の要素の座標を、実空間内の座標を有する候補関節に変換する。最後に、システムは候補関節のコンステレーションを識別し、コンステレーションは、実空間内の座標を有する候補関節のそれぞれのセットを、実空間における多関節被写体として含む。

【００２１】

一実施形態では、画像認識エンジンが畳み込みニューラル・ネットワークを含む。画像認識エンジンによる画像の処理は、画像の要素に対する信頼度配列を生成することを含む

。画像の特定の要素についての信頼度配列は、特定の要素についての複数の関節タイプについての信頼値を含む。信頼度配列は、信頼度配列に基づいて、特定の要素の関節データ構造の関節タイプを選択するために使用される。

【００２２】

多関節被写体を追跡するためのシステムの一実施形態では、候補関節のセットを識別することは、候補関節のセットを多関節被写体として識別するために、実空間における被写体の関節間の物理的関係に基づいてヒューリスティック関数を適用することを含む。この処理は、多関節被写体として識別された関節のセットを記憶することを含む。候補関節のセットを識別することは、特定の時間に撮影された画像において識別された候補関節が先行する画像において多関節被写体として識別された候補関節のセットのうちの１つのメンバーに対応するかどうかを判定することを含む。

【００２３】

一実施形態では、複数のカメラによって取得された画像シーケンスのそれぞれの画像が、空間を通る被写体の移動の時間スケール上の単一の時点で実空間を表すように、画像シーケンスが同期される。

【００２４】

多関節被写体として識別された候補関節のセットのメンバーの実空間内の座標は、多関節被写体のエリア内の位置を識別する。いくつかの実施形態では、処理が、実空間のエリア内の複数の多関節被写体の位置の同時追跡を含む。いくつかの実施形態では、処理が、複数の多関節被写体内の多関節被写体が実空間のエリアを離れるときを特定することを含む。いくつかの実施形態では、処理が、多関節被写体が所与の時点で向いている方向を判定することを含む。本明細書で説明される実施形態では、システムが、複数のカメラを使用して、実空間内の対応する視野のそれぞれの画像シーケンスを生成する。各カメラの視野は、複数のカメラのうちの少なくとも１つの他のカメラの視野と重なる。システムは、複数のカメラから受け取った画像シーケンス内の画像を処理して、画像内に表された被写体を識別し、識別された被写体の分類を生成する。最後に、システムは、識別された被写体による在庫商品を取ることと、識別された被写体による棚に在庫商品を置くことを検出するために、画像シーケンス内の画像のセットに対する識別された被写体の分類を処理する。

【００２５】

一実施形態では、分類が、識別された被写体が在庫商品を保持しているかどうかを識別する。分類はまた、識別された被写体の手が棚の近くにあるかどうか、または識別された被写体の手が識別された被写体の近くにあるかどうかを識別する。手が識別された被写体の近くにあるかどうかの分類は、識別された被写体の手が識別された被写体に関連付けられたバスケットの近くにあり、識別された被写体の身体の近くにあるかどうかを含むことができる。

【００２６】

複数の画像内の被写体の手の分類を時系列で生成するために、視野内の被写体の手を表す画像を処理できる技術が記載されている。画像シーケンスからの手の分類は、いくつかの実施形態では、被写体による行為を識別するために、畳み込みニューラル・ネットワークを使用して処理することができる。行為は、本明細書に記載された実施形態に記載されているように、在庫商品を置くこと及び取ること、或いは、手の画像を処理することによって解読可能な他のタイプの行為であり得る。

【００２７】

画像を処理して視野内の被写体を識別し、被写体の関節の位置を見つける技術が記載されている。被写体の関節の位置は、被写体の手を含む対応する画像内の有界ボックスを識別するために、本明細書で説明するように処理することができる。有界ボックス内のデータは、対応する画像内の被写体の手の処理された分類とすることができる。画像シーケンスからこのようにして生成された識別された被写体からの手の分類は、被写体による行為を識別するために処理することができる。

【００２８】
前景と背景の画像認識エンジンのような複数の画像認識エンジンを含むシステムにおいて、該システムは、識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの第１の検出セット、及び、識別された被写体による在庫商品を取ることと識別された被写体による在庫陳列構造上に在庫商品を置くことの第２の検出セットを、作成できる。第１及び第２の検出セットを処理するための選択ロジックを使用して、ログ・データ構造を生成することができる。ログ・データ構造は、識別された被写体に関する在庫商品のリストを含む。
【００２９】
本明細書で説明する実施形態では、複数のカメラにおいて、カメラからの画像シーケンスが同期される。１つの好ましい実施態様では、同じカメラ及び同じ画像シーケンスが前景及び背景イメージプロセッサの両方によって使用される。その結果、同じ入力データを用いて、在庫商品を置くこと及び取ることの冗長な検出が行われ、結果として得られるデータにおいて高い信頼性と高い精度を可能にする。
【００３０】
本明細書で説明される１つの技術では、システムが、画像シーケンスで表されるジェスチャに関連付けられた被写体のジェスチャ及び在庫商品を識別することによって、在庫商品を置くこと及び取ることを検出するロジックを備える。これは、本明細書に記載されるように、被写体画像認識エンジンと協調して前景画像認識エンジンを使用して行うことができる。
【００３１】
本明細書で説明される別の技術では、システムは、棚のような在庫陳列構造上の在庫商品の意味的に重要な変化を経時的に識別し、意味的に重要な変化を画像シーケンスで表される被写体に関連付けることによって、在庫商品を置くこと及び取ることを検出するロジックを備える。これは、本明細書で説明するように、背景画像認識エンジンを被写体画像認識エンジンと協調させて使用して行うことができる。
【００３２】
本明細書で説明するテクノロジを適用するシステムでは、ジェスチャ分析と意味的差異分析の両方を組み合わせて、カメラの配列からの同期画像の同じシーケンスに対して実行することができる。
【００３３】
コンピュータ・システムによって実行することができる方法及びコンピュータ・プログラム製品も、本明細書に記載されている。
【００３４】
本発明の他の実施態様及び利点は、以下の図面、詳細な説明、及び特許請求の範囲を検討することによって理解することができる。
【図面の簡単な説明】
【００３５】
【図１】追跡エンジンが画像認識エンジンによって生成された関節データを使用して被写体を追跡するシステムのアーキテクチャレベル概略図を示す。
【００３６】
【図２】カメラ配置を示すショッピングストアの通路の側面図である。
【００３７】
【図３】カメラ配置を示すショッピングストアにおける図２の通路の上面図である。
【００３８】
【図４】図１の画像認識エンジンをホストするように構成されたカメラ及びコンピュータ・ハードウェア構成である。
【００３９】
【図５】図１の画像認識エンジンにおける関節の識別を示す畳み込みニューラル・ネットワークを示す。

【００４０】
【図６】関節情報を記憶するための例示的なデータ構造を示す。
【００４１】
【図７】グローバル・メトリック計算器を有する図１の追跡エンジンを示す。
【００４２】
【図８】関連する関節の情報を含む被写体を記憶するための例示的なデータ構造を示す。
【００４３】
【図９】図１のシステムによって被写体を追跡するための処理ステップを示すフローチャートである。
【００４４】
【図１０】図９のカメラ較正ステップのより詳細な処理ステップを示すフローチャートである。
【００４５】
【図１１】図９のビデオ処理ステップのより詳細な処理ステップを示すフローチャートである。
【００４６】
【図１２Ａ】図９のシーン処理のためのより詳細な処理ステップの第１の部分を示すフローチャートである。
【００４７】
【図１２Ｂ】図９のシーン処理のためのより詳細な処理ステップの第２の部分を示すフローチャートである。
【００４８】
【図１３】図１のシステムの実施形態が使用される環境の図である。
【００４９】
【図１４】図１のシステムの一実施形態におけるビデオ処理及びシーン処理の図である。
【００５０】
【図１５Ａ】実空間において被写体毎にショッピングカート・データ構造を生成するための関節ＣＮＮ、ＷｈａｔＣＮＮ、及びＷｈｅｎＣＮＮを含む複数の畳み込みニューラル・ネットワーク（ＣＮＮ）を有するパイプラインを示す概略図である。
【００５１】
【図１５Ｂ】複数のカメラからの複数の画像チャネルと、被写体及びそれらのそれぞれのショッピングカート・データ構造のための調整ロジックとを示す。
【００５２】
【図１６】実空間内の被写体を識別して更新する処理ステップを示すフローチャートである。
【００５３】
【図１７】在庫商品を識別するために被写体の手関節を処理するための処理ステップを示すフローチャートである。
【００５４】
【図１８】被写体毎のショッピングカート・データ構造を作成するための、手関節毎の在庫商品の時系列分析のための処理ステップを示すフローチャートである。
【００５５】
【図１９】図１５Ａのシステムの実施形態におけるＷｈａｔＣＮＮモデルの図である。
【００５６】
【図２０】図１５Ａのシステムの一実施形態におけるＷｈｅｎＣＮＮモデルの図である。
【００５７】
【図２１】畳み込み層の次元を識別するＷｈａｔＣＮＮモデルの例示的なアーキテクチャを示す。
【００５８】
【図２２】手画像の分類のためのＷｈａｔＣＮＮモデルの実施形態の高レベルブロック図

を示す。
【００５９】
【図２３】図２２に示されるＷｈａｔＣＮＮモデルの高レベルブロック図の第１のブロックの詳細を示す。
【００６０】
【図２４】図２２に提示された例示的ＷｈａｔＣＮＮモデルにおける全結合層における演算子を提示する。
【００６１】
【図２５】ＷｈａｔＣＮＮモデルのためのトレーニング・データセットの一部として記憶される画像ファイルの例示的なファイル名である。
【００６２】
【図２６】背景意味的差分抽出を使用する第１の検出と、前景領域提案を使用する冗長検出との間で選択ロジックが選択する、実空間のエリア内の被写体による変化を追跡するためのシステムの高レベルアーキテクチャである。
【００６３】
【図２７】図２６のシステムを実施するサブシステムの構成要素を示す。
【００６４】
【図２８Ａ】在庫イベントを決定し、ショッピングカート・データ構造を生成するための詳細な処理ステップの第１の部分を示すフローチャートである。
【００６５】
【図２８Ｂ】在庫イベントを決定し、ショッピングカート・データ構造を生成するための詳細な処理ステップの第２の部分を示すフローチャートである。
【発明を実施するための形態】
【００６６】
以下の説明は、当業者が本発明を作成し使用することを可能にするために提示され、特定の用途及びその要件に即して提供される。開示された実施態様に対する様々な修正は、当業者には容易に明らかであり、本明細書で定義される一般原則は、本発明の精神及び範囲から逸脱することなく、他の実施態様及び用途に適用され得る。従って、本発明は、示された実施態様に限定されることを意図するものではなく、本明細書に開示された原理及び特徴と一致する最も広い範囲が与えられるべきである。
［システム概要］
【００６７】
図１～図２８Ａ／２８Ｂを参照して、対象技術のシステム及び様々な実施態様を説明する。システム及び処理は、本実施態様によるシステムのアーキテクチャレベル概略図である図１を参照して説明される。図１は、アーキテクチャ図であるため、説明の明確性を向上させるために、特定の詳細は省略されている。
【００６８】
図１の説明は、以下のように編成される。最初に、システムの要素を説明し、次にそれらの相互接続を説明する。次に、システムにおける要素の使用についてより詳細に説明する。
【００６９】
図１は、システム１００のブロック図レベルの説明図を提供する。本システム１００は、カメラ１１４、ネットワーク・ノードがホスティングする画像認識エンジン１１２ａ、１１２ｂ及び１１２ｎ、ネットワーク上の１つまたは複数のネットワーク・ノードに配置される追跡エンジン１１０、較正器１２０、被写体データベース１４０、トレーニング・データベース１５０、関節ヒューリスティックス用、置く及び取るヒューリスティックス用、及び、後述する複数の画像認識エンジンの出力を調整し、結合するための他のヒューリスティックス用のヒューリスティックス・データベース１６０、較正データベース１７０、及び、１または複数の通信ネットワーク１８１を含む。ネットワーク・ノードは、１つの画像認識エンジンのみ、または本明細書で説明されるように、複数の画像認識エンジ

ンをホストすることができる。システムはまた、在庫データベース及び他のサポートデータを含むことができる。

【００７０】

本明細書で使用されるように、ネットワーク・ノードは、ネットワークに接続され、通信チャネルを介して他のネットワーク・ノードとの間で情報を送信、受信、または転送することができる、アドレス可能なハードウェア・デバイスまたは仮想デバイスである。ハードウェア・ネットワーク・ノードとして配置することができる電子デバイスの例には、あらゆる種類のコンピュータ、ワークステーション、ラップトップ・コンピュータ、ハンドヘルド・コンピュータ、及びスマートフォンが含まれる。ネットワーク・ノードは、クラウドベースのサーバ・システムで実施することができる。ネットワーク・ノードとして構成された複数の仮想デバイスを、単一の物理デバイスを使用して実施することができる。

【００７１】

明確性のために、画像認識エンジンをホストする３つのネットワーク・ノードのみがシステム１００に示されている。しかしながら、画像認識エンジンをホストする任意の数のネットワーク・ノードを、ネットワーク１８１を介して追跡エンジン１１０に接続することができる。また、本明細書で説明する画像認識エンジン、追跡エンジン、及び他の処理エンジンは、分散アーキテクチャ内の複数のネットワーク・ノードを使用して実行することができる。

【００７２】

次に、システム１００の要素の相互接続について説明する。ネットワーク１８１は、画像認識エンジン１１２ａ、１１２ｂ、及び１１２ｎをそれぞれホストするネットワーク・ノード１０１ａ、１０１ｂ、及び１０１ｃ、追跡エンジン１１０をホストするネットワーク・ノード１０２、較正器１２０、被写体データベース１４０、トレーニング・データベース１５０、関節ヒューリスティックス・データベース１６０、及び較正データベース１７０を結合する。カメラ１１４は、画像認識エンジン１１２ａ、１１２ｂ、及び１１２ｎをホストするネットワーク・ノードを介して追跡エンジン１１０に接続される。一実施形態では、カメラ１１４がショッピングストア（スーパーマーケットなど）に設置され、重なり合う視野を有するカメラ１１４のセット（２つ以上）が各通路の上に配置されて、店舗内の実空間の画像を取得する。図１では、２つのカメラが通路１１６ａの上に配置され、２つのカメラが通路１１６ｂの上に配置され、３つのカメラが通路１１６ｎの上に配置されている。カメラ１１４は、重なり合う視野を有する通路上に設置される。斯かる実施形態では、カメラは、ショッピングストアの通路内を移動する顧客がいつの時点でも２つ以上のカメラの視野内に存在することを目標として構成される。

【００７３】

カメラ１１４は互いに時間的に同期させることができ、その結果、画像は、同時にまたは時間的に近く、かつ同じ画像キャプチャレートで取得される。カメラ１１４は、画像認識エンジン１１２ａ～１１２ｎをホストするネットワーク・ノードに、所定のレートでそれぞれの継続的な画像ストリームを送ることができる。同時にまたは時間的に近くに、実空間のエリアをカバーする全てのカメラにおいて取得された画像は、同期された画像が実空間において固定された位置を有する被写体の異なる光景を表すものとして処理エンジンにおいて識別され得るという意味で、同期している。例えば、一実施形態では、カメラが、３０フレーム／秒（ｆｐｓ）のレートで、画像認識エンジン１１２ａ～１１２ｎをホストするそれぞれのネットワーク・ノードに画像フレームを送信する。各フレームは、画像データと共に、タイムスタンプ、カメラの識別情報（「カメラＩＤ」と略される）、及びフレーム識別情報（「フレームＩＤ」と略される）を有する。

【００７４】

通路上に設置されたカメラは、それぞれの画像認識エンジンに接続される。例えば、図１において、通路１１６ａ上に設置された２つのカメラは、画像認識エンジン１１２ａをホストするネットワーク・ノード１０１ａに接続される。同様に、通路１１６ｂ上に設置された２つのカメラは、画像認識エンジン１１２ｂをホストするネットワーク・ノード１

０１ｂに接続される。ネットワーク・ノード１０１ａ～１０１ｎ内でホストされる各画像認識エンジン１１２ａ～１１２ｎは、図示の例ではそれぞれ１つのカメラから受信した画像フレームを別々に処理する。

【００７５】

一実施形態では、各画像認識エンジン１１２ａ、１１２ｂ、及び１１２ｎは、畳み込みニューラル・ネットワーク（ＣＮＮと略す）などの深層学習アルゴリズムとして実装される。斯かる実施形態では、ＣＮＮがトレーニング・データベース１５０を使用してトレーニングされる。本明細書で説明される実施形態では、実空間内の被写体の画像認識が、画像内で認識可能な関節を識別しグループ化することに基づいており、関節のグループは個々の被写体に帰属することができる。この関節ベースの分析のために、トレーニング・データベース１５０は、被写体のための異なるタイプの関節の各々に対して膨大な画像を収集している。ショッピングストアの例示的な実施形態では、被写体は、棚の間の通路を移動する顧客である。例示的な実施形態では、ＣＮＮのトレーニング中に、システム１００は「トレーニング・システム」と呼ばれる。トレーニング・データベース１５０を使用してＣＮＮをトレーニングした後、ＣＮＮは、プロダクション・モードに切り替えられ、ショッピングストア内の顧客の画像をリアルタイムで処理する。例示的な実施形態では、プロダクション中に、システム１００はランタイム・システムと呼ばれる（推論システムとも呼ばれる）。それぞれの画像認識装置のＣＮＮは、それぞれの画像ストリーム中の画像に対して関節データ構造の配列を生成する。本明細書に記載される実施形態では、関節データ構造の配列が、各処理された画像に対して生成されることで、各画像認識エンジン１１２ａ～１１２ｎが、関節データ構造の配列の出力ストリームを生成する。重なり合う視野を有するカメラからの関節データ構造のこれらの配列は、関節のグループを形成し、斯かる関節のグループを被写体として識別するために、更に処理される。

【００７６】

カメラ１１４は、ＣＮＮをプロダクション・モードに切り替える前に較正される。キャリブレータ１２０はカメラを較正し、較正データを較正データベース１７０に格納する。

【００７７】

追跡エンジン１１０は、ネットワーク・ノード１０２上でホストされ、画像認識エンジン１１２ａ～１１２ｎから被写体の関節データ構造の配列の継続的なストリームを受信する。追跡エンジン１１０は、関節データ構造の配列を処理し、様々なシーケンスの画像に対応する関節データ構造の配列内の要素の座標を、実空間内の座標を有する候補関節に変換する。同期画像の各セットについて、実空間全体にわたって識別された候補関節の組み合わせは、類推目的のために、候補関節の銀河に似ていると考えることができる。後続の各時点において、銀河が経時的に変化するように、候補関節の動きが記録される。追跡エンジン１１０の出力は、被写体データベース１４０に格納される。

【００７８】

追跡エンジン１１０は、実空間内の座標を有する候補関節のグループまたはセットを、実空間内の被写体として識別するロジックを使用する。類推目的のために、候補点の各セットは、各時点における候補関節の星座（コンステレーション）に似ている。候補関節のコンステレーションは、時間とともに移動することができる。

【００７９】

候補関節のセットを識別するロジックは、実空間における被写体の関節間の物理的関係に基づくヒューリスティック関数を含む。これらのヒューリスティック関数は、候補関節のセットを被写体として識別するために使用される。ヒューリスティック関数はヒューリスティックス・データベース１６０に格納される。追跡エンジン１１０の出力は、被写体データベース１４０に格納される。従って、候補関節のセットは、他の個々の候補関節とヒューリスティックス・パラメータに従った関係を有する個々の候補関節、及び、個々の被写体として識別された、または識別することができる所与のセット内の候補関節のサブセットを含む。

【００８０】

ネットワーク１８１を通る実際の通信経路は、公衆ネットワーク及び／またはプライベート・ネットワーク上のポイント・ツー・ポイントとすることができる。通信は、プライベート・ネットワーク、ＶＰＮ、ＭＰＬＳ回路、またはインターネットなどの様々なネットワーク１８１を介して行うことができ、適切なアプリケーション・プログラミング・インターフェース（ＡＰＩ）及びデータ交換フォーマット、例えば、ＲＥＳＴ（Representational State Transfer）、ＪＳＯＮ（JavaScript（商標）Object Notation）、ＸＭＬ（Extensible Markup Language）、ＳＯＡＰ（Simple Object Access Protocol）、ＪＭＳ（Java（商標）Message Service）、及び／またはＪａｖａプラットフォーム・モジュール・システムなどを使用することができる。すべての通信は、暗号化することができる。通信は、一般に、ＥＤＧＥ、３Ｇ、４Ｇ LTE、Ｗｉ－Ｆｉ、及びＷｉＭＡＸなどのプロトコルを介して、ＬＡＮ(ローカル・エリア・ネットワーク）、ＷＡＮ(ワイド・エリア・ネットワーク）、電話ネットワーク（公衆交換電話網（ＰＳＴＮ））、セッション開始プロトコル（ＳＩＰ）、無線ネットワーク、ポイント・ツー・ポイント・ネットワーク、星型ネットワーク，トークンリング型ネットワーク，ハブ型ネットワーク、インターネット（モバイルインターネットを含む）などのネットワーク上で行われる。更に、ユーザ名／パスワード、オープン許可（ＯＡｕｔｈ）、Ｋｅｒｂｅｒｏｓ、ＳｅｃｕｒｅＩＤ、デジタル証明書などの様々な承認及び認証技術を使用して、通信を保護することができる。

【００８１】

本明細書に開示される技術は、データベースシステム、マルチテナント環境、または、Ｏｒａｃｌｅ（商標）と互換性のあるデータベース実施態様、ＩＢＭ ＤＢ２ Ｅｎｔｅｒｐｒｉｓｅ Ｓｅｒｖｅｒ（商標）と互換性のあるリレーショナル・データベース実施態様、ＭｙＳＱＬ（商標）またはＰｏｓｔｇｒｅＳＱＬ（商標）と互換性のあるリレーショナル・データベース実施態様またはＭｉｃｒｏｓｏｆｔ ＳＱＬ Ｓｅｒｖｅｒ（商標）と互換性のあるリレーショナル・データベース実施態様等のリレーショナル・データベース実施態様、または、Ｖａｍｐｉｒｅ（商標）と互換性のある非リレーショナル・データベース実施態様、Ａｐａｃｈｅ Ｃａｓｓａｎｄｒａ（商標）と互換性のある非リレーショナル・データベース実施態様、 ＢｉｇＴａｂｌｅ（商標）と互換性のある非リレーショナル・データベース実施態様、またはＨＢａｓｅ（商標）またはＤｙｎａｍｏＤＢ（商標）と互換性のある非リレーショナル・データベース実施態様、等のＮｏＳＱＬ（商標）の非リレーショナル・データベース実施態様を含む何かのコンピュータ実装システムという状況下で実施され得る。更に、開示された技術は、ＭａｐＲｅｄｕｃｅ（商標）、バルク同期プログラミング、ＭＰＩプリミティブ等の様々なプログラミングモデル、または、Ａｐａｃｈｅ Ｓｔｏｒｍ（商標）、Ａｐａｃｈｅ Ｓｐａｒｋ（商標）、Ａｐａｃｈｅ Ｋａｆｋａ（商標）、Ａｐａｃｈｅ Ｆｌｉｎｋ（商標）、Ｔｒｕｖｉｓｏ（商標）、Ａｍａｚｏｎ Ｅｌａｓｔｉｃｓｅａｒｃｈ Ｓｅｒｖｉｃｅ（商標）、Ａｍａｚｏｎ Ｗｅｂ Ｓｅｒｖｉｃｅｓ（ＡＷＳ）（商標）、ＩＢＭ Ｉｎｆｏ‐Ｓｐｈｅｒｅ（商標）、Ｂｏｒｅａｌｉｓ（商標）、及びＹａｈｏｏ！ Ｓ４（商標）等の様々なスケーラブルなバッチ及びストリーム管理システムを使用して実施され得る。

［カメラ配置］

【００８２】

カメラ１１４は、３次元（３Ｄと略される）実空間において多関節存在物（または被写体）を追跡するように配置される。ショッピングストアの例示的な実施形態では、実空間は、販売用の商品が棚に積み重ねられるショッピングストアのエリアを含むことができる。実空間内の点は、（ｘ，ｙ，ｚ）座標系で表すことができる。システムが適用される実空間のエリア内の各点は、２つ以上のカメラ１１４の視野によってカバーされる。

【００８３】

ショッピングストアでは、棚及び他の在庫陳列構造は、ショッピングストアの側壁に沿って、または通路を形成する列に、または２つの構成の組合せでなど、様々な方法で配置することができる。図２は、通路１１６ａの一端から見た、通路１１６ａを形成する棚の

配置を示す。２つのカメラ、カメラＡ ２０６及びカメラＢ ２０８は、棚のような在庫陳列構造の上のショッピングストアの天井２３０及びフロア２２０から所定の距離で通路１１６ａの上に配置される。カメラ１１４は、実空間内の在庫陳列構造及びフロアエリアのそれぞれの部分を包含する視野を有し、その上に配置されたカメラを備える。被写体として識別された候補関節のセットのメンバーの実空間内の座標は、被写体のフロアエリア内の位置を識別する。ショッピングストアの例示的な実施形態では、実空間は、在庫にアクセスできるショッピングストア内のフロア２２０のすべてを含むことができる。カメラ１１４は、フロア２２０及び棚のエリアが少なくとも２つのカメラによって見えるように配置され、配向される。カメラ１１４はまた、棚２０２及び２０４の少なくとも一部と、棚２０２及び２０４の前のフロアスペースとを覆う。カメラの角度は急峻な視点、真っ直ぐな視点及び角度の付いた視点の両方を有するように選択され、これにより、顧客のより完全な身体画像が得られる。一実施形態では、カメラ１１４が、ショッピングストア全体を通して、８フィート高さ以上で構成される。図１３に、斯かる実施形態の説明図を示す。

【００８４】

図２では、カメラ２０６及び２０８が重なり合う視野を有し、それぞれ重なり合う視野２１６及び２１８で棚Ａ ２０２と棚Ｂ ２０４との間の空間をカバーする実空間内の位置は、実空間座標系の（ｘ，ｙ，ｚ）点として表される。「ｘ」及び「ｙ」は、ショッピングストアのフロア２２０とすることができる２次元（２Ｄ）平面上の位置を表し、値「ｚ」は、１つの構成ではフロア２２０における２Ｄプレーン上の点の高さである。

【００８５】

図３は、図２の上から見た通路１１６ａを示し、通路１１６ａ上のカメラ２０６及び２０８の位置の例示的な配置を更に示す。カメラ２０６及び２０８は、通路１１６ａの両端の近くに配置される。カメラＡ ２０６は棚Ａ ２０２から所定の距離に配置され、カメラＢ ２０８は棚Ｂ ２０４から所定の距離に配置される。３つ以上のカメラが通路上に配置される別の実施形態では、カメラは互いに等しい距離に配置される。このような実施形態では、２つのカメラが両端の近くに配置され、第３のカメラが通路の中央に配置される。多数の異なるカメラ配置が可能であることが理解される。

［カメラ較正］

【００８６】

カメラ較正器１２０は２つのタイプの較正、即ち、内部及び外部較正を実行する。内部較正では、カメラ１１４の内部パラメータが較正される。内部カメラパラメータの例には、焦点距離、主点、スキュー、魚眼係数などが含まれる。内部カメラ較正のための様々な手法を使用することができる。斯かる手法の１つは、Ｚｈａｎｇによって、ＩＥＥＥ Ｔｒａｎｓａｃｔｉｏｎｓ ｏｎ Ｐａｔｔｅｒｎ Ａｎａｌｙｓｉｓ ａｎｄ Ｍａｃｈｉｎｅ Ｉｎｔｅｌｌｉｇｅｎｃｅ、Ｖｏｌｕｍｅ ２２、Ｎｏ．１１、Ｎｏｖｅｍｂｅｒ ２０００に発行された「Ａ ｆｌｅｘｉｂｌｅ ｎｅｗ ｔｅｃｈｎｉｑｕｅ ｆｏｒ ｃａｍｅｒａ ｃａｌｉｂｒａｔｉｏｎ」に示されている。

【００８７】

外部較正では、外部カメラパラメータが、２Ｄ画像データを実空間の３Ｄ座標に変換するためのマッピング・パラメータを生成するために較正される。一実施形態では、人物などの１つの被写体が実空間に導入される。被写体は、各カメラ１１４の視野を通過する経路上で実空間を移動する。実空間内の任意の所与の点において、被写体は、３Ｄシーンを形成する少なくとも２つのカメラの視野内に存在する。しかしながら、２つのカメラは、それぞれの２次元（２Ｄ）画像平面において同じ３Ｄシーンの異なるビューを有する。被写体の左手首などの３Ｄシーン内の特徴は、それぞれの２Ｄ画像平面内の異なる位置にある２つのカメラによって見られる。

【００８８】

点対応は、所与のシーンについて重複する視野を有する全てのカメラ・ペアの間で確立される。各カメラは同じ３Ｄシーンの異なる視野を有するので、点対応は３Ｄシーンにおける同じ点の投影を表す２つのピクセル位置（重なり合う視野を有する各カメラからの１

つの位置）である。外部較正のために、画像認識エンジン１１２ａ～１１２ｎの結果を使用して、各３Ｄシーンについて多くの点対応が識別される。画像認識エンジンは関節の位置を、それぞれのカメラ１１４の２Ｄ画像平面内のピクセルの（ｘ，ｙ）座標、例えば、行及び列番号として識別する。一実施形態では、関節は、被写体の１９の異なるタイプの関節のうちの１つである。被写体が異なるカメラの視野を通って移動するとき、追跡エンジン１１０は、較正に使用される被写体の１９の異なるタイプの関節の各（ｘ，ｙ）座標を、画像毎にカメラ１１４から受け取る。

【００８９】

例えば、カメラＡからの画像と、カメラＢからの画像との両方が同じ時点に、重なり合う視野で撮影された場合を考える。カメラＡからの画像には、カメラＢからの同期画像のピクセルに対応するピクセルがあり、カメラＡとカメラＢの両方の視野内の或る物体または表面の特定の点があり、その点が両方の画像フレームのピクセルに取り込まれていると考える。外部カメラ較正では、多数のそのような点が識別され、対応点と呼ばれる。較正中にカメラＡ及びカメラＢの視野内に１つの被写体があるので、この被写体の主要な関節、例えば左手首の中心が識別される。これらの主要な関節がカメラＡ及びカメラＢの両方からの画像フレーム内に見える場合、これらは対応点を表すと仮定される。この処理は、多くの画像フレームについて繰り返され、重なり合う視野を有する全てのカメラ・ペアについて対応点の大きな集合を構築する。一実施形態では、画像が３０ ＦＰＳ(フレーム／秒）以上のレートで、フルＲＧＢ(赤、緑、及び青）カラーで７２０ピクセルの解像度で、すべてのカメラからストリーミングされる。これらの画像は、一次元配列（フラット配列とも呼ばれる）の形態である。

【００９０】

被写体について上記で収集された多数の画像を使用して、重なり合う視野を有するカメラ間の対応点を決定することができる。重なり合う視野を有する２つのカメラＡ及びＢを考える。カメラＡ、Ｂのカメラ中心と３Ｄシーンの関節位置（特徴点ともいう）を通る平面を「エピポーラ平面」と呼び、エピポーラ平面とカメラＡ、Ｂの２Ｄ画像平面との交差箇所を「エピポーラ線」と定義する。これらの対応点が与えられると、カメラＡからの対応点を、カメラＢの画像フレーム内の対応点と交差することが保証されるカメラＢの視野内のエピポーラ線に正確にマッピングすることができる変換が決定される。被写体について上記で収集された画像フレームを使用して、変換が生成される。この変換は非線形であることが当技術分野で知られている。更に、一般形態では、投影された空間へ及び投影された空間から移動する非線形座標変換と同様に、それぞれのカメラのレンズの半径方向の歪み補正が必要であることが知られている。外部カメラ較正では、理想的な非線形変換への近似が非線形最適化問題を解くことによって決定される。この非線形最適化機能は、重なり合う視野を有するカメラ１１４の画像を処理する様々な画像認識エンジン１１２ａ～１１２ｎの出力（関節データ構造の配列）内の同じ関節を識別するために、追跡エンジン１１０によって使用される。内部カメラ較正及び外部カメラ較正の結果は、較正データベース１７０に格納される。

【００９１】

実空間におけるカメラ１１４の画像内の点の相対位置を決定するための様々な手法を使用することができる。例えば、Ｌｏｎｇｕｅｔ－Ｈｉｇｇｉｎｓが、「Ａ ｃｏｍｐｕｔｅｒ ａｌｇｏｒｉｔｈｍ ｆｏｒ ｒｅｃｏｎｓｔｒｕｃｔｉｎｇ ａ ｓｃｅｎｅ ｆｒｏｍ ｔｗｏ ｐｒｏｊｅｃｔｉｏｎｓ」（Ｎａｔｕｒｅ、第２９３巻、１９８１年９月１０日）を公表している。本論文では、２つの投影間の空間的関係が未知であるとき、遠近投影の相関ペアからシーンの３次元構造を計算することが提示されている。Ｌｏｎｇｕｅｔ－Ｈｉｇｇｉｎｓの論文は、実空間での各カメラの他のカメラに対する位置を決定する手法を提示する。更に、その手法は、実空間における被写体の三角測量を可能にし、重なり合う視野を有するカメラ１１４からの画像を使用してｚ座標の値（フロアからの高さ）を識別する。実空間の任意の点、例えば、実空間の一角の棚の端を、実空間の（ｘ，ｙ，ｚ）座標系上の（０，０，０）点とする。

【００９２】
本技術の一実施形態では、外部較正のパラメータが２つのデータ構造に格納される。第１のデータ構造は、固有パラメータを格納する。固有パラメータは、３Ｄ座標から２Ｄ画像座標への射影変換を表す。第１のデータ構造は以下に示すように、カメラ毎の固有パラメータを含む。データ値はすべて浮動小数点数値である。このデータ構造は、「Ｋ」及び歪み係数として表される３×３固有行列を格納する。歪み係数は、６つの半径方向歪み係数と２つの接線方向歪み係数とを含む。半径方向の歪みは、光線がその光学的中心よりも、レンズの縁部の近傍でより大きく屈曲するときに生じる。接線方向の歪みは、レンズと像平面が平行でないときに生じる。以下のデータ構造は、第１のカメラのみの値を示す。同様のデータが全てのカメラ１１４に対して記憶される。

```
{
        1: {
                K: [[x, x, x], [x, x, x], [x, x, x]],
                distortion _coefficients: [x, x, x, x, x, x, x, x]
        },
        ……
}
```

【００９３】
第２のデータ構造はカメラ・ペア毎に、３×３基本行列（Ｆ）、３×３必須行列（Ｅ）、３×４投影行列（Ｐ）、３×３回転行列（Ｒ）、及び３×１平行移動ベクトル（ｔ）を記憶する。このデータは、１つのカメラの基準フレーム内の点を別のカメラの基準フレームに変換するために使用される。カメラの各ペアについて、１つのカメラから別のカメラへフロア２２０の平面をマッピングするために、８つのホモグラフィ係数も記憶される。基本行列は、同じシーンの２つの画像間の関係であり、シーンからの点の投影が両方の画像において起こり得る場所を制約する。必須行列は、カメラが較正されている状態での、同じシーンの２つの画像間の関係でもある。投影行列は、３Ｄ実空間から部分空間へのベクトル空間投影を与える。回転行列は、ユークリッド空間における回転を実行するために使用される。平行移動ベクトル「ｔ」は、図形または空間のすべての点を所与の方向に同じ距離だけ移動させる幾何学的変形を表す。ホモグラフィ・フロア係数は、重なり合う視野を有するカメラによって見られるフロア２２０上の被写体の特徴の画像を結合するために使用される。第２のデータ構造を以下に示す。同様のデータが、全てのカメラ・ペアについて記憶される。前述のように、ｘは浮動小数点数値を表す。

```
{
        1: {
                2: {
                        F: [[x, x, x], [x, x, x], [x, x, x]],
                        E: [[x, x, x], [x, x, x], [x, x, x]],
                        P: [[x, x, x, x], [x, x, x, x], [x, x, x, x]],
                        R: [[x, x, x], [x, x, x], [x, x, x]],
                        t: [x, x, x],
                        homography_floor_coefficients: [x, x, x, x, x, x, x, x]
                }
        },
        …….
}
```

［ネットワーク構成］
【００９４】
図４は、画像認識エンジンをホストするネットワークのアーキテクチャ４００を示す。システムは、図示する実施形態では、複数のネットワーク・ノード１０１ａ～１０１ｎを含む。該実施形態では、ネットワーク・ノードは、処理プラットフォームとも呼ばれる。

処理プラットフォーム１０１ａ～１０１ｎ及びカメラ４１２、４１４、４１６、４１８は、ネットワーク４８１に接続される。

【００９５】

図４は、ネットワークに接続された複数のカメラ４１２、４１４、４１６、４１８を示す。多数のカメラを特定のシステムに配備することができる。一実施形態では、カメラ４１２～４１８が、それぞれイーサネット（登録商標）ベースのコネクタ４２２、４２４、４２６、及び４２８を使用してネットワーク４８１に接続される。該実施形態では、イーサネット（登録商標）ベースのコネクタがギガビットイーサネット（登録商標）とも呼ばれる１ギガビット／秒のデータ転送速度を有する。他の実施形態では、カメラ１１４が、ギガビットイーサネット（登録商標）よりも高速または低速のデータ転送速度を有することができる他のタイプのネットワーク接続を使用してネットワークに接続されると理解される。また、代替の実施形態では、１組のカメラを各処理プラットフォームに直接接続することができ、処理プラットフォームをネットワークに結合することができる。

【００９６】

記憶サブシステム４３０は、本発明の特定の実施形態の機能を提供する基本的なプログラミング及びデータ構成を記憶する。例えば、複数の画像認識エンジンの機能を実施する様々なモジュールを記憶サブシステム４３０に格納することができる。記憶サブシステム４３０は、非一時的なデータ記憶媒体を備えるコンピュータ可読メモリの一例であり、コンピュータによって実行可能なメモリに記憶されたコンピュータ命令を有し、本明細書で説明されるデータ処理機能及び画像処理機能のすべてまたは任意の組合せを実行し、これには、実空間の変化を識別し、被写体を追跡し、本明細書で説明されるような処理によって実空間のエリア内において在庫商品を置くこと及び取ることを検出するためのロジックが含まれる。他の例では、コンピュータ命令は、１つまたは複数のコンピュータ可読非一時的データ記憶媒体を備えるポータブルメモリを含む他のタイプのメモリに格納することができる。

【００９７】

これらのソフトウェアモジュールは、一般に、プロセッサ・サブシステム４５０によって実行される。ホスト・メモリ・サブシステム４３２は、通常、プログラム実行中に命令及びデータを記憶するためのメイン・ランダム・アクセス・メモリ（ＲＡＭ）４３４と、固定命令が記憶される読取り専用メモリ（ＲＯＭ）４３６とを含むいくつかのメモリを含む。一実施形態では、ＲＡＭ ４３４がプラットフォーム１０１ａに接続されたカメラ１１４からのビデオストリームを格納するためのバッファとして使用される。

【００９８】

ファイル記憶サブシステム４４０は、プログラム・ファイル及びデータ・ファイルのための永続的記憶を提供する。例示的な一実施形態では、記憶サブシステム４４０が番号４４２で識別されるＲＡＩＤ０（独立ディスクの冗長配列）構成内に４つの１２０ギガバイト（ＧＢ）ソリッド・ステート・ディスク（ＳＳＤ）を有する。ＣＮＮが被写体の関節を識別するために使用される例示的な実施形態では、ＲＡＩＤ０ ４４２が訓練データを記憶するために使用される。訓練中、ＲＡＭ ４３４にないトレーニング・データはＲＡＩＤ０ ４４２から読み出される。同様に、画像がトレーニングのために記録されているとき、ＲＡＭ ４３４にないデータはＲＡＩＤ０ ４４２に記憶される。例示的な実施形態では、ハードディスク・ドライブ（ＨＤＤ）４４６が１０テラバイトのストレージである。これは、ＲＡＩＤ０ ４４２ストレージよりもアクセス速度が遅い。ソリッド・ステート・ディスク（ＳＳＤ）４４４は、画像認識エンジン１１２ａのためのオペレーティング・システム及び関連ファイルを格納する。

【００９９】

例示的な構成では、３つのカメラ４１２、４１４、及び４１６が処理プラットフォーム１０１ａに接続される。各カメラは、カメラによって送られた画像を処理するために、専用グラフィックス処理ユニットＧＰＵ１ ４６２、ＧＰＵ２ ４６４、及びＧＰＵ３ ４６６を有する。１つの処理プラットフォームにつき、３つより少ないまたは多いカメラを接続

することできると理解される。従って、各カメラがカメラから受信した画像フレームを処理するための専用ＧＰＵを有するように、より少ないまたはより多いＧＰＵがネットワーク・ノード内に構成される。プロセッサ・サブシステム４５０、記憶サブシステム４３０、及びＧＰＵ４６２、４６４、４６６は、バス・サブシステム４５４を使用して通信する。

【０１００】

ネットワーク・インターフェース・サブシステム、ユーザ・インターフェース出力デバイス、及びユーザ・インターフェース入力デバイスなどのいくつかの周辺デバイスも、処理プラットフォーム１０１ａの一部を形成するバス・サブシステム４５４に接続される。これらのサブシステム及びデバイスは説明の明確性を改善するために、図４には意図的に示されていない。バス・サブシステム４５４は単一のバスとして概略的に示されているが、バス・サブシステムの代わりの実施形態では複数のバスを使用することができる。

【０１０１】

一実施形態では、カメラ４１２が、１２８８×９６４の解像度、３０ＦＰＳのフレームレート、及び１．３メガピクセル／イメージで、３００mm～無限大の作動距離を有する可変焦点レンズ、９８．２°～２３．８°の１／３インチセンサによる視野を有するＣｈａｍｅｌｅｏｎ３ １．３ ＭＰ Ｃｏｌｏｒ ＵＳＢ３ Ｖｉｓｉｏｎ(Ｓｏｎｙ ＩＣＸ４４５）を使用して実装することができる。

［畳み込みニューラル・ネットワーク］

【０１０２】

処理プラットフォーム内の画像認識エンジンは、所定のレートで継続的な画像ストリームを受信する。一実施形態では、画像認識エンジンが畳み込みニューラル・ネットワーク（ＣＮＮと略す）を含む。

【０１０３】

図５は、符号５００で示されるＣＮＮによる画像フレームの処理を示す。入力画像５１０は、行列状に配置された画像ピクセルからなる行列である。一実施形態では、入力画像５１０が１２８０ピクセルの幅、７２０ピクセルの高さ、及びＲＧＢとも呼ばれる３チャネルの赤、青、及び緑を有する。チャネルは、互いに積み重ねられた３つの１２８０×７２０の２次元画像として想像することができる。従って、入力画像は図５に示すように、１２８０×７２０×３の寸法を有する。

【０１０４】

２×２フィルタ５２０は、入力画像５１０と畳み込まれる。この実施形態では、フィルタが入力と畳み込まれるとき、パディングは適用されない。これに続いて、非線形関数が畳み込み画像に適用される。本実施形態では、正規化線形ユニット（ＲｅＬＵ）活性化を用いる。非線形関数の他の例には、シグモイド、双曲正接（ｔａｎｈ）、及びリーキーＲｅＬＵなどのＲｅＬＵの変形が含まれる。探索は、ハイパー・パラメータ値を見つけるために実行される。ハイパー・パラメータは、C_1、C_2、・・・、C_Nであり、C_Nは、畳み込み層「Ｎ」に対するチャネル数を意味する。Ｎ及びＣの典型的な値を図５に示す。Ｎ＝２５で表されるＣＮＮには２５層がある。Ｃの値は、層１～２５の各畳み込み層におけるチャネルの数である。他の実施形態では、残留接続、スクイズ励起モジュール、及び複数の解像度などの追加の特徴がＣＮＮ５００に追加される。

【０１０５】

画像分類に使用される典型的なＣＮＮでは、画像が畳み込み層を介して処理されるにつれて、画像のサイズ（幅及び高さ）が低減される。これは、入力画像のクラスを予測することを目的とするので、特徴識別に役立つ。しかし、図示の実施形態では、画像フレーム内の関節（特徴とも呼ばれる）を識別するだけでなく、実空間内の座標にマッピングできるように画像内のその位置を識別することも目標とするので、入力画像のサイズ（すなわち、画像の幅及び高さ）は縮小されない。従って、図５に示すように、この例では、ＣＮＮの畳み込み層を介して処理が進行することにつれて、画像の幅及び高さの寸法は変化しないままである。

【０１０６】

一実施形態では、ＣＮＮ５００が画像の各要素における被写体の１９個の可能な関節のうちの１つを識別する。可能な関節は、足関節と非足関節の２つのカテゴリに分類することができる。関節分類の１９番目のタイプは、被写体の全ての非関節特徴（すなわち、関節として分類されない画像の要素）に対するものである。

足関節：

足首関節（左右）

非足関節：

首

鼻

眼（左右）

耳（左右）

肩（左右）

肘（左右）

手首（左右）

尻（左右）

膝（左右）

非関節

【０１０７】

以上のように、本説明の目的のための「関節」は、実空間における被写体の追跡可能な特徴である。関節は、被写体の生理学的関節、または眼もしくは鼻などの他の特徴に対応し得る。

【０１０８】

入力画像のストリーム上の第１の分析セットは、実空間内の被写体の追跡可能な特徴を識別する。一実施形態では、これは「関節分析」と呼ばれる。このような実施形態では、関節分析に使用されるＣＮＮは「関節ＣＮＮ」と呼ばれる。一実施形態では、関節分析は、対応するカメラから受信される毎秒３０フレームにわたって毎秒３０回実行される。分析は時間的に同期され、すなわち、実空間における全ての被写体の関節を識別するために、１／３０秒で、全てのカメラ１１４からの画像が、対応する関節ＣＮＮにおいて分析される。複数のカメラからの一時点での画像の分析の結果は、「スナップショット」として記憶される。

【０１０９】

スナップショットは、システムによってカバーされる実空間のエリア内の候補関節のコンステレーションを表す、ある時点の全てのカメラ１１４の画像からの関節データ構造の配列を含む辞書形式であり得る。一実施形態では、スナップショットは被写体データベース１４０に格納される。

【０１１０】

このＣＮＮの例では、ソフトマックス関数が畳み込み層５３０の最終層内の画像のすべての要素に適用される。ソフトマックス関数は、任意の実数値のＫ次元ベクトルを、合計で１になる範囲［０，１］の実数値のＫ次元ベクトルに変換する。一実施形態では、画像の要素は単一のピクセルである。ソフトマックス関数は、ピクセル毎の任意の実数値の１９次元配列（１９次元ベクトルとも呼ばれる）を、合計で１になる［０，１］の実数値の１９次元信頼度配列に変換する。画像フレーム内のピクセルの１９次元は、被写体の１９タイプの関節に更に対応するＣＮＮの最終層内の１９個のチャネルに対応する。

【０１１１】

多数の画素は、その画像に対するソースカメラの視野内の被写体の数に応じて、１つの画像内の１９タイプの関節の各々の１つとして分類することができる。

【０１１２】

画像認識エンジン１１２ａ～１１２ｎは、画像を処理して、画像の要素に対する信頼度配列を生成する。画像の特定の要素についての信頼度配列は、その特定の要素についての複数の関節タイプについての信頼値を含む。画像認識エンジン１１２ａ～１１２ｎの各々

は、それぞれ、画像毎に信頼度配列の出力行列５４０を生成する。最後に、各画像認識エンジンは、画像当たりの信頼度配列の各出力行列５４０に対応する関節データ構造の配列を生成する。特定の画像に対応する関節データ構造の配列は、関節タイプ、特定の画像の時間、及び特定の画像内の要素の座標によって、特定の画像の要素を分類する。信頼度配列の値に基づいて、各イメージ内の特定の要素の関節データ構造の関節タイプが選択される。

【０１１３】

被写体の各関節は、ヒートマップとして出力行列５４０に分布していると考えることができる。ヒートマップは、各関節タイプについて最高値（ピーク）を有する画像要素を示すように分解することができる。理想的には、特定の関節タイプの高い値を有する所与の画素について、所与の画素からの範囲外の周囲の画素はその関節タイプについてより低い値を有し、その結果、その関節タイプを有する特定の関節の位置を画像空間座標において識別することができる。それに対応して、その画像要素に対する信頼度配列はその関節に対して最も高い信頼値を有し、残りの１８種類の関節に対してより低い信頼値を有する。

【０１１４】

一実施形態では、各カメラ１１４からの画像のバッチがそれぞれの画像認識エンジンによって処理される。例えば、６つの連続的にタイムスタンプされた画像は、キャッシュ・コヒーレンスを有効に利用するためにバッチで連続的に処理される。ＣＮＮ５００の１つの層に対するパラメータは、メモリにロードされ、６つの画像フレームのバッチに適用される。次に、次の層のパラメータがメモリにロードされ、６つの画像のバッチに適用される。これは、ＣＮＮ５００内のすべての畳み込み層５３０について繰り返される。キャッシュ・コヒーレンスは処理時間を短縮し、画像認識エンジンの性能を改善する。

【０１１５】

３次元（３Ｄ）畳み込みと呼ばれる１つの斯かる実施形態では、ＣＮＮ５００の性能の更なる改善がバッチ内の画像フレームにわたって情報を共有することによって達成される。これは、関節のより正確な識別に役立ち、誤検知を減少させる。例えば、所与のバッチ内の複数の画像フレームにわたってピクセル値が変化しない画像フレーム内の特徴は、シェルフなどの静的物体である可能性が高い。所与のバッチ内の画像フレームにわたる同じピクセルの値の変化は、このピクセルが関節である可能性が高いことを示す。従って、ＣＮＮ５００はそのピクセルによって識別された関節を正確に識別するために、そのピクセルの処理により焦点を当てることができる。

［関節データ構造］

【０１１６】

ＣＮＮ５００の出力は、カメラ当たりの各画像に対する信頼度配列の行列である。信頼度配列の行列は、関節データ構造の配列に変換される。図６に示すような関節データ構造６００は、各関節の情報を記憶するために使用される。関節データ構造６００は、画像が受信されるカメラの２Ｄ画像空間内の特定の画像内の要素のｘ位置及びｙ位置を識別する。関節番号は、識別された関節のタイプを識別する。例えば、一実施形態では、値は１～１９の範囲である。値１は関節が左足首であることを示し、値２は関節が右足首であることを示し、以下同様である。関節のタイプは、出力行列５４０内のその要素に対する信頼度配列を使用して選択される。例えば、一実施形態では、左足首関節に対応する値がその画像要素の信頼度配列において最も高い場合、関節番号の値は「１」である。

【０１１７】

信頼度数は、その関節を予測する際のＣＮＮ５００の信頼度の程度を示す。信頼度数の値が高ければ、ＣＮＮは自身の予想に確信していることになる。関節データ構造を一意に識別するために、関節データ構造に整数ＩＤが割り当てられる。上記マッピングに続いて、画像毎の信頼度配列の出力行列５４０は、画像毎の関節データ構造の配列に変換される。

【０１１８】

画像認識エンジン１１２ａ～１１２ｎはカメラ１１４から画像のシーケンスを受信し、画像を処理して、上述のように関節データ構造の対応する配列を生成する。特定の画像の

関節データ構造の配列は、関節タイプ、特定の画像の時間、及び特定の画像内の要素の座標によって、特定の画像の要素を分類する。一実施形態では画像認識エンジン１１２ａ～１１２ｎが畳み込みニューラル・ネットワークＣＮＮ５００であり、関節タイプは被写体の１９種類の関節のうちの１つ、特定の画像の時間は特定の画像についてソースカメラ１１４によって生成された画像のタイムスタンプであり、座標（ｘ，ｙ）は２Ｄ画像平面上の要素の位置を特定する。

【０１１９】

一実施形態では、関節分析が、各入力画像に対して、ｋ最近傍、ガウス混合、様々な画像形態変換、及び関節ＣＮＮの組み合わせを実行することを含む。この結果は、各時点において画像数をビットマスクにマッピングするリング・バッファ内にビットマスクの形式で格納することができる関節データ構造の配列を含む。

［追跡エンジン］

【０１２０】

追跡エンジン１１０は、重なり合う視野を有するカメラからの画像のシーケンス内の画像に対応する、画像認識エンジン１１２ａ～１１２ｎによって生成された関節データ構造の配列を受信するように構成される。画像当たりの関節データ構造の配列は、図７に示すように、画像認識エンジン１１２ａ～１１２ｎによってネットワーク１８１を介して追跡エンジン１１０に送られる。追跡エンジン１１０は、様々なシーケンスの画像に対応する関節データ構造の配列内の要素の座標を、実空間内の座標を有する候補関節に変換する。追跡エンジン１１０は、実空間における座標（関節のコンステレーション）を有する候補関節のセットを、実空間における被写体として識別するためのロジックを備える。一実施形態では、追跡エンジン１１０が、所与の時点におけるすべてのカメラについて、画像認識エンジンからの関節データ構造の配列を蓄積し、候補関節のコンステレーションを識別するために使用されるように、この情報を辞書として被写体データベース１４０に格納する。辞書は、キー値ペアの形式で編成することができ、ここで、キーはカメラＩＤであり、値はカメラからの関節データ構造の配列である。斯かる実施形態では、この辞書が候補関節を決定し、関節を被写体に割り当てるために、ヒューリスティックス・ベースの分析で使用される。斯かる実施形態では、追跡エンジン１１０の高レベル入力、処理、及び出力が表１に示されている。

表１：例示的な実施形態における追跡エンジン１１０からの入力、処理、及び出力。

入力	処理	出力
画像毎、及び、関節データ構造毎の関節データ構造の配列 - 個有ＩＤ - 信頼度数 - 関節番号 - 画像空間内の（ｘ，ｙ）の位置	- 関節辞書の作成 - 重なり合う視野を有するカメラの視野における関節の位置を候補関節に再投影	- ある時点での実空間内の被写体のリスト

［関節の候補関節へのグループ化］

【０１２１】

追跡エンジン１１０は２つの次元、すなわち、時間及び空間に沿った関節データ構造の配列を受け取る。時間次元に沿って、追跡エンジンは、カメラ当たり画像識別エンジン１１２ａ～１１２ｎによって処理された関節データ構造のタイムスタンプ付き配列を連続的

に受け取る。関節データ構造は、重なり合う視野を有するカメラからの画像において、ある期間にわたる同じ被写体の同じ関節の複数のインスタンスを含む。特定の画像内の要素の（ｘ，ｙ）座標は、通常、特定の関節が属する被写体の動きのために、関節データ構造の連続的にタイムスタンプされた配列において異なっている。例えば、左手首関節として分類された２０個の画素は、特定のカメラからの多くの連続的にタイムスタンプされた画像に現れることができ、各左手首関節は、画像毎に変化していること或いは変化しないでいることができる実空間内の位置を有する。その結果、多くの連続的にタイムスタンプされた関節データ構造の配列内の２０個の左手首関節データ構造６００は、経時的に実空間内の同じ２０個の関節を表すことができる。

【０１２２】

重なり合う視野を有する複数のカメラは実空間内の各位置をカバーするので、任意の所与の時点に、カメラ１１４のうちの２つ以上の画像に同じ関節が現れる可能性がある。カメラ１１４は時間的に同期され、従って、追跡エンジン１１０は任意の所与の時点に、重なり合う視野を有する複数のカメラから、特定の関節の関節データ構造を受信する。これは空間次元であり、２つの次元、すなわち、時間及び空間のうちの第２の次元であり、追跡エンジン１１０は、空間次元に沿って関節データ構造の配列内のデータを受け取る。

【０１２３】

追跡エンジン１１０は、ヒューリスティックス・データベース１６０に格納されたヒューリスティックスの最初の組を使用して、関節データ構造の配列から関節データ構造の候補を識別する。目標は、ある期間にわたってグローバル・メトリックを最小化することである。グローバル・メトリック計算器７０２は、グローバル距離を計算する。グローバル・メトリックは、以下に説明する複数の値の合計である。直観的には、追跡エンジン１１０によって時間次元と空間次元に沿って受信された関節データ構造の配列における関節がそれぞれの被写体に正しく割り当てられる場合、グローバル・メトリックの値は最小である。例えば、顧客が通路内を移動するショッピングストアの実施形態を考える。顧客Ａの左手首が顧客Ｂに誤って割り当てられた場合、グローバル・メトリックの値は増加する。従って、各顧客に対する各関節のグローバル・メトリックを最小化することは、最適化問題である。この問題を解決する１つの選択肢は、関節の全ての可能な接続を試みることである。しかしながら、これは、顧客の数が増加することにつれて、扱いにくくなる可能性がある。

【０１２４】

この問題を解決するための第２のアプローチは、ヒューリスティックスを使用して、単一の被写体に対する候補関節のセットのメンバーとして識別される関節の可能な組み合わせを低減することである。例えば、関節の相対位置の既知の生理学的特性のために、左手首関節は被写体の他の関節から空間的に遠く離れた被写体に属することができない。同様に、画像から画像への位置の変化が小さい左手首関節は、被写体が非常に高速で動くことが期待されないため、時間的に遠く離れた画像から同じ位置に同じ関節を有する被写体に属する可能性が低い。これらの初期ヒューリスティックスは、特定の被写体として分類され得る候補関節のコンステレーションのための時間及び空間における境界を構築するために使用される。特定の時間及び空間境界内の関節データ構造内の関節は、実空間内に存在する被写体としての候補関節のセットに割り当てるための「候補関節」と見なされる。これらの関節候補は、ある期間（時間次元）にわたる同じカメラからの多数の画像からの関節データ構造の配列において、重なり合う視野（空間次元）を有する様々なカメラにわたって識別された関節を含む。

［足関節］

【０１２５】

関節は、関節のリストで上述したように、関節をコンステレーションに、足関節及び非足関節にグループ化するための手順を目的として分割することができる。本実施例における左及び右足首関節タイプは、この手順の目的として足関節と考えられる。追跡エンジン１１０は、足関節を使用して、特定の被写体の候補関節のセットの識別を開始することが

できる。ショッピングストアの実施形態では、顧客の足が図２に示すようにフロア２２０上にある。カメラ１１４のフロア２２０までの距離は既知である。従って、重なり合う視野を有するカメラの画像に対応するデータ関節データ構造の配列からの足関節の関節データ構造を組み合わせる場合、追跡エンジン１１０は、既知の深さ（ｚ軸に沿った距離）を仮定することができる。足関節の深さの値はゼロ、すなわち、実空間の（ｘ，ｙ，ｚ）座標系において（ｘ，ｙ，０）である。この情報を使用して、画像追跡エンジン１１０は候補足関節を識別するために、重なり合う視野を有するカメラからの足関節の関節データ構造を組み合わせるために、ホモグラフィック・マッピングを適用する。このマッピングを使用して、画像空間における（ｘ，ｙ）座標における関節の位置が、実空間における（ｘ，ｙ，ｚ）座標における位置に変換され、候補足関節が得られる。この処理は、それぞれの関節データ構造を使用して候補左足関節及び候補右足関節を識別するために別々に実行される。

【０１２６】

これに続いて、追跡エンジン１１０は、候補左足関節及び候補右足関節を組み合わせて（候補関節のセットにそれらを割り当てて）、被写体を作成することができる。候補関節の銀河からの他の関節は、作成された被写体の関節タイプの幾つかまたは全てのコンステレーションを構築するために、被写体にリンクすることができる。

【０１２７】

左側候補足関節が１つしかなく、右側候補足関節が１つしかない場合、特定の時点で特定の空間に１つの被写体しか存在しないことを意味する。追跡エンジン１１０は、関節のセットに属する候補左足関節及び候補右足関節を有する新しい被写体を作成する。被写体は、被写体データベース１４０に保存される。複数の候補左足関節及び候補右足関節がある場合、グローバル・メトリック計算器７０２はグローバル・メトリックの値が最小化されるように、各候補左足関節を各候補右足関節に結合して被写体を作成することを試みる。

［非足関節］

【０１２８】

特定の時間及び空間境界内の関節データ構造の配列から候補非足関節を識別するために、追跡エンジン１１０は任意の所与のカメラＡから、重なり合う視野を有するその隣接するカメラＢへの非線形変換（基本行列とも呼ばれる）を使用する。非線形変換は、単一の多関節被写体を使用して計算され、上述のように較正データベース１７０に格納される。例えば、重なり合う視野を有する２つのカメラＡ及びＢについて、候補非足関節は、以下のように識別される。カメラＡからの画像フレーム内の要素に対応する関節の配列データ構造内の非足関節はカメラＢからの同期画像フレーム内のエピポーラ線にマッピングされる。カメラＡの特定の画像の関節データ構造の配列内の関節データ構造によって識別される関節（マシンビジョン文献では特徴とも呼ばれる）が、カメラＢの画像内に現れる場合、対応するエピポーラ線上に現れる。例えば、カメラＡからの関節データ構造内の関節が左手関節である場合、カメラＢの画像内のエピポーラ線上の左手関節はカメラＢの視点から見て同じ左手関節を表す。カメラＡ及びＢの画像内のこれら２つの点が実空間内の３Ｄ場面内の同じ点の投影であり、「共役ペア」と呼ばれる。

【０１２９】

Ｎａｔｕｒｅ、Ｖｏｌｕｍｅ ２９３、１９８１年９月１０日号に、「Ａ ｃｏｍｐｕｔｅｒ ａｌｇｏｒｉｔｈｍ ｆｏｒ ｒｅｃｏｎｓｔｒｕｃｔｉｎｇ ａ ｓｃｅｎｅ ｆｒｏｍ ｔｗｏ ｐｒｏｊｅｃｔｉｏｎｓ」という表題の論文に掲載されたＬｏｎｇｕｅｔ－Ｈｉｇｇｉｎｓによる手法などのマシンビジョン技術は、実空間におけるフロア２２０からの関節の高さを決定するために、対応点の共役ペアに適用される。上記の方法を適用するには、重なり合う視野を有するカメラ間の所定のマッピングが必要である。そのデータは、上述のカメラ１１４の較正中に決定された非線形関数として較正データベース１７０に格納される。

【０１３０】

追跡エンジン１１０は、重なり合う視野を有するカメラからの画像シーケンス内の画像

に対応する関節データ構造の配列を受信し、様々なシーケンス内の画像に対応する関節データ構造の配列内の要素の座標を、実空間内の座標を有する候補非足関節に変換する。識別された候補非足関節は、グローバル・メトリック計算器７０２を使用して、実空間内の座標を有する被写体の集合にグループ化される。グローバル・メトリック計算器７０２は、グローバル・メトリック値を計算し、非足関節の異なる組み合わせをチェックすることによって値を最小化することを試みる。一実施形態では、グローバル・メトリックが４つのカテゴリに編成されたヒューリスティックスの合計である。候補関節のセットを識別するロジックは、候補関節のセットを被写体として識別するために、実空間における被写体の関節間の物理的関係に基づくヒューリスティック関数を含む。関節間の物理的関係の例は、以下に記載されるようなヒューリスティックスにおいて考慮される。

［第１カテゴリのヒューリスティックス］

【０１３１】

第１カテゴリのヒューリスティックスは、同じまたは異なる時点における同じカメラ視野内の２つの提案された被写体関節位置間の類似性を確認するためのメトリックを含む。一実施形態ではこれらのメトリックは浮動小数点値であり、より高い値は関節の２つのリストが同じ被写体に属する可能性が高いことを意味する。ショッピングストアの例示的な実施形態を考えると、メトリックは、時間次元に沿った１つの画像から次の画像までの、１つのカメラ内の顧客の同じ関節間の距離を決定する。カメラ４１２の視野内の顧客Ａが与えられると、メトリックの第１のセットは、カメラ４１２からの１つの画像からカメラ４１２からの次の画像までの人物Ａの関節の各々の間の距離を決定する。メトリックは、カメラ１１４からの画像当たりの関節データ構造の配列において関節データ構造６００に適用される。

【０１３２】

一実施形態では、第１カテゴリのヒューリスティックスにおける２つの例示的なメトリックを以下に列挙する：

１．フロア上の２人の被写体の左足首関節とフロア上の２人の被写体の右足首関節との間の合計されたユークリッド２Ｄ座標距離の逆数（特定のカメラからの特定の画像のｘ、ｙ座標値を使用する）。

２．画像フレーム内の被写体の非足関節のすべてのペアの間のユークリッド２Ｄ座標距離の逆数の合計。

［第２カテゴリのヒューリスティックス］

【０１３３】

第２カテゴリのヒューリスティックスは、同じ時点に複数のカメラの視野から２つの提案された被写体関節位置間の類似性を確認するためのメトリックを含む。一実施形態では、これらのメトリックは浮動小数点値であり、より高い値は関節の２つのリストが同じ被写体に属する可能性が高いことを意味する。ショッピングストアの例示的な実施形態を考えると、第２のセットのメトリックは、同じ時点で（重なり合う視野を有する）２つ以上のカメラからの画像フレーム内の顧客の同じ関節間の距離を決定する。

【０１３４】

一実施形態では、第２カテゴリのヒューリスティックスにおける２つの例示的なメトリックを以下に列挙する：

【０１３５】

フロア上の２人の被写体の左足首関節とフロア上の２人の被写体の右足首関節との間のユークリッド２Ｄ座標距離の逆数（特定のカメラからの特定の画像のｘ、ｙ座標値を使用する）を合計した。第１の被写体の足首関節位置は、第２の被写体がホモグラフィック・マッピングを通して見えるカメラに投影される。

【０１３６】

線と点との間のユークリッド２Ｄ座標距離の逆数の関節の全てのペアの和であり、ここで、線は、視野内に第１の被写体を有する第１のカメラから、視野内に第２の被写体を有する第２のカメラまでの画像の関節のエピポーラ線であり、点は、第２のカメラからの画

像内の第２の被写体の関節である。
［第３カテゴリのヒューリスティックス］
【０１３７】
第３カテゴリのヒューリスティックスは、同じカメラビュー内の提案された被写体関節位置のすべての関節間の類似性を同じ時点に確認するためのメトリックを含む。ショッピングストアの例示的な実施形態を考えると、このカテゴリのメトリックは、１つのカメラからの１つのフレームにおける顧客の関節間の距離を決定する。
［第４カテゴリのヒューリスティックス］
【０１３８】
第４カテゴリのヒューリスティックスは、提案された被写体関節位置間の相違を確認するためのメトリックを含む。一実施形態では、これらのメトリックは浮動小数点値である。より高い値は、関節の２つのリストが同じ被写体ではない可能性がより高いことを意味する。一実施形態では、このカテゴリにおける２つの例示的なメトリックが以下を含む：
１．　２人の提案された被写体の頚部関節間の距離。
２．　２人の被写体間の関節のペア間の距離の合計。
【０１３９】
一実施形態では、経験的に決定され得る様々な閾値が以下に記載されるように、上記に列挙されたメトリックに適用される：
１．　メトリック値が、関節が既知の被写体に属すると考えるのに十分に小さい場合を判定するための閾値。
２．　メトリック類似性スコアが良好すぎる状態で関節が属する可能性がある潜在的な候補被写体が多すぎる場合を判定するための閾値。
３．　関節の集合が、経時的に、以前は実空間には存在しなかった新しい被写体と見なされるのに十分に高いメトリック類似性を有する場合を判定するための閾値。
４．　被写体が既に実空間にいない場合を判定するための閾値。
５．　追跡エンジン１１０が、間違って２つの被写体を混同した場合を判定するための閾値。
【０１４０】
追跡エンジン１１０は、被写体として識別された関節のセットを記憶するロジックを含む。候補関節のセットを識別するロジックは、特定の時間に撮影された画像において識別された候補関節が先行する画像において被写体として識別された候補関節のセットのうちの１つのメンバーに対応するかどうかを判定するロジックを含む。一実施形態では、追跡エンジン１１０が被写体の現在の関節位置を、同じ被写体の以前に記録された関節位置と、定期的に比較する。この比較により、追跡エンジン１１０は、実空間内の被写体の関節位置を更新することができる。更に、これを使用して、追跡エンジン１１０は誤検知（すなわち、誤って識別された被写体）を識別し、実空間に既に存在しない被写体を除去する。
【０１４１】
ショッピングストアの実施形態の例を考えると、追跡エンジン１１０はより早い時点に、顧客（被写体）を生成したが、ある時間の後、追跡エンジン１１０はその特定の顧客に対して現在の関節位置を有していない。それは、顧客が誤って生成されたことを意味する。追跡エンジン１１０は、誤って生成された被写体を被写体データベース１４０から削除する。一実施形態では、追跡エンジン１１０はまた、上述の処理を用いて、実空間から積極的に識別された被写体を除去する。ショッピングストアの例を考えると、顧客がショッピングストアを離れると、追跡エンジン１１０は、被写体データベース１４０から対応する顧客レコードを削除する。斯かる一実施形態では、追跡エンジン１１０が「顧客が店を出た」ことを示すために、被写体データベース１４０内のこの顧客レコードを更新する。
【０１４２】
一実施形態では、追跡エンジン１１０が足ヒューリスティックスと非足ヒューリスティックスを同時に適用することによって、被写体を識別しようと試みる。これにより、被写体の連結関節の「アイランド」が生成される。追跡エンジン１１０が、時間次元と空間次

元に沿って関節データ構造の配列を更に処理すると、アイランドの大きさが増加する。最終的に、関節のアイランドは被写体を形成する関節の他のアイランドと融合し、そして、被写体データベース１４０に格納される。一実施形態では、追跡エンジン１１０が所定の期間、未割り当ての関節の記録を維持する。この間、追跡エンジンは、未割り当ての関節を既存の被写体に割り当てるか、またはこれらの未割り当ての関節から新しい多関節存在物を作成しようと試みる。追跡エンジン１１０は、所定の期間の後、未割り当ての関節を破棄する。他の実施形態では、被写体を識別し追跡するために、上述の列挙したものとは異なるヒューリスティックスが使用されることを理解されたい。

【０１４３】

一実施形態では、追跡エンジン１１０をホストするノード１０２に接続されたユーザ・インターフェース出力デバイスが、実空間内の各被写体の位置を表示する。斯かる一実施形態では、出力デバイスの表示が、被写体の新しい位置でもって、定期的にリフレッシュされる。

［被写体データ構造］

【０１４４】

被写体の関節は、上述のメトリックを使用して互いに接続される。その際、追跡エンジン１１０は新しい被写体を生成し、それぞれの関節位置を更新することによって既存の被写体の位置を更新する。図８は、被写体を格納するための被写体データ構造８００を示す。該データ構造８００は、被写体関連データをキー値辞書として格納する。キーはフレーム番号であり、値は別のキー値辞書であり、ここでは、キーはカメラＩＤであり、値は（被写体の）１８個の関節と実空間内のそれらの位置のリストである。被写体データは、被写体データベース１４０に格納される。新しい被写体毎に、被写体データベース１４０内の被写体のデータにアクセスするために使用される固有識別子も割り当てられる。

【０１４５】

一実施形態では、システムが被写体の関節を識別し、被写体の骨格を作成する。骨格は、実空間に投影され、実空間における被写体の位置及び向きを示す。これは、マシンビジョンの分野では「姿勢推定」とも呼ばれる。一実施形態では、システムがグラフィカル・ユーザ・インターフェース（ＧＵＩ）上に実空間内の被写体の向き及び位置を表示する。一実施形態では、画像分析は匿名であり、すなわち、関節分析によって作成された被写体に割り当てられた固有識別子は実空間内の任意の特定被写体の詳細な個人識別情報（名前、電子メールアドレス、郵送先住所、クレジットカード番号、銀行口座番号、運転免許証番号など）を識別しない。

［被写体追跡の処理フロー］

【０１４６】

本明細書では、ロジックを示す幾つかのフローチャートを説明する。ロジックは、プロセッサによってアクセス可能かつ実行可能なメモリに格納されたコンピュータ・プログラムを使用してプログラムされ、上述のように構成されたプロセッサを使用して、及び他の構成では、フィールドプログラマブル集積回路を含む専用ロジックハードウェアによって、及び専用ロジックハードウェアとコンピュータ・プログラムとの組合せによって実装され得る。本明細書のすべてのフローチャートでは、達成される機能に影響を及ぼすことなく、ステップの多くを組み合わせること、並列に実行すること、または異なる順序で実行することができることが理解されよう。幾つか場合では、読者が理解するように、ステップの再編は、特定の他の変更が同様に行われる場合にのみ、同じ結果を達成する。他の場合には、読者が理解するように、ステップの再編は特定の条件が満たされる場合にのみ、同じ結果を達成する。更に、本明細書のフローチャートは実施形態の理解に関連するステップのみを示し、他の機能を達成するための多数の追加のステップが、示されたステップの前、後、及びそれらの間で実行され得ることが理解されるのであろう。

【０１４７】

図９は、被写体を追跡するための処理ステップを示すフローチャートである。処理はステップ９０２で開始する。実空間のエリア内に視野を有するカメラ１１４は、ステップ９

０４の進行中に較正される。ビデオ処理は、ステップ９０６において、画像認識エンジン１１２ａ～１１２ｎによって実行される。一実施形態では、ビデオ処理がそれぞれのカメラから受信された画像フレームのバッチを処理するために、カメラ毎に実行される。それぞれの画像認識エンジン１１２ａ～１１２ｎからのすべてのビデオ処理の出力は、ステップ９０８で追跡エンジン１１０によって実行されるシーン処理への入力として与えられる。シーン処理は新しい被写体を識別し、既存の被写体の共同位置を更新する。ステップ９１０では、処理すべき画像フレームがまだあるかどうかがチェックされる。更に画像フレームがある場合、処理はステップ９０６に進み、なければ、処理はステップ９１４で終了する。

【０１４８】

処理ステップ９０４「実空間でカメラを較正する」のより詳細な処理ステップが、図１０のフローチャートに示されている。較正処理は、ステップ１００２で、実空間の（ｘ，ｙ，ｚ）座標に対する（０，０，０）点を識別することによって開始する。ステップ１００４において、視野内に位置（０，０，０）を有する第１のカメラが較正される。カメラ較正の更なる詳細は、本出願において以前に提示されている。ステップ１００６において、第１のカメラと重なり合う視野を有する次のカメラが較正される。ステップ１００８では、較正すべきカメラがまだあるかどうかがチェックされる。この処理は、全てのカメラ１１４が較正されるまで、ステップ１００６で繰り返される。

【０１４９】

次の処理ステップ１０１０では、被写体が実空間に導入され、重なり合う視野を有するカメラ間の対応点の共役ペアが識別される。この処理のいくつかの詳細は上述されている。この処理は、ステップ１０１２で、重なり合うカメラのすべてのペアについて繰り返される。カメラがこれ以上存在しない場合、処理は終了する（ステップ１０１４）。

【０１５０】

図１１のフローチャートは、「ビデオ処理」ステップ９０６のより詳細なステップを示す。ステップ１１０２では、カメラ当たりｋ個の連続してタイムスタンプされた画像が更なる処理のためのバッチとして選択される。一実施形態では、ｋの値＝６で、画像認識エンジン１１２ａ～１１２ｎをそれぞれホストするネットワーク・ノード１０１ａ～１０１ｎにおけるビデオ処理のための利用可能なメモリに基づいて計算される。次のステップ１１０４では、画像のサイズが適切な寸法に設定される。一実施形態では、画像が１２８０ピクセルの幅、７２０ピクセルの高さ、及び３つのチャネルＲＧＢ(赤色、緑色、及び青色を表す）を有する。ステップ１１０６では、複数のトレーニングされた畳み込みニューラル・ネットワーク（ＣＮＮ）が画像を処理し、画像当たりの関節データ構造の配列を生成する。ＣＮＮの出力は、画像当たりの関節データ構造の配列である（ステップ１１０８）。この出力は、ステップ１１１０でシーン処理に送られる。

【０１５１】

図１２Ａは、図９「シーン処理」ステップ９０８のより詳細なステップの第１の部分を示すフローチャートである。シーン処理は、ステップ１２０２において、複数のビデオ処理からの出力を結合する。ステップ１２０４では、関節データ構造が足関節または非足関節を識別するかどうかがチェックされる。関節データ構造が足関節のものである場合、ステップ１２０６において、重なり合う視野を有するカメラからの画像に対応する関節データ構造を結合するために、ホモグラフィック・マッピングが適用される。この処理は、候補足関節（左足関節及び右足関節）を識別する。ステップ１２０８で、ステップ１２０６で識別された候補足関節にヒューリスティックスを適用して、候補足関節のセットを被写体として識別する。ステップ１２１０において、候補足関節のセットが既存の被写体に属するかどうかがチェックされる。属さない場合には、ステップ１２１２で、新しい被写体が作成される。属する場合は、ステップ１２１４で、既存の被写体が更新される。

【０１５２】

図１２Ｂは、「シーン処理」ステップ９０８のより詳細なステップの第２の部分を示すフローチャートである。ステップ１２４０では、重なり合う視野を有するカメラからの画

像シーケンス内の画像に対応する関節データ構造の複数の配列から、非足関節のデータ構造が組み合わされる。これは、第１のカメラからの第１の画像からの対応点を、重なり合う視野を有する第２のカメラからの第２の画像にマッピングすることによって実行される。この処理の幾つかの詳細は上述されている。ヒューリスティックスは、ステップ１２４２において、候補非足関節に適用される。ステップ１２４６では、候補非足関節が既存の被写体に属するかどうかが判定される。属する場合、ステップ１２４８において、既存の被写体が更新される。属さない場合は、ステップ１２５０において、所定の時間の後に、候補非足関節を既存の被写体と一致させるために、候補非足関節が再び処理される。ステップ１２５２において、非足関節が既存の被写体に属するかどうかがチェックされる。属するのであれば、ステップ１２５６で被写体が更新される。属さない場合は、ステップ１２５４で関節は破棄される。

【０１５３】

例示的な実施形態では、新しい被写体を識別し、被写体を追跡し、被写体（実空間を離れたか、または間違って生成された）を削除する処理はランタイムシステム（推論システムとも呼ばれる）によって実行される「存在物結束アルゴリズム」の一部として実装される。存在物は、上記の被写体と呼ばれる関節のコンステレーションである。存在物結束アルゴリズムは、実空間内の存在物を識別し、実空間内の関節の位置を更新して、存在物の移動を追跡する。

【０１５４】

図１４は、ビデオ処理１４１１及びシーン処理１４１５を示す。図示の実施形態では、４つのビデオ処理が示されており、それぞれが、１または複数のカメラ１１４からの画像を処理する。ビデオ処理は、上述のように画像を処理し、フレーム毎に関節を識別する。一実施形態では、それぞれのビデオ処理が、フレーム当たりの関節毎に、２Ｄ座標、信頼度数、関節番号、及び固有ＩＤを識別する。すべてのビデオ処理の出力１４５２は、入力１４５３としてシーン処理１４１５に与えられる。一実施形態では、シーン処理が、キーがカメラＩＤであり、値が関節の配列である時点毎の関節キー値辞書を作成する。関節は、重なり合う視野を有するカメラの視点に再投影される。再投影された関節はキー値辞書として記憶され、後述するように、各カメラ内の各画像について前景被写体マスクを生成するために使用することができる。この辞書のキーは、関節ＩＤとカメラＩＤの組み合わせである。辞書内の値は、対象のカメラの視野に再投影された関節の２Ｄ座標である。

【０１５５】

シーン処理１４１５は、ある時点での実空間内のすべての被写体のリストを含む出力１４５７を生成する。リストは、被写体毎にキー値辞書を含む。キーは、被写体の固有識別子であり、値は、キーをフレーム番号とし、値をカメラ?被写体関節キー値辞書とする別のキー値辞書である。カメラ－被写体関節キー値辞書は、キーがカメラ識別子であり、値が関節のリストである被写体毎の辞書である。

［被写体毎に在庫商品を識別し追跡するための画像分析］

【０１５６】

図１５～図２５を参照して、実空間のエリア内の被写体による在庫商品を置くこと及び取ることを追跡するシステム及び様々な実施態様について説明する。システム及び処理は、一実施態様によるシステムのアーキテクチャレベル概略図である図１５Ａを参照して説明される。図１５Ａはアーキテクチャ図であるため、説明の明確性を向上させるために、特定の詳細は省略される。

［マルチＣＮＮパイプラインのアーキテクチャ］

【０１５７】

図１５Ａは、カメラ１１４から受信した画像フレームを処理して、実空間内の各被写体についてショッピングカート・データ構造を生成する畳み込みニューラル・ネットワークのパイプライン（マルチＣＮＮパイプラインとも呼ばれる）の高レベルアーキテクチャである。本明細書に記載のシステムは、多関節被写体を識別し追跡するための、上述のカメラ毎の画像認識エンジンを含む。個人毎に１つの「関節」のみが認識され追跡される例、

または空間及び時間にわたる他の特徴または他のタイプの画像データが、処理されている実空間内の被写体を認識し追跡するために利用される例を含む、代替の画像認識エンジンを使用することができる。

【０１５８】

マルチＣＮＮパイプラインは、カメラ毎に並列に作動し、各カメラからの画像を、カメラ毎に循環バッファ１５０２を介して画像認識エンジン１１２ａ～１１２ｎに移動させる。一実施形態では、システムが第１の画像プロセッサ・サブシステム２６０２、第２の画像プロセッサ・サブシステム２６０４、及び第３の画像プロセッサ・サブシステム２６０６の３つのサブシステムから構成される。一実施形態では、第１の画像プロセッサ・サブシステム２６０２が、畳み込みニューラル・ネットワーク（ＣＮＮ）として実装され、関節ＣＮＮ１１２ａ～１１２ｎと呼ばれる画像認識エンジン１１２ａ～１１２ｎを含む。図１に関連して説明したように、カメラ１１４は互いに時間的に同期させることができ、その結果、画像は同時に、または時間的に近く、かつ同じ画像キャプチャレートで取得される。同時にまたは時間的に近い実空間のエリアをカバーする全てのカメラにおいて取得された画像は、同期された画像が実空間において固定された位置を有する被写体のある時点での様々な光景を表すものとして処理エンジンにおいて識別されることができるという意味で同期される。

【０１５９】

一実施形態では、カメラ１１４がショッピングストア（スーパーマーケットなど）に設置され、重なり合う視野を有するカメラのセット（２つ以上）が各通路の上に配置されて、店舗内の実空間の画像を取得する。実空間にはＮ台のカメラがあるが、簡略化のために、図１７Ａではカメラ（ｉ）として１台のカメラしか示されておらず、ｉの値は１からＮまでの範囲である。各カメラは、それぞれの視野に対応する実空間の画像シーケンスを生成する。

【０１６０】

一実施形態では、各カメラからの画像シーケンスに対応する画像フレームが毎秒３０フレーム（ｆｐｓ）のレートでそれぞれの画像認識エンジン１１２ａ～１１２ｎに送られる。各画像フレームは画像データと共に、タイムスタンプ、カメラの識別情報（「カメラＩＤ」と略される）、及びフレーム識別情報（「フレームＩＤ」と略される）を有する。画像フレームは、カメラ１１４毎に循環バッファ１５０２（リング・バッファとも呼ばれる）に格納される。循環バッファ１５０２は、それぞれのカメラ１１４からの連続的にタイムスタンプされた画像フレームのセットを格納する。

【０１６１】

関節ＣＮＮはカメラ当たりの画像フレームのシーケンスを処理し、それぞれの視野に存在する各被写体の１８個の異なるタイプの関節を識別する。重なり合う視野を有するカメラに対応する関節ＣＮＮ１１２ａ～１１２ｎの出力は、各カメラの２Ｄ画像座標から実空間の３Ｄ座標に関節の位置をマッピングするために組み合わされる。ｊが１～ｘに等しい被写体（ｊ）毎の関節データ構造８００は、実空間における被写体（ｊ）の関節の位置を識別する。被写体データ構造８００の詳細を図８に示す。１つの例示的な実施形態では、関節データ構造８００が各被写体の関節の２レベルのキー値辞書である。第１のキーはフレーム番号であり、値は、キーがカメラＩＤであり、値が被写体に割り当てられた関節のリストである第２のキー値辞書である。

【０１６２】

関節データ構造８００によって識別される被写体と、カメラ当たりの画像フレームのシーケンスからの対応する画像フレームとを含むデータセットは、第３の画像プロセッサ・サブシステム２６０６内の有界ボックス生成器１５０４への入力として与えられる。第３の画像プロセッサ・サブシステムは、前景画像認識エンジンを更に備える。一実施形態では、前景画像認識エンジンが、例えば、前景における意味的に重要な物体（すなわち、買物客、その手及び在庫商品）が、各カメラからの画像において経時的に、在庫商品を置くこと及び取ることに関連するときに、当該物体を認識する。図１５Ａに示される例示的な

実施態様では、前景画像認識エンジンがＷｈａｔＣＮＮ１５０６及びＷｈｅｎＣＮＮ１５０８として実装される。有界ボックス生成器１５０４は、データセットを処理して、画像シーケンス内の画像内の識別された被写体の手の画像を含む有界ボックスを指定するロジックを実装する。有界ボックス生成器１５０４は、それぞれのソース画像フレームに対応する関節データ構造８００内の手関節の位置を使用して、カメラ毎に各ソース画像フレーム内の手関節の位置を識別する。被写体データ構造内の関節の座標が３Ｄ実空間座標内の関節の位置を示す一実施形態では、有界ボックス生成器が、関節位置を３Ｄ実空間座標からそれぞれのソース画像の画像フレーム内の２Ｄ座標にマッピングする。

【０１６３】

有界ボックス生成器１５０４は、カメラ１１４毎に循環バッファ内の画像フレーム内の手関節のための有界ボックスを作成する。一実施形態では有界ボックスが、画像フレームの１２８ピクセル（幅）×１２８ピクセル（高さ）部分であり、手関節は有界ボックスの中心に位置する。他の実施形態では、有界ボックスのサイズが６４ピクセル×６４ピクセルまたは３２ピクセル×３２ピクセルである。カメラからの画像フレーム内のｍ個の被写体について、最大２ｍ個の手関節、従って２ｍ個の有界ボックスが存在し得る。しかしながら、実際には、他の被写体または他の物体による遮蔽のために、２ｍより少ない手が画像フレーム内で見える。１つの例示的な実施形態では、被写体の手の位置が肘関節及び手首関節の位置から推測される。例えば、被写体の右手の位置は、右肘の位置（ｐ１として識別される）及び右手首の位置（ｐ２として識別される）を用いて、外挿量×(ｐ２－ｐ１)+ ｐ２として外挿される。ここで外挿量は０．４である。別の実施形態では、関節ＣＮＮ１１２ａ～１１２ｎが左手画像及び右手画像を使用してトレーニングされる。従って、斯かる実施形態では、関節ＣＮＮ１１２ａ～１１２ｎがカメラ当たりの画像フレーム内の手関節の位置を直接識別する。画像フレーム当たりの手の位置は、識別された手関節当たりの有界ボックスを生成するために有界ボックス生成器１５０４によって使用される。

【０１６４】

ＷｈａｔＣＮＮ１５０６は、識別された被写体の手の分類を生成するために、画像内の指定された有界ボックスを処理するようにトレーニングされた畳み込みニューラル・ネットワークである。１つの訓練されたＷｈａｔＣＮＮ１５０６は、１つのカメラからの画像フレームを処理する。ショッピングストアの例示的な実施形態では、各画像フレーム内の各手関節について、ＷｈａｔＣＮＮ１５０６は手関節が空であるかどうかを識別する。ＷｈａｔＣＮＮ１５０６はまた、手関節内の在庫商品のＳＫＵ(在庫管理単位）番号、手関節内の商品を示す信頼値が非ＳＫＵ商品（すなわち、ショッピングストア在庫に属さない）、及び画像フレーム内の手関節位置の状況を識別する。

【０１６５】

すべてのカメラ１１４のＷｈａｔＣＮＮモデル１５０６の出力は、所定の時間帯の間、単一のＷｈｅｎＣＮＮモデル１５０８によって処理される。ショッピングストアの例では、ＷｈｅｎＣＮＮ１５０８が被写体の両手について時系列分析を実行して、被写体が棚から店舗在庫商品を取るか、または店舗在庫商品を棚に置くかを識別する。ショッピングカート・データ構造１５１０（在庫商品のリストを含むログ・データ構造とも呼ばれる）は、被写体に関連するショッピングカート（またはバスケット）内の店舗在庫商品の記録を保持するために、被写体毎に作成される。

【０１６６】

第２の画像プロセッサ・サブシステム２６０４は、関節データ構造８００によって識別される被写体と、第３の画像プロセッサへの入力として与えられるカメラ当たりの画像フレームのシーケンスからの対応する画像フレームとを含む同じデータセットを受信する。サブシステム２６０４は、背景画像認識エンジンを含み、背景（すなわち、棚のような在庫陳列構造）における意味的に重要な差異を、例えば、当該差異が、各カメラからの画像において経時的に、在庫商品を置くこと及び取ることに関連するときに認識する。選択ロジック・コンポーネント（図１５Ａには図示せず）は信頼度スコアを使用して、第２の画像プロセッサまたは第３の画像プロセッサのいずれかからの出力を選択し、ショッピング

カート・データ構造１５１０を生成する。
【０１６７】
図１５Ｂは、複数のＷｈａｔＣＮＮモデルの結果を結合し、それを単一のＷｈｅｎＣＮＮモデルへの入力として与える調整ロジック・モジュール１５２２を示す。上述したように、重なり合う視野を有する２つ以上のカメラは、実空間における被写体の画像を取得する。単一の被写体の関節は、それぞれの画像チャネル１５２０内の複数のカメラの画像フレーム内に現れることができる。別個のＷｈａｔＣＮＮモデルは、被写体の手（手関節によって表される）における在庫商品のＳＫＵを識別する。調整ロジック・モジュール１５２２は、ＷｈａｔＣＮＮモデルの出力を結合して、ＷｈｅｎＣＮＮモデルのための単一の統合入力とする。ＷｈｅｎＣＮＮモデル１５０８は被写体のショッピングカートを生成するために、統合された入力に基づいて作動する。
【０１６８】
図１５ＡのマルチＣＮＮパイプラインを含むシステムの詳細な実施態様は、図１６、１７、及び１８に提示される。ショッピングストアの例では、システムが、実空間のエリア内の被写体による在庫商品を置くこと及び取ることを追跡する。実空間のエリアは、図２及び図３に示すように通路に設置された棚に配置された在庫商品を有するショッピングストアである。在庫商品を含む棚は、様々な異なる配置で構成され得ることを理解されたい。例えば、棚はそれらの背面がショッピングストアの側壁に当接し、前面が実空間の開放エリアに面した状態で一列に配置することができる。実空間において重なり合う視野を有する複数のカメラ１１４は、それらの対応する視野の画像シーケンスを生成する。図２及び図３に示すように、１つのカメラの視野は、少なくとも１つの他のカメラの視野と重なる。
［関節ＣＮＮ－被写体の識別と更新］
【０１６９】
図１６は、関節ＣＮＮ１１２ａ～１１２ｎが実空間内の被写体を識別するために実行する処理ステップのフローチャートである。ショッピングストアの例では、被写体は、棚と他のオープンスペースとの間の通路内で店舗内を移動する顧客である。処理はステップ１６０２で開始する。上述したように、カメラは、被写体を識別するためにカメラからの画像シーケンスが処理される前に較正されることに留意されたい。カメラ較正の詳細は、上述されている。重なり合う視野を有するカメラ１１４は、被写体が存在する実空間の画像を取得する（ステップ１６０４）。一実施形態では、カメラは同期された画像シーケンスを生成するように構成される。各カメラの画像シーケンスは、カメラ毎にそれぞれの循環バッファ１５０２に保存される。循環バッファ（リング・バッファとも呼ばれる）は、スライドする時間帯に画像のシーケンスを格納する。一実施形態では、循環バッファが対応するカメラからの画像フレームを格納する（１１０）。別の実施形態では、各循環バッファ１５０２が３．５秒間、画像フレームを格納する。他の実施形態では、画像フレーム（または期間）の数が上記の列挙した例示的な値よりも大きくても小さくてもよいことを理解されたい。
【０１７０】
関節ＣＮＮ１１２ａ～１１２ｎは、対応するカメラ１１４から画像フレームのシーケンスを受信する（ステップ１６０６）。各関節ＣＮＮは対応するカメラからの画像のバッチを複数の畳み込みネットワーク層を介して処理し、対応するカメラからの画像フレーム内の被写体の関節を識別する。例示的な畳み込みニューラル・ネットワークによる画像のアーキテクチャ及び処理を図５に示す。カメラ１１４は重なり合う視野を有するので、被写体の関節は、２つ以上の関節ＣＮＮによって識別される。関節ＣＮＮによって生成される関節データ構造６００の２次元（２Ｄ）座標は、実空間の３次元（３Ｄ）座標にマッピングされ、実空間における関節位置を識別する。このマッピングの詳細は、追跡エンジン１１０が様々な画像シーケンス内の画像に対応する関節データ構造の配列内の要素の座標を、実空間内の座標を有する候補関節に変換する、図７の説明において提示される。
【０１７１】

被写体の関節は上述のように、関節をコンステレーションにグループ化するために、２つのカテゴリ（足関節及び非足関節）に編成される。本実施例における左及び右足首関節タイプは、この手順の目的として足関節と考えられる。ステップ１６０８で、ヒューリスティックスを適用して、候補左足関節及び候補右足関節を候補関節のセットに割り当てて、被写体を作成する。これに続いて、ステップ１６１０において、新たに識別された被写体が既に実空間に存在するかどうかが判定される。存在していない場合には、ステップ１６１４で、新しい被写体が生成され、存在している場合は、ステップ１６１２で既存の被写体が更新される。

【０１７２】

候補関節の銀河からの他の関節は、作成された被写体の関節タイプのいくつかまたはすべてのコンステレーションを構築するために、被写体にリンクすることができる。ステップ１６１６において、ヒューリスティックスが非足関節に適用され、それらが識別された被写体に割り当てられる。グローバル・メトリック計算器７０２はグローバル・メトリック値を計算し、非足関節の異なる組み合わせをチェックすることによって値を最小化することを試みる。一実施形態では、グローバル・メトリックは上述のように４つのカテゴリに編成されたヒューリスティックスの合計である。

【０１７３】

候補関節のセットを識別するロジックは、候補関節のセットを被写体として識別するために、実空間における被写体の関節間の物理的関係に基づくヒューリスティック関数を含む。ステップ１６１８において、既存の被写体は、対応する非足関節を使用して更新される。処理する画像がまだある場合（ステップ１６２０）、ステップ１６０６～１６１８が繰り返され、なければ、処理はステップ１６２２で終了する。第１のデータセットは、上述の処理の終わりに生成される。第１のデータセットは、被写体と、実空間における識別された被写体の位置とを識別する。一実施形態では、第１のデータセットが図１５Ａに関連して、被写体毎の関節データ構造８００として上述される。

［ＷｈａｔＣＮＮ－手関節の分類］

【０１７４】

図１７は、実空間で特定された被写体の手の中の在庫商品を特定する処理ステップを示すフローチャートである。ショッピングストアの例では、被写体はショッピングストア内の顧客である。顧客が通路及びオープンスペースを移動すると、顧客は棚に貯蔵された在庫商品を取り上げ、その商品をショッピングカートまたはバスケット内に置く。画像認識エンジンは、複数のカメラから受け取った画像シーケンス内の画像セット内の被写体を識別する。このシステムは、識別された被写体によって在庫商品を取ることと、識別された被写体によって棚に在庫商品を置くことを検出するために、識別された被写体を含む画像シーケンス内の画像のセットを処理するロジックを含む。

【０１７５】

一実施形態では、画像のセットを処理するロジックが、識別された被写体に対して、識別された被写体の画像の分類を生成するために画像を処理するロジックを含む。分類は、識別された被写体が在庫商品を保持しているかどうかを含む。分類は、棚との相対的な識別された被写体の手の位置を示す第１の近似度分類を含む。分類は、識別された被写体の身体との相対的な識別された被写体の手の位置を示す第２の近似度分類を含む。分類は、識別された被写体に関連するバスケットとの相対的な識別された被写体の手の位置を示す第３の近似度分類を更に含む。最後に、分類は、可能性のある在庫商品の識別子を含む。

【０１７６】

別の実施形態では、画像のセットを処理するロジックが、識別された被写体について、識別された被写体の画像のセット内の画像内の手を表すデータの有界ボックスを識別するロジックを含む。有界ボックス内のデータは、識別された被写体の有界ボックス内のデータの分類を生成するために処理される。斯かる実施形態では、分類は識別された被写体が在庫商品を保持しているかどうかを含む。分類は、棚との相対的な識別された被写体の手の位置を示す第１の近似度分類を含む。分類は、識別された被写体の身体との相対的な識

別された被写体の手の位置を示す第２の近似度分類を含む。分類は、識別された被写体に関連するバスケットとの相対的な識別された被写体の手の位置を示す第３の近似度分類を含む。最後に、分類は、可能性のある在庫商品の識別子を含む。

【０１７７】

処理はステップ１７０２で開始する。ステップ１７０４では、画像フレーム内の被写体の手（手関節によって表される）の位置が識別される。有界ボックス生成器１５０４は、図１８で説明したように、関節ＣＮＮ１１２ａ～１１２ｎによって生成された第１のデータセット内で識別された関節位置を使用して、各カメラからフレーム当たりの被写体の手の位置を識別する。これに続いて、ステップ１７０６で、有界ボックス生成器１５０４は、第１のデータセットを処理して、画像シーケンス内の画像内の識別された多関節被写体の手の画像を含む有界ボックスを指定する。有界ボックス生成器の詳細は、図１５Ａの議論において上述されている。

【０１７８】

第２の画像認識エンジンは複数のカメラから画像シーケンスを受け取り、画像内の指定された有界ボックスを処理して、識別された被写体の手の分類を生成する（ステップ１７０８）。一実施形態では、手の画像に基づいて被写体を分類するために使用される画像認識エンジンのそれぞれは、ＷｈａｔＣＮＮ１５０６と呼ばれるトレーニングされた畳み込みニューラル・ネットワークを備える。ＷｈａｔＣＮＮは、図１５Ａに関連して上述したように、マルチＣＮＮパイプラインに配置される。一実施形態では、ＷｈａｔＣＮＮへの入力が多次元配列Ｂ×Ｗ×Ｈ×Ｃ（Ｂ×Ｗ×Ｈ×Ｃテンソルとも呼ばれる）である。「Ｂ」はＷｈａｔＣＮＮによって処理される画像のバッチ内の画像フレームの数を示すバッチサイズであり、「Ｗ」及び「Ｈ」は有界ボックスの幅及び高さをピクセルで示し、「Ｃ」は、チャネルの数である。一実施形態では、バッチ内に３０個の画像があり（Ｂ＝３０）、それで、有界ボックスのサイズは３２ピクセル（幅）×３２ピクセル（高さ）である。赤、緑、青、前景マスク、前腕マスク、及び上腕マスクをそれぞれ表す６つのチャネルが存在し得る。前景マスク、前腕マスク、及び上腕マスクは、この例ではＷｈａｔＣＮＮのための追加の任意的な入力データソースであり、ＣＮＮは、これをＲＧＢ画像データ内の情報を分類する処理に含めることができる。前景マスクは、例えば、ガウス・アルゴリズムの混合を使用して生成することができる。前腕マスクは、関節データ構造内の情報を使用して生成される状況を提供する、手首と肘との間の線とすることができる。同様に、上腕マスクは、関節データ構造内の情報を使用して生成される肘と肩との間の線とすることができる。他の実施形態では、Ｂ、Ｗ、Ｈ、及びＣパラメータの異なる値を使用することができる。例えば、別の実施形態では、有界ボックスのサイズはより大きく、例えば、６４ピクセル（幅）×６４ピクセル（高さ）または１２８ピクセル（幅）×１２８ピクセル（高さ）である。

【０１７９】

各ＷｈａｔＣＮＮ１５０６は、画像のバッチを処理して、識別された被写体の手の分類を生成する。分類は、識別された主題が在庫商品を保持しているかどうかを含む。分類は、置くこと及び取ることを検出するために使用可能な、棚及び被写体に対する相対的な手の位置を示す１または複数の分類を含む。この例では、第１の近似度分類が棚との相対的な識別された被写体の手の位置を示す。分類は、この例では、識別された被写体の身体との相対的な識別された被写体の手の位置を示す第２の近似度分類を含み、その場合に、被写体は買い物中に在庫商品を保持することができる。この例における分類は、識別された被写体に関連するバスケットとの相対的な識別された被写体の手の位置を示す第３の近似度分類を更に含み、この状況における「バスケット」は、買い物中に在庫商品を保持するために被写体によって使用されるバッグ、バスケット、カート、または他の物体である。最後に、分類は、可能性のある在庫商品の識別子を含む。ＷｈａｔＣＮＮ１５０６の最終レイヤは、未加工の予測値であるロジットを生成する。ロジットは浮動小数点値として表され、以下に説明するように、分類結果を生成するために更に処理される。一実施形態では、ＷｈａｔＣＮＮモデルの出力が多次元配列Ｂ×Ｌ（Ｂ×Ｌテンソルとも呼ばれる）を

含む。「Ｂ」はバッチサイズであり、「Ｌ＝Ｎ＋５」は画像フレーム当たりのロジット出力数であり、「Ｎ」は、ショッピングストアで販売される「Ｎ」個の固有在庫商品を表すＳＫＵの数である。

【０１８０】

１フレーム当たりの出力「Ｌ」は、ＷｈａｔＣＮＮ１５０６からの生の活性化である。ロジット「Ｌ」がステップ１７１０で処理され、在庫商品及び状況を識別する。最初の「Ｎ」個のロジットは被写体が「Ｎ」個の在庫商品の１つを保持していることの信頼度を表す。ロジット「Ｌ」が以下に説明する追加の５つのロジットを含む。第１のロジットは、被写体の手の中にある商品の画像が店舗ＳＫＵ商品（非ＳＫＵ商品とも呼ばれる）の１つでないという信頼度を表す。第２のロジットは、被写体が商品を保持しているか否かの信頼度を示す。大きな正の値は、ＷｈａｔＣＮＮモデルが、被写体が商品を保持しているという高いレベルの信頼度を有することを示す。大きな負の値は、モデルが、被写体が商品を保持していないことを確信していることを示す。第２のロジットのゼロに近い値は、ＷｈａｔＣＮＮモデルが、被写体が商品を保持しているか否かを予測することに確信がないことを示す。

【０１８１】

次の３つのロジットは第１、第２、及び第３の近似度分類を表す。第１の近似度分類は、棚との相対的な識別された被写体の手の位置を示し、第２の近似度分類は、識別された被写体の身体との相対的な識別された被写体の手の位置を示し、第３の近似度分類は、識別された被写体に関連するバスケットとの相対的な識別された被写体の手の位置を示す。従って、３つのロジットは手の位置の状況を表し、１つのロジットはそれぞれ、手の状況が棚の近く、バスケット（またはショッピングカート）の近く、または被写体の身体の近くにあるという信頼度を示す。一実施形態では、ＷｈａｔＣＮＮが棚の近く、バスケット（またはショッピングカート）の近く、及び被写体の身体の近くの３つの状況で手の画像を含むトレーニング・データセットを使用してトレーニングされる。別の実施形態では、「近似度分類」パラメータが手の状況を分類するためにシステムによって使用される。斯かる実施形態では、システムが状況を分類するために、棚、バスケット（またはショッピングカート）、及び被写体の身体までの識別された被写体の手の距離を決定する。

【０１８２】

ＷｈａｔＣＮＮの出力は上述したように、Ｎ個のＳＫＵロジット、１個の非ＳＫＵロジット、１個の保持ロジット、及び３個の状況ロジットから構成される「Ｌ」個のロジットである。ＳＫＵロジット（最初のＮロジット）及び非ＳＫＵロジット（Ｎロジットに続く最初のロジット）は、ｓｏｆｔｍａｘ関数によって処理される。図５を参照して上述したように、ｓｏｆｔｍａｘ関数は、任意の実数値のＫ次元ベクトルを、合計で１になる範囲［０，１］の実数値のＫ次元ベクトルに変換する。ｓｏｆｔｍａｘ関数は、Ｎ＋１個の商品にわたる商品の確率分布を計算する。出力値は０と１の間であり、すべての確率の合計は１に等しい。（複数クラス分類のための）ｓｏｆｔｍａｘ関数は、各クラスの確率を返す。最高の確率を有するクラスは、予測クラス（目標クラスとも呼ばれる）である。

【０１８３】

保持ロジットは、シグモイド関数によって処理される。シグモイド関数は入力として実数値をとり、０～１の範囲の出力値を生成する。シグモイド関数の出力は、手が空であるか、商品を保持しているかを識別する。３つの状況ロジットは、手関節位置の状況を識別するためにｓｏｆｔｍａｘ関数によって処理される。ステップ１７１２では、処理すべき画像がまだあるかどうかがチェックされる。処理すべき画像がまだあれば、ステップ１７０４～１７１０が繰り返され、なければ、処理はステップ１７１４で終了する。

［ＷｈｅｎＣＮＮ－商品を置くこと及び取ることを識別するための時系列分析］

【０１８４】

一実施形態では、システムが被写体の前景画像処理に基づいて、識別された被写体による置くこと及び取ることを検出するために、被写体の分類にわたって時系列分析を実行するロジックを実装する。時系列分析は、被写体のジェスチャと、画像シーケンスで表され

るジェスチャに関連する在庫商品とを識別する。

【０１８５】

マルチＣＮＮパイプラインにおけるＷｈａｔＣＮＮ１５０６の出力は、ＷｈｅｎＣＮＮ１５０８への入力として与えられ、ＷｈｅｎＣＮＮ１５０８は、識別された被写体による置くこと及び取ることを検出するために、これらの入力を処理する。最後に、システムは、検出された置くこと及び取ることに応答して、識別された各被写体に対して在庫商品のリストを含むログ・データ構造を生成するロジックを含む。ショッピングストアの例では、ログ・データ構造は、被写体毎のショッピングカート・データ構造１５１０とも呼ばれる。

【０１８６】

図１８は、被写体毎にショッピングカート・データ構造を生成するためのロジックを実施する処理を示す。処理はステップ１８０２で開始する。ＷｈｅｎＣＮＮ１５０８への入力は、ステップ１８０４で準備される。ＷｈｅｎＣＮＮへのインプットは多次元配列Ｂ×Ｃ×Ｔ×Ｃａｍｓであり、ここで、Ｂはバッチサイズであり、Ｃはチャネルの数であり、Ｔは時間帯の間考慮されるフレームの数であり、Ｃａｍｓはカメラ１１４の数である。一実施形態では、バッチサイズ「Ｂ」は６４であり、「Ｔ」の値は１１０画像フレームまたは３．５秒の時間内の画像フレームの数である。

【０１８７】

画像フレーム毎に識別された各被写体に対して、カメラ毎に、手関節毎に１０ロジット（両手に対して２０ロジット）のリストが生成される。保持ロジット及び状況ロジットは、上述のようにＷｈａｔＣＮＮ１５０６によって生成される「Ｌ」ロジットの一部である。

```
[
        holding,                          # 1 logit
        context,                          # 3 logits
        slice_dot(sku, log_sku),               # 1 logit
        slice_dot(sku, log_other_sku),            # 1 logit
        slice_dot(sku, roll(log_sku, -30)),     # 1 logit
        slice_dot(sku, roll(log_sku, 30)),      # 1 logit
        slice_dot(sku, roll(log_other_sku, -30)),      # 1 logit
        slice_dot(sku, roll(log_other_sku, 30))        # 1 logit
]
```

【０１８８】

上記のデータ構造は、画像フレーム内の手ごとに生成され、同じ被写体の他方の手に関するデータも含む。例えば、データが被写体の左手関節に対するものである場合、右手に対する対応する値は「他の」ロジットとして含まれる。５番目のロジット（log_skuと呼ばれる上記リストの項目番号３）は、上述の「Ｌ」ロジットにおけるＳＫＵロジットのログである。６番目のロジットが他の手に対するＳＫＵロジットのログである。「ｒｏｌｌ」関数が現在のフレームの前後で同じ情報を生成する。例えば、第７のロジット（roll(log_sku, -30)と呼ばれる）は、現在のフレームより３０フレーム早いＳＫＵロジットのログである。８番目のロジットは手のＳＫＵロジットのログであり、現在のフレームより３０フレーム遅い。リスト内の第９及び第１０のデータ値は、現在のフレームよりも３０フレーム前及び３０フレーム後の他方の手についての類似データである。他方の手についての同様のデータ構造も生成され、その結果、カメラ当たり画像フレーム当たり被写体当たり合計２０ロジットとなる。従って、ＷｈｅｎＣＮＮへの入力におけるチャネル数は２０である（すなわち、多次元配列Ｂ×Ｃ×Ｔ×ＣａｍｓにおいてＣ＝２０）。

【０１８９】

各カメラからの画像フレームのバッチ（例えば、Ｂ＝６４）内のすべての画像フレームについて、画像フレーム内で識別される、被写体当たり２０個の手ロジットの同様のデータ構造が生成される。時間帯(Ｔ＝３．５秒または１１０画像フレーム）を使用して、被写体の手関節に対して画像フレームのシーケンス内の前方及び後方画像フレームを探索する

。ステップ１８０６では、フレーム当たりの被写体当たり２０個の手ロジットがマルチＣＮＮパイプラインから統合される。一実施形態では、画像フレームのバッチ（６４）が、前方及び後方探索のための追加の画像フレームを両側に有する、画像フレーム１１０のより大きなウィンドウの中央に配置された画像フレームのより小さなウィンドウとして想像することができる。ＷｈｅｎＣＮＮ１５０８への入力Ｂ×Ｃ×Ｔ×Ｃａｍｓは、全てのカメラ１１４（「Ｃａｍｓ」と呼ばれる）からの画像フレームのバッチ「Ｂ」で識別された被写体の両手に対する２０個のロジットから構成される。統合された入力は、ＷｈｅｎＣＮＮモデル１５０８と呼ばれる単一のトレーニングされた畳み込みニューラル・ネットワークに与えられる。

【０１９０】

ＷｈｅｎＣＮＮモデルの出力は３つのロジットで構成され、識別された被写体の３つの可能な行為、すなわち棚から在庫商品を取ること、在庫商品を棚に置くこと、及び行為を行わないことに対する信頼度を表す。３つの出力ロジットは、実行される行為を予測するためにｓｏｆｔｍａｘ関数によって処理される。３つの分類ロジットは各被写体に対して一定の間隔で生成され、結果はタイムスタンプと共に個人毎に記憶される。一実施形態では、３つのロジットが被写体当たり２０フレーム毎に生成される。斯かる実施形態では、カメラ当たり２０画像フレーム毎の間隔で、１１０画像フレームのウィンドウが現在の画像フレームの周りに形成される。

【０１９１】

ある期間にわたる被写体当たりのこれら３つのロジットの時系列分析が実行されて（ステップ１８０８）、真のイベント及びそれらの発生時間に対応するジェスチャが識別される。この目的のために、非最大抑制（ＮＭＳ）アルゴリズムが使用される。１つのイベント（すなわち、被写体による商品を置くことまたは取ること）がＷｈｅｎＣＮＮ１５０８によって複数回（同じカメラ及び複数のカメラの両方から）検出されると、ＮＭＳは、被写体に対する余分なイベントを除去する。ＮＭＳは、２つの主要なタスク、すなわち、余分な検出にペナルティを課す「マッチングロス」と、より良好な検出が手近に存在するかどうかを知るための近隣の「ジョイント処理」とを含む再スコアリング技術である。

【０１９２】

各被写体に対する取ること及び置くことの真のイベントは、真のイベントを有する画像フレームの前の３０画像フレームに対するＳＫＵロジットの平均を計算することによって更に処理される。最後に、最大値の引数(arg maxまたはargmaxと略す）を使用して、最大値を決定する。ａｒｇｍａｘ値によって分類された在庫商品は、棚に置かれたまたは棚から取られた在庫商品を識別するために使用される。在庫商品は、ステップ１８１０で、それぞれの被写体のＳＫＵ（ショッピングカートまたはバスケットとも呼ばれる）のログに追加される。分類データが更にある場合（ステップ１８１２でチェックされる）、処理ステップ１８０４～１８１０年が繰り返される。ある期間にわたって、この処理の結果、各被写体のショッピングカートまたはバスケットが更新される。処理はステップ１８１４で終了する。

［シーン処理とビデオ処理を伴うＷｈａｔＣＮＮ］

【０１９３】

図１９は、シーン処理１４１５及びビデオ処理１４１１からのデータがＷｈａｔＣＮＮモデル１５０６に入力として与えられ、手の画像分類を生成するシステムの実施形態を示す。各ビデオ処理の出力は、別個のＷｈａｔＣＮＮモデルに与えられることに留意されたい。シーン処理１４１５からの出力は関節辞書である。この辞書ではキーは固有関節識別子であり、値は関節が関連付けられる固有被写体識別子である。関節に関連する被写体がない場合、それは辞書に含まれない。各ビデオ処理１４１１はシーン・処理から関節辞書を受け取り、フレーム番号を返された辞書にマッピングするリング・バッファにそれを格納する。返されたキー値辞書を使用して、ビデオ処理は、識別された被写体に関連付けられた手の近くにある各時点における画像のサブセットを選択する。手の関節の周りの画像フレームのこれらの部分は、領域提案と呼ぶことができる。

【０１９４】
ショッピングストアの事例では、領域提案が１または複数のカメラからの手の位置のフレームイメージであり、被写体は対応する視野にある。領域提案は、システム内のすべてのカメラによって生成される。これには、空の手だけでなく、ショッピングストア在庫商品及びショッピングストア在庫に属さない商品を持ち運ぶ手も含まれる。ビデオ処理は、時点毎に手の関節を含む画像フレームの部分を選択する。前景マスクの同様のスライスが生成される。上記（手関節の画像部分、前景マスク）を関節辞書（各手関節が属する被写体を示す）に連結して多次元配列を作成する。ビデオ処理からのこの出力は、ＷｈａｔＣＮＮモデルへの入力として与えられる。
【０１９５】
ＷｈａｔＣＮＮモデルの分類結果は、領域提案データ構造（ビデオ処理によって生成される）に格納される。ある時点での全ての領域は、その後、シーン処理への入力として戻される。シーン処理は結果をキー値辞書に格納する。但し、キーは被写体識別子であり、値はキー値辞書であり、但し、キーはカメラ識別子であり、値は領域のロジットである。次に、この集約されたデータ構造は、フレーム番号を時点毎に集約された構造にマッピングするリング・バッファに格納される。
［シーン処理とビデオ処理を伴うＷｈｅｎＣＮＮ］
【０１９６】
図２０は、ＷｈｅｎＣＮＮ１５０８が、図１９で説明したように、ビデオ処理毎にＷｈａｔＣＮＮモデルによって実行される手画像分類に続くシーン処理から出力を受け取るシステムの実施形態を示す。ある期間、例えば、１秒間の領域提案データ構造が、シーン処理への入力として与えられる。カメラが毎秒３０フレームの速度で画像を撮影している一実施形態では、入力が３０の期間と、対応する領域提案とを含む。シーン処理は、３０個の領域提案（手当たり）を、在庫商品ＳＫＵを表す単一の整数に縮小する。シーン処理の出力は、キーが被写体識別子であり、値がＳＫＵ整数であるキー値辞書である。
【０１９７】
ＷｈｅｎＣＮＮモデル１５０８は、時系列分析を実行して、この辞書の経時変化を判定する。この結果、棚から取り出され、ショッピングストアの棚に置かれた商品が識別される。ＷｈｅｎＣＮＮモデルの出力は、キーが被写体識別子であり、値がＷｈｅｎＣＮＮによって生成されたロジットであるキー値辞書である。一実施形態では、１組のヒューリスティックス２００２を使用して、被写体毎のショッピングカート・データ構造１５１０を決定する。ヒューリスティックスは、ＷｈｅｎＣＮＮの出力、それぞれの関節データ構造によって示される被写体の関節位置、及びプラノグラムに適用される。プラノグラムは、棚上の在庫商品の予め計算されたマップである。ヒューリスティックス２００２は、在庫商品が棚に置かれているか棚から取られているか、在庫商品がショッピングカート（またはバスケット）に置かれているか、またはショッピングカート（またはバスケット）から取られているか、または在庫商品が識別された被写体の身体に近いかを、取ることまたは置くことの夫々に対して判定する。
［Ｗｈａｔ－ＣＮＮモデルのアーキテクチャ例］
【０１９８】
図２１は、ＷｈａｔＣＮＮモデル１５０６の例示的なアーキテクチャを示す。この例示的なアーキテクチャでは、合計２６の畳み込み層がある。それぞれの幅（ピクセル単位）、高さ（ピクセル単位）、及びチャネル数に関する異なる層の次元も提示される。第１の畳み込み層２１１３は入力２１１１を受け取り、６４ピクセルの幅、６４ピクセルの高さ、及び６４チャネル（６４×６４×６４と記載）を有する。ＷｈａｔＣＮＮへの入力の詳細は、上述されている。矢印の方向は、１つの層から次の層へのデータの流れを示す。第２の畳み込み層２１１５は、３２×３２×６４の次元を有する。続いて第２の層があり、それぞれ３２×３２×６４の次元を有する８つの畳み込み層（ボックス２１１７に示される）がある。２つの層２１１９及び２１２１のみが、例示の目的のためにボックス２１１７に示されている。この後に、１６×１６×１２８の次元の別の８つの畳み込み層２１２

３が続く。このような２つの畳み込み層２１２５及び２１２７が図２１に示されている。最後に、最後の８つの畳み込み層２１２９は、それぞれ８×８×２５６の次元数を有する。２つの畳み込み層２１３１及び２１３３が、説明のためにボックス２１２９に示されている。

【０１９９】

Ｎ＋５個の出力を生成する最後の畳み込み層２１３３からの２５６個の入力を有する１つの全結合層２１３５がある。上述したように、「Ｎ」は、ショッピングストアで販売される「Ｎ」個の固有在庫商品を表すＳＫＵの数である。５つの追加のロジットは、画像内の商品が非ＳＫＵ商品であるという信頼性を表す第１のロジットと、被写体が商品を保持しているかどうかの信頼度を表す第２のロジットとを含む。次の３つのロジットは上述したように、第１、第２及び第３の近似度分類を表す。ＷｈａｔＣＮＮの最終出力は２１３７に示されている。例示的なアーキテクチャは、バッチ正規化（ＢＮ）を使用する。畳み込みニューラル・ネットワーク（ＣＮＮ）における各層の分布はトレーニング中に変化し、層別に変化する。これは、最適化アルゴリズムの収束速度を低下させる。バッチ正規化(Ｉｏｆｆｅ及びＳｚｅｇｅｄｙの２０１５年の論文）は、この問題を克服するための技術である。ＲｅＬＵ (正規化線形ユニット)活性化は、ｓｏｆｔｍａｘが使用される最終出力を除いて、各層の非線形性のために使用される。

【０２００】

図２２、図２３、及び図２４は、ＷｈａｔＣＮＮ１５０６の実施態様の様々な部分の図式的な視覚化である。これらの図は、ＴｅｎｓｏｒＢｏａｒｄ（商標）によって生成されたＷｈａｔＣＮＮモデルの図式的視覚化から編集された図である。ＴｅｎｓｏｒＢｏａｒｄ（商標）は、深層学習モデル、例えば、畳み込みニューラル・ネットワークを検査し、理解するための一連の視覚化ツールである。

【０２０１】

図２２は、片手（「片手」モデル２２１０）を検出する畳み込みニューラル・ネットワークモデルの高レベルアーキテクチャを示す。ＷｈａｔＣＮＮモデル１５０６は、それぞれ左手及び右手を検出するための２つの当該畳み込みニューラル・ネットワークを備える。図示の実施形態では、アーキテクチャが、ブロック０ ２２１６、ブロック１ ２２１８、ブロック２ ２２２０、及びブロック３ ２２２２と呼ばれる４つのブロックを含む。ブロックはより高レベルの抽象化であり、畳み込み層を表す複数のノードを含む。ブロックは１つのブロックからの出力が次のブロックに入力されるように、下から上への順序で配置される。このアーキテクチャは、プーリング層２２１４及び畳み込み層２２１２も含む。ブロック間では、異なる非線形性を使用することができる。図示の実施形態では、上述のようにＲｅＬＵ非線形性が使用される。

【０２０２】

図示の実施形態では、片手モデル２２１０への入力がＷｈａｔＣＮＮ１５０６の説明において上記で定義したＢ×Ｗ×Ｈ×Ｃテンソルである。「Ｂ」はバッチサイズであり、「Ｗ」及び「Ｈ」は入力画像の幅及び高さを示し、「Ｃ」はチャネル数である。片手モデル２２１０の出力は、第２の片手モデルと結合され、全結合ネットワークに転送される。

【０２０３】

トレーニング中、片手モデル２２１０の出力は、グラウンドトゥルースと比較される。出力とグランドトゥルースとの間で計算された予測誤差は、畳み込み層の重みを更新するために使用される。図示の実施形態では、ＷｈａｔＣＮＮ１５０６をトレーニングするために確率的勾配降下法（ＳＧＤ）が使用される。

【０２０４】

図２３は、図２２の片手畳み込みニューラル・ネットワークモデルのブロック０ ２２１６の更なる詳細を示す。これは、ボックス２３１０内のｃｏｎｖ０、ｃｏｎｖ１ ２３１８、ｃｏｎｖ２ ２３２０、及びｃｏｎｖ３ ２３２２とラベル付けされた４つの畳み込み層を含む。畳み込み層ｃｏｎｖ０の更なる詳細は、ボックス２３１０内に提示されている。入力は、畳み込み層２３１２によって処理される。畳み込み層の出力は、バッチ正規化層

２３１４によって処理される。ＲｅＬＵ非線形性２３１６は、バッチ正規化層２３１４の出力に適用される。畳み込み層ｃｏｎｖ０の出力は、次の層ｃｏｎｖ１ ２３１８に転送される。最終的な畳み込み層ｃｏｎｖ３の出力は、加算演算２３２４を介して処理される。この演算は、層ｃｏｎｖ３ ２３２２からの出力を、スキップ接続２３２６を介して到来する修正されていない入力に合計する。Ｈｅらの論文「深層残余ネットワークにおけるアイデンティティ・マッピング」（２０１６年７月２５日にｈｔｔｐｓ://ａｒｘｉｖ.ｏｒｇ/ｐｄｆ/１６０３．０５０２７.ｐｄｆで公開）では、順方向信号及び逆方向信号が１つのブロックから任意の他のブロックに直接的に伝播することができることが示されている。信号は、畳み込みニューラル・ネットワークを通って変化せずに伝播する。この技術は、深い畳み込みニューラル・ネットワークのトレーニング及び試験性能を改善する。

【０２０５】

図２１で説明したように、ＷｈａｔＣＮＮの畳み込み層の出力は、全結合層によって処理される。２つの片手モデル２２１０の出力は結合され、入力として全結合層に転送される。図２４は、全結合層（ＦＣ）２４１０の例示的な実施態様である。ＦＣ層への入力は、再整形演算子２４１２によって処理される。再整形演算子は、テンソルを次の層２４２０に転送する前にテンソルの形状を変更する。再整形は、畳み込み層からの出力を平坦化すること、すなわち、多次元行列からの出力を１次元行列またはベクトルに再整形することを含む。再構築演算子２４１２の出力はＭａｔＭｕｌ ２４２２と表示される行列乗算演算子にパスされ、ＭａｔＭｕｌ ２４２２からの出力はｘｗ＿ｐｌｕｓ＿ｂ ２４２４と表示される行列加算演算子に転送される。入力「ｘ」毎に、演算子２４２４は入力に行列「ｗ」及びベクトル「ｂ」を乗算して出力を生成する。「ｗ」が入力「ｘ」に関連するトレーニング可能なパラメータであり、「ｂ」がバイアスまたはインターセプトと呼ばれる別のトレーニング可能なパラメータである。全結合層２４１０からの出力２４２６が、ＷｈａｔＣＮＮ１５０６の説明において上述したように、Ｂ×Ｌテンソルである。「Ｂ」はバッチサイズであり、「Ｌ＝Ｎ＋５」は画像フレーム当たりの出力ロジット数である。「Ｎ」がショッピングストアで販売するための「Ｎ」個の固有在庫商品を表すＳＫＵの数である。

［ＷｈａｔＣＮＮモデルのトレーニング］

【０２０６】

様々な状況における空の手のみならず、様々な状況における様々な在庫商品を保持する手の画像のトレーニング・データセットが作成される。これを達成するために、人間の行為者が、試験環境の様々な場所で、多数の異なる方法で、各々の固有のＳＫＵ在庫商品を保持する。彼らの手の状況は、行為者の身体に近いこと、店舗の棚に近いこと、及び行為者のショッピングカートまたはバスケットに近いことに及ぶ。行為者は、空の手でも上記の行為を行う。この手順は、左手及び右手の両方について完了する。複数の行為者が、実際のショッピングストアで起こる自然な閉塞をシミュレーションするために、同じテスト環境でこれらの行為を同時に実行する。

【０２０７】

カメラ１１４は、上記行為を実行する行為者の画像を撮影する。一実施形態では、２０台のカメラがこの処理で使用される。関節ＣＮＮ１１２ａ～１１２ｎ及び追跡エンジン１１０は、関節を識別するために画像を処理する。有界ボックス生成器１５０４は、プロダクションまたは推論に類似した手領域の有界ボックスを作成する。ＷｈａｔＣＮＮ１５０６を介してこれらの手領域を分類する代わりに、画像は記憶ディスクに保存される。保存された画像は、精査され、ラベル付けされる。画像には、在庫商品ＳＫＵ、状況、及び手が何かを保持しているか否かという３つのラベルが割り当てられる。この処理は、多数の画像（数百万枚までの画像）に対して行われる。

【０２０８】

画像ファイルは、データ収集シーンに従って編成される。画像ファイルの命名規則は、画像のコンテンツ及び状況を識別する。図２５は、一実施形態における画像ファイル名を示す図である。数表示２５０２によって参照されるファイル名の第１の部分は、データ収

集シーンを識別し、画像のタイムスタンプも含む。ファイル名の第２の部分２５０４は、ソースカメラを識別する。図２５に示す例では、「カメラ４」で撮像されている。ファイル名の第３の部分２５０６は、ソースカメラからのフレーム番号を識別する。図示の例では、ファイル名が、それがカメラ４からの９４，６００番目の画像フレームであることを示す。ファイル名の第４の部分２５０８は、この手領域画像が取得されるソース画像フレーム内のｘ座標領域及びｙ座標領域の範囲を識別する。図示の例では、領域がピクセル１１７から３７０までのｘ座標値と、ピクセル３７０から４９８までのｙ座標値との間で定義される。ファイル名の第５の部分２５１０は、シーン内の行為者の個人ＩＤを識別する。図示の例では、シーン内の人物がＩＤ「３」を有する。最後に、ファイル名の第６の部分２５１２は、画像内で識別された在庫商品のＳＫＵ数（商品＝６８）を識別する。

【０２０９】

ＷｈａｔＣＮＮ１５０６のトレーニング・モードでは、順方向パスのみが実行されるプロダクション・モードとは対照的に、順方向パスと逆方向伝播が実行される。トレーニング中、ＷｈａｔＣＮＮは、順方向パスにおいて識別された被写体の手の分類を生成する。ＷｈａｔＣＮＮの出力は、グランドトゥルースと比較される。逆伝播では、１または複数のコスト関数の勾配が計算される。次いで、勾配は、畳み込みニューラル・ネットワーク（ＣＮＮ）及び全結合（ＦＣ）ニューラル・ネットワークに伝播され、その結果、予測誤差が低減され、出力がグランドトゥルースに近づく。一実施形態では、ＷｈａｔＣＮＮ１５０６をトレーニングするために、確率的勾配降下法（ＳＧＤ）が使用される。

【０２１０】

一実施形態では、６４個の画像がトレーニング・データからランダムに選択され、増強される。画像増強の目的はトレーニング・データを多様化し、モデルの性能を向上させることである。画像増強は、画像のランダムフリッピング、ランダム回転、ランダム色相シフト、ランダムガウスノイズ、ランダムコントラスト変化、及びランダムクロッピングを含む。増強の量はハイパー・パラメータであり、ハイパー・パラメータ探索によって調整される。増強された画像は、トレーニング中にＷｈａｔＣＮＮ１５０６によって分類される。分類はグラウンドトゥルースと比較され、ＷｈａｔＣＮＮ１５０６の係数または重みは、勾配損失関数を計算し、勾配に学習レートを乗算することによって更新される。上記処理は、エポックを形成するために何度も（例えば、約１０００回）繰り返される。５０から２００のエポックが実行される。各エポックの間、学習速度は、余弦アニーリングスケジュールに従ってわずかに減少する。

［ＷｈｅｎＣＮＮモデルのトレーニング］

【０２１１】

ＷｈｅｎＣＮＮ１５０８のトレーニングは、予測誤差を低減するために逆伝播を使用する、上述のＷｈａｔＣＮＮ１５０６のトレーニングと同様である。行為者は、トレーニング環境において様々な行為を実行する。例示的な実施形態では、トレーニングは、在庫商品が貯蔵された棚を有するショッピングストアで実行される。行為者によって実行される行為の例には、棚から在庫商品を取り出すこと、在庫商品を棚に置いて戻すこと、在庫商品をショッピングカート（またはバスケット）に置くこと、ショッピングカートから在庫商品を取り戻すこと、商品を左手と右手との間で交換すること、在庫商品を行為者のヌックに入れることが含まれる。ヌックとは、左手及び右手以外の在庫商品を保持することができる行為者の身体上の位置を指す。ヌックの幾つかの例は、在庫商品を、前腕と上腕との間で挟み込むこと、前腕と胸との間で挟み込むこと、首と肩との間で挟み込むことが含まれる。

【０２１２】

カメラ１１４は、トレーニング中に上述した全ての行為のビデオを記録する。ビデオは精査され、全ての画像フレームはタイムスタンプ及び実行された行為を示すラベルが付される。これらのラベルは、それぞれの画像フレームに対する行為ラベルと呼ばれる。画像フレームはプロダクションまたは推論のために、上述したように、ＷｈａｔＣＮＮ１５０６までのマルチＣＮＮパイプラインを介して処理される。次に、関連付けられた行為ラベ

ルに沿ったＷｈａｔＣＮＮの出力を、グラウンドトゥルースとして作用する行為ラベルとともに使用して、ＷｈｅｎＣＮＮ１５０８をトレーニングする。ＷｈａｔＣＮＮ１５０６のトレーニングについて上述したように、余弦アニーリングスケジュールを有する確率的勾配降下法（ＳＧＤ）がトレーニングのために使用される。

【０２１３】

画像増強（ＷｈａｔＣＮＮのトレーニングに使用される）に加えて、時間増強は、ＷｈｅｎＣＮＮのトレーニング中の画像フレームにも適用される。幾つかの例は、ミラーリング、ガウスノイズの追加、左手及び右手に関連するロジットの交換、時間の短縮、画像フレームをドロップすることによる時系列の短縮、フレームを複製することによる時系列の延長、及びＷｈｅｎＣＮＮのための入力を生成する基礎となるモデルにおけるスポッティ性をシミュレーションするための時系列におけるデータポイントのドロップを含む。ミラーリングは時系列及びそれぞれのラベルを反転させることを含み、例えば、置く行為は、反転されると取る行為になる。

［背景画像処理を使用した在庫イベントの予測］

【０２１４】

図２６～図２８Ｂを用いて、実空間のエリアにおける被写体による変化を追跡するシステム及び各種実施態様について説明する。

［システム・アーキテクチャ］

【０２１５】

図２６は、本実施態様に係るシステムの高レベル概略図である。図２６はアーキテクチャ図であるため、説明の明確性を向上させるために、特定の詳細は省略されている。

【０２１６】

図２６に示すシステムは、複数のカメラ１１４から画像フレームを受信する。上述のように、一実施形態では、カメラ１１４が、画像が同時に、または時間的に近く、かつ同じ画像キャプチャレートで取得されるように、互いに時間的に同期させることができる。同時にまたは時間的に近い実空間のエリアをカバーする全てのカメラにおいて取得された画像は、同期された画像が実空間において固定された位置を有する被写体のある時点での様々な光景を表すものとして処理エンジンにおいて識別されることができるという意味で同期される。

【０２１７】

一実施形態では、カメラ１１４がショッピングストア（スーパーマーケットなど）に設置され、重なり合う視野を有するカメラのセット（２つ以上）が各通路の上に配置されて、店舗内の実空間の画像を取得する。実空間には「ｎ」台のカメラがある。各カメラは、それぞれの視野に対応する実空間の画像シーケンスを生成する。

【０２１８】

被写体識別サブシステム２６０２（第１の画像プロセッサとも呼ばれる）は、カメラ１１４から受け取った画像フレームを処理して、実空間内の被写体を識別し追跡する。第１の画像プロセッサは、被写体画像認識エンジンを含む。被写体画像認識エンジンは、複数のカメラから対応する画像シーケンスを受け取り、画像を処理して、対応する画像シーケンス内の画像に表される被写体を識別する。一実施形態では、システムが多関節被写体を識別し追跡するための、上述したようなカメラ毎の画像認識エンジンを含む。個人毎に１つの「関節」のみが認識され追跡される例、または空間及び時間にわたる他の特徴または他のタイプの画像データが、処理されている実空間内の被写体を認識し追跡するために利用される例を含む、代替の画像認識エンジンを使用することができる。

【０２１９】

「意味的差分抽出」サブシステム２６０４（第２の画像プロセッサとも呼ばれる）は背景画像認識エンジンを含み、複数のカメラから対応する画像シーケンスを受信し、例えば、背景（すなわち棚のような在庫陳列構造）内の意味的に重要な差異が、各カメラからの画像において経時的に在庫商品を置くこと及び取ることに関連するときに、当該差異を認識する。第２の画像プロセッサは、被写体識別サブシステム２６０２の出力と、カメラ１

１４からの画像フレームとを入力として受け取る。第２の画像プロセッサは、前景内の識別された被写体をマスクして、マスクされた画像を生成する。マスクされた画像は、前景被写体に対応する有界ボックスを背景画像データに置き換えることによって生成される。これに続いて、背景画像認識エンジンはマスクされた画像を処理して、対応する画像シーケンス内の画像に表される背景変化を識別し且つ分類する。一実施形態では、背景画像認識エンジンが畳み込みニューラル・ネットワークを含む。

【０２２０】

最後に、第２の画像プロセッサは、識別された背景変化を処理して、識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの第１の検出セットを行う。第１の検出セットは、在庫商品を置くこと及び取ることの背景検出とも呼ばれる。ショッピングストアの例では、第１の検出が店舗の顧客または従業員によって棚から取られた、または棚に置かれた在庫商品を識別する。意味的差分抽出サブシステムは、識別された背景変化を識別された被写体に関連付けるロジックを含む。

【０２２１】

領域提案サブシステム２６０６（第３の画像プロセッサとも呼ばれる）は前景画像認識エンジンを含み、複数のカメラ１１４から対応する画像シーケンスを受信し、例えば、前景（すなわち、買物客、買物客の手、及び在庫商品）内の意味的に重要な物体が、各カメラからの画像において経時的に、在庫商品を置くこと及び取ることに関連するときに、当該物体を認識する。サブシステム２６０６はまた、被写体識別サブシステム２６０２の出力を受信する。第３の画像プロセッサは、カメラ１１４からの画像シーケンスを処理して、対応する画像シーケンス内の画像に表される前景変化を識別し且つ分類する。第３の画像プロセッサは、識別された前景変化を処理して、識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの第２の検出セットを行う。第２の検出セットは、在庫商品を置くこと及び取ることの前景検出とも呼ばれる。ショッピングストアの例では、第２の検出セットが、在庫商品を取ることと、店舗の顧客及び従業員による在庫陳列構造上に在庫商品を置くこととを識別する。

【０２２２】

図２６に記載されるシステムは、第１及び第２の検出セットを処理して、識別された被写体についての在庫商品のリストを含むログ・データ構造を生成するための選択ロジック・コンポーネント２６０８を含む。実空間内の置くこと及び取ることのために、選択ロジック２６０８は、意味的差分抽出サブシステム２６０４または領域提案サブシステム２６０６の何れかからの出力を選択する。一実施形態では、選択ロジック２６０８が、第１の検出セットについて意味的差分抽出サブシステムによって生成された信頼度スコアと、第２の検出セットについて領域提案サブシステムによって生成された信頼度スコアとを使用して、選択を行う。特定の検出に対するより高い信頼度スコアを有するサブシステムの出力が選択され、識別された前景被写体に関連付けられた在庫商品のリストを含むログ・データ構造１５１０（ショッピングカート・データ構造とも呼ばれる）を生成するために使用される。

［サブシステム・コンポーネント］

【０２２３】

図２７は、実空間のエリア内の被写体による変化を追跡するためのシステムを実施するサブシステム・コンポーネントを示す。システムは、実空間における対応する視野のそれぞれの画像シーケンスを生成する複数のカメラ１１４を備える。各カメラの視野は上述したように、複数のカメラのうちの少なくとも１つの他のカメラの視野と重なる。一実施形態では、複数のカメラ１１４によって生成された画像に対応する画像フレームのシーケンスがカメラ１１４毎に循環バッファ１５０２（リング・バッファとも呼ばれる）に格納される。各画像フレームは、画像データと共に、タイムスタンプ、カメラの識別情報（「カメラＩＤ」と略される）、及びフレーム識別情報（「フレームＩＤ」と略される）を有する。循環バッファ１５０２は、それぞれのカメラ１１４からの連続的にタイムスタンプされた画像フレームのセットを格納する。一実施形態では、カメラ１１４が同期された画像

シーケンスを生成するように構成される。

【０２２４】

１つの好ましい実施態様では、同じカメラ及び同じ画像シーケンスが前景及び背景画像プロセッサの両方によって使用される。その結果、同じ入力データを用いて、在庫商品を置くこと及び取ることの冗長な検出が行われ、結果として得られるデータにおいて高い信頼度と高い精度を可能にする。

【０２２５】

被写体識別サブシステム２６０２（第１の画像プロセッサとも呼ばれる）は、複数のカメラ１１４から対応する画像シーケンスを受信する被写体画像認識エンジンを含む。被写体画像認識エンジンは、画像を処理して、対応する画像シーケンス内の画像に表される被写体を識別する。一実施形態では、被写体画像認識エンジンが関節ＣＮＮ１１２ａ～１１２ｎと呼ばれる畳み込みニューラル・ネットワーク（ＣＮＮ）として実装される。重なり合う視野を有するカメラに対応する関節ＣＮＮ１１２ａ～１１２ｎの出力は、各カメラの２Ｄ画像座標から実空間の３Ｄ座標に関節の位置をマッピングするために組み合わされる。ｊが１～ｘに等しい被写体（ｊ）毎の関節データ構造８００は、各画像について実空間及び２Ｄ空間における被写体（ｊ）の関節の位置を識別する。被写体データ構造８００の幾つかの詳細を図８に示す。

【０２２６】

背景画像格納装置２７０４は、意味的差分抽出サブシステム２６０４において、カメラ１１４からの対応する画像シーケンスのためのマスクされた画像（前景被写体がマスクによって除去された背景画像とも呼ばれる）を記憶する。背景画像格納装置２７０４は、背景バッファとも呼ばれる。一実施形態では、マスクされた画像のサイズが循環バッファ１５０２内の画像フレームのサイズと同じである。一実施形態では、マスクされた画像が、カメラ当たりの画像フレームのシーケンス内の各画像フレームに対応する背景画像格納装置２７０４に格納される。

【０２２７】

意味的差分抽出サブシステム２６０４（または第２の画像プロセッサ）は、カメラからの対応する画像シーケンス内の画像に表される前景被写体のマスクを生成するマスク生成器２７２４を含む。一実施形態では、１つのマスク生成器がカメラ毎に画像シーケンスを処理する。ショッピングストアの例では、前景被写体が、販売用の商品を含む背景棚の前の顧客または店舗の従業員である。

【０２２８】

一実施形態では、関節データ構造８００及び循環バッファ１５０２からの画像フレームがマスク生成器２７２４への入力として与えられる。関節データ構造は、各画像フレームにおける前景被写体の位置を識別する。マスク生成器２７２４は、画像フレーム内で識別された前景被写体毎に有界ボックスを生成する。斯かる実施形態では、マスク生成器２７２４が、２Ｄ画像フレーム内の関節位置のｘ座標及びｙ座標の値を使用して、有界ボックスの４つの境界を決定する。ｘの最小値（被写体の関節のすべてのｘ値からの）は、被写体の有界ボックスの左側垂直境界を定義する。ｙの最小値（被写体に対する関節の全てのｙ値からの）は、有界ボックスの下側水平境界を定義する。同様に、ｘ座標及びｙ座標の最大値は、有界ボックスの右側垂直境界及び上側水平境界を識別する。第２の実施形態では、マスク生成器２７２４が畳み込みニューラル・ネットワークベースの人物検出及び位置特定アルゴリズムを使用して、前景被写体の有界ボックスを生成する。斯かる実施形態では、マスク生成器２７２４が前景被写体のための有界ボックスを生成するために関節データ構造８００を使用しない。

【０２２９】

意味的差分抽出サブシステム２６０４（または第２の画像プロセッサ）は、識別された被写体を表す前景画像データを、対応する画像シーケンスに対する背景画像からの背景画像データで置き換えるための、画像シーケンス内の画像を処理するマスクロジックを含み、処理用の新しい背景画像となるマスクされた画像を提供する。循環バッファがカメラ１

１４から画像フレームを受け取ると、マスクロジックは、画像マスクによって定義された前景画像データを背景画像データで置き換えるために、画像シーケンス内の画像を処理する。背景画像データは、対応するマスクされた画像を生成するために、対応する画像シーケンスの背景画像から取得される。

【０２３０】

ショッピングストアの例を考える。最初に時間ｔ＝０において、店舗内に顧客がいない場合、背景画像格納装置２７０４内の背景画像は、カメラ当たりの画像シーケンス内の対応する画像フレームと同じである。次に、時間ｔ＝１において、顧客が棚の前を移動して棚内の商品を購入する場合を考える。マスク生成器２７２４は、顧客の有界ボックスを作成し、それをマスクロジック・コンポーネント２７０２に送る。マスクロジック・コンポーネント２７０２は、有界ボックス内のｔ＝１における画像フレーム内のピクセルを、ｔ＝０における背景画像フレーム内の対応するピクセルで置き換える。この結果、循環バッファ１５０２内のｔ＝１における画像フレームに対応するｔ＝１におけるマスクされた画像が得られる。マスクされた画像は、t＝０で背景画像フレームからのピクセルによって置き換えられる前景被写体（または顧客）のピクセルを含まない。ｔ＝１におけるマスクされた画像は、背景画像格納装置２７０４に格納され、対応するカメラからの画像シーケンス内のｔ＝２における次の画像フレームに対する背景画像として作用する。

【０２３１】

一実施形態では、マスクロジック・コンポーネント２７０２が、ピクセルによる平均化または加算などによって、画像シーケンス内のＮ個のマスクされた画像のセットを組み合わせて、各カメラのファクタ化画像のシーケンスを生成する。斯かる実施形態では、第２の画像プロセッサが、ファクタ化画像のシーケンスを処理することによって背景変化を識別し且つ分類する。ファクタ化画像は、例えば、カメラ当たりのマスクされた画像シーケンスにおけるＮ個のマスクされた画像内のピクセルの平均値をとることによって生成することができる。一実施形態では、Ｎの値がカメラ１１４のフレームレートに等しく、例えば、フレームレートが３０ＦＰＳ(フレーム／秒）である場合、Ｎの値は３０である。斯かる実施形態では、１秒の期間に対してマスクされた画像がファクタ化画像を生成するために組み合わされる。ピクセル値の平均をとることにより、実空間のエリアでのセンサノイズ及び明度変化によるピクセル変動が最小限に抑えられる。

【０２３２】

第２の画像プロセッサはファクタ化画像のシーケンスを処理することによって、背景変化を識別し且つ分類する。ファクタ化画像のシーケンス内のファクタ化画像は、ビットマスク計算器２７１０によって、同じカメラに対する先行するファクタ化画像と比較される。ファクタ化画像２７０６のペアは、２つのファクタ化画像の対応するピクセルの変化を識別するビットマスクを生成するために、ビットマスク計算器２７１０への入力として与えられる。ビットマスクは、対応するピクセル（現在及び前のファクタ化画像）のＲＧＢ（赤、緑及び青チャネル）値間の差が「差閾値」よりも大きいピクセル位置に１を有する。差閾値の値は調整可能である。一実施形態では、差閾値の値は０．１に設定される。

【０２３３】

ビットマスクと、カメラ当たりのファクタ化画像のシーケンスからのファクタ化画像のペア（現在及び前）は、背景画像認識エンジンへの入力として与えられる。一実施形態では、背景画像認識エンジンが畳み込みニューラル・ネットワークを含み、変化ＣＮＮ２７１４ａ～２７１４ｎと呼ばれる。単一の変化ＣＮＮは、カメラ毎にファクタ化画像のシーケンスを処理する。別の実施形態では、対応する画像シーケンスからのマスクされた画像は結合されない。ビットマスクは、マスクされた画像のペアから計算される。この実施形態では、マスクされた画像とビットマスクのペアが次に、変化ＣＮＮへの入力として与えられる。

【０２３４】

この例での変化ＣＮＮモデルへの入力は、ファクタ化画像毎に３つの画像チャンネル（赤、緑、青）とビットマスクの１つのチャンネルを含む７チャンネルから構成されている

。変化ＣＮＮは、複数の畳み込み層と、１または複数の全結合（ＦＣ）層とを含む。一実施形態では、変化ＣＮＮが、図５に示す関節ＣＮＮ１１２ａ～１１２ｎと同じ数の畳み込み層及びＦＣ層を含む。

【０２３５】

背景画像認識エンジン（変化ＣＮＮ ２７１４ａ－２７１４ｎ）は、ファクタ化画像の変化を識別し且つ分類し、対応する画像シーケンスに対して変化データ構造を生成する。変化データ構造は、識別された背景変化のマスクされた画像内の座標、識別された背景変化の在庫商品被写体の識別子、及び識別された背景変化の分類を含む。変化データ構造における識別された背景変化の分類は、識別された在庫商品が背景画像に対して追加されたか除去されたかを分類する。

【０２３６】

複数の商品が１つまたは複数の被写体によって同時に棚上で取られ、または置かれ得るので、変化ＣＮＮは出力位置毎に数「Ｂ」の重複有界ボックス予測を生成する。有界ボックス予測はファクタ化画像の変化に対応する。ショッピングストアが固有のＳＫＵによって識別される数「Ｃ」の固有の在庫商品を有すると考える。変化ＣＮＮは、変化の在庫商品被写体のＳＫＵを予測する。最後に、変化ＣＮＮは識別された商品が棚から取られるか、または棚に置かれるかを示す、出力内のすべての位置（ピクセル）についての変化（または在庫イベントタイプ）を識別する。変化ＣＮＮからの出力の上記３つの部分は式「５×Ｂ＋Ｃ＋１」によって記述される。各有界ボックス「Ｂ」予測が５つの数字を含むので、「Ｂ」は５で乗算される。これらの５つの数字は、有界ボックスの中心の「ｘ」及び「ｙ」座標、有界ボックスの幅及び高さを表す。５番目の数字は有界ボックスの予測のための変化ＣＮＮモデルの信頼度スコアを表す。「Ｂ」は変化ＣＮＮモデルの性能を改善するために調整可能なハイパー・パラメータである。一実施形態では、「Ｂ」の値が４に等しい。変化ＣＮＮからの出力の幅及び高さ（ピクセル単位）がそれぞれ、Ｗ及びＨによって表されると考える。変化ＣＮＮの出力は「Ｗ×Ｈ×（５×Ｂ＋Ｃ＋１）」として表される。有界ボックス出力モデルは、論文「ＹＯＬＯ９０００：Ｂｅｔｔｅｒ、Ｆａｓｔｅｒ、Ｓｔｒｏｎｇｅｒ」（２０１６年１２月２５日発行）においてＲｅｄｍｏｎ及びＦａｒｈａｄｉによって提案された物体検出システムに基づく。この論文はhttps://arxiv.org/pdf/1612.08242.pdfで入手可能である。

【０２３７】

重なり合う視野を有するカメラからの画像シーケンスに対応する変化ＣＮＮ２７１４ａ～２７１４ｎの出力は、調整ロジック・コンポーネント２７１８によって結合される。調整ロジック・コンポーネントは、重なり合う視野を有するカメラのセットからの変化データ構造を処理して、実空間内での識別された背景変化の位置を確認する。調整ロジック・コンポーネント２７１８は重なり合う視野を有する複数のカメラから、同じＳＫＵ及び同じ在庫イベントタイプ（取るまたは置く）を有する在庫商品を表す有界ボックスを選択する。次いで、選択された有界ボックスは３Ｄ実空間における在庫商品の位置を識別するために、上述の三角測量技法を使用して３Ｄ実空間において三角測量される。実空間における棚の位置は、３Ｄ実空間における在庫商品の三角測量された位置と比較される。誤検知予測は廃棄される。例えば、有界ボックスの三角測量された位置が実空間内の棚の位置にマッピングされない場合、出力は破棄される。棚にマップする３Ｄ実空間内の有界ボックスの三角測量された位置は、在庫イベントの真の予測と考えられる。

【０２３８】

一実施形態では、第２の画像プロセッサによって生成された変化データ構造における識別された背景変化の分類が、識別された在庫商品が背景画像に対して追加されたか除去されたかを分類する。別の実施形態では、変化データ構造における識別された背景変化の分類が、識別された在庫商品が背景画像に対して追加されたか除去されたかを示し、システムは背景変化を識別された被写体に関連付けるロジックを含む。システムは、識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの検出を行う。

【０２３９】
ログ生成器２７２０は、変化の真の予測によって識別された変化を、変化の位置付近の識別された被写体に関連付けるためのロジックを実施する。関節識別エンジンを利用して被写体を識別する実施形態では、ログ生成器２７２０が関節データ構造８００を使用して３Ｄ実空間内の被写体の手関節の位置を決定する。手関節位置が、変化時の変化の位置までの閾値距離内にある被写体が識別される。ログ生成器は、変化を識別された被写体に関連付ける。
【０２４０】
一実施形態では、上述のように、Ｎ個のマスクされた画像が組み合わされてファクタ化画像が生成され、次いで、ファクタ化画像が変化ＣＮＮへの入力として与えられる。Ｎはカメラ１１４のフレームレート（フレーム／秒）に等しいと考える。従って、斯かる実施形態では、１秒の期間中の被写体の手の位置を変化の位置と比較して、変化を識別された被写体に関連付ける。２つ以上の被写体の手関節位置が変化の位置までの閾値距離内にある場合、被写体との変化の関連付けは、前景画像処理サブシステム２６０６の出力に対して保留される。
【０２４１】
前景画像処理（領域提案）サブシステム２６０６（第３の画像プロセッサとも呼ばれる）は、複数のカメラからの画像シーケンスから画像を受信する前景画像認識エンジンを含む。第３の画像プロセッサは、対応する画像シーケンス内の画像に表される前景変化を識別し且つ分類するロジックを含む。領域提案サブシステム２６０６は、識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの第２の検出セットを生成する。図２７に示すように、サブシステム２６０６は、有界ボックス生成器１５０４、ＷｈａｔＣＮＮ１５０６、及びＷｈｅｎＣＮＮ１５０８を含む。循環バッファ１５０２からのカメラ当たりの関節データ構造８００及び画像フレームは、有界ボックス生成器１５０４への入力として与えられる。有界ボック生成器１５０４、ＷｈａｔＣＮＮ１５０６、及びＷｈｅｎＣＮＮ１５０８の詳細は、以前に提示されている。
【０２４２】
図２７に記載されたシステムは、識別された被写体に対する在庫商品のリストを含むログ・データ構造を生成するために、第１及び第２の検出セットを処理する選択ロジックを含む。識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの第１の検出セットは、ログ生成器２７２０によって生成される。第１の検出セットは、上述したように、第２の画像プロセッサの出力及び関節データ構造８００を使用して決定される。識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことが、第３の画像処理装置の出力を用いて決定される。各真の在庫イベント（取るまたは置く）について、選択ロジック・コントローラ２６０８は、第２の画像プロセッサ（意味的差分抽出サブシステム２６０４）または第３の画像プロセッサ（領域提案サブシステム２６０６）の何れかからの出力を選択する。一実施形態では、選択ロジックが、その在庫イベントの予測のために、より高い信頼度スコアを有する画像プロセッサからの出力を選択する。
［背景画像意味的差分抽出の処理フロー］
【０２４３】
図２８Ａ及び図２８Ｂは、実空間のエリア内の被写体による変化を追跡するために意味的差分抽出サブシステム２６０４によって実行される詳細なステップを示す。ショッピングストアの例では、被写体が棚と他の空きスペースとの間の通路内で店舗内を移動する顧客及び店舗の従業員である。処理はステップ２８０２で開始する。上述のように、カメラ１１４は、被写体を識別するためにカメラからの画像シーケンスが処理される前に較正される。カメラ較正の詳細は、上述されている。重なり合う視野を有するカメラ１１４は、被写体が存在する実空間の画像を取得する。一実施形態では、カメラが毎秒Ｎフレームの速度で同期された画像シーケン

スは、ステップ２８０４において、カメラ毎にそれぞれの循環バッファ１５０２に格納される。循環バッファ（リング・バッファとも呼ばれる）は、スライドする時間帯に画像シーケンスを格納する。背景画像格納装置２７０４は、前景被写体のないカメラ当たりの画像フレームのシーケンス内の初期画像フレームで初期化される（ステップ２８０６）。

【０２４４】

被写体が棚の前を移動することにつれて、被写体当たりの有界ボックスが上述のように、それらの対応する関節データ構造８００を使用して生成される（ステップ２８０８）。ステップ２８１０では、画像フレーム当たりの有界ボックス内のピクセルを、背景画像格納装置２７０４からの背景画像からの同じ位置のピクセルで置き換えることによって、マスクされた画像が作成される。カメラ毎の画像のシーケンス内の各画像に対応するマスクされた画像は、背景画像格納装置２７０４に格納される。ｉ番目のマスクされた画像は、カメラ当たりの画像フレームのシーケンス内の次の（ｉ＋１）画像フレーム内のピクセルを置換するための背景画像として使用される。

【０２４５】

ステップ２８１２において、Ｎ個のマスクされた画像が組み合わされて、ファクタ化画像が生成される。ステップ２８１４では、ファクタ化画像のペアのピクセル値を比較することによって、差異ヒートマップが生成される。一実施形態では２つのファクタ化画像（ｆｉ１及びｆｉ２）の２Ｄ空間内の位置（ｘ，ｙ）におけるピクセル間の差は以下の式１に示すように計算される：

$$\sqrt{\begin{array}{c}((fi1[x,y][red]-fi2[x,y][red])^2+(fi1[x,y][green]-fi2[x,y][green])^2\\ +(fi1[x,y][blue]-fi2[x,y][blue])^2)\end{array}} \quad (1)$$

【０２４６】

２Ｄ空間内の同じｘ及びｙ位置におけるピクセル間の差は式に示されるように、赤、緑及び青（ＲＧＢ）チャネルのそれぞれの強度値を使用して決定される。上記の式は、２つのファクタ化画像における対応するピクセル間の差（ユークリッドノルムとも呼ばれる）の大きさを与える。

【０２４７】

差異ヒートマップは、実空間のエリアにおけるセンサノイズ及び明度変化によるノイズを含み得る。図２８Ｂでは、ステップ２８１６で、差異ヒートマップのためのビットマスクが生成される。意味的に重要な変化は、ビットマスク内の１のクラスタによって識別される。これらのクラスタは、棚から取られた、または棚に置かれた在庫商品を識別する変化に対応する。しかしながら、差異ヒートマップのノイズは、ビットマスクにランダムな１を導入する可能性がある。更に、複数の変化（複数の商品が棚から取り出されるか、または棚に置かれる）は、１の重なり合うクラスタを導入し得る。処理フローの次のステップ（２８１８）では、画像形態操作がビットマスクに適用される。画像形態操作はノイズ（望ましくない１）を除去し、また、１の重なり合うクラスタを分離しようと試みる。この結果、意味的に重要な変更に対応する１のクラスタを含む、よりクリーンなビットマスクが得られる。

【０２４８】

形態的操作には２つの入力が与えられる。第１の入力はビットマスクであり、第２の入力は構造化要素またはカーネルと呼ばれる。２つの基本的な形態的操作は、「収縮」及び「膨張」である。カーネルは、様々なサイズの矩形行列に配置された１からなる。異なる形状（例えば、円形、楕円形、または十字形）のカーネルは、行列内の特定の位置に０を加えることによって生成される。異なる形状のカーネルがビットマスクをクリーニングする際に所望の結果を達成するために、画像形態操作に使用される。収縮操作では、カーネルはビットマスク上をスライド（または移動）する。カーネルの下のすべてのピクセルが１である場合、ビットマスク内のピクセル（１または０の何れか）は１と見なされる。そ

れ以外では、それは収縮される（０に変化する）。収縮操作は、ビットマスク内の孤立した１を除去するのに有用である。しかしながら、収縮はまた、エッジを収縮することによって１のクラスタを収縮させる。

【０２４９】

膨張操作は、収縮とは逆である。この操作では、カーネルがビットマスク上をスライドするとき、カーネルの下の少なくとも１つのピクセルの値が１である場合、カーネルによってオーバーラップされたビットマスクエリア内のすべてのピクセルの値が１に変更される。１のサイズクラスタを増大させるために、収縮後にビットマスクに膨張が適用される。ノイズが収縮において除去されるので、膨張は、ビットマスクにランダムノイズを導入しない。よりクリーンなビットマスクを達成するために、収縮操作と膨張操作との組み合わせが適用される。例えば、コンピュータ・プログラム・コードの以下の行は、ビットマスクに１の３×３フィルタを適用して、「オープン」操作を実行し、この「オープン」操作では、収縮操作と、それに続く膨張操作とを適用して、ノイズを除去し、上述のようにビットマスク内の１のクラスタのサイズを復元する。上記のコンピュータ・プログラム・コードは、リアルタイム・コンピュータ・ビジョン・アプリケーション用のプログラミング機能のＯｐｅｎＣＶ(オープンソース・コンピュータ・ビジョン）ライブラリを使用する。ライブラリは、https://opencv.org/で入手できる。

```
_bit_mask = cv2.morphologyEx(bit_mask, cv2.MORPH_OPEN, self.kernel_3x3, dst=_bit_mask)
```

【０２５０】

「クローズ」操作は、膨張操作に続いて収縮操作を適用する。これは、１のクラスタの内側の小さな穴を閉じるのに有用である。以下のプログラム・コードは、３０×３０十字形（クロス・シェープ）フィルタを使用してビットマスクにクローズ操作を適用する。

```
_bit_mask = cv2.morphologyEx(bit_mask, cv2.MORPH_CLOSE, self.kernel_30x30_cross, dst=_bit_mask)
```

【０２５１】

ビットマスク及び２つのファクタ化画像（前後）は、カメラ毎に畳み込みニューラル・ネットワーク（上記の変化ＣＮＮと呼ばれる）への入力として与えられる。変化ＣＮＮの出力は、変化データ構造である。ステップ２８２２では、重なり合う視野を有する変化ＣＮＮからの出力が前述の三角測量技法を使用して結合される。３Ｄ実空間における変化の位置は、棚の位置と一致する。在庫イベントの位置が棚上の位置にマップされる場合、変化は真のイベントと見なされる（ステップ２８２４）。マップされない場合は、変化は誤検知であり、廃棄される。真のイベントは、前景被写体に関連付けられる。ステップ２８２６において、前景被写体が識別される。一実施形態では、関節データ構造８００が変化の閾値距離内の手関節の位置を決定するために使用される。ステップ２８２８で前景被写体が識別された場合、ステップ２８３０で、その変化が識別された被写体に関連付けられる。例えば、変化の閾値距離内の複数の被写体の手の関節位置のために、ステップ２８２８で前景被写体が識別されない場合は、次に、ステップ２８３２において、領域提案サブシステムによる変化の冗長検出が選択される。処理はステップ２８３４で終了する。

［変化ＣＮＮのトレーニング］

【０２５２】

７つのチャネル入力のトレーニング・データセットが、変化ＣＮＮをトレーニングするために作成される。顧客として行為する１または複数の被写体が、ショッピングストアで買い物をするふりをすることによって、取る及び置く行為を実行する。被写体が通路を移動し、棚から在庫商品を取り、在庫商品を棚に置き戻す。取る行為及び置く行為を実行する行為者の画像は、循環バッファ１５０２に収集される。画像は上述のように、ファクタ化画像を生成するために処理される。２つのファクタ化画像間の変化を視覚的に識別するために、ファクタ化画像２７０６のペア及びビットマスク計算器２７１０によって出力された対応するビットマスクが手動で精査される。変化を有するファクタ化画像については、有界ボックスが変化の周りに手動で描かれる。これは、ビットマスクの変化に対応する

１のクラスタを含む最小の有界ボックスである。変化における在庫商品のＳＫＵ数が識別され、有界ボックスと共に画像のラベルに含まれる。在庫商品の取ることまたは置くことを識別するイベントタイプも、有界ボックスのラベルに含まれる。従って、各有界ボックスのラベルは、ファクタ化画像上のその位置、商品のＳＫＵ、及びイベントタイプを識別する。ファクタ化画像は、２つ以上の有界ボックスを有することができる。上記の処理は、トレーニング・データセット内の全ての収集されたファクタ化画像における全ての変化について繰り返される。１対のファクタ化画像はビットマスクと共に、変化ＣＮＮへの７チャネル入力を形成する。

【０２５３】

変化ＣＮＮのトレーニング中に、順方向パス及び逆方向伝播が実行される。順方向パスでは、変化ＣＮＮが、トレーニング・データセット内の画像の対応するシーケンス内のファクタ化画像内に表される背景変化を識別し且つ分類する。変化ＣＮＮは、識別さてた背景変化を処理し、識別された被写体による在庫商品を取ることと、識別された被写体による在庫陳列構造上に在庫商品を置くことの第１の検出セットを生成する。逆方向伝播の間、変化ＣＮＮの出力はトレーニング・データセットのラベルに示されるように、グラウンドトゥルースと比較される。１または複数のコスト関数に対する勾配が計算される。次いで、勾配は、畳み込みニューラル・ネットワーク（ＣＮＮ）及び全結合（ＦＣ）ニューラル・ネットワークに伝播され、その結果、予測誤差が低減され、出力がグランドトゥルースに近づく。一実施形態では、ソフトマックス関数及びクロスエントロピー損失関数が、出力のクラス予測部分に対する変化ＣＮＮのトレーニングのために使用される。出力のクラス予測部分は、在庫商品のＳＫＵ識別子及びイベントタイプ、すなわち取ることまたは置くことを含む。

【０２５４】

第２の損失関数は、有界ボックスの予測のために変化ＣＮＮをトレーニングするために使用される。この損失関数は、予測されたボックスとグラウンドトゥルース・ボックスとの間の共通集合／和集合（ＩＯＵ）を計算する。変化ＣＮＮによって予測された有界ボックスと真の有界ボックスラベルとの共通集合の面積が、同じ有界ボックスの和集合の面積によって割り算される。ＩＯＵの値は、予測ボックスとグラウンドトゥルース・ボックスとの間のオーバーラップが大きい場合に高い。２つ以上の予測された有界ボックスがグラウンドトゥルース有界ボックスとオーバーラップする場合、最も高いＩＯＵ値を有するものが選択され、損失関数が計算される。損失関数の詳細は、２０１６年５月９日に発行されたＲｅｄｍｏｎらの論文「Ｙｏｕ Ｏｎｌｙ Ｌｏｏｋ Ｏｎｃｅ：Ｕｎｉｆｉｅｄ、Ｒｅａｌ-Ｔｉｍｅ Ｏｂｊｅｃｔ Ｄｅｔｅｃｔｉｏｎ」に記載されている。この論文はhttps://arxiv.org/pdf/1506.02640.pdfで入手可能である。

［特定の実施態様］

【０２５５】

様々な実施形態において、上述の実空間のエリア内において被写体による在庫商品を置くこと及び取ることを追跡するためのシステムは、また、以下の特徴の１または複数を含む。

［１．領域提案］

【０２５６】

領域提案は、人物をカバーする全ての異なるカメラからの手の位置のフレーム画像である。領域提案は、システム内のすべてのカメラによって生成される。それは、店舗の商品を持っている手だけでなく、空の手も含む。

［１．１　ＷｈａｔＣＮＮモデル］

【０２５７】

領域提案は、深層学習アルゴリズムを使用して画像分類への入力として使用することができる。この分類エンジンは、「ＷｈａｔＣＮＮ」モデルと呼ばれる。それは、インハンド分類モデルである。それは手の中の物を分類する。インハンド画像分類は、物体の部分が手によって遮蔽されていても、動作することができる。より小さい商品は、手で９０％

まで遮閉することができる。ＷｈａｔＣＮＮモデルによる画像分析のための領域は、計算コストが高いため、幾つかの実施形態では意図的に小さく保たれる。各カメラは、専用のＧＰＵを有することができる。これは、すべてのカメラからのすべての手の画像について、すべてのフレームについて実行される。ＷｈａｔＣＮＮモデルによる上記の画像分析に加えて、信頼度重みもその画像（１つのカメラ、１つの時点）に割り当てられる。分類アルゴリズムは、在庫管理単位（ＳＫＵ）のリスト全体にわたるロジットを出力して、ｎ個の商品について店舗の製品及びサービス識別コードリストを生成し、空の手（ｎ＋１）について１つの追加を生成する。

【０２５８】

シーン処理は、キー値辞書を各ビデオに送ることによって、その結果を各ビデオ処理に送り返す。ここで、キーは固有関節ＩＤであり、値は、関節が関連付けられた固有の個人ＩＤである。関節に関連する人物が見つからなかった場合、それは辞書に含まれない。

【０２５９】

各ビデオ処理はシーン処理からキー値辞書を受け取り、フレーム番号を返された辞書にマッピングするリング・バッファにそれを格納する。

【０２６０】

返されたキー値辞書を使用して、ビデオは、既知の人々に関連付けられた手の近くにある時点毎の画像のサブセットを選択する。これらの領域は、ｎｕｍｐｙスライスである。また、前景マスクと関節ＣＮＮの生の出力特徴配列の周りに同様のスライスを取る。これらの組み合わされた領域は一緒に連結されて単一の多次元ｎｕｍｐｙ配列になり、領域が関連付けられている個人ＩＤと、その領域が人物からのどの手から来たかと同様に、ｎｕｍｐｙ配列を保持するデータ構造に格納される。

【０２６１】

次に、全ての提案された領域がＦＩＦＯキューに供給される。このキューは領域を取り込み、それらのｎｕｍｐｙ配列をＧＰＵ上のメモリにプッシュする。

【０２６２】

配列がＧＰＵに到着すると、それらは、ＷｈａｔＣＮＮと呼ばれる、分類専用のＣＮＮに供給される。このＣＮＮの出力は、Ｎ＋１の大きさのフロートの平坦な配列であり、ここで、Ｎは店舗内の固有のＳＫＵの数であり、最終クラスは、空クラス、すなわち空の手を表す。この配列のフロートは、ロジットと呼ばれる。

【０２６３】

ＷｈａｔＣＮＮの結果は、領域データ構造に記憶される。

【０２６４】

その後、ある時点での全ての領域が、各ビデオ処理からシーン処理に送り返される。

【０２６５】

シーン処理は、ある時点ですべてのビデオからすべての領域を受け取り、その結果をキー値辞書に格納し、そこでは、キーは個人ＩＤであり、値はキー値辞書であり、そこでは、キーはカメラＩＤであり、値は領域のロジットである。

【０２６６】

次に、この集約されたデータ構造は、時点毎にフレーム番号を集約された構造にマッピングするリング・バッファに格納される。

［１．２　ＷｈｅｎＣＮＮモデル］

【０２６７】

ＷｈａｔＣＮＮモデルによって処理された様々なカメラからの画像は、ある期間にわたって結合される（ある期間にわたって複数のカメラ）。このモデルへの追加の入力は、複数のカメラから三角測量された３Ｄ空間における手の位置である。このアルゴリズムへのもう１つの入力は、店舗のプラノグラムからの手の距離である。いくつかの実施形態では、プラノグラムを使用して、手が特定の商品を含む棚（例えば、チェリオボックス）に近いかどうかを識別することができる。このアルゴリズムへの別の入力は、店舗上の足の位置である。

【０２６８】
ＳＫＵを使用する物体分類に加えて、第２の分類モデルは、時系列分析を使用して、物体が棚から取り出されたか、または棚上に置かれたかを判定する。画像は、以前の画像フレームにおいて手の中にあった物体が棚に戻されたか、または、棚から取り出されたか否かの判定を行うために、ある期間にわたって分析される。
【０２６９】
１秒間（３０フレーム／秒）及び３台のカメラについて、システムは、信頼度の付加された同じ手に対して９０の分類出力を有する。この組み合わされた画像分析は、手の中の物体を正確に識別する確率を劇的に増加させる。時間分析は、個々のフレームの幾つかの非常に低い信頼度レベル出力にもかかわらず、出力の品質を改善する。このステップは例えば、８０％の精度から９５％の精度までの出力信頼度を取ることができる。
【０２７０】
このモデルはまた、棚モデルからの出力をその入力として含み、この人物がどの物体を選択したかを識別する。
【０２７１】
シーン処理は、３０以上の集約された構造が蓄積されるのを待ち、少なくとも１秒のリアルタイムを表し、次いで、集約された構造を、個人ＩＤと手のペア毎に単一の整数に縮小するために、更なる分析を実行し、ここで、整数は、店舗内のＳＫＵを表す固有ＩＤである。一時点において、この情報はキーが個人ＩＤと手のペアであり、値がＳＫＵ整数であるキー値辞書に記憶される。この辞書は、その時点においてフレーム番号を各辞書にマッピングするリング・バッファに経時的に格納される。
【０２７２】
次に、人が何かを何時取ったか、取られた物が何かを識別するために、この辞書が時間の経過と共にどのように変化するかを見て、追加の分析を実行することができる。このモデル（ＷｈｅｎＣＮＮ）は、何かが取られたか？、何かが置かれたか？というブール式の各質問に対するロジットのみならず、ＳＫＵロジットも発する。
【０２７３】
ＷｈｅｎＣＮＮの出力は、フレーム番号を、キーが個人ＩＤであり、値がＷｈｅｎＣＮＮによって発せられる拡張ロジットであるキー値辞書にマッピングするリング・バッファに格納される。
【０２７４】
次いで、ヒューリスティックスの更なる集合が、ＷｈｅｎＣＮＮ及び格納された人々の関節位置の両方の格納された結果、並びに、店舗の棚上の商品の事前に計算されたマップ上で実行される。このヒューリスティックスの集合は、取ること及び置くことの結果、商品がどこに追加されるか、またはどこから除去されるかを決定する。各取ること／置くことについて、ヒューリスティックスは、取ることまたは置くことが、棚からまたは棚へ、バスケットからまたはバスケットへ、または、人からまたは人へ、であったかどうかを決定する。出力は、ＳＫＵの指数における配列の値が個人の有するＳＫＵの数である配列として格納された、個人毎の在庫である。
【０２７５】
買物客が店舗の出口に近づくと、システムは、物品明細リストを買物客の電話に送ることができる。次に、電話はユーザの物品明細を表示し、格納されたクレジットカード情報に課金するための確認を求める。ユーザが了承すると、そのクレジットカードに課金される。システム内で知られているクレジットカードを有していない場合、クレジットカード情報を提供するように要求される。
【０２７６】
代替的に、買物客は店舗内キオスクに接近することもできる。システムは買物客が何時キオスクの近くにいるかを識別し、その買物客の物品明細を表示するためにキオスクにメッセージを送信する。キオスクは、買物客に物品明細の料金を受け入れるように求める。買物客が了承した場合、買物客は、自分のクレジットカードを通すか、または支払いのた

めに現金を投入することができる。図１６は、領域提案のためのＷｈｅｎＣＮＮモデルの図を示す。

［２．置き間違えられた商品］

【０２７７】

この特徴では、 置き間違えられた商品を、それらが人によってでたらめな棚に置き戻されたときに識別する。これにより、プラノグラムに対する足及び手の位置が不正確になるので、物体識別に問題が生じる。従って、システムは、経時的に修正されたプラノグラムを構築する。以前の時系列分析に基づいて、システムは、人が商品を棚に戻したかどうかを判定することができる。次に、物体がその棚位置から取り出されると、システムは、その手の位置に少なくとも１つの置き間違えられた商品があることを知る。それに対応して、アルゴリズムは、人がその棚から置き間違えられた商品を取り上げることができるというある程度の信頼度を有する。置き間違えられた商品が棚から取り出される場合、システムはその商品をその位置から取り去り、従って、棚は、もはやその商品を有していない。システムはまた、アプリを介して置き間違えられた商品について店員に通知し、店員がその商品をその正しい棚に移動させることができるようにすることができる。

［３．意味的差分抽出（棚モデル）］

【０２７８】

背景画像処理のための代替技術は、棚上の商品（取り除かれた、または、置かれた商品）に対する変化を識別するための背景減算アルゴリズムを含む。これは、ピクセルレベルでの変化に基づく。棚の前に人がいる場合、人の存在によるピクセル変化を考慮しないようにアルゴリズムは停止する。背景減算はノイズの多い処理である。従って、クロスカメラ分析が行われる。シェルフに「意味的に重要な」変化があることを十分な数のカメラが認める場合、システムは、棚のその部分に変化があることを記録する。

【０２７９】

次のステップは、その変化が「置く」または「取る」変化であるかどうかを確認することである。このために、第２の分類モデルの時系列分析が使用される。棚のその特定の部分に対する領域提案が生成され、深層学習アルゴリズムを通過する。これは、物体が手の中で遮蔽されないので、手の中の画像分析よりも容易である。第４の入力は、３つの典型的なＲＧＢ入力に加えてアルゴリズムに与えられる。第４のチャネルは背景情報である。棚または意味的差分抽出の出力は、第２の分類モデル（時系列分析モデル）に再び入力される。

【０２８０】

このアプローチにおける意味的差分抽出は、以下のステップを含む：

１．カメラからの画像は、同じカメラからの以前の画像と比較される。

２．２つの画像間の対応する各ピクセルは、ＲＧＢ空間におけるユークリッド距離を介して比較される。

３．ある閾値を超える距離がマーキングされ、その結果、マーキングされたばかりのピクセルの新しい画像が得られる。

４．マーキングされた画像からノイズを除去するために、画像形態フィルタの集合が使用される。

５．次に、マーキングされたピクセルの大きな集合を探索し、それらの周りに有界ボックスを形成する。

６．次に、各有界ボックスについて、２つの画像中の元のピクセルを見て、２つの画像スナップショットを得る。

７．次に、これらの２つの画像スナップショットは、画像領域が取られている商品を表すか、または置かれている商品を表すか、及び商品が何であるかを分類するようにトレーニングされたＣＮＮにプッシュされる。

［３．店舗監査］

【０２８１】

各棚の在庫は、システムによって維持される。それは、商品が顧客によって取り出され

ると更新される。任意の時点で、システムは、店舗在庫の監査報告書を生成することができる。

［４．手の中の複数の商品］

【０２８２】

複数の商品に対して異なる画像が使用される。手の中の２つの商品は、１つの商品と比較して別に扱われる。幾つかのアルゴリズムは、１つの商品のみを予測することができるが、複数の商品を予測することはできない。従って、ＣＮＮは、商品の「２つの」量のためのアルゴリズムが手の中の単一の商品とは別個に実行され得るようにトレーニングされる。

［５．データ収集システム］

【０２８３】

所定のショッピングスクリプトが、画像の良質なデータを収集するために使用される。これらの画像は、アルゴリズムのトレーニングに使用される。

［５．１　ショッピングスクリプト］

【０２８４】

データ収集は、以下のステップを含む：

１．スクリプトが自動的に生成され、どのような行為を取るべきかを人間の行為者に伝える。

２．これらの行為は、商品Ｘを取る、商品Ｘの置く、商品ＸをＹ秒間保持することを含む行為の集合から任意に抽出される。

３．これらの行為を実行しながら、行為者は所与の行為を持続させながら、可能な限り多くの方法で、自分自身を移動させ、向きを合わせる。

４．一連の行為の間、カメラの集合体は、多くの視点から行為者を記録する。

５．行為者がスクリプトを終了した後、カメラビデオは一緒に束ねられ、元のスクリプトと共に保存される。

６．スクリプトは、行為者のビデオでトレーニングする機械学習モデル（ＣＮＮなど）への入力ラベルとして機能する。

［６．製品ライン］

【０２８５】

本システム及びその一部は、以下のアプリでサポートされるレジレス・チェックアウトに使用できる。

［６．１　店舗アプリ］

【０２８６】

店舗アプリは、幾つかの主要な機能を有しており、データ分析視覚化を提供し、損失防止をサポートし、人々が店舗内のどこにいるか、及びどの商品を収集したかを小売業者に示すことによって顧客を支援するプラットフォームを提供する。従業員に対する許可レベル及びアプリアクセスは、小売業者の裁量で指示することができる。

［６．１．１　標準分析］

【０２８７】

データは、プラットフォームによって収集され、様々な方法で使用することができる。

【０２８８】

１．デリバティブデータは、店舗、店舗が提供するショッピング体験、及び、製品、環境、及び他の人々との顧客の交流に関する様々な種類の分析を実行するために使用される。

ａ．データは、店舗と顧客の交流の分析を実行するために、格納され背景で使用される。店舗アプリは、このデータの視覚化の一部を小売業者に表示する。他のデータは、データポイントが要求されるときに格納され、照会される。

２．ヒートマップ：

プラットフォームは、小売業者のフロアプラン、棚レイアウト、及び他の店舗環境を、様々な種類の活動のレベルを示すオーバーレイとともに視覚化する。

１．例：

１．人は通り過ぎるが、どんな製品も扱っていない場所のマップ。
２．製品と接触するときに、人がフロアのどこに立っているかのマップ。
３．置き間違えられた商品：
プラットフォームは、店舗のすべてのＳＫＵを追跡する。商品が間違った場所に置かれると、プラットフォームはその商品がどこにあるかを知り、ログを構築する。ある閾値で、または直ちに、店舗の従業員は、置き間違えられた商品に対し注意喚起され得る。或いは、スタッフは、店舗アプリ内の置き間違えられた商品マップにアクセスすることができる。都合の良いときに、スタッフが置き間違えられた商品を迅速に見つけ出し、修正することができる。

［６．１．２　標準アシスト］

・　店舗アプリは店舗のフロアプランを表示する。
・　それは、店舗内の各人物を表す図画を表示する。
・　タッチ、クリック、または他の手段を介して図画が選択されると、店舗の従業員に対して適切な情報が表示される。例えば、ショッピングカートの商品（収集した商品）がリストに表示される。
・　プラットフォームが、特定の商品（単数または複数）に対する、及び個人が所有（ショッピングカート）している期間に対する、所定の閾値より低い信頼度レベルを有する場合、その図画（現在はドット）は差を示す。アプリは色の変化を使用する。緑は高い信頼度を示し、黄色／オレンジは低い信頼度を示す。
・　店舗アプリを所持した店舗従業員には、より低い信頼度を通知することができる。店舗従業員は、顧客のショッピングカートが正確であることを確認することができる。
・　店舗アプリを介して、小売業者の従業員は、顧客のショッピングカート商品を調整（追加または削除）することができる。
・

［６．１．３　標準ＬＰ］

・　買物客が買物客アプリを使用している場合、買物客は単に店舗を出て、課金される。しかし、買物客が買物客アプリを使用していない場合、買物客は、彼らのショッピングカート内の商品に対する支払い用のゲストアプリを使用する必要があるだろう。
・　買物客が、店舗から出る途中でゲストアプリを迂回する場合、買物客の図画は、買物客が店舗を出る前にアプローチしなければならないことを示す。このアプリでは、色を赤色に変更する。スタッフはまた、潜在的な損失の通知を受ける。
・　店舗アプリを介して、小売業者の従業員は、顧客のショッピングカート商品を調整（追加または削除）することができる。

［６．２　非店舗アプリ］

【０２８９】
以下の分析の特徴は、プラットフォームの追加機能を表す。

［６．２．１　標準分析］

１．製品交流：
次のような製品交流の細かな内訳：
a．各製品の交流時間対転化率。
b．Ａ／Ｂ比較（色、スタイル等）。ディスプレイ上のより小さい製品の一部は、色、風味などのような複数の選択肢を有する。
・　バラの金は銀よりも取扱いが多いか？
・　青い缶は赤い缶よりも多くの交流を招くのだろうか
２．方向性インプレッション：
位置ベースのインプレッションと、買物客の注視がどこにあるかの差を知る。もし買物客が１５フィート離れた製品を２０秒間見ているならば、インプレッションは、彼らがどこにいるかが重要ではなく、彼らがどこを見ているかが重要である。
３．顧客認識：

リピータ買物客とそれに付随する電子メールアドレス（小売業者によって様々な方法で収集された）及び買物プロフィールを記憶する。
４．グループダイナミックス：
買物客が、他の誰かが製品と接触するのを何時見ているかを判定する。
・ その後、その人が製品と接触するかどうかを答える。
・ その人たちは一緒に店舗に入ったのか、或いは、他人同士だろうか？
・ 個人或いは集団が、店舗でより多くの時間を費やしているか？
５．顧客タッチバック：
顧客に対象情報、店舗後体験の提供。この特徴は、特定の慣行及び方針に応じて、それぞれの小売業者とわずかに異なった実施態様を有することができる。この特徴を採用するためには、小売業者からの統合及び／または開発が必要となる場合がある。
・ 買物客は、関心のある製品に関する通知を受領したいかどうかを尋ねられる。そのステップは、電子メールを収集する店舗の方法と統合されてもよい。
・ 店舗を出た後、顧客は、店舗で時間を費やした製品を伴う電子メールを受け取ることができる。持続時間、接触、及び視界（方向インプレッション）に対する交流閾値が決定される。閾値が満たされると、製品はそれを顧客のリストに送り、店舗を出た後すぐに顧客に送る。
【０２９０】
追加的に、または代替的に、買物客に、販売中の製品または他の特別な情報を提供した後の期間に電子メールを送ることができる。これらの製品は興味を表明した商品であるが、購入しなかった商品である。
［６．３ ゲストアプリ］
【０２９１】
買物客アプリは、店舗を出るときに自動的に人々をチェックアウトする。しかしながら、プラットフォームは、買物客が店舗を使用するために買物客アプリを有することも使用することも必要としない。
【０２９２】
買物客／個人が買物客アプリを持っていないか、または使用していないとき、買物客はキオスク（ｉＰａｄ（登録商標）／タブレットまたは他の画面）まで歩いて行くか、または予めインストールされたセルフ・チェックアウト・マシンまで歩いて行く。プラットフォームと一体化されたディスプレイは、顧客のショッピングカートを自動的に表示する。
【０２９３】
買物客は、何が表示されているかを見直す機会を有する。買物客がディスプレイ上の情報に同意する場合、買物客は、現金をマシンに投入するか（その機能がハードウェア（例えば、セルフ・チェックアウト・マシン）に装備されている場合）、または、買物客のクレジットカードまたはデビットカードを通すことができる。そして、店舗を出ることができる。
【０２９４】
買物客が、ディスプレイに同意しない場合に、タッチスクリーン、ボタン、または他の手段を介して、異議を申し立てることを選択することで、店員に通知される（店舗アプリの「店舗アシスト」を参照）。
［６．４ 買物客アプリ］
【０２９５】
アプリ、買物客アプリを使用することにより、顧客は商品と共に店舗を出ることができ、自動的に課金され、デジタルレシートが与えられる。買物客は、店舗のショッピングエリア内にいる間は常に自分のアプリを開かなければならない。プラットフォームは、買物客のデバイスに表示される固有画像を認識する。プラットフォームは、それらを買物客のアカウントに結びつけ（顧客関連付け）、買物客がアプリを開いたままにしているかどうかにかかわらず、店舗のショッピングエリア内での時間中、誰がいるかを覚えておくことができる。

【０２９６】
買物客が商品を集めると、買物客アプリは、買物客のショッピングカートに商品を表示する。買物客が望む場合、買物客は取り出した（すなわち、ショッピングカートに追加された）各商品に関する製品情報を見ることができる。製品情報は、店舗のシステムに格納されるか、またはプラットフォームに追加される。製品販売を提供すること、または価格を表示することなど、その情報を更新する能力は、小売業者が要求／購入または開発することができるオプションである。
【０２９７】
買物客が商品を下に置くと、バックエンド及び買物客アプリ上のショッピングカートから商品が取り除かれる。
【０２９８】
買物客アプリが開かれ、顧客関連付けが完了した後に閉じられると、プラットフォームは、買物客のショッピングカートを維持し、買物客が店舗を出ると、それらに正しく課金する。
【０２９９】
買物客アプリはまた、開発ロードマップに関するマッピング情報を有する。それは、顧客が、捜している商品をタイプ入力することによって情報を要求する場合、店舗内の商品をどこで見つけるべきかを顧客に伝えることができる。後日、買物客のショッピングリスト（手動で、または他のインテリジェントシステムを介してアプリに入力された）を取得し、すべての所望の商品を収集する店舗を通る最速ルートを表示する。「袋詰め傾向」などの他のフィルタを追加することができる。袋詰め傾向フィルタにより、買物客は最も速いルートをたどらず、最初に頑丈な商品を収集し、その後、より壊れやすい商品を収集することができる。
［７．顧客のタイプ］
【０３００】
メンバー顧客： 最初のタイプの顧客が、アプリを使用してシステムにログインする。顧客は画面で促され、クリックすると、システムはそれをその顧客の内部ＩＤにリンクする。顧客がアカウントを有する場合、顧客が店舗から出るときにアカウントに自動的に課金される。これは、会員制店舗である。
【０３０１】
ゲスト顧客： すべての店舗が会員権を持っているわけではない。或いは、顧客がスマートフォンやクレジットカードを持っていないこともある。このタイプの顧客はキオスクまで歩いていくだろう。キオスクは顧客が有する商品を表示し、顧客にお金を入れるように依頼する。キオスクは、顧客が購入した全ての商品について既に知っている。このタイプの顧客の場合、システムは顧客がショッピングカート内の商品に対して支払っていないかどうかを識別し、顧客がそこに到着する前に、チェッカーに未支払い商品について知らせるようにドアのチェッカーに促すことができる。システムはまた、支払いが行われていないか、システムがその商品について低い信頼度を有する１つの商品に対してプロンプトを表示することもできる。これは、予測経路探索と呼ばれる。
【０３０２】
システムは、信頼度レベルに基づいて、店舗内を歩いている顧客にカラーコード（緑色及び黄色）を割り当てる。緑色で色分けされた顧客は、システムにログインされているか、またはシステムがそれら顧客について高い信頼度を有している顧客である。黄色の色分けされた顧客は、高い信頼度で予測されない１または複数の商品を有している顧客である。店員は黄色の点を見て、それらをクリックして、問題商品を特定し、顧客まで歩いて行き、問題を修正することができる。
［８．分析］
【０３０３】
顧客が特定の棚の前でどれだけの時間を費やしたかといった、顧客に関する多くの分析情報が収集される。更に、システムは、顧客が見ている場所（システム上のインプレッシ

ョン）と、顧客が取り出して棚に戻した商品とを追跡する。このような分析は現在、電子商取引で利用可能であるが、小売店では利用可能ではない。

［９．機能モジュール］

【０３０４】

以下は、機能モジュールのリストである：

１．同期カメラを使用して、店舗内の画像の配列を取得するシステム。

２．画像中の関節を識別し、個々の人物の関節のセットを識別するシステム。

３．関節セットを使用して新しい人物を作成するシステム。

４．関節セットを使用してゴースト人物を削除するシステム。

５．関節セットを追跡することによって、経時的に個々の人物を追跡するシステム。

６．店舗内にいる各人に対して手の中の商品のＳＫＵ数を示す領域提案を生成するシステム（ＷｈａｔＣＮＮ）。

７．手の中の商品が棚上において取り出されたか、または置かれたかを示す領域提案のための取ること／置くこと分析を実行するシステム（ＷｈｅｎＣＮＮ）。

８．領域提案及び取ること／置くこと分析を用いて、１人当たりの在庫配列を生成するためのシステム（ヒューリスティックスと人物の保存された関節位置と店舗の棚上の事前に計算された商品のマップを組み合わせたＷｈｅｎＣＮＮの出力）。

９．棚上の置き間違えられた商品の位置を識別し、追跡し、更新するシステム。

１０．ピクセルベースの分析を使用して、棚上の商品に対する変化（取る／置く）を追跡するシステム。

１１．店舗の在庫監査を実施するシステム。

１２．手の中の複数の商品を識別するシステム。

１３．ショッピングスクリプトを用いて店舗から商品画像データを収集するシステム。

１４．会員顧客からのチェックアウトを実行し、集金を行うシステム。

１５．ゲスト顧客からのチェックアウトを実行し、集金を行うシステム。

１６．カート内の未払商品を特定し、損失防止を行うシステム。

１７．顧客のカート内で誤って識別された商品を店員が識別するのを支援するために、例えばカラーコードを使用して顧客を追跡するシステム。

１８．位置ベースのインプレッション、方向性インプレッション、Ａ／Ｂ分析、顧客認識、グループダイナミクス等を含む顧客ショッピング分析を生成するシステム。

１９．ショッピング分析を使用して目標顧客タッチバックを生成するシステム。

２０．様々な活動を視覚化するために店舗のヒートマップオーバーレイを生成するシステム。

【０３０５】

本明細書に記載されている技術は、レジレス・チェックアウトをサポートすることができる。店舗に行く。商品を取る。去る。

【０３０６】

レジレス・チェックアウトは、純粋なマシンビジョンと深層学習に基づくシステムである。買物客は、列に並ばず、より早くより簡単に欲しいものを得る。ＲＦＩＤタグは不要。店舗のバックエンドシステムに対する変更は不要。第三者の販売時点在庫管理システムと統合することができる。

各ビデオフィードのリアルタイム３０ ＦＰＳ分析。

最先端の構内ＧＰＵクラスタ。

買物客と彼らが交流する商品を認識する。

例示的な実施形態では、インターネットに依存しない。

マシンビジョン技術のギャップを初めて解決するために、独自のカスタムアルゴリズムを含む複数の最先端の深層学習モデル。

【０３０７】

技術と機能には以下が含まれる：

１．スタンダード・コグニションの機械学習パイプラインは、以下を解決する：

a．人物検出。
b．存在物追跡。
c．マルチカメラ人物一致。
d．手検出。
e．商品分類。
f．商品所有権決定。

【０３０８】
これらの技術を組み合わせると、以下のことができる：
１．買い物体験を通じて、すべての人々をリアルタイムで追跡する。
２．買物客が手に持っているもの、どこに立っているか、どんな商品を戻すかを知る。
３．買物客が、どの方向にどれだけ長く向いているのかを知る。
４．置き間違えられた商品を認識し、２４／７の目視商品化監査を実施する。

【０３０９】
買物客が手に持っているものとバスケットに持っているものを正確に検出することができる。

店舗の学習：

【０３１０】
特定の店舗や商品についてトレーニングされたカスタムニューラル・ネットワーク。トレーニング・データは、全ての店舗位置にわたって再利用可能である。

標準配備：

【０３１１】
天井カメラは、店舗の全エリアを二重にカバーするように設置しなければならない。典型的な通路には２～６台のカメラが必要である。

【０３１２】
構内ＧＰＵクラスタは、バックオフィス内の１つまたは２つのサーバラックに収容できる。

【０３１３】
例示的なシステムは、販売時点在庫管理システムと統合することができ、またはそれらを含むことができる。

【０３１４】
同期カメラを使用して店舗内の画像配列を取得する第１のシステム、方法、コンピュータ・プログラム製品。

【０３１５】
画像内の関節、及び個々の人物の関節のセットを識別する第２のシステム、方法、及びコンピュータ・プログラム製品。

【０３１６】
関節のセットを使用して新しい人物を作成する第３のシステム、方法、及びコンピュータ・プログラム製品。

【０３１７】
関節のセットを使用してゴースト人物を削除する第４のシステム、方法、及びコンピュータ・プログラム製品。

【０３１８】
関節のセットを追跡することにより経時的に個々の人物を追跡する第５のシステム、方法、及びコンピュータ・プログラム製品。

【０３１９】
手の中の商品のＳＫＵ数を示す、店舗内にいる各人物のための領域提案を生成する第６のシステム、方法、及びコンピュータ・プログラム製品（ＷｈａｔＣＮＮ）。

【０３２０】
手の中の商品が棚上に取り出されたか、または置かれたかを示す領域提案のために取る／置く分析を実行する第７のシステム、方法、及びコンピュータ・プログラム製品（Ｗｈ

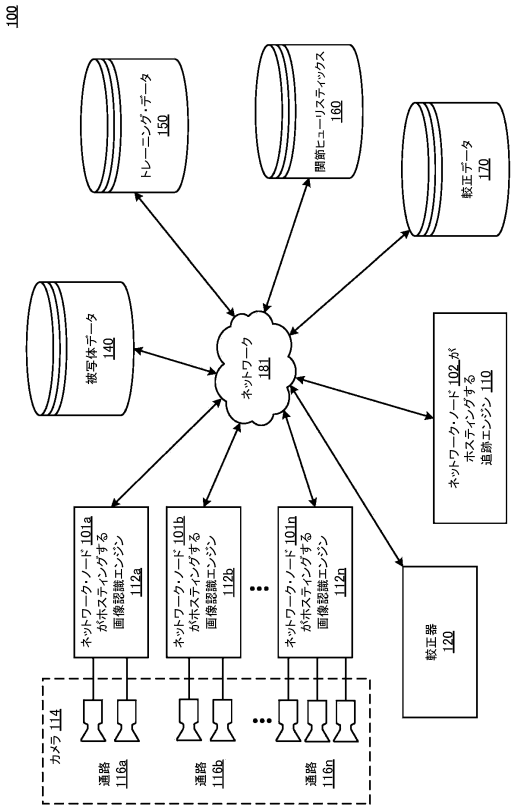
ｅｎＣＮＮ）。
【０３２１】
領域提案と取る／置く分析（例えば、ヒューリスティックス、格納された個人の関節位置、及び、店舗棚上の商品の予め計算されたマップと組み合わされたＷｈｅｎＣＮＮの出力）を使用して個人当たりの在庫配列を生成する第８のシステム、方法、及びコンピュータ・プログラム製品。
【０３２２】
棚上に置き間違えられた商品の位置を識別し、追跡し、更新するための第９のシステム、方法、及びコンピュータ・プログラム製品。
【０３２３】
ピクセルベースの分析を使用して棚上の商品に対する変化（取る／置く）を追跡する第１０のシステム、方法、及びコンピュータ・プログラム製品。
【０３２４】
店舗の在庫監査を実行する第１１のシステム、方法、及びコンピュータ・プログラム製品。
【０３２５】
手の中の複数の商品を識別する第１２のシステム、方法、及びコンピュータ・プログラム製品。
【０３２６】
ショッピングスクリプトを使用して店舗から商品画像データを収集する第１３のシステム、方法、及びコンピュータ・プログラム製品。
【０３２７】
会員顧客からチェックアウトを実行し、集金を行う第１４のシステム、方法、及びコンピュータ・プログラム製品。
【０３２８】
ゲスト顧客からのチェックアウトを実行し、集金を行う第１５のシステム、方法、及びコンピュータ・プログラム製品。
【０３２９】
カート内の未払商品を特定し、損失防止を行う第１６のシステム、方法、及びコンピュータ・プログラム製品。
【０３３０】
顧客のカート内で誤って識別された商品を店員が識別するのを支援するために、例えばカラーコードを使用して顧客を追跡する第１７のシステム、方法、及びコンピュータ・プログラム製品。
【０３３１】
位置ベースのインプレッション、方向性インプレッション、Ａ／Ｂ分析、顧客認識、グループダイナミクス等を含む顧客ショッピング分析を生成する第１８のシステム、方法、及びコンピュータ・プログラム製品。
【０３３２】
ショッピング分析を使用して目標顧客タッチバックを生成する第１９のシステム、方法、及びコンピュータ・プログラム製品。
【０３３３】
様々な活動を視覚化するために店舗のヒートマップオーバーレイを生成する第２０のシステム、方法、及びコンピュータ・プログラム製品。
【０３３４】
手検出のための第２１のシステム、方法、及びコンピュータ・プログラム。
【０３３５】
商品分類のための第２２のシステム、方法、及びコンピュータ・プログラム。
【０３３６】
商品所有権決定のための第２３のシステム、方法、及びコンピュータ・プログラム。

【０３３７】
商品人物検出のための第２４のシステム、方法、及びコンピュータ・プログラム。
【０３３８】
商品存在物追跡のための第２５のシステム、方法、及びコンピュータ・プログラム。
【０３３９】
商品マルチカメラ人物一致のための第２６の方法及びコンピュータ・プログラム。
【０３４０】
実質的に本明細書に記載されているレジレス・チェックアウトのための第２７のシステム、方法、及びコンピュータ・プログラム製品。
【０３４１】
第１～第２６のシステムの何れかと、上記列挙した第１～第２６のシステムの何れかの他の１つまたは複数のシステムとの組合せ。
【０３４２】
ここに記載されるのは、実空間のエリア内の被写体による在庫商品を置くこと及び取ることを追跡する方法であって：
【０３４３】
各カメラの視野が少なくとも１つの他のカメラの視野と重なり合う複数のカメラを使用して、実空間内の対応する視野のそれぞれの画像シーケンスを生成すること；
【０３４４】
複数のカメラから画像シーケンスを受け取り、第１の画像認識エンジンを使用して画像を処理し、被写体及び識別された被写体の実空間内の位置を識別する第１のデータセットを生成すること；
【０３４５】
画像シーケンス内の画像内の識別された被写体の手の画像を含む有界ボックスを指定するために第１のデータセットを処理すること；
【０３４６】
複数のカメラからの画像シーケンスを受信し、第２の画像認識エンジンを使用して識別された被写体の手の分類を生成するために、画像内の有界ボックスを処理すること、但し、前記分類は、識別された被写体が在庫商品を保持しているかどうか、棚に対する識別された被写体の手の位置を示す第１の近似度分類と、識別された被写体の身体に対する識別された被写体の手の位置を示す第２の近似度分類と、識別された被写体に関連するバスケットに対する識別された被写体の手の位置を示す第３の近似度分類と、可能性のある在庫商品の識別子とを含み； 及び、
【０３４７】
識別された被写体の画像シーケンス内の画像セットの手の分類を処理し、識別された被写体による在庫商品を取ること、及び、識別された被写体による在庫陳列構造上に在庫商品を置くことを検出すること、を有する。
【０３４８】
この説明された方法では、第１のデータセットが、識別された各被写体について、実空間内の座標を有する候補関節のセットを含むことができる。
【０３４９】
この説明された方法は、有界ボックスを指定するために第１のデータセットを処理することを含むことができ、各被写体の候補関節のセット内の関節の位置に基づいて有界ボックスを指定することを含む。
【０３５０】
この説明された方法では、第１及び第２の画像認識エンジンの一方または両方が畳み込みニューラル・ネットワークを備えることができる。
【０３５１】
この説明された方法は、畳み込みニューラル・ネットワークを使用して有界ボックスの分類を処理することを含むことができる。

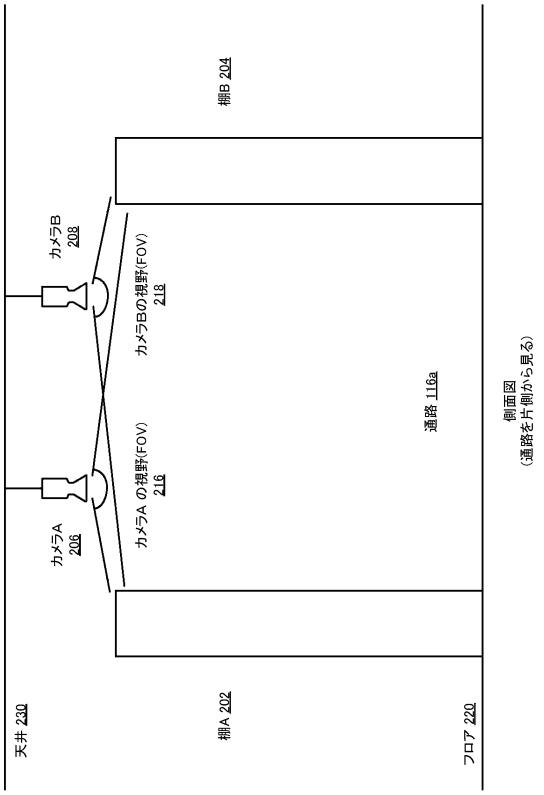
【０３５２】
非一時的データ記憶媒体を備えるコンピュータ可読メモリと、本明細書に記載の処理の何れかによって実空間のエリア内の被写体による在庫商品を置くこと及び取ることを追跡するために、コンピュータによって実行可能なメモリに記憶されたコンピュータ命令とを含むコンピュータ・プログラム製品が記載されている。
【０３５３】
被写体の手を含む画像シーケンスを生成する複数のカメラと、複数のカメラに結合された処理システムであって、画像シーケンスを受信して時系列に手の分類を生成する手画像認識エンジンと、画像シーケンスから手の分類を処理して被写体による、在庫商品を置くこと及び取ることのうちの１つである行為を識別するロジックとを含む処理システムと、を含むシステムが記載されている。
【０３５４】
このシステムは画像シーケンス内の画像内の被写体の関節の位置を識別し、識別された関節に基づいて被写体の手を含む対応する画像内の有界ボックスを識別するロジックを含むことができる。
【０３５５】
追記に列挙するコンピュータ・プログラムは、本明細書に添付され、本願において提供されるシステムの特定の部分を実装するためのコンピュータ・プログラムの一例の一部を含む。追記には、被写体の関節及び在庫商品を識別するためのヒューリスティックスの例が含まれる。追記は、被写体のショッピングカート・データ構造を更新するためのコンピュータ・プログラム・コードを提示する。追記はまた、畳み込みニューラル・ネットワークのトレーニング中に学習率を計算するためのコンピュータ・プログラム・ルーチンを含む。追記には、各カメラからの画像フレーム毎、被写体毎、手毎のデータ構造における畳み込みニューラル・ネットワークから、被写体の手の分類結果を保存するためのコンピュータ・プログラム・ルーチンが含まれている。

【図面】

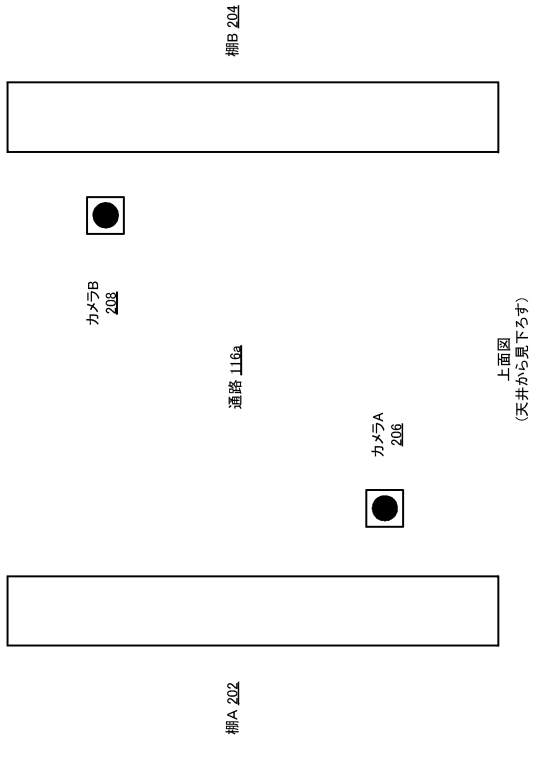
【図１】


100
トレーニング・データ 150
関節ヒューリスティックス 160
較正データ 170
被写体データ 140
ネットワーク 181
ネットワーク・ノード 102 がホスティングする 追跡エンジン 110
ネットワーク・ノード 101a がホスティングする 画像認識エンジン 112a
ネットワーク・ノード 101b がホスティングする 画像認識エンジン 112b
ネットワーク・ノード 101n がホスティングする 画像認識エンジン 112n
較正器 120
カメラ 114
通路 116a
通路 116b
通路 116n

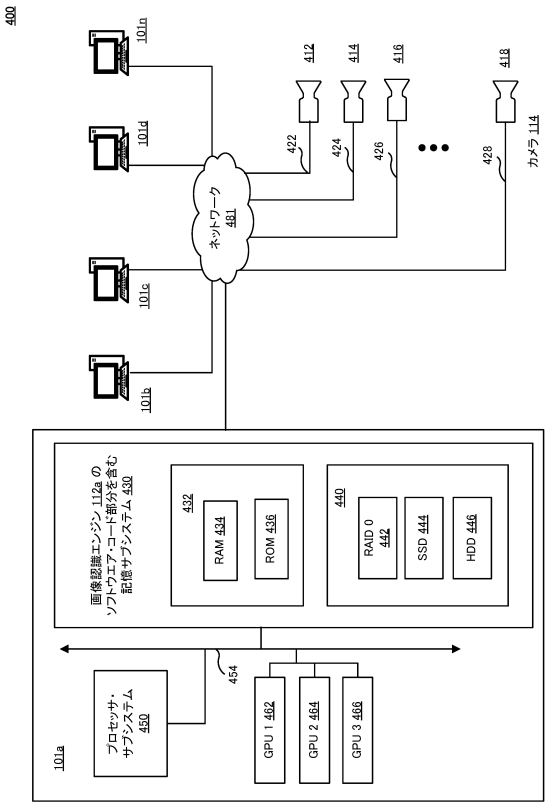
【図２】


棚B 204
カメラB 208
カメラBの視野(FOV) 218
カメラA の視野(FOV) 216
カメラA 208
通路 116a
側面図
（通路を片側から見る）
天井 230
棚A 202
フロア 220

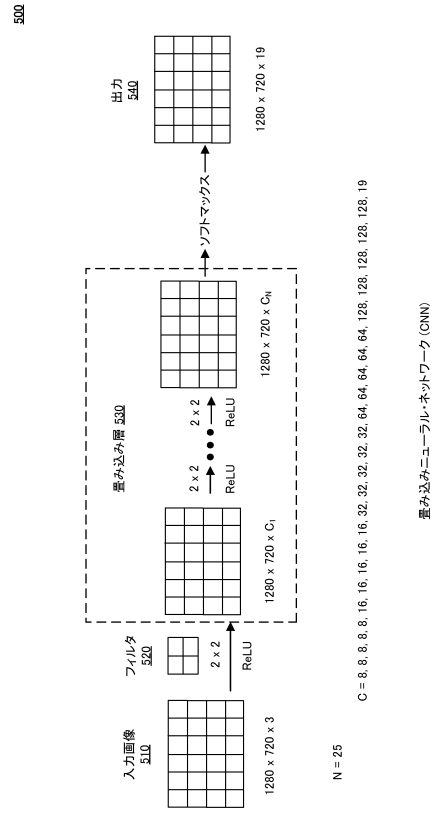
【図３】


棚B 204
カメラB 208
通路 116a
カメラA 206
上面図
（天井から見下ろす）
棚A 202

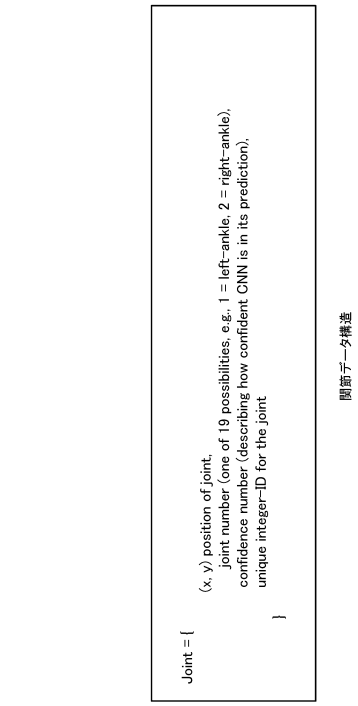
【図４】


400
101n
101d
101c
101b
ネットワーク 481
412
414
416
418
422
424
426
428
カメラ 114
画像認識エンジン 112a のソフトウエア・コード部分を含む記憶サブシステム 430
432
RAM 434
ROM 436
440
RAID 0 442
SSD 444
HDD 446
454
プロセッサ・サブシステム 450
GPU 1 462
GPU 2 464
GPU 3 466
101a

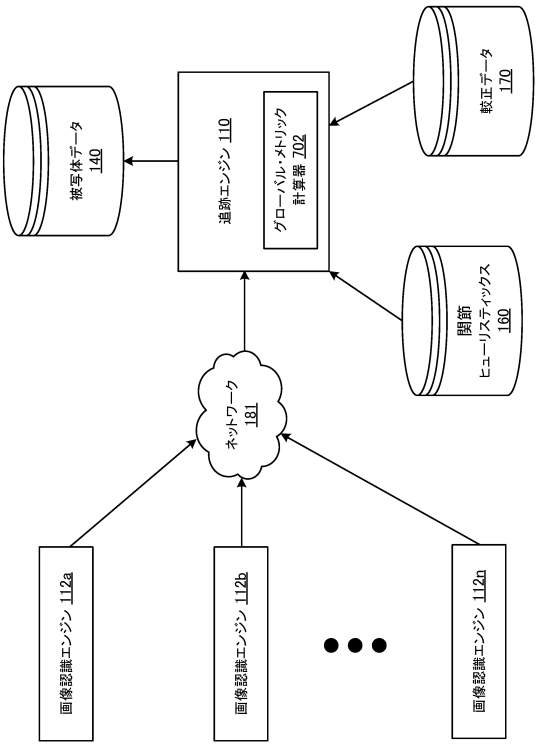
【図５】


500
入力画像
510
1280 x 720 x 3
フィルタ
520
2 x 2
ReLU
1280 x 720 x C1
畳み込み層 530
2 x 2
ReLU
2 x 2
ReLU
1280 x 720 x CN
ソフトマックス
出力
540
1280 x 720 x 19
N = 25
C = 8, 8, 8, 8, 16, 16, 16, 16, 16, 16, 32, 32, 32, 32, 32, 64, 64, 64, 64, 64, 128, 128, 128, 128, 128, 19
畳み込みニューラル・ネットワーク (CNN)

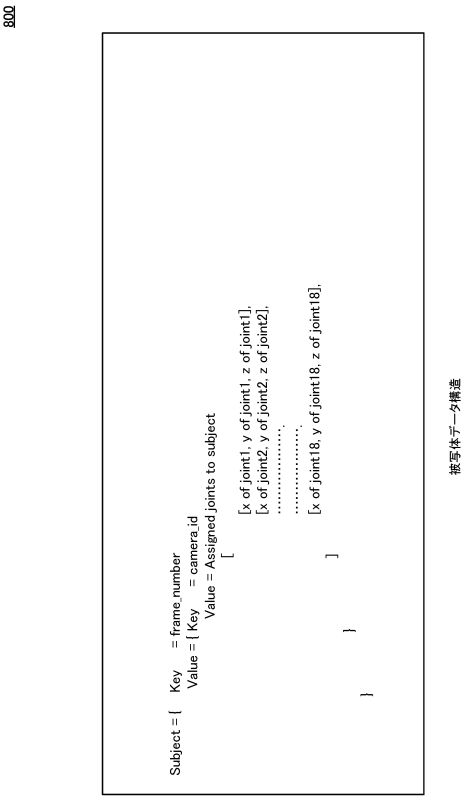
【図６】


600
Joint = {
〈x, y〉 position of joint,
joint number (one of 19 possibilities, e.g., 1 = left-ankle, 2 = right-ankle),
confidence number (describing how confident CNN is in its prediction),
unique integer-ID for the joint
}
関節データ構造

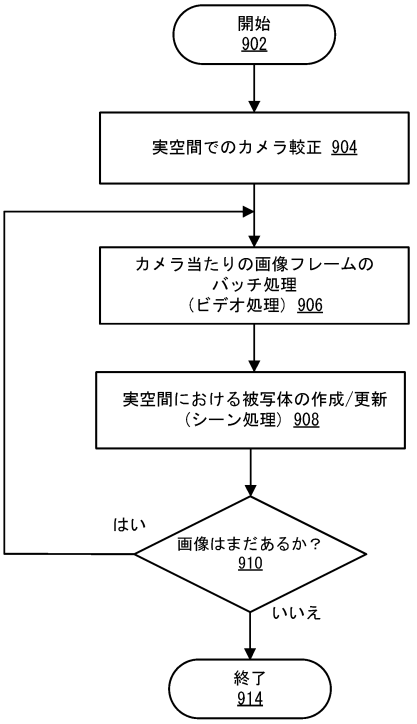
【図７】


被写体データ
140
追跡エンジン 110
グローバル・メトリック
計算器 702
較正データ
170
関節
ヒューリスティックス
160
ネットワーク
181
画像認識エンジン 112a
画像認識エンジン 112b
画像認識エンジン 112n

【図８】

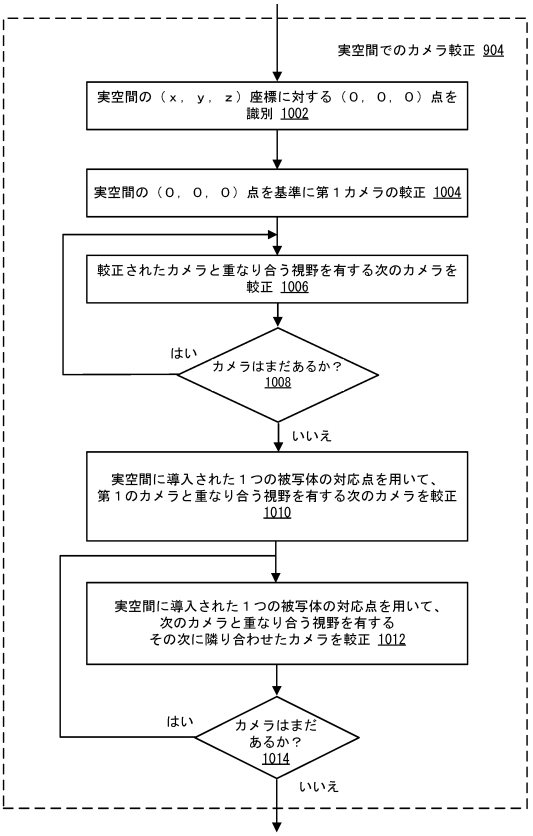

800
Subject = { Key = frame_number
Value = { Key = camera_id
Value = Assigned joints to subject
[
[x of joint1, y of joint1, z of joint1],
[x of joint2, y of joint2, z of joint2],
.....,
.....,
[x of joint18, y of joint18, z of joint18],
]
}
}
被写体データ構造

【図９】

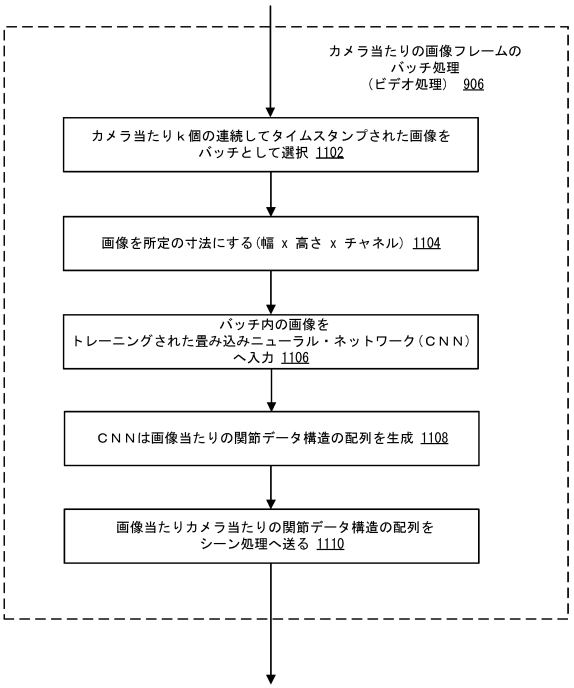


主なデータ・フロー

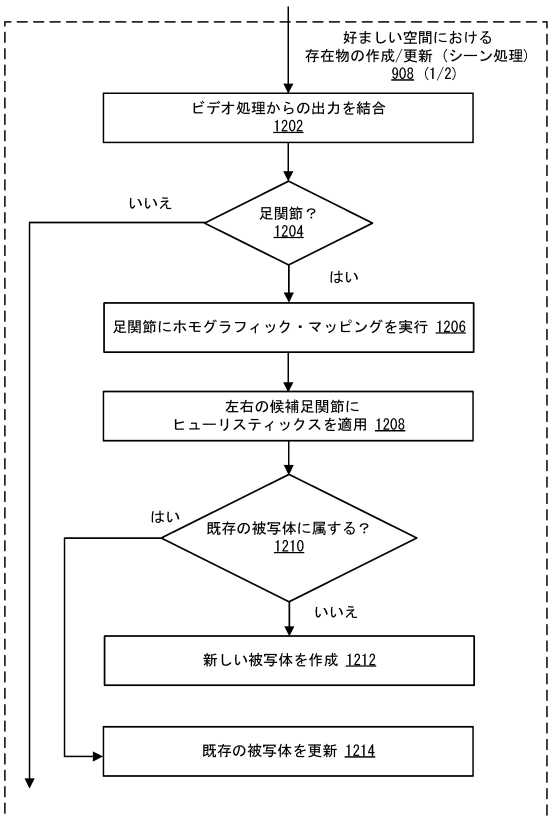
【図１０】



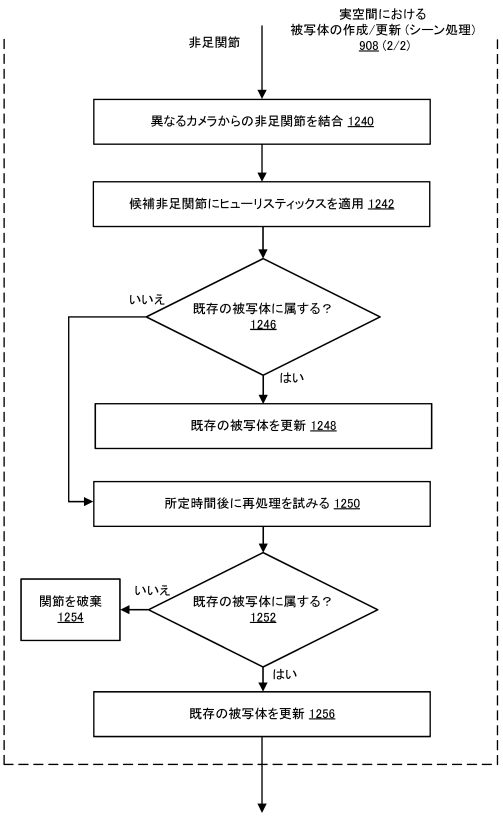
【図１１】



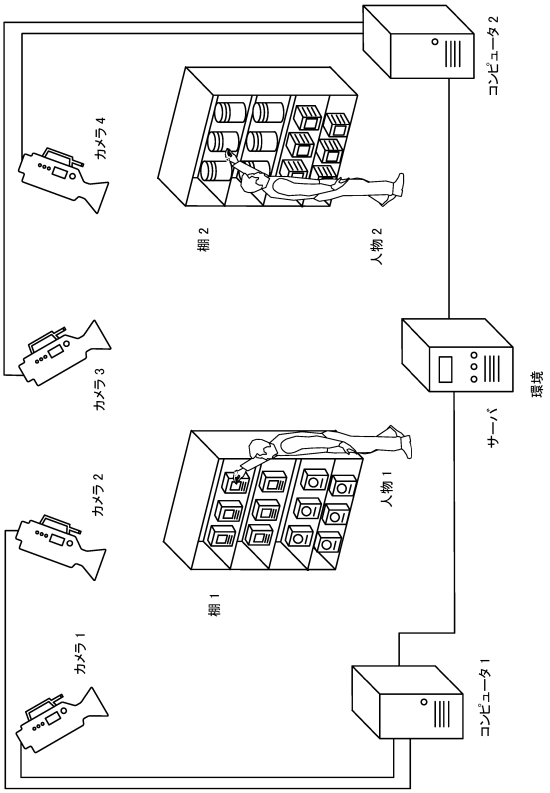
【図１２Ａ】



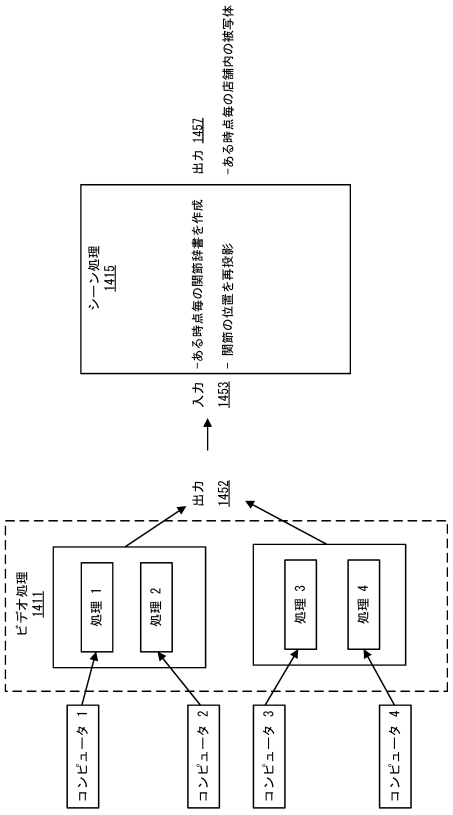
【図１２Ｂ】



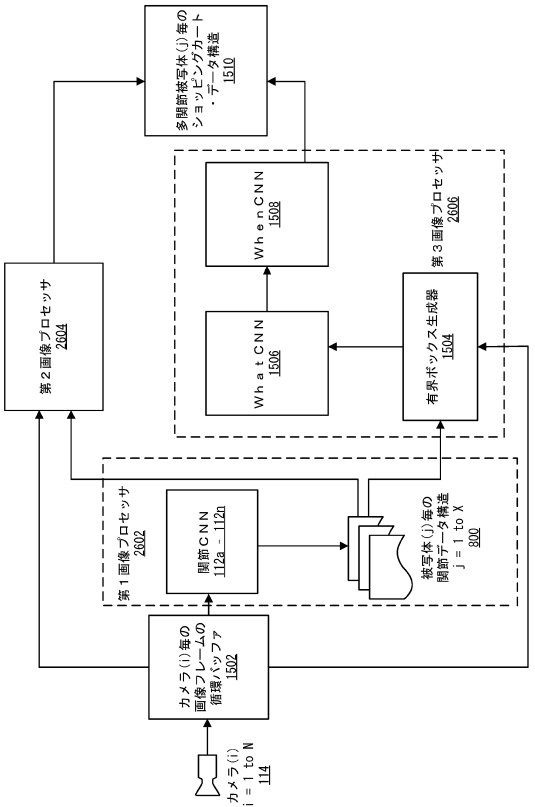
【図１３】



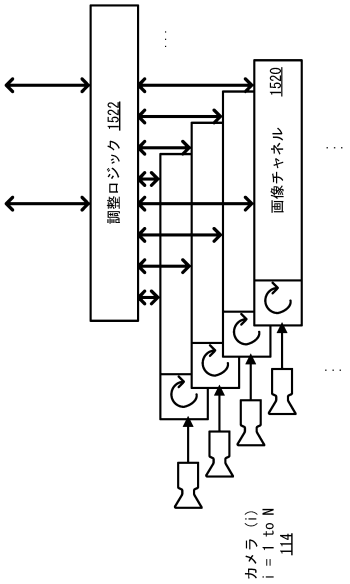
【図１４】



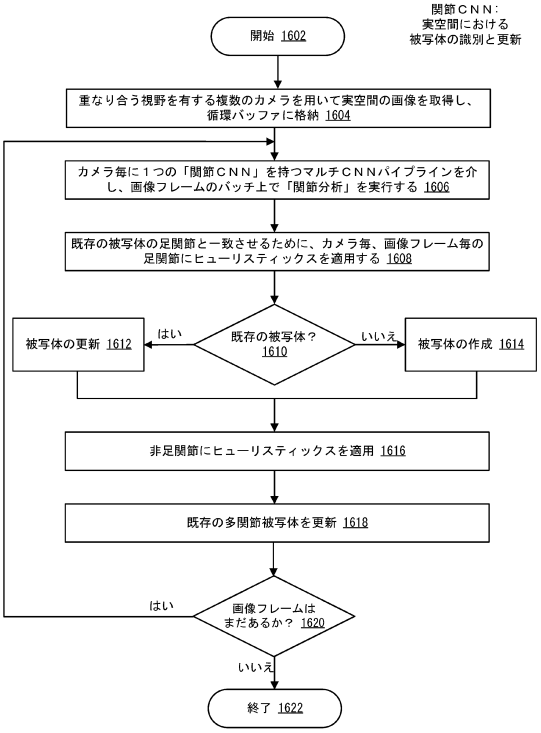
【図１５Ａ】



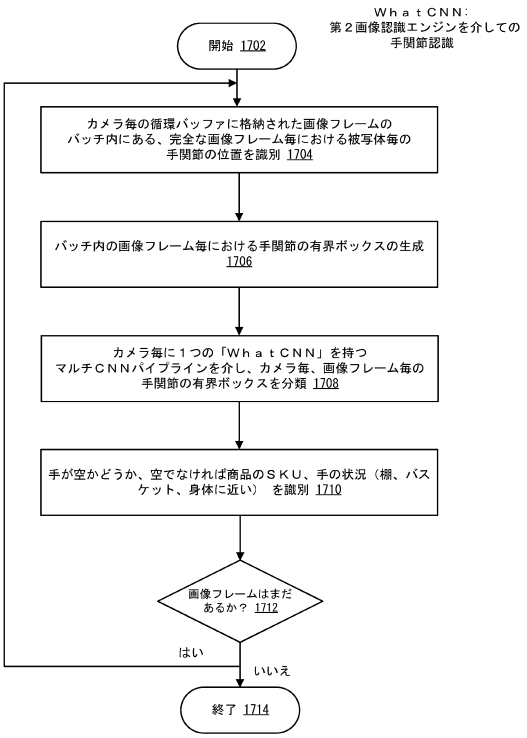
【図１５Ｂ】



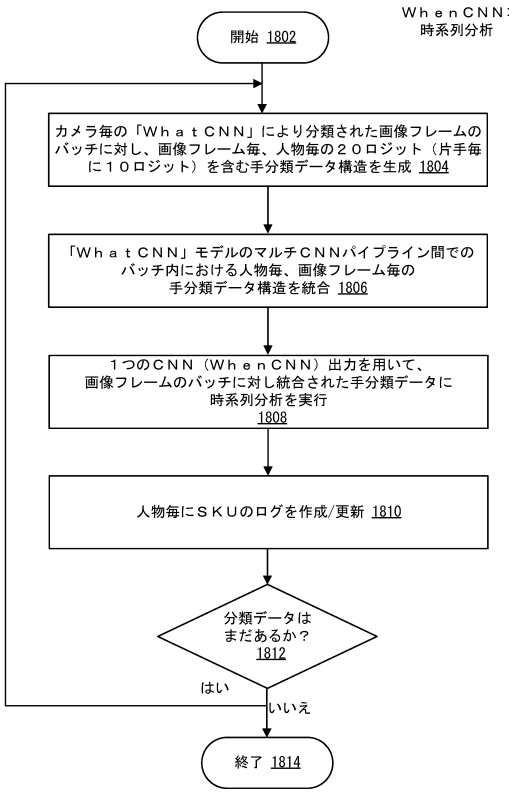
【図１６】



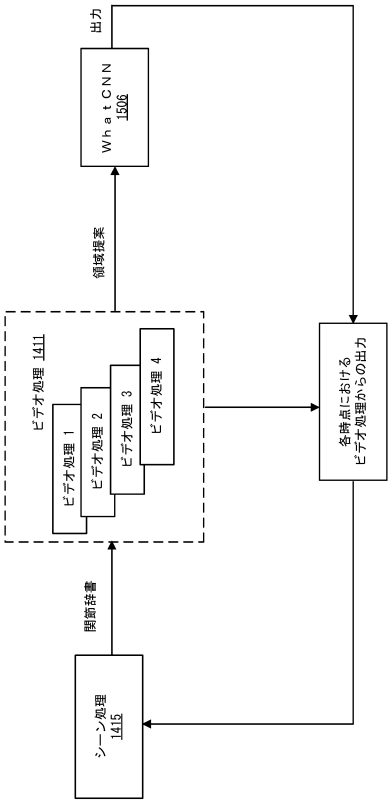
【図１７】



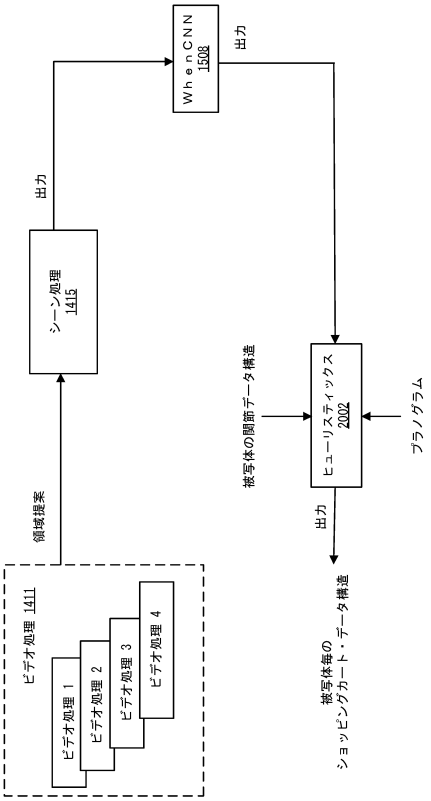
【図１８】



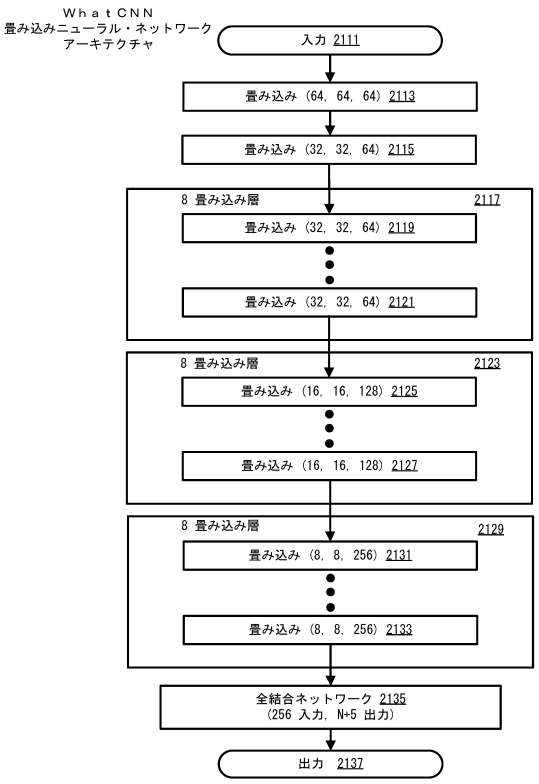
【図１９】



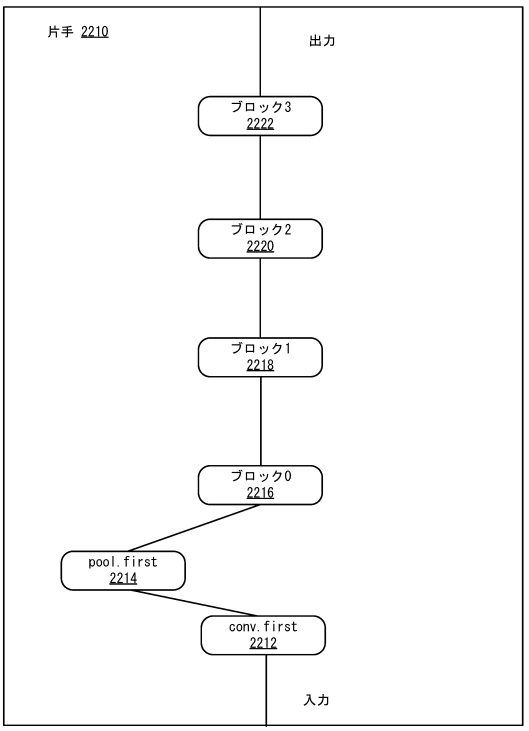
【図２０】



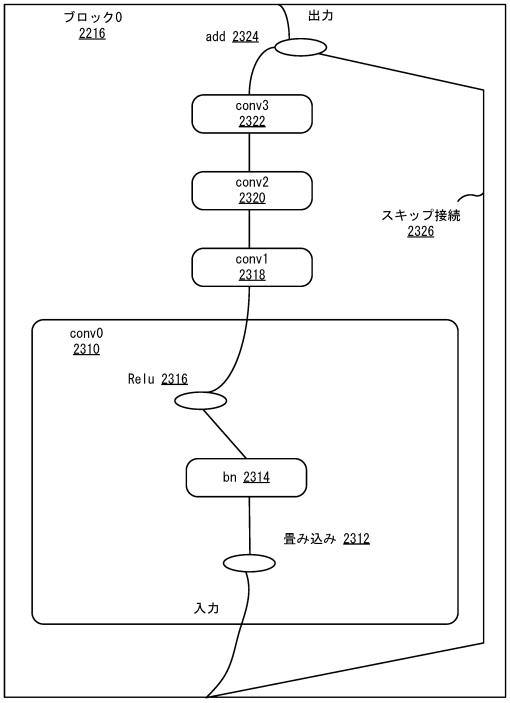
【図２１】



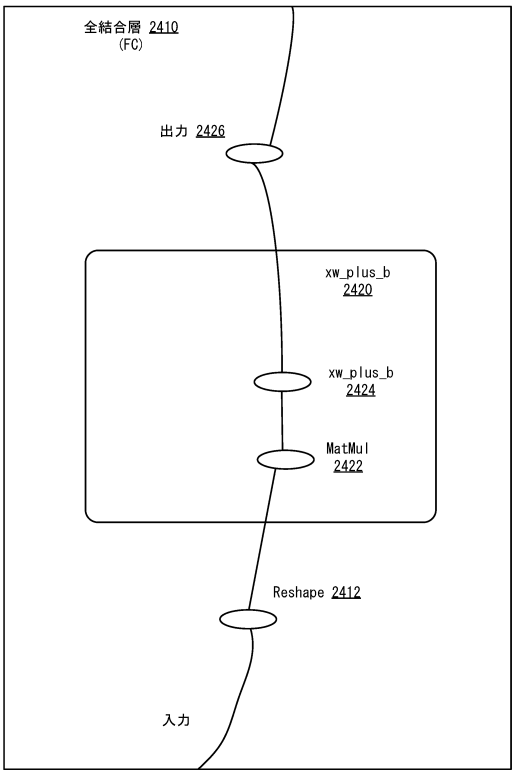
【図２２】



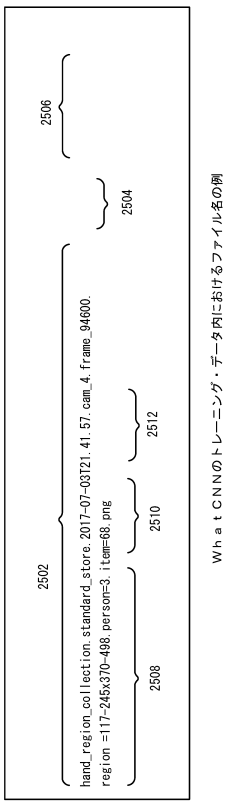
【図２３】



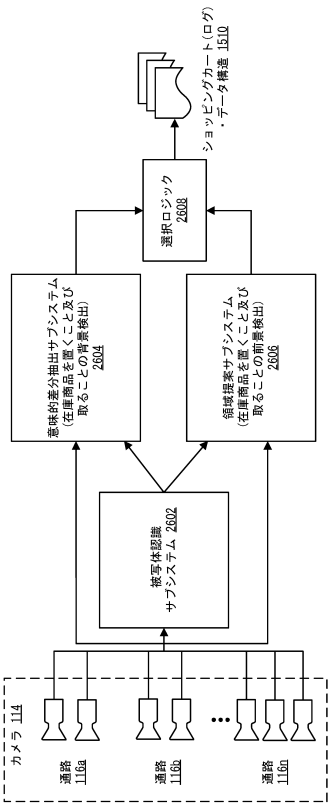
【図２４】



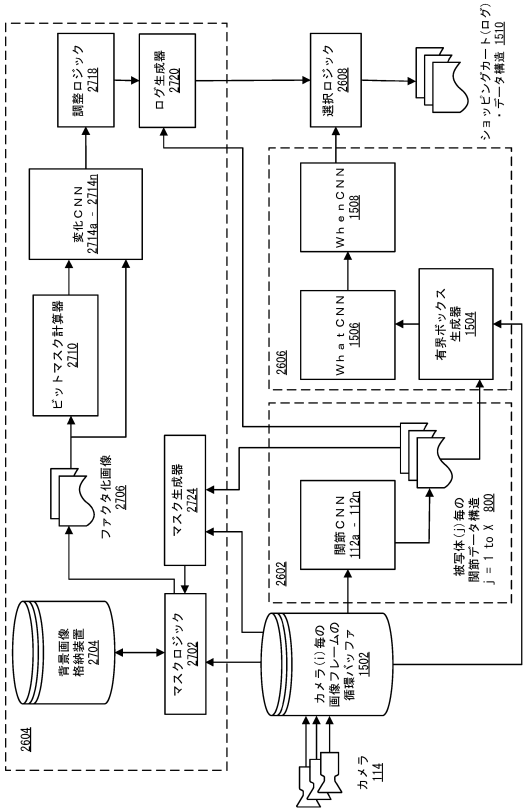
【図２５】



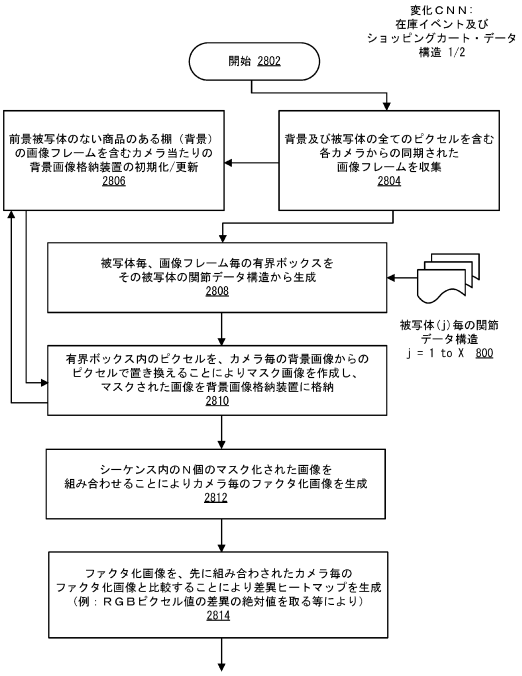
【図２６】



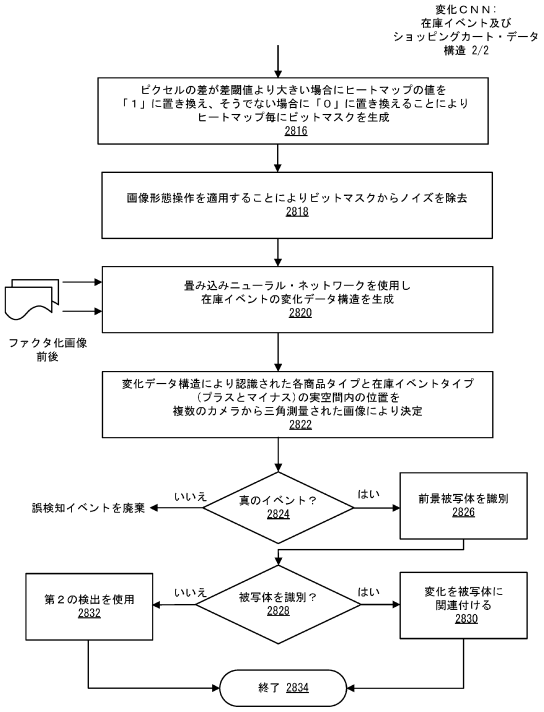
【図２７】



【図２８Ａ】



【図２８Ｂ】



フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　15/907,112
(32)優先日　平成30年2月27日(2018.2.27)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　15/945,466
(32)優先日　平成30年4月4日(2018.4.4)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　15/945,473
(32)優先日　平成30年4月4日(2018.4.4)
(33)優先権主張国・地域又は機関
米国(US)
８２８　８番ストリート　３３
(72)発明者　ノヴァク，　ジョン　エフ．
アメリカ合衆国　カリフォルニア州　９４１３１　サンフランシスコ，　バーンサイド　アヴェニュー　１５５
(72)発明者　オグル，　ブランドン　エル．
アメリカ合衆国　カリフォルニア州　９４１５８　サンフランシスコ，　ナンバー３２６　４番ストリート　１１５５
審査官　片岡　利延
(56)参考文献　米国特許出願公開第２０１４／０２１９５５０（ＵＳ，Ａ１）
特開２０１２－１２３６６７（ＪＰ，Ａ）
特開２００９－０５５１３９（ＪＰ，Ａ）
特開２００４－３４８６１８（ＪＰ，Ａ）
特開２０１０－００２９９７（ＪＰ，Ａ）
米国特許出願公開第２０１３／０１８２１１４（ＵＳ，Ａ１）
米国特許出願公開第２０１７／０２０６６６９（ＵＳ，Ａ１）
米国特許第０８０９８８８８（ＵＳ，Ｂ１）
Marcel Germann et al.，Space-time Body Pose Estimation in Uncontrolled Environments，[online]，2011年，https://ieeexplore.ieee.org/document/5955367
Umar Iqbal et al.，PoseTrack: Joint Multi-Person Pose Estimation and Tracking，[online]，2017年07月26日，https://ieeexplore.ieee.org/document/8099978
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　７／２０
Ｇ０６Ｔ　７／００