

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7310927号
(P7310927)

(45)発行日 令和5年7月19日(2023.7.19)

(24)登録日 令和5年7月10日(2023.7.10)

(51)国際特許分類 F I
 G 0 6 T 7/246(2017.01) G 0 6 T 7/246
 G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 3 5 0 B

請求項の数 9 (全25頁)

(21)出願番号	特願2021-566677(P2021-566677)	(73)特許権者	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(86)(22)出願日	令和1年12月26日(2019.12.26)	(74)代理人	100104765 弁理士 江上 達夫
(86)国際出願番号	PCT/JP2019/051088	(74)代理人	100107331 弁理士 中村 聡延
(87)国際公開番号	WO2021/130951	(74)代理人	100131015 弁理士 三輪 浩誉
(87)国際公開日	令和3年7月1日(2021.7.1)	(72)発明者	福井 宏 東京都港区芝五丁目7番1号 日本電気株式会社内
審査請求日	令和4年6月8日(2022.6.8)	審査官	小太刀 慶明

最終頁に続く

(54)【発明の名称】 物体追跡装置、物体追跡方法及び記録媒体

(57)【特許請求の範囲】

【請求項1】

第1時刻に撮影された第1画像中の物体の位置に関する第1位置情報及び前記第1時刻とは異なる第2時刻に撮影された第2画像中の物体の位置に関する第2位置情報に基づいて、前記第1位置情報の特徴量を示す第1特徴ベクトルと、前記第2位置情報の特徴量を示す第2特徴ベクトルとを生成する第1生成手段と、

前記第1及び第2特徴ベクトルを用いた演算処理によって得られる情報を、前記第1画像中の物体と前記第2画像中の物体との対応関係を示す対応情報として生成する第2生成手段と、

前記対応情報を用いて前記第2位置情報を補正する補正手段と

を備える物体追跡装置。

【請求項2】

前記演算処理は、前記第1特徴ベクトルと第2特徴ベクトルとの行列積を算出する処理を含む

請求項1に記載の物体追跡装置。

【請求項3】

前記対応情報は、行列を用いて前記第1画像中の物体と前記第2画像中の物体との対応関係を示す

請求項1又は2に記載の物体追跡装置。

【請求項4】

前記第 2 生成手段は、前記行列の一の行のベクトル成分及び前記行列の一の列のベクトル成分の夫々をソフトマックス関数で正規化することで、前記行列を正規化する
請求項 3 に記載の物体追跡装置。

【請求項 5】

前記対応情報は、前記第 1 及び第 2 画像のいずれか一方中の一の物体に対応する行ベクトル成分と前記第 1 及び第 2 画像のいずれか他方中の前記一の物体に対応する列ベクトル成分とが交差する位置において要素が反応する行列を用いて、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示し、

前記第 2 生成手段は、前記行ベクトル成分及び各列ベクトル成分の夫々をソフトマックス関数で正規化することで、前記行列を正規化する

請求項 1 から 4 のいずれか一項に記載の物体追跡装置。

【請求項 6】

前記補正手段は、前記対応情報を重みとして用いる注意機構を用いて、前記第 2 位置情報を補正する

請求項 1 から 5 のいずれか一項に記載の物体追跡装置。

【請求項 7】

前記対応情報は、行列を用いて前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示し、

前記第 2 位置情報は、前記第 2 画像中の物体の位置に関する情報を示す位置マップを含み、

前記注意機構は、前記位置マップと前記対応情報との行列積を算出する処理を行うことで、前記第 2 位置情報である前記位置マップを補正する

請求項 6 に記載の物体追跡装置。

【請求項 8】

第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、

前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することと、
前記対応情報を用いて前記第 2 位置情報を補正することと

を含む物体追跡方法。

【請求項 9】

コンピュータに物体追跡方法を実行させるコンピュータプログラムであって、

前記物体追跡方法は、

第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、

前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することと、
前記対応情報を用いて前記第 2 位置情報を補正することと

を含むコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、時系列データに相当する複数の画像中に映り込んでいる物体を追跡する物体追跡装置、物体追跡方法及び記録媒体の技術分野に関する。

【背景技術】

【0002】

10

20

30

40

50

カメラ等で撮影された時系列データとしての複数の画像を用いて、画像中に映り込んでいる物体を追跡する物体追跡装置が知られている。物体追跡装置の一例が、特許文献1から特許文献3に記載されている。更に、物体を追跡するためのアルゴリズムの一例が、非特許文献1から非特許文献2に記載されている。その他、本願発明に関連する先行技術文献として、非特許文献3があげられる。

【先行技術文献】

【特許文献】

【0003】

【文献】特開2009-265732号公報

特開2012-181710号公報

10

特開2016-71830号公報

米国特許第9,858,496号

【非特許文献】

【0004】

【文献】C. Feichtenhofer et al., "Detect to Track and Track to Detect", ICCV2017, 2017年10月

Yihong Xu et al., "DeepMOT: A Differentiable Framework for Training Multiple Object Trackers", arXiv 1906.06618, 2019年1月

Xingyi Zhou et al., "Object as Points", arXiv 2019.04.07850, 2019年4月

Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" IEEE Transactions on Pattern Analysis and Machine Intelligence, pp1137-1149, 2016年6月6日

Wei Liu et al., "SSD: Single Shot MultiBox Detector", European Conference on Computer Vision, pp21-37, 2016年9月

【発明の概要】

30

【発明が解決しようとする課題】

【0005】

従来の物体追跡装置は、画像に映り込んでいる物体を検出する物体検出動作と、時刻 t - (尚、 T は、基準周期を示す) の画像 IM_t に映り込んでいる物体 O_t と時刻 t の画像 IM_t に映り込んでいる物体 O_t とを照合する物体照合動作とを、互いに独立した別個の動作として行う。このため、従来の物体追跡装置は物体検出動作が行われた後に物体照合動作を行うために、前処理等を行う必要がある。その結果、従来の物体追跡装置では、物体を追跡するための処理コストが相対的に高くなってしまいう可能性がある。

【0006】

本発明は、上述した技術的問題を解決可能な物体追跡装置、物体追跡方法及び記録媒体を提供することを課題とする。一例として、本発明は、相対的に低い処理コストで物体を追跡可能な物体追跡装置、物体追跡方法及び記録媒体を提供することを課題とする。

40

【課題を解決するための手段】

【0007】

物体追跡装置の一態様は、第1時刻に撮影された第1画像中の物体の位置に関する第1位置情報及び前記第1時刻とは異なる第2時刻に撮影された第2画像中の物体の位置に関する第2位置情報に基づいて、前記第1位置情報の特徴量を示す第1特徴ベクトルと、前記第2位置情報の特徴量を示す第2特徴ベクトルとを生成する第1生成手段と、前記第1及び第2特徴ベクトルを用いた演算処理によって得られる情報を、前記第1画像中の物体と前記第2画像中の物体との対応関係を示す対応情報として生成する第2生成手段とを備

50

える。

【 0 0 0 8 】

物体追跡方法の一態様は、第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することを含む。

【 0 0 0 9 】

記録媒体の一態様は、コンピュータに物体追跡方法を実行させるコンピュータプログラムが記録された、一時的でない記録媒体であって、前記物体追跡方法は、第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することを含む。

【 発明の効果 】

【 0 0 1 0 】

上述した物体追跡装置、物体追跡方法及び記録媒体のそれぞれの一態様によれば、第 1 及び第 2 特徴ベクトルを用いた演算処理により対応情報を生成することができるため、後に詳述するように、相対的に低い処理コストで物体を追跡することができる。

【 図面の簡単な説明 】

【 0 0 1 1 】

【 図 1 】 図 1 は、本実施形態の物体追跡装置の構成を示すブロック図である。

【 図 2 】 図 2 は、物体照合動作及びリファイン動作を行うために物体追跡装置内に実現される論理的な機能ブロックの構成を示すブロック図である。

【 図 3 】 図 3 は、物体検出動作によって検出された物体位置情報を概念的に示す平面図である。

【 図 4 】 図 4 は、物体照合動作の流れを示すフローチャートである。

【 図 5 】 図 5 は、特徴ベクトルと類似性行列との関係を概念的に示す平面図である。

【 図 6 】 図 6 は、リファイン動作の流れを示すフローチャートである。

【 図 7 】 図 7 は、第 1 変形例の物体追跡装置の構成を示すブロック図である。

【 図 8 】 図 8 は、学習 DB のデータ構造を示すデータ構造図である。

【 図 9 】 図 9 は、第 2 変形例の物体追跡装置の構成を示すブロック図である。

【 図 1 0 】 図 1 0 は、ソフトマックス関数を用いて類似性行列を正規化する様子を示す平面図である。

【 発明を実施するための形態 】

【 0 0 1 2 】

以下、図面を参照しながら、物体追跡装置、物体追跡方法及び記録媒体の実施形態について説明する。以下では、物体追跡装置、物体追跡方法及び記録媒体の実施形態が適用された物体追跡装置 1 について説明する。物体追跡装置 1 は、時系列データに相当する複数の画像 $I M$ が入力された場合に各画像 $I M$ に映り込んでいる少なくとも一つの物体 O を追跡するための物体追跡動作を行う。物体追跡動作は、例えば、画像 $I M$ に映り込んでいる物体 O を検出する物体検出動作を含む。物体追跡動作は、例えば、時刻 $t -$ (尚、 $t -$ は、基準周期を示す) に取得された (例えば、撮影された) 画像 $I M_{t -}$ に映り込んでいる少なくとも一つの物体 $O_{t -}$ と、時刻 t に取得された画像 $I M_t$ に映り込んでいる少なくとも一つの物体 O_t とを照合することで、画像 $I M_{t -}$ に映り込んでいた物体 $O_{t -}$ の画像 $I M_t$ 内での位置を追跡する物体照合動作を含む。物体追跡動作は、例えば、物体照合動作の結果を用いて、画像 $I M_t$ に映り込んでいる物体 O_t の検出結果 (つまり、物体

10

20

30

40

50

検出動作の結果)を補正するリファイン動作を含む。

【0013】

尚、以下の説明において、「X及び/又はY」という表現は、「X及びY」という表現及び「X又はY」という表現の双方を包含する表現として用いられている。

【0014】

(1) 物体追跡装置1の構成

初めに、図1及び図2を参照しながら、本実施形態の物体追跡装置1の構成について説明する。図1は、本実施形態の物体追跡装置1の構成を示すブロック図である。図2は、物体照合動作及びリファイン動作を行うために物体追跡装置1内に実現される論理的な機能ブロックの構成を示すブロック図である。

10

【0015】

図1に示すように、物体追跡装置1は、演算装置2と、記憶装置3とを備えている。更に、物体追跡装置1は、入力装置4と、出力装置5とを備えていてもよい。但し、物体追跡装置1は、入力装置4及び出力装置5の少なくとも一方を備えていなくてもよい。演算装置2と、記憶装置3と、入力装置4と、出力装置5は、データバス6を介して接続されている。

【0016】

演算装置2は、例えば、CPU(Central Processing Unit)及びGPU(Graphic Processing Unit)の少なくとも一方を含む。演算装置2は、コンピュータプログラムを読み込む。例えば、演算装置2は、記憶装置3が記憶しているコンピュータプログラムを読み込んでよい。例えば、演算装置2は、コンピュータで読み取り可能であって且つ一時的でない記録媒体が記憶しているコンピュータプログラムを、図示しない記録媒体読み取り装置を用いて読み込んでよい。演算装置2は、不図示の通信装置を介して、物体追跡装置1の外部に配置される不図示の装置からコンピュータプログラムを取得してもよい(つまり、ダウンロードしてもよい又は読み込んでよい)。演算装置2は、読み込んだコンピュータプログラムを実行する。その結果、演算装置2内には、物体追跡装置1が行うべき動作(具体的には、上述した物体追跡動作)を実行するための論理的な機能ブロックが実現される。つまり、演算装置2は、物体追跡動作を実行するための論理的な機能ブロックを実現するためのコントローラとして機能可能である。

20

30

【0017】

図1には、物体追跡動作を実行するために演算装置2内に実現される論理的な機能ブロックの一例が示されている。図1に示すように、演算装置2内には、論理的な機能ブロックとして、物体検出部21と、物体照合部22と、リファイン部23とが実現される。物体検出部21は、物体検出動作を行う。物体照合部22は、物体照合動作を行う。物体照合動作を行うために、物体照合部22は、図2に示すように、特徴マップ変換部221と、特徴ベクトル変換部222と、特徴マップ変換部223と、特徴ベクトル変換部224と、行列演算部225とを備えている。リファイン部23は、リファイン動作を行う。リファイン動作を行うために、リファイン部23は、図2に示すように、行列演算部231と、特徴ベクトル変換部232と、特徴マップ変換部233と、残差処理部234と、特徴マップ変換部235と、特徴マップ変換部236と、特徴ベクトル変換部237とを備えている。尚、物体検出動作、物体照合動作及びリファイン動作については、後に詳述する。

40

【0018】

記憶装置3は、所望のデータを記憶可能である。例えば、記憶装置3は、演算装置2が実行するコンピュータプログラムを一時的に記憶していてもよい。記憶装置3は、演算装置2がコンピュータプログラムを実行している際に演算装置2が一時的に使用するデータを一時的に記憶してもよい。記憶装置3は、物体追跡装置1が長期的に保存するデータを記憶してもよい。本実施形態では特に、記憶装置3には、不図示のカメラが撮影した時系列データとしての複数の画像IMを格納するための画像DB(Data Base)31が

50

記憶されていてもよい。更に、記憶装置 3 には、物体検出動作の結果を示す物体検出情報（つまり、画像 I M に映り込んでいる物体 O の検出結果に関する情報）を格納するための物体検出 D B 3 2 が記憶されていてもよい。更に、記憶装置 3 には、物体照合動作の結果を示す物体照合情報（つまり、画像 I M_t に映り込んでいる物体 O_t と画像 I M_t に映り込んでいる物体 O_t との照合結果に関する情報であり、典型的には、物体 O_t と物体 O_t との対応付けを示す情報）を格納するための物体照合 D B 3 3 が記憶されていてもよい。尚、記憶装置 3 は、R A M (R a n d o m A c c e s s M e m o r y)、R O M (R e a d O n l y M e m o r y)、ハードディスク装置、光磁気ディスク装置、S S D (S o l i d S t a t e D r i v e) 及びディスクアレイ装置のうち少なくとも一つを含んでいてもよい。つまり、記憶装置 3 は、一時的でない記録媒体を含んでいてもよい。

10

【 0 0 1 9 】

入力装置 4 は、物体追跡装置 1 の外部からの物体追跡装置 1 に対する情報の入力を受け付ける装置である。本実施形態では、入力装置 4 には、不図示のカメラが撮影した時系列データとしての複数の画像 I M が入力される。入力装置 4 に入力された複数の画像 I M は、記憶装置 3 が記憶している画像 D B 3 1 に格納される。

【 0 0 2 0 】

出力装置 5 は、物体追跡装置 1 の外部に対して情報を出力する装置である。例えば、出力装置 5 は、物体追跡動作の結果に関する情報を出力してもよい。例えば、出力装置 5 は、物体追跡動作の結果に関する情報を、画像として出力してもよい。この場合、出力装置 5 は、画像を表示する表示装置を含んでいてもよい。例えば、出力装置 5 は、物体追跡動作の結果に関する情報を、データとして出力してもよい。この場合、出力装置 5 は、データを出力するデータ出力装置を含んでいてもよい。

20

【 0 0 2 1 】

(2) 物体追跡装置 1 が行う物体追跡動作

続いて、物体追跡装置 1 が行う物体追跡動作について説明する。上述したように、物体追跡動作は、物体検出動作と、物体照合動作と、リファイン動作とを行う。従って、以下では、物体追跡動作は、物体検出動作と、物体照合動作と、リファイン動作とについて順に説明する。

【 0 0 2 2 】

(2 - 1) 物体検出動作

初めに、物体検出部 2 1 が行う物体検出動作について説明する。物体検出部 2 1 は、画像 D B 3 1 に格納されている画像 I M を読み出し、読み出した画像 I M に対して物体検出動作を行う。物体検出部 2 1 は、画像に映り込んだ物体を検出するための既存の方法を用いて、画像 I M に映り込んだ物体 O を検出してもよい。但し、物体検出部 2 1 は、画像 I M に映り込んだ物体 O を検出することで画像 I M 内での物体 O の位置に関する情報（以降、“物体位置情報 P I ”と称する）を取得可能な方法を用いて、物体検出動作を行うことが好ましい。物体検出部 2 1 が取得した物体位置情報 P I は、物体検出部 2 1 による物体検出動作の結果を示す物体検出情報として、物体検出 D B 3 2 に格納される。

30

【 0 0 2 3 】

本実施形態では、上述した非特許文献 3 に記載された方法（いわゆる、C e n t e r N e t ）を用いて物体検出部 2 1 が物体 O を検出する例について説明する。この場合、物体検出部 2 1 は、物体位置情報 P I として、画像 I M 内での物体 O の中心位置（K e y P o i n t）K P を示すヒートマップ（いわゆるスコアマップ）を生成する。より具体的には、物体検出部 2 1 は、画像 I M 内での物体 O の中心位置 K P を示すヒートマップを、物体 O のクラス毎に生成する。このため、画像 I M 内での物体 O の中心位置 K P を示す情報は、画像 I M の縦方向の画素数を H とし、画像 I M の横方向の画素数を W とし、クラス数を K とした場合には、 $H \times W \times K$ のサイズを有するマップ情報となる。尚、中心位置 K P を示すヒートマップは、位置に関するマップであるがゆえに、位置マップと称されてもよい。

40

50

【 0 0 2 4 】

非特許文献 3 に記載された方法を用いて物体検出部 2 1 が物体 O を検出する場合には、物体検出部 2 1 は、物体位置情報 P I として、物体 O の検出枠 (B o u n d i n g B o x) B B のサイズをスコアマップとして示す情報を生成してもよい。物体 O の検出枠 B B のサイズを示す情報は、実質的には、物体 O のサイズを示す情報であるとみなしてもよい。物体 O の検出枠 B B のサイズを示す情報は、例えば、 $H \times W \times 2$ のサイズを有するマップ情報となる。尚、検出枠 B B のサイズを示すマップ情報もまた、位置に関するマップであるがゆえに、位置マップと称されてもよい。

【 0 0 2 5 】

非特許文献 3 に記載された方法を用いて物体検出部 2 1 が物体 O を検出する場合には、物体検出部 2 1 は、物体位置情報 P I として、物体 O の検出枠 B B の補正量 (L o c a l O f f s e t) をスコアマップとして示す情報を生成してもよい。物体 O の検出枠 B B の補正量を示す情報は、 $H \times W \times 2$ のサイズを有するマップ情報となる。尚、検出枠 B B の補正量を示すマップ情報もまた、位置に関するマップであるがゆえに、位置マップと称されてもよい。

【 0 0 2 6 】

図 3 は、物体検出動作によって検出された物体位置情報 P I を概念的に示す平面図である。図 3 は、画像 I M に四つの物体 O (具体的には、物体 O # 1、物体 O # 2、物体 O # 3 及び物体 O # 4) が映り込んでいる例を示している。この場合、物体検出部 2 1 は、物体位置情報 P I として、四つの物体 O の夫々の中心位置 K P を示す情報、四つの物体の夫々の検出枠 B B のサイズを示す情報、及び、四つの物体 O の夫々の検出枠 B B の補正量を示す情報を生成する。

【 0 0 2 7 】

物体検出部 2 1 は、画像 I M が入力された場合に物体位置情報 P I を出力する演算モデルを用いて、物体検出動作を行ってもよい。このような演算モデルの一例として、ニューラルネットワーク (例えば、C N N : C o n v o l u t i o n a l N e u r a l N e t w o r k) を用いた演算モデルがあげられる。この場合、演算モデルは、後述するように、学習用データを用いて学習されてもよい。つまり、演算モデルのパラメータは、適切な物体位置情報 P I を出力するように最適化されていてもよい。

【 0 0 2 8 】

但し、物体検出部 2 1 は、非特許文献 3 に記載された方法 (いわゆる、C e n t e r N e t) とは異なる他の方法を用いて、物体検出動作を行ってもよい。他の方法の一例として、非特許文献 4 に記載された F a s t e r R - C N N と称される方法及び非特許文献 5 に記載された S S D と称される方法があげられる。

【 0 0 2 9 】

(2 - 2) 物体照合動作

続いて、物体照合部 2 2 が行う物体照合動作について説明する。物体照合部 2 2 は、物体検出 D B 3 2 に格納されている物体位置情報 P I を読み出し、読み出した物体位置情報 P I を用いて物体照合動作を行う。以下、時刻 $t -$ に撮影された画像 $I M_{t -}$ に映り込んでいる物体 $O_{t -}$ と、時刻 t に撮影された画像 $I M_t$ に映り込んでいる物体 O_t とを照合する物体照合動作について、図 4 を参照しながら説明する。図 4 は、物体照合動作の流れを示すフローチャートである。

【 0 0 3 0 】

図 4 に示すように、物体照合部 2 2 は、物体位置情報 P I から特徴ベクトル C V を生成する (ステップ S 2 2 1 からステップ S 2 2 6) 。本実施形態では、物体照合部 2 2 は、物体位置情報 P I から特徴ベクトル C V を生成するために、物体位置情報 P I から特徴マップ C M を生成し、その後、特徴マップ C M から特徴ベクトル C V を生成する。特徴マップ C M は、物体位置情報 P I の特徴量を、任意のチャンネル毎に示す特徴マップである。つまり、チャンネル数を C とすると、物体照合部 2 2 は、 $H \times W \times (K + 2 + 2) = H \times W \times (K + 4)$ (K は、上述したように物体 O のクラス数であり、典型的には、物体検出

DB32又は物体照合DB33に登録されているクラス数)のサイズを有するマップ情報である物体位置情報PIから、 $H \times W \times C$ のサイズを有する特徴マップCMを生成し、その後、 $H \times W \times C$ のサイズを有する特徴マップCMから、 $H \times W \times C$ のサイズを有する特徴ベクトルCVを生成する。但し、物体照合部22は、特徴マップCMを生成することなく、物体位置情報PIから特徴ベクトルCVを直接生成してもよい。

【0031】

具体的には、「第1生成手段」の一具体例である特徴マップ変換部221は、物体検出DB32から、時刻 t に撮影された画像 IM_t に映り込んでいる物体 O_t に関する物体位置情報 PI_t を取得する(つまり、読み出す)(ステップS221)。尚、物体位置情報 PI_t は、第1位置情報の一具体例である。例えば、図3に示すように、画像 IM_t に四つの物体 O_t が映り込んでいる場合には、特徴マップ変換部221は、物体検出DB32から、四つの物体 O_t に関する物体位置情報 PI_t を取得する。その後、特徴マップ変換部221は、ステップS221で取得した物体位置情報 PI_t から、特徴マップ CM_t を生成する(ステップS222)。特徴マップ変換部221は、 $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 PI_t から、 $H \times W \times C$ のサイズを有する特徴マップ CM_t を生成する。

10

【0032】

例えば、特徴マップ変換部221は、物体位置情報PIが入力された場合に特徴マップCMを出力する演算モデルを用いて、特徴マップ CM_t を生成してもよい。このような演算モデルの一例として、ニューラルネットワーク(例えば、CNN: Convolutional Neural Network)を用いた演算モデルがあげられる。この場合、演算モデルは、後述するように、学習用データを用いて学習されてもよい。つまり、演算モデルのパラメータは、適切な特徴マップCM(特に、後述する類似性行列AMを生成するのに適した特徴マップCM)を出力するように最適化されていてもよい。

20

【0033】

その後、「第1生成手段」の一具体例である特徴ベクトル変換部222は、ステップS222で生成された特徴マップ CM_t から、特徴ベクトル CV_t を生成する(ステップS223)。特徴ベクトル変換部222は、 $H \times W \times C$ のサイズを有する特徴マップ CM_t から、 $H \times W \times C$ のサイズを有する特徴ベクトル CV_t を生成する。尚、特徴ベクトル CV_t は、第1特徴ベクトルの一具体例である。

30

【0034】

ステップS221からステップS223までの処理と並行して又は相前後して、「第1生成手段」の一具体例である特徴マップ変換部223は、物体検出DB32から、時刻 t に撮影された画像 IM_t に映り込んでいる物体 O_t に関する物体位置情報 PI_t を取得する(ステップS224)。尚、物体位置情報 PI_t は、第2位置情報の一具体例である。その後、特徴マップ変換部223は、ステップS224で取得した物体位置情報 PI_t から、特徴マップ CM_t を生成する(ステップS225)。特徴マップ変換部223は、 $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 PI_t から、 $H \times W \times C$ のサイズを有する特徴マップ CM_t を生成する。尚、ステップS224からステップS225の処理の内容は、上述したステップS221からステップS222の処理の内容と同一であってもよい。このため、特徴マップ変換部223は、特徴マップ変換部221と同様に、物体位置情報PIが入力された場合に特徴マップCMを出力する演算モデルを用いて、特徴マップ CM_t を生成してもよい。

40

【0035】

その後、「第1生成手段」の一具体例である特徴ベクトル変換部224は、ステップS225で生成された特徴マップ CM_t から、特徴ベクトル CV_t を生成する(ステップS226)。特徴ベクトル変換部224は、 $H \times W \times C$ のサイズを有する特徴マップ CM_t から、 $H \times W \times C$ のサイズを有する特徴ベクトル CV_t を生成する。尚、ステップS226の処理の内容は、上述したステップS223の処理の内容と同一であってもよい。尚、特徴ベクトル CV_t は、第2特徴ベクトルの一具体例である。

50

【 0 0 3 6 】

その後、「第2生成手段」の一具体例である行列演算部225は、ステップS223で生成された特徴ベクトル $C V_{t-}$ とステップS226で生成された特徴ベクトル $C V_t$ とを用いて、類似性行列(Affinity Matrix)AMを生成する(ステップS227)。具体的には、行列演算部225は、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ とを用いた演算処理によって得られる情報を、類似性行列AMとして生成する。

【 0 0 3 7 】

例えば、行列演算部225は、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ との行列積を算出する演算処理によって得られる情報(つまり、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ との行列積)を、類似性行列AMとして生成してもよい。ここで言う行列積は、典型的には、テンソル積(言い換えれば、直積)であってもよい。一例として、行列積は、クロネッカー積であってもよい。この場合、類似性行列AMのサイズは、 $H W \times H W$ となる。

10

【 0 0 3 8 】

その結果、特徴ベクトル $C V_{t-}$ 及び $C V_t$ と類似性行列AMとの関係を概念的に示す平面図である図5に示すように、類似性行列AMは、物体 O_{t-} と物体 O_t との対応関係を示す情報となる。例えば、類似性行列AMは、(1)複数の物体 O_{t-} のうちの第1の物体 O_{t-} が、複数の物体 O_t のうちの第1の物体 O_t に対応しており(つまり、両者が同一の物体であり)、(2)複数の物体 O_{t-} のうちの第2の物体 O_{t-} が、複数の物体 O_t のうちの第2の物体 O_t に対応しており、 \dots 、(N)複数の物体 O_{t-} のうちの第Nの物体 O_{t-} が、複数の物体 O_t のうちの第Nの物体 O_t に対応していることを示す情報となる。尚、類似性行列AMは、物体 O_{t-} と物体 O_t との対応関係を示す情報であるがゆえに、対応情報と称してもよい。

20

【 0 0 3 9 】

具体的には、図5に示すように、類似性行列AMは、その縦軸が特徴ベクトル $C V_{t-}$ のベクトル成分に対応しており且つその横軸が特徴ベクトル $C V_t$ のベクトル成分に対応している行列であるとみなすことができる。このため、類似性行列AMの縦軸のサイズは、 $H W$ (つまり、特徴ベクトル $C V_{t-}$ のサイズであり、時刻 $t-$ に撮影された画像 $I M_{t-}$ のサイズ(つまり、画素数)に対応するサイズ)になる。同様に、類似性行列AMの横軸のサイズは、 $H W$ (つまり、特徴ベクトル $C V_t$ のサイズであり、時刻 t に撮影された画像 $I M_t$ のサイズ(つまり、画素数)に対応するサイズ)になる。言い換えれば、類似性行列AMは、その縦軸が時刻 $t-$ の画像 $I M_{t-}$ に映り込んでいる物体 O_{t-} の検出結果(つまり、物体 O_{t-} の検出位置)に対応しており且つその横軸が時刻 t の画像 $I M_t$ に映り込んでいる物体 O_t の検出結果(つまり、物体 O_t の検出位置)に対応している行列であるとみなすことができる。この場合、縦軸上のある物体 O_{t-} に対応するベクトル成分と横軸上の同じ物体 O_t に対応するベクトル成分とが交差する位置において、類似性行列AMの要素が反応する(典型的には、0でない値を有する)。言い換えれば、縦軸上の物体 O_{t-} の検出結果と横軸上の物体 O_t の検出結果とが交差する位置において、類似性行列AMの要素が反応する。つまり、類似性行列AMは、典型的には、特徴ベクトル $C V_{t-}$ に含まれるある物体 O_{t-} に対応するベクトル成分と特徴ベクトル $C V_t$ に含まれる同じ物体 O_t に対応するベクトル成分とが交差する位置の要素の値が、両ベクトル成分を掛け合わせることで得られる値(つまり、0ではない値)となる一方で、それ以外の要素の値が0になる行列となる。

30

40

【 0 0 4 0 】

例えば、図5に示す例では、特徴ベクトル $C V_{t-}$ に含まれる物体 $O \# k$ (但し、 k は、検出された物体 O の数であり、図5に示す例では、 $k = 1, 2, 3$ 又は 4)に対応するベクトル成分と特徴ベクトル $C V_t$ に含まれる同じ物体 $O \# k$ に対応するベクトル成分とが交差する位置において、類似性行列AMの要素が反応する。つまり、画像 $I M_{t-}$ に映り込んだ物体 $O \# k$ の検出結果と画像 $I M_t$ に映り込んだ物体 $O \# k$ の検出結果とが交差する位置において、類似性行列AMの要素が反応する。

50

【 0 0 4 1 】

逆に、特徴ベクトル CV_{t-} に含まれるある物体 O_{t-} に対応するベクトル成分と特徴ベクトル CV_t に含まれる同じ物体 O_t に対応するベクトル成分とが交差する位置において類似性行列 AM の要素が反応しない（典型的には、0 になる）場合には、画像 IM_{t-} に映り込んでいた物体 O_{t-} は、画像 IM_t には映り込んでいない（例えば、カメラの撮影画角外へ出てしまった）と推定される。

【 0 0 4 2 】

このように、類似性行列 AM は、物体 O_{t-} と物体 O_t との対応関係を示す情報として利用可能である。つまり、類似性行列 AM は、画像 IM_{t-} に映り込んでいる物体 O_{t-} と画像 IM_t に映り込んでいる物体 O_t との照合結果を示す情報として利用可能である。類似性行列 AM は、画像 IM_{t-} に映り込んでいた物体 O_{t-} の画像 IM_t 内での位置を追跡するための情報として利用可能である。

10

【 0 0 4 3 】

行列演算部 2 2 5 が生成した類似性行列 AM を示す情報は、物体照合部 2 2 による物体照合動作の結果を示す物体照合情報として、物体照合 DB 3 3 に格納される。或いは、行列演算部 2 2 5 は、類似性行列 AM に基づいて、物体 O_{t-} と物体 O_t との対応関係を示す他の種類の情報を生成し、物体 O_{t-} と物体 O_t との対応関係を示す他の種類の情報が物体照合情報として物体照合 DB 3 3 に格納されてもよい。

【 0 0 4 4 】

(2 - 3) リファイン動作

続いて、「補正手段」の一具体例であるリファイン部 2 3 が行うリファイン動作について説明する。リファイン部 2 3 は、物体検出 DB 3 2 に格納されている物体位置情報 PI を読み出し、物体照合部 2 2 から類似性行列 AM を取得し、取得した類似性行列 AM を用いて、読み出した物体位置情報 PI を補正する。以下、物体位置情報 PI_{t-} 及び PI_t に基づいて生成された類似性行列 AM を用いて物体位置情報 PI_t を補正するリファイン動作について、図 6 を参照しながら説明する。図 6 は、リファイン動作の流れを示すフローチャートである。

20

【 0 0 4 5 】

図 6 に示すように、特徴マップ変換部 2 3 6 は、物体検出 DB 3 2 に格納されている物体位置情報 PI_t を取得する（つまり、読み出す）（ステップ S 2 3 1）。その後、特徴マップ変換部 2 3 6 は、ステップ S 2 3 1 で取得した物体位置情報 PI_t から、特徴マップ CM_t を生成する（ステップ S 2 3 2）。ここで、上述したように、物体照合部 2 2 が備える特徴マップ変換部 2 2 3 もまた、物体位置情報 PI_t から、特徴マップ CM_t を生成している。しかしながら、特徴マップ変換部 2 2 3 は、類似性行列 AM を生成する目的（つまり、物体照合動作を行う目的）で、特徴マップ CM_t を生成している。特徴マップ変換部 2 2 3 は、類似性行列 AM を生成する目的で、特徴マップ CM_t を生成している。一方で、特徴マップ変換部 2 3 6 は、物体照合動作において生成された類似性行列 AM を用いて物体位置情報 PI_t を補正する目的（つまり、リファイン動作を行う目的）で、特徴マップ CM_t' を生成している。このため、後述する学習動作によって夫々学習される（つまり、更新される）特徴マップ変換部 2 2 3 及び 2 3 6 は、特徴マップ変換部 2 2 3 が、類似性行列 AM を生成するのにより適した特徴マップ CM_t を生成する一方で、特徴マップ変換部 2 3 6 が、物体位置情報 PI_t を補正するのにより適した特徴マップ CM_t' を生成するという点で異なっている。つまり、特徴マップ CM_t 及び CM_t' は、特徴マップ CM_t が、類似性行列 AM を生成するのにより適している一方で、特徴マップ CM_t' が、物体位置情報 PI_t を補正するのにより適しているという点で異なっている。このため、リファイン部 2 3 は、物体照合部 2 2 が備える特徴マップ変換部 2 2 3 とは別個に、特徴マップ変換部 2 3 6 を備えている。尚、このような特徴マップ変換部 2 2 3 及び 2 3 6 は、後述する学習動作の結果として構築される。つまり、後述する学習動作により、特徴マップ変換部 2 2 3 が、類似性行列 AM を生成するのにより適した特徴マップ CM_t を生成するように学習される一方で、特徴マップ変換部 2 3 6 が、物体位置情報 PI_t を補正す

30

40

50

るのにより適した特徴マップ CM_t' を生成するように学習される。

【0046】

特徴マップ変換部 236 は、物体位置情報 PI が入力された場合に特徴マップ CM を出力する演算モデルを用いて、特徴マップ CM_t' を生成してもよい。このような演算モデルの一例として、ニューラルネットワーク（例えば、CNN: Convolutional Neural Network）を用いた演算モデルがあげられる。この場合、演算モデルは、後述するように、学習用データを用いて学習されてもよい。つまり、演算モデルのパラメータは、（特に、物体位置情報 PI_t を補正するのに）適切な特徴マップ CM_t' を出力するように最適化されていてもよい。

【0047】

その後、特徴ベクトル変換部 237 は、ステップ S 232 で生成された特徴マップ CM_t' から、特徴ベクトル CV_t' を生成する（ステップ S 233）。

【0048】

ステップ S 231 からステップ 233 までの処理と並行して又は相前後して、行列演算部 231 は、物体照合部 22 から、物体照合部 22 が生成した類似性行列 AM を取得する（ステップ S 234）。或いは、行列演算部 231 は、物体照合 DB 33 から、物体照合部 22 が生成した類似性行列 AM を取得してもよい（ステップ S 234）。

【0049】

その後、行列演算部 231 は、ステップ S 233 で生成された特徴ベクトル CV_t' とステップ S 234 で取得された類似性行列 AM とを用いて、特徴ベクトル CV_res を生成する（ステップ S 235）。具体的には、行列演算部 231 は、特徴ベクトル CV_t' と類似性行列 AM とを用いた演算処理によって得られる情報を、特徴ベクトル CV_res として生成する。

【0050】

例えば、行列演算部 231 は、特徴ベクトル CV_t' と類似性行列 AM との行列積を算出する演算処理によって得られる情報（つまり、特徴ベクトル CV_t' と類似性行列 AM との行列積）を、特徴ベクトル CV_res として生成してもよい。

【0051】

その後、特徴ベクトル変換部 232 は、ステップ S 235 で生成された特徴ベクトル CV_res から、特徴マップ CM と同じサイズを有する特徴マップ CM_res を生成する（ステップ S 236）。つまり、特徴ベクトル変換部 232 は、任意のサイズを有する特徴ベクトル CV_res から、 $H \times W \times C$ のサイズを有する特徴マップ CM_res を生成する。例えば、特徴ベクトル変換部 232 は、特徴ベクトル CV_res を特徴マップ CM_res に変換することで、特徴マップ CM_res を生成してもよい。

【0052】

その後、特徴マップ変換部 233 は、ステップ S 236 で生成された特徴マップ CM_res から、物体位置情報 PI_t と同じサイズを有する物体位置情報 PI_t_res を生成する（ステップ S 237）。つまり、特徴マップ変換部 233 は、ステップ S 236 で生成された特徴マップ CM_res から、 $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 PI_t_res を生成する。特徴マップ変換部 233 は、ステップ S 236 で生成された特徴マップ CM_res から、(i) $H \times W \times K$ のサイズを有する、画像 IM_t 内での物体 O_t の中心位置 KP を示すマップ情報と、(ii) $H \times W \times 2$ のサイズを有する、画像 IM_t 内での物体 O_t の検出枠 BB のサイズを示すマップ情報と、(iii) $H \times W \times 2$ のサイズを有する、画像 IM_t 内での物体 O_t の検出枠 BB の補正量を示すマップ情報とを含む物体位置情報 PI_t_res を生成する。つまり、特徴マップ変換部 233 は、ステップ S 236 で生成された特徴マップ CM_res の次元を変換することで、特徴マップ CM_res から物体位置情報 PI_t_res を生成する。

【0053】

例えば、特徴マップ変換部 233 は、特徴マップ CM_res を物体位置情報 PI_t_res に変換することで、物体位置情報 PI_t_res を生成してもよい。特徴マップ変

10

20

30

40

50

換部 233 は、特徴マップ CM が入力された場合に物体位置情報 PI を出力する演算モデルを用いて、物体位置情報 PI_t_res を生成してもよい。このような演算モデルの一例として、ニューラルネットワーク（例えば、CNN: Convolutional Neural Network）を用いた演算モデルがあげられる。この場合、演算モデルは、後述するように、学習用データを用いて学習されてもよい。つまり、演算モデルのパラメータは、適切な物体位置情報 PI_t_res を出力するように最適化されていてもよい。

【0054】

ステップ S237 の処理は、実質的には、類似性行列 AM を重みとして用いる注意機構 (Attention Mechanism) を用いて、物体位置情報 PI_t_res を生成する処理と等価であるとみなしてもよい。つまり、リファイン部 23 は、行列演算部 231、特徴ベクトル変換部 232 及び特徴マップ変換部 233 を含む注意機構の少なくとも一部を構成しているとみなしてもよい。物体位置情報 PI_res は、リファインされた物体位置情報 PI_t として用いられてもよい。この場合、ステップ S237 の処理は、実質的には、類似性行列 AM を重みとして用いる注意機構を用いて物体位置情報 PI_t を補正する（言い換えれば、更新する、調整する又は改善する）処理と等価であるとみなしてもよい。

【0055】

但し、ステップ S235 で生成された物体位置情報 PI_t_res は、オリジナルの物体位置情報 PI_t （つまり、リファイン動作が施されていない物体位置情報 PI_t ）に含まれていた情報が消失している可能性がある。なぜならば、物体位置情報 PI_t_res は、注意機構において注意を払うべき部分（本実施形態では、物体 O の検出位置）を示す類似性行列 AM が重みとして用いられるがゆえに、物体検出情報のうちの物体 O の検出位置に関する情報とは異なる情報部分が消失してしまう可能性があるからである。そこで、本実施形態では、リファイン部 23 は、オリジナルの物体位置情報 PI_t に含まれていた情報の消失を抑制するための処理を更に行ってもよい。具体的には、残差処理部 234 は、ステップ S237 で生成された物体位置情報 PI_t_res をオリジナルの物体位置情報 PI_t に加算することで、物体位置情報 PI_t_ref を生成してもよい（ステップ S238）。物体位置情報 PI_t_ref は、物体位置情報 PI_t と同じサイズを有する。このため、残差処理部 234 は、 $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 PI_t_ref を生成する。このような $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 PI_t_ref を生成するために、残差処理部 234 は、(i) 物体位置情報 PI_t_res に含まれる物体 O_t の中心位置 KP を示すマップ情報と、オリジナルの物体位置情報 PI_t に含まれる物体 O_t の中心位置 KP を示すマップ情報とを加算し、(ii) 物体位置情報 PI_t_res に含まれる物体 O_t の検出枠 BB のサイズを示すマップ情報と、オリジナルの物体位置情報 PI_t に含まれる物体 O_t の検出枠 BB のサイズを示すマップ情報とを加算し、(iii) 物体位置情報 PI_t_res に含まれる検出枠 BB の補正量を示すマップ情報と、オリジナルの物体位置情報 PI_t に含まれる検出枠 BB の補正量を示すマップ情報とを加算する。

【0056】

この場合、ステップ S238 の処理は、実質的には、残差処理部 234 を含む残差処理機構 (Residual Attention Mechanism) を用いて、物体位置情報 PI_t_ref を生成する処理と等価であるとみなしてもよい。つまり、リファイン部 23 は、行列演算部 231、特徴ベクトル変換部 232、特徴マップ変換部 233 及び残差処理部 234 を含む残差注意機構の少なくとも一部を構成しているとみなしてもよい。物体位置情報 PI_t_ref は、リファインされた物体位置情報 PI_t であると共に、オリジナルの物体位置情報 PI_t に含まれていた情報をも含む。この場合、物体位置情報 PI_t_ref は、リファインされた物体位置情報 PI_t として用いられてもよい。但し、リファイン部 23 は、オリジナルの物体位置情報 PI_t に含まれていた情報の消失を抑制するための処理（ステップ S238 の処理）を行わなくてもよい。この場合、リファ

10

20

30

40

50

イン部 23 は、残差処理部 234 を備えていなくてもよい。

【0057】

更に、特徴マップ変換部 235 は、物体位置情報 $P I_t_ref$ から、物体位置情報 $P I_t_ref$ と同じサイズ： $H \times W \times (K + 4)$ を有する物体位置情報 $P I_t_ref'$ を生成してもよい（ステップ S239）。物体位置情報 $P I_t_ref'$ は、物体位置情報 $P I_t$ と同じサイズを有する。このため、特徴マップ変換部 235 は、 $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 $P I_t_ref'$ を生成する。典型的には、特徴マップ変換部 235 は、畳み込み処理（例えば、ニューラルネットワークを構成する畳み込み層で行われる処理）を物体位置情報 $P I_t_ref$ に施すことで、物体位置情報 $P I_t_ref'$ を生成してもよい。特徴マップ変換部 235 は、ニューラルネットワークを構成する畳み込み層に物体位置情報 $P I_t_ref$ を入力することで、物体位置情報 $P I_t_ref$ を物体位置情報 $P I_t_ref'$ に変換してもよい。この場合、物体位置情報 $P I_t_ref'$ は、リファインされた物体位置情報 $P I_t$ として用いられてもよい。但し、リファイン部 23 は、特徴マップ変換部 235 を備えていなくてもよい。

【0058】

物体位置情報 $P I_t_res$ 、物体位置情報 $P I_t_ref$ 又は物体位置情報 $P I_t_ref'$ がリファインされた物体位置情報 $P I_t$ として用いられる場合には、物体検出 DB32 に格納されている物体位置情報 $P I_t$ が、物体位置情報 $P I_t_res$ 、物体位置情報 $P I_t_ref$ 又は物体位置情報 $P I_t_ref'$ に置き換えられてもよい。この場合、物体照合部 22 は、リファインされる前の物体位置情報 $P I_t$ に代えて、リファインされた物体位置情報 $P I_t$ を用いて、時刻 t に撮影された画像 $I M_t$ に映り込んでいる物体 O_t と、時刻 $t + 1$ に撮影された画像 $I M_{t+1}$ に映り込んでいる物体 O_{t+1} とを照合する物体照合動作を行うことができる。更には、物体 O_t と物体 O_{t+1} とを照合する物体照合動作の結果を用いて、物体 O_{t+1} に関する物体位置情報 $P I_{t+1}$ がリファインされる。このため、物体照合部 22 は、リファインされる前の物体位置情報 $P I_{t+1}$ に代えて、リファインされた物体位置情報 $P I_{t+1}$ を用いて、時刻 $t + 1$ に撮影された画像 $I M_{t+1}$ に映り込んでいる物体 O_{t+1} と、時刻 $t + 2$ に撮影された画像 $I M_{t+2}$ に映り込んでいる物体 O_{t+2} とを照合する物体照合動作を行うことができる。つまり、物体検出 DB32 に格納されている物体位置情報 $P I_t$ が、物体位置情報 $P I_t_res$ 、物体位置情報 $P I_t_ref$ 又は物体位置情報 $P I_t_ref'$ に置き換えられる場合には、リファインされた物体位置情報 $P I_t$ が連鎖的に物体照合部 22 に入力される。このため、リファインされる前の物体位置情報 $P I_t$ を用いて物体照合動作が行われる場合と比較して、物体 O の照合の精度の向上と、物体追跡動作に要する処理コストの低減とが期待される。

【0059】

（3）物体追跡装置 1 の技術的効果

以上説明したように、本実施形態の物体追跡装置 1 は、物体位置情報 $P I_t$ 及び $P I_t$ から類似性行列 $A M$ を生成し、類似性行列 $A M$ を用いて物体位置情報 $P I_t$ をリファインすることができる。つまり、物体追跡装置 1 は、物体検出部 21 の出力である物体位置情報 $P I_t$ 及び $P I_t$ に対して前処理等を行うことなく、物体照合動作を行うことができる。言い換えれば、物体追跡装置 1 は、物体検出部 21 の出力である物体位置情報 $P I_t$ 及び $P I_t$ を、物体照合動作を行う物体照合部 22 の入力としてそのまま用いることができる。同様に、物体追跡装置 1 は、物体照合部 22 の出力である類似性行列 $A M$ に対して前処理等を行うことなく、リファイン動作を行うことができる。言い換えれば、物体追跡装置 1 は、物体照合部 22 の出力である類似性行列 $A M$ 及び物体検出部 21 の出力である物体位置情報 $P I_t$ を、リファイン動作を行うリファイン部 23 の入力としてそのまま用いることができる。このため、物体検出動作、物体照合動作及びリファイン動作を互いに独立した別個の三つの動作として行う比較例の物体追跡装置と比較して、物体追跡装置 1 は、物体検出動作、物体照合動作及びリファイン動作の夫々を行うための前処理等を行わなくともよくなる。その結果、物体追跡装置 1 は、相対的に低い処理コストで物体 O を追跡することができる。

10

20

30

40

50

【 0 0 6 0 】

尚、物体追跡装置 1 は、実質的には、物体追跡動作と一般的な注意機構で行われる動作とが似ていることに着目し、物体照合動作において生成される情報を用いてリファイン動作を行っていると言える。具体的には、物体追跡動作では、上述したように、物体 O を検出する処理、物体 O を照合する処理及び物体 O の検出結果をリファインする処理が行われる。一方で、一般的な注意機構では、物体 O の特徴を抽出する処理、重みを算出する処理及び物体 O の特徴の抽出結果をリファインする処理が行われる。物体追跡装置 1 は、注意機構において重みを算出する処理を、実質的には、物体追跡動作において物体 O を照合する処理としても流用している。言い換えれば、物体追跡装置 1 は、物体追跡動作において物体 O を照合する処理を、実質的には、注意機構において重みを算出する処理としても流用している。従って、物体追跡装置 1 は、物体検出動作、物体照合動作及びリファイン動作を、注意機構を用いて実現しているとも言える。具体的には、物体追跡装置 1 は、物体位置情報 $P I_{t-}$ をクエリとし、物体位置情報 $P I_t$ をキー及びバリューとし、且つ、類似性行列 $A M$ を重みとして用いる注意機構を用いて、物体追跡動作を行っているとも言える。このため、物体追跡装置 1 は、物体検出動作、物体照合動作及びリファイン動作を含む物体追跡動作を、一連の動作として行うことができる。つまり、物体追跡装置 1 は、物体検出動作、物体照合動作及びリファイン動作を行う単一のネットワーク構造（いわゆる、エンドツーエンドなシングルステージ型のネットワーク構造）を用いて、物体追跡動作を行うことができる。このように注意機構を用いて物体追跡動作（特に、リファイン動作）を行うことができるがゆえに、物体追跡装置 1 は、注意機構を用いることなく物体追跡動作が行われる場合（つまり、物体検出動作を行うネットワーク構造、物体照合動作を行うネットワーク構造及びリファイン動作を行うネットワーク構造が別個独立に用いられる場合）と比較して、物体位置情報 $P I$ をより適切にリファインすることができる。

【 0 0 6 1 】

また、比較例の物体追跡装置（特に、類似性行列 $A M$ を生成することなく物体照合動作を行う物体追跡装置）は、時刻 $t -$ に撮影された画像に N_{t-} 個の物体が映り込んでおり且つ時刻 t に撮影された画像に N_t 個の物体が映り込んでいる場合には、 N_{t-} 個の物体の夫々と N_t 個の物体の夫々とを個別に照合する必要がある。つまり、比較例の物体追跡装置は、二つの物体を照合する動作を、 $N_{t-} \times N_t$ 回も繰り返す必要がある。このため、物体を追跡するための処理コストが高くなってしまいう可能性があるという技術的問題が存在する。一方で、本実施形態の物体追跡装置 1 は、特徴ベクトル $C V_{t-}$ 及び $C V_t$ を用いて、時刻 $t -$ に撮影された画像 $I M_{t-}$ に映り込んでいる物体 O_{t-} と時刻 t に撮影された画像 $I M_t$ に映り込んでいる物体 O_t とを照合する物体照合動作を行うことができる。このため、物体追跡装置 1 は、相対的に低い処理コストで物体 O を追跡することができる。具体的には、物体追跡装置 1 は、画像 $I M_{t-}$ に N_{t-} 個（尚、 N_{t-} は、1 以上の整数）の物体 O_{t-} が映り込んでおり且つ画像 $I M_t$ に N_t 個（尚、 N_t は、1 以上の整数）の物体 O_t が映り込んでいる場合であっても、 N_{t-} 個の物体 O_{t-} の夫々と N_t 個の物体 O_t の夫々とを個別に照合しなくてもよくなる。つまり、物体追跡装置 1 は、二つの物体 O_{t-} 及び O_t を照合する動作を、 $N_{t-} \times N_t$ 回も繰り返さなくてもよくなる。つまり、物体追跡装置 1 は、特徴ベクトル $C V_{t-}$ 及び $C V_t$ を用いて類似性行列 $A M$ を生成する処理を一度行えば、 N_{t-} 個の物体 O_{t-} の夫々と N_t 個の物体 O_t の夫々との照合を完了することができる。このため、比較例の物体追跡装置と比較して、相対的に低い処理コストで物体 O を追跡することができる。特に、比較例の物体追跡装置では、画像 $I M$ に映り込んでいる物体 O の数が増えるほど処理コストが指数関数的に増加する一方で、本実施形態の物体追跡装置 1 では、処理コストが画像 $I M$ に映り込んでいる物体 O の数に依存しにくくなる。このため、画像 $I M$ に映り込んでいる物体 O の数が多くなるほど、物体追跡装置 1 による処理コストの低減効果が大きくなる。

【 0 0 6 2 】

(4) 変形例

(4 - 1) 第 1 変形例

10

20

30

40

50

初めに、図7を参照しながら、第1変形例の物体追跡装置1aについて説明する。図7は、第1変形例の物体追跡装置1aの構成を示すブロック図である。

【0063】

図7に示すように、第1変形例の物体追跡装置1aは、演算装置2内に実現される論理的な機能ブロックとして、物体検出部21、物体照合部22及びリファイン部23の学習動作を行う学習部24aを更に備えているという点で異なっている。更に、物体追跡装置1aは、学習動作のための学習用データ341aを格納するための学習DB34aが記憶装置3によって記憶されているという点で異なっている。物体追跡装置1aのその他の特徴は、物体追跡装置1のその他の特徴と同一であってもよい。尚、学習動作は、典型的には、物体追跡装置1aが実際に物体追跡動作を行う前に予め行われる。但し、学習動作は、物体追跡装置1aが物体追跡動作を開始した後の所望のタイミングで行われてもよい。

10

【0064】

学習用データ341aは、例えば、図8に示すように、ある時刻に取得された(例えば、撮影された)画像IMを含む。更に、学習用データ341aは、画像IMに映り込んでいる物体Oの位置の正解ラベルを示す物体位置情報PI_labelを含む。図8に示す例では、学習DB34aには、時刻sに取得された画像IM_s及び画像IM_sに映り込んでいる物体O_sの位置に関する物体位置情報PI_s_labelを含む学習用データ341aと、時刻s+に取得された画像IM_{s+}と、画像IM_{s+}に映り込んでいる物体O_{s+}の位置に関する物体位置情報PI_{s+}_labelとを含む学習用データ341aとを含む複数の学習用データ341aが格納されている。尚、各画像IMに複数の物体Oが映り込んでいる場合には、学習用データ341aは、複数の物体Oの位置の正解ラベルを夫々示す複数の物体位置情報PI_labelを含んでいてもよい。

20

【0065】

学習部24aは、学習動作を行うために、学習用データ341aに含まれる画像IM_s及びIM_{s+}を物体検出部21に入力する。その結果、物体検出部21は、物体O_sの位置に関する物体位置情報PI_sと、物体O_{s+}の位置に関する物体位置情報PI_{s+}とを出力する。更に、物体照合部22は、物体位置情報PI_sと物体位置情報PI_{s+}とを用いて、類似性行列AMを生成する。更に、リファイン部23は、類似性行列AMを用いて、物体位置情報PI_{s+}をリファインする。

【0066】

学習部24aは、少なくともリファインされた物体位置情報PI_{s+}に関する損失関数L₁に基づいて、物体検出部21が物体検出動作を行うために用いる、物体照合部22が物体照合動作を行うために用いる及び/又はリファイン部23がリファイン動作を行うために用いる学習可能な演算モデルを更新する。つまり、学習部24aは、損失関数L₁に基づいて、物体検出部21、物体照合部22及びリファイン部23の少なくとも一つの動作内容を規定する演算モデルを更新する。一例として、損失関数L₁が、リファインされた物体位置情報PI_{s+}と正解ラベルとしての物体位置情報PI_{s+}_labelとの誤差が小さくなるほど小さくなる損失関数である場合には、学習部24aは、損失関数L₁が小さくなるように(典型的には、最小になるように)演算モデルを更新してもよい。

30

【0067】

本実施形態では、上述したように、物体検出部21、物体照合部22及びリファイン部23が単一のネットワーク構造(つまり、単一の演算モデル)によって実現可能である。このため、物体検出部21、物体照合部22及びリファイン部23が単一のネットワーク構造(つまり、単一の演算モデル)によって実現されている場合には、学習部24aは、当該単一の演算モデルを更新してもよい。尚、このような演算モデルの一例として、ニューラルネットワーク(例えば、CNN: Convolutional Neural Network)を用いた演算モデルがあげられる。この場合、演算モデルを更新する動作は、演算モデルのパラメータを更新する、決定する又は調整する動作を含んでいてもよい。演算モデルのパラメータは、ニューラルネットワークが備えるノード間の重み、各ノードで付与されるバイアス及びノード間の接続経路の少なくとも一つを含んでいてもよい。

40

50

【 0 0 6 8 】

学習部 2 4 a は、学習 DB 3 4 a に格納されている複数の学習データ 3 4 1 a に夫々含まれる複数の画像 IM を、複数の画像 IM に対応する時刻の順に物体検出部 2 1 に順次入力することで、学習動作を繰り返す。つまり、学習部 2 4 a は、画像 IM_s 及び IM_{s+} を物体検出部 2 1 に入力し、リファインされた物体位置情報 PI_{s+} に関する損失関数 L_1 に基づいて演算モデルを更新する。その後、学習部 2 4 a は、画像 IM_{s+} 及び IM_{s+2} を物体検出部 2 1 に入力し、リファインされた物体位置情報 PI_{s+2} に関する損失関数 L_1 に基づいて演算モデルを更新する。以降、学習部 2 4 a は、同様の動作を繰り返す。その結果、物体検出部 2 1、物体照合部 2 2 及びリファイン部 2 3 の少なくとも一つの動作内容を規定する演算モデルが適切に更新される（つまり、学習される）。

10

【 0 0 6 9 】

学習部 2 4 a は、物体検出部 2 1 に入力される二つの画像 IM に夫々対応する二つの時刻の間の時間間隔を変更してもよい。例えば、上述した説明では、学習部 2 4 a は、の時間間隔を隔てる二つの時刻に夫々取得された二つの画像 IM（例えば、画像 IM_s 及び IM_{s+} ）を物体検出部 2 1 に入力している。しかしながら、学習部 2 4 a は、 $m \times$ （尚、 m は、学習部 2 4 a が変更可能な係数であり、例えば、1、2、3、・・・等の整数）の時間間隔を隔てる二つの時刻に夫々取得された二つの画像 IM（例えば、画像 IM_s 及び IM_{s+m} ）を物体検出部 2 1 に入力してもよい。この場合、学習部 2 4 a は、画像 IM_s 及び IM_{s+m} を物体検出部 2 1 に入力し、リファインされた物体位置情報 PI_{s+m} に関する損失関数 L_1 に基づいて演算モデルを更新してもよい。その後、学習部 2 4 a は、画像 IM_{s+m} 及び IM_{s+2m} を物体検出部 2 1 に入力し、リファインされた物体位置情報 PI_{s+2m} に関する損失関数 L_1 に基づいて演算モデルを更新してもよい。以降、学習部 2 4 a は、同様の動作を繰り返してもよい。この場合、物体検出部 2 1 に入力される二つの画像 IM の間での物体 O の移動量が、係数 m に応じて変わる。1 回の学習動作ごと、すなわち、学習部 2 4 a が二つの画像 IM（例えば、画像 IM_s 及び IM_{s+} ）を物体検出部 2 1 に入力するごとに、係数 m は例えば乱数によって決定されてもよい。その結果、物体検出部 2 1、物体照合部 2 2 及びリファイン部 2 3 の少なくとも一つの動作内容を規定する演算モデルは、様々な移動速度で移動する物体を追跡できるように更新される。

20

【 0 0 7 0 】

学習部 2 4 a は、リファインされた物体位置情報 PI_{s+} に関する損失関数 L_1 に加えて、物体検出部 2 1 が出力した物体位置情報 PI_s に関する損失関数 L_2 及び物体検出部 2 1 が出力した物体位置情報 PI_{s+} に関する損失関数 L_3 に基づいて、演算モデルを更新してもよい。損失関数 L_2 は、例えば、物体検出部 2 1 が出力した物体位置情報 PI_s と正解ラベルとしての物体位置情報 PI_{s_label} との誤差が小さくなるほど小さくなる損失関数であってもよい。損失関数 L_3 は、例えば、物体検出部 2 1 が出力した物体位置情報 PI_{s+} と正解ラベルとしての物体位置情報 PI_{s+_label} との誤差が小さくなるほど小さくなる損失関数であってもよい。この場合、学習部 2 4 a は、損失関数 L_1 から L_3 の総和が小さくなるように（典型的には、最小になるように）演算モデルを更新してもよい。

30

【 0 0 7 1 】

損失関数 L_1 から L_3 を用いる場合には、学習部 2 4 a は、損失関数 L_1 から L_3 に対して重み付け処理を施し、重み付け処理が施された損失関数 L_1 から L_3 に基づいて演算モデルを更新してもよい。つまり、学習部 2 4 a は、 1 、 2 及び 3 という重み付け係数を用いて損失関数 L_1 から L_3 に対して夫々重み付け処理を施し、 $1 \times L_1 + 2 \times L_2 + 3 \times L_3$ という数式で特定される損失関数に基づいて演算モデルを更新してもよい。

40

【 0 0 7 2 】

一例として、損失関数 L_1 及び L_3 の夫々は、物体位置情報 PI_{s+} に関する損失関数である一方で、損失関数 L_2 は、物体位置情報 PI_s に関する損失関数である。演算モデ

50

ルを適切に更新するという観点から言えば、演算モデルの更新に対する物体位置情報 PI_s の寄与度（つまり、損失関数 L_1 及び L_3 の寄与度）と、演算モデルの更新に対する物体位置情報 PI_{s+} の寄与度（つまり、損失関数 L_2 の寄与度）とは、大きく異なっていないことが好ましい。典型的には、演算モデルの更新に対する物体位置情報 PI_s の寄与度は、演算モデルの更新に対する物体位置情報 PI_{s+} の寄与度と同じであることが好ましい。このため、学習部 24a は、損失関数 L_1 と損失関数 L_3 との総和に対する重みと、損失関数 L_2 に対する重みとが同じになるように、重み付け処理を行ってもよい。一例として、学習部 24a は、 $0.5 \times (L_1 + L_3) + 0.5 \times L_2$ という数式を用いて損失関数 L_1 から L_3 に対して重み付け処理を施し、 $0.5 \times (L_1 + L_3) + 0.5 \times L_2$ という数式で特定される損失関数に基づいて演算モデルを更新してもよい。このような重み付け処理は、後述する第 3 変形例においてソフトマックス関数を用いて類似性行列 AM が正規化される場合に特に有益である。以下、その理由である。まず、学習動作の初期においてソフトマックス関数を用いて類似性行列 AM が正規化されると、類似性行列 AM の殆どの要素がゼロになってしまう。その結果、リファインされた物体位置情報 PI_{s+} と物体検出部 21 が検出した物体位置情報 PI_{s+} とが同一になってしまう。このような状況下において損失関数 L_1 から L_3 の重みが全て同一のままである（例えば、全て 1 である）と仮定すると、演算モデルの学習効果が薄れてしまう。なぜならば、損失関数 L_1 と損失関数 L_3 の値が同じであるがゆえに、見かけ上、演算モデルの更新に対する損失関数 L_2 の寄与度が、演算モデルの更新に対する損失関数 L_1 及び L_3 の寄与度に対して小さくなってしまふからである。そこで、学習部 24a は、学習動作の初期の期間（例えば、正規化処理によって似性行列 AM の殆どの要素がゼロになってしまう期間）は、損失関数 L_1 と損失関数 L_3 との総和に対する重みと、損失関数 L_2 に対する重みとが同じになるように、重み付け処理を行ってもよい。その結果、学習部 24a は、学習動作の初期の期間（例えば、正規化処理によって似性行列 AM の殆どの要素がゼロになってしまう期間）においても、学習効果が薄れにくくなる。

【0073】

尚、演算モデルの更新に対する損失関数 L_1 の寄与度と、演算モデルの更新に対する損失関数 L_3 の寄与度とを同じにするために、学習部 24a は、 $0.25 \times L_1 + 0.25 \times L_3 + 0.5 \times L_2$ という数式を用いて損失関数 L_1 から L_3 に対して重み付け処理を施し、 $0.25 \times L_1 + 0.25 \times L_3 + 0.5 \times L_2$ という数式で特定される損失関数に基づいて演算モデルを更新してもよい。

【0074】

（4-2）第 2 変形例

上述した説明では、物体追跡装置 1 は、物体検出動作を含む物体追跡動作を行っている。しかしながら、第 2 変形例では、物体追跡動作は、物体検出動作を含んでいなくてもよい。つまり、物体追跡装置 1 は、物体検出動作を行わなくてもよい。この場合、物体追跡装置 1 の第 2 変形例である物体追跡装置 1b の構成を示す図 9 に示すように、物体追跡装置 1b は、物体検出部 21 を備えていなくてもよい。更には、記憶装置 3 には、物体検出動作を行うために用いられる画像 IM を格納するための画像 $DB31$ 及び物体検出動作の結果を示す物体検出情報を格納するための物体検出 $DB32$ が記憶されていなくてもよい。物体追跡装置 1b（特に、物体照合部 22）は、物体追跡装置 1b とは異なる装置が行った物体検出動作の結果を示す物体検出情報を用いて、物体照合動作を行ってもよい。

【0075】

上述した説明では、物体追跡装置 1 は、リファイン動作を含む物体追跡動作を行っている。しかしながら、第 2 変形例では、物体追跡動作は、リファイン動作を含んでいなくてもよい。つまり、物体追跡装置 1 は、リファイン動作を行わなくてもよい。この場合、物体追跡装置 1 の第 2 変形例である物体追跡装置 1b の構成を示す図 9 に示すように、物体追跡装置 1b は、リファイン部 23 を備えていなくてもよい。リファイン動作は、物体追跡装置 1b とは異なる装置によって行われてもよい。この場合、物体追跡装置 1b（特に、物体照合部 22）は、物体照合動作の結果を示す物体照合情報を、物体追跡装置 1b と

は異なる、リファイン動作を行う装置に対して出力してもよい。

【0076】

(4-3) 第3変形例

行列演算部225は、特徴ベクトル CV_t と特徴ベクトル CV_t を用いた演算処理によって得られる類似性行列 AM を正規化してもよい。例えば、行列演算部225は、特徴ベクトル CV_t と特徴ベクトル CV_t との行列積を正規化することで、類似性行列 AM を正規化してもよい。

【0077】

行列演算部225は、類似性行列 AM に対して任意の正規化処理を行ってもよい。例えば、行列演算部225は、類似性行列 AM に対して、シグモイド関数を用いた正規化処理を行ってもよい。この場合、類似性行列 AM の各要素が、シグモイド関数を用いて正規化される。

10

【0078】

或いは、例えば、行列演算部225は、類似性行列 AM に対して、ソフトマックス($softmax$)関数を用いた正規化処理を行ってもよい。具体的には、図10に示すように、行列演算部225は、類似性行列 AM の各行の複数の要素から構成される列ベクトル成分と、類似性行列 AM の各列の複数の要素から構成される列ベクトル成分との夫々に対して、ソフトマックス関数を用いた正規化処理を行ってもよい。この場合、行列演算部225は、列ベクトル成分を構成する複数の要素の総和が1になるように、列ベクトル成分に対してソフトマックス関数を用いた正規化処理を行い、行ベクトル成分を構成する複数の要素の総和が1になるように、行ベクトル成分に対してソフトマックス関数を用いた正規化処理を行ってもよい。その後、正規化処理が行われたれベクトル成分と行ベクトル成分とを掛け合わせることで得られる要素を含む行列が、正規化処理が行われた類似性行列 AM となる。一例として、行列演算部225は、類似性行列 AM の物体 O_t に対応する一の行の複数の要素から構成される列ベクトル成分と、類似性行列 AM の同じ物体 O_t に対応する一の列の複数の要素から構成される列ベクトル成分との夫々に対して、ソフトマックス関数を用いた正規化処理を行ってもよい。この場合、行列演算部225は、物体 O_t に対応する列ベクトル成分を構成する複数の要素の総和が1になるように、列ベクトル成分に対してソフトマックス関数を用いた正規化処理を行い、物体 O_t に対応する行ベクトル成分を構成する複数の要素の総和が1になるように、行ベクトル成分に対してソフトマックス関数を用いた正規化処理を行ってもよい。その後、正規化処理が行われたれベクトル成分と行ベクトル成分とを掛け合わせることで得られる要素を含む行列が、正規化処理が行われた類似性行列 AM となる。

20

30

【0079】

(4-4) その他の変形例

上述した説明では、物体検出動作を行うことで物体検出部21が生成する物体位置情報 PI は、物体 O の中心位置 KP を示す情報、物体 O の検出枠 BB のサイズを示す情報及び検出枠 BB の補正量を示す情報を含んでいる。しかしながら、物体位置情報 PI は、物体 O の検出枠 BB のサイズを示す情報及び検出枠 BB の補正量を示す情報の少なくとも一方を含んでいなくてもよい。物体位置情報 PI は、物体 O の中心とは異なる部分の位置に関する情報を含んでいてもよい。

40

【0080】

上述した説明では、物体照合部22は、物体検出動作の結果を示す物体位置情報 PI を用いて、特徴ベクトル CV を生成している。ここで、上述したように物体検出動作においてニューラルネットワーク等の演算モデルが用いられる場合には、物体照合部22は、演算モデルの最終出力に相当する物体位置情報 PI に加えて、演算モデルの中間出力を用いて、特徴ベクトル CV を生成してもよい。例えば、物体照合部22は、演算モデルとして用いられるニューラルネットワークの出力層の出力に相当する物体位置情報 PI に加えて、演算モデルとして用いられるニューラルネットワークの中間層の出力を用いて特徴マップ CM を生成し、生成した特徴マップ CM から特徴ベクトル CV を生成してもよい。

50

【 0 0 8 1 】

上述した説明では、行列演算部 2 2 5 は、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ との行列積を算出する演算処理によって得られる情報を、類似性行列 $A M$ として生成している。しかしながら、行列演算部 2 2 5 は、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ とを用いた任意の演算処理によって得られる情報を、類似性行列 $A M$ として生成してもよい。例えば、行列演算部 2 2 5 は、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ との行列和を算出する演算処理によって得られる情報を、類似性行列 $A M$ として生成してもよい。例えば、行列演算部 2 2 5 は、特徴ベクトル $C V_{t-}$ と特徴ベクトル $C V_t$ とが入力された場合に類似性行列 $A M$ を出力する任意の演算モデルを用いて、類似性行列 $A M$ を生成してもよい。このような演算モデルの一例として、ニューラルネットワーク（例えば、 $C N N : C o n v o l u t i o n a l N e u r a l N e t w o r k$ ）を用いた演算モデルがあげられる。要は、行列演算部 2 2 5 は、物体 O_{t-} と物体 O_t との対応関係を示す類似性行列 $A M$ を生成可能な演算処理を行う限りは、類似性行列 $A M$ を生成するためにどのような演算処理を行ってもよい。

10

【 0 0 8 2 】

上述した説明では、物体照合部 2 2 は、 $H \times W \times (K + 4)$ のサイズを有するマップ情報である物体位置情報 $P I$ から、 $H W \times H W$ のサイズを有する類似性行列 $A M$ を生成している。しかしながら、物体照合部 2 2 は、 $H \times W \times (K + 4)$ のサイズを有する物体位置情報 $P I$ から、 $H W \times H W$ のサイズよりも小さいサイズを有する類似性行列 $A M$ を生成している。つまり、物体照合部 2 2 は、ダウンスケーリングされた類似性行列 $A M$ を生成してもよい。例えば、物体照合部 2 2 の特徴マップ変換部 2 2 1 及び 2 2 3 は、 $H \times W \times (K + 4)$ のサイズを有する物体位置情報 $P I_{t-}$ 及び $P I_t$ から、 $H \times W \times C$ のサイズよりも小さいサイズを有する特徴マップ $C M_{t-}$ 及び $C M_t$ を夫々生成してもよい。この場合、物体照合部 2 2 の行列演算部 2 2 5 は、 $H \times W \times C$ のサイズよりも小さいサイズを有する特徴マップ $C M$ から、 $H W \times H W$ のサイズよりも小さいサイズを有する類似性行列 $A M$ を生成することができる。尚、 $H \times W \times C$ のサイズよりも小さいサイズを有する特徴マップ $C M_{t-}$ 及び $C M_t$ を夫々生成するために、物体照合部 2 2 の特徴マップ変換部 2 2 1 及び 2 2 3 は、特徴マップ $C M_{t-}$ 及び $C M_t$ を生成するための畳み込み処理を行う畳み込み層で用いられるカーネル（つまり、畳み込みフィルタ）のストライド量（つまり、移動量）を調整してもよい。カーネルのストライド量が大きくなればなるほど、生成される特徴マップ $C M_{t-}$ 及び $C M_t$ のサイズが小さくなる。

20

30

【 0 0 8 3 】

(5) 付記

以上説明した実施形態に関して、更に以下の付記を開示する。

[付記 1]

第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成する第 1 生成手段と、

前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成する第 2 生成手段と

40

を備える物体追跡装置。

[付記 2]

前記演算処理は、前記第 1 特徴ベクトルと第 2 特徴ベクトルとの行列積を算出する処理を含む

付記 1 に記載の物体追跡装置。

[付記 3]

前記対応情報は、行列を用いて前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す

50

付記 1 又は 2 に記載の物体追跡装置。

[付記 4]

前記第 2 生成手段は、前記行列の一の行のベクトル成分及び前記行列の一の列のベクトル成分の夫々をソフトマックス関数で正規化することで、前記行列を正規化する

付記 1 から 3 のいずれか一項に記載の物体追跡装置。

[付記 5]

前記対応情報は、前記第 1 及び第 2 画像のいずれか一方中の一の物体に対応する行ベクトル成分と前記第 1 及び第 2 画像のいずれか他方中の前記一の物体に対応する列ベクトル成分とが交差する位置において要素が反応する行列を用いて、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示し、

前記第 2 生成手段は、前記行ベクトル成分及び各列ベクトル成分の夫々をソフトマックス関数で正規化することで、前記行列を正規化する

付記 1 から 4 のいずれか一項に記載の物体追跡装置。

[付記 6]

前記対応情報を用いて前記第 2 位置情報を補正する補正手段を更に備える

付記 1 から 4 のいずれか一項に記載の物体追跡装置。

[付記 7]

前記補正手段は、前記対応情報を重みとして用いる注意機構を用いて、前記第 2 位置情報を補正する

付記 6 に記載の物体追跡装置。

[付記 8]

前記対応情報は、行列を用いて前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示し、

前記第 2 位置情報は、前記第 2 画像中の物体の位置に関する情報を示す位置マップを含み、

前記注意機構は、前記位置マップと前記対応マップとの行列積を算出する処理を行うことで、前記第 2 位置情報である前記位置マップを補正する

付記 7 に記載の物体追跡装置。

[付記 9]

前記注意機構は、前記位置マップと前記対応情報との行列積を算出することで得られる補正マップを前記位置マップに加算する処理を行うことで、前記第 2 位置情報である前記位置マップを補正する

付記 8 に記載の物体追跡装置。

[付記 10]

前記第 1 生成手段は、前記第 1 及び第 2 画像が入力されると前記第 1 及び第 2 位置情報を夫々出力する演算モデルから、前記第 1 及び第 2 位置情報と、前記演算モデルの中間出力に相当する中間出力情報とを取得し、

前記第 1 生成手段は、前記第 1 及び第 2 位置情報と前記中間出力情報とに基づいて前記第 1 及び第 2 特徴ベクトルを算出する

付記 1 から 9 のいずれか一項に記載の物体追跡装置。

[付記 11]

前記第 1 及び第 2 画像から前記第 1 及び第 2 位置情報を生成する第 3 生成手段と、

前記対応情報を用いて前記第 2 位置情報を補正する補正手段と、

学習用の前記第 1 及び第 2 画像を含む学習用データを前記第 3 生成手段に入力し、前記学習用データが前記第 3 生成手段に入力された場合に前記補正手段が補正した前記第 2 位置情報に関する第 1 損失関数に基づいて、前記第 1 から第 3 生成手段及び前記補正手段のうちの少なくとも一つの動作内容を規定する学習モデルを更新する学習手段と

を更に備える付記 1 から 10 のいずれか一項に記載の物体追跡装置。

[付記 12]

前記学習手段は、前記第 1 損失関数と、前記学習用データが前記第 3 生成手段に入力さ

10

20

30

40

50

れた場合に前記第 3 生成手段が生成した前記第 1 位置情報に関する第 2 損失関数と、前記学習用データが前記第 3 生成手段に入力された場合に前記第 3 生成手段が生成した前記第 2 位置情報に関する第 3 損失関数とに基づいて、前記学習モデルを更新し、

前記学習手段は、前記第 1 及び第 3 損失関数全体の重みと、前記第 2 損失関数の重みとが等しくなるように、前記第 1 から第 3 損失関数に対して重み付け処理を施し、前記重み付け処理が施された前記第 1 から第 3 損失関数に基づいて、前記学習モデルを生成する
付記 1 1 に記載の物体追跡装置。

[付記 1 3]

前記学習手段は、前記第 1 時刻と前記第 2 時刻との間隔が異なる複数の前記学習用データを前記第 1 生成手段に入力する

付記 1 1 又は 1 2 に記載の物体追跡装置。

[付記 1 4]

第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、

前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することとを含む物体追跡方法。

[付記 1 5]

コンピュータに物体追跡方法を実行させるコンピュータプログラムが記録された、一時的でない記録媒体であって、

前記物体追跡方法は、

第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、

前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することとを含む記録媒体。

[付記 1 6]

コンピュータに物体追跡方法を実行させるコンピュータプログラムであって、

前記物体追跡方法は、

第 1 時刻に撮影された第 1 画像中の物体の位置に関する第 1 位置情報及び前記第 1 時刻とは異なる第 2 時刻に撮影された第 2 画像中の物体の位置に関する第 2 位置情報に基づいて、前記第 1 位置情報の特徴量を示す第 1 特徴ベクトルと、前記第 2 位置情報の特徴量を示す第 2 特徴ベクトルとを生成することと、

前記第 1 及び第 2 特徴ベクトルを用いた演算処理によって得られる情報を、前記第 1 画像中の物体と前記第 2 画像中の物体との対応関係を示す対応情報として生成することとを含むコンピュータプログラム。

【 0 0 8 4 】

本発明は、請求の範囲及び明細書全体から読み取るこのできる発明の要旨又は思想に反しない範囲で適宜変更可能であり、そのような変更を伴う物体追跡装置、物体追跡方法及び記録媒体もまた本発明の技術思想に含まれる。

【 符号の説明 】

【 0 0 8 5 】

- 1 物体追跡装置
- 2 演算装置
 - 2 1 物体検出部
 - 2 2 物体照合部

10

20

30

40

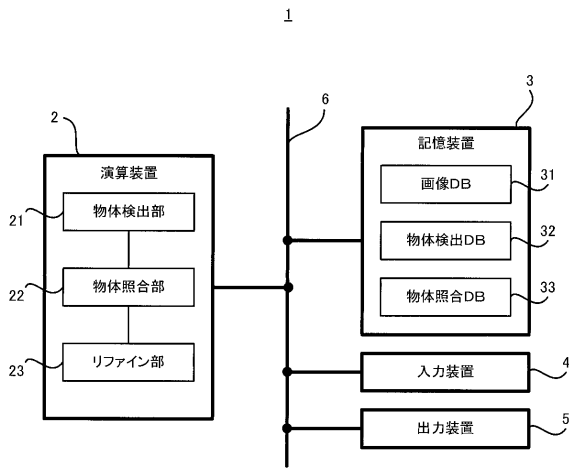
50

- 2 2 1、2 2 3 特徴マップ変換部
- 2 2 2、2 2 4 特徴ベクトル変換部
- 2 2 5 行列演算部
- 2 3 リファイン部
- 2 3 1 行列演算部
- 2 3 2 特徴ベクトル変換部
- 2 3 3 特徴マップ変換部
- 2 3 4 残差処理部
- 2 3 5 特徴マップ変換部
- 3 記憶装置
- 3 1 画像 D B
- 3 2 物体検出 D B
- 3 3 物体照合 D B

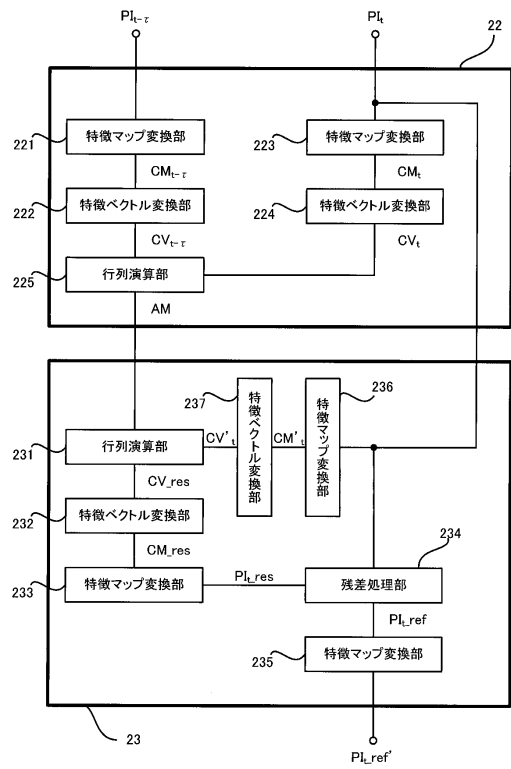
10

【図面】

【図 1】



【図 2】



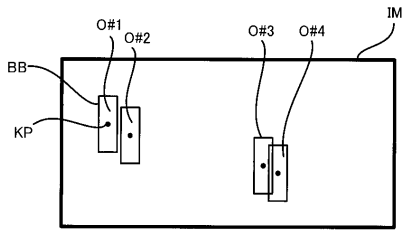
20

30

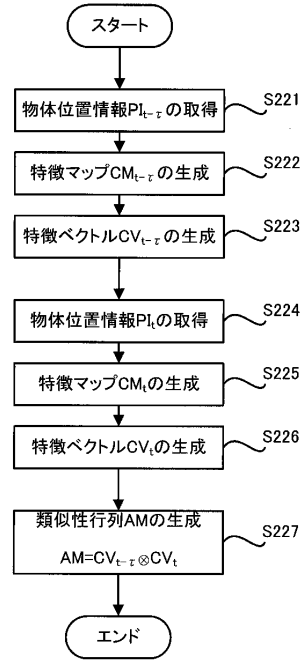
40

50

【図3】



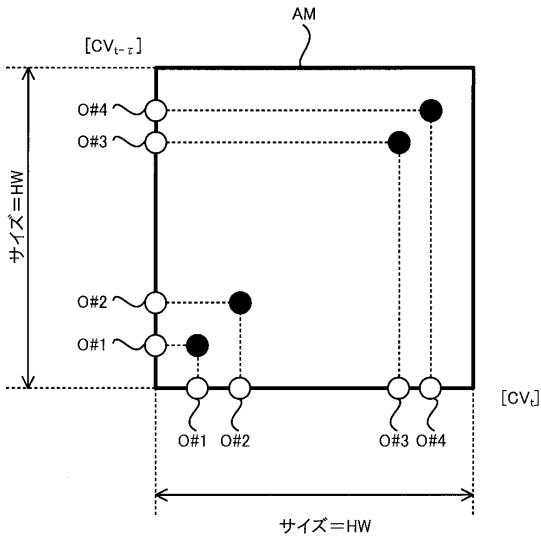
【図4】



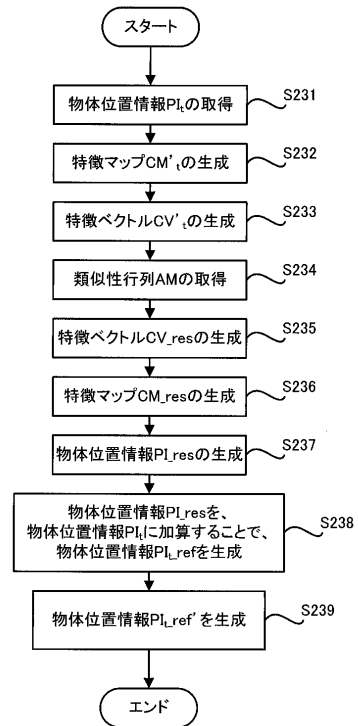
10

20

【図5】



【図6】

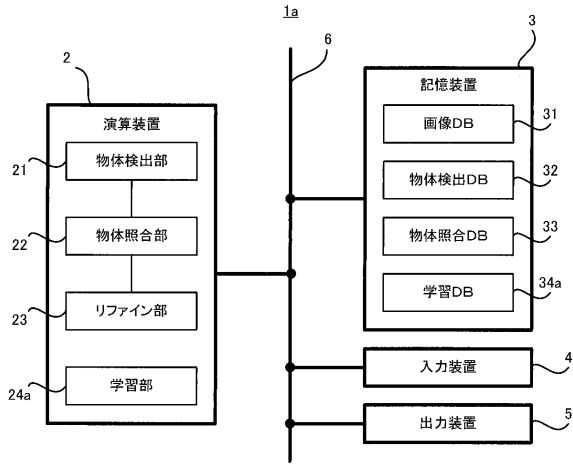


30

40

50

【図7】



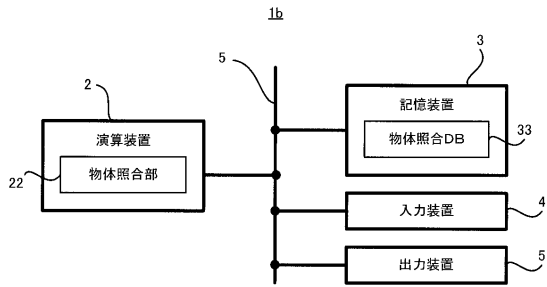
【図8】

34a

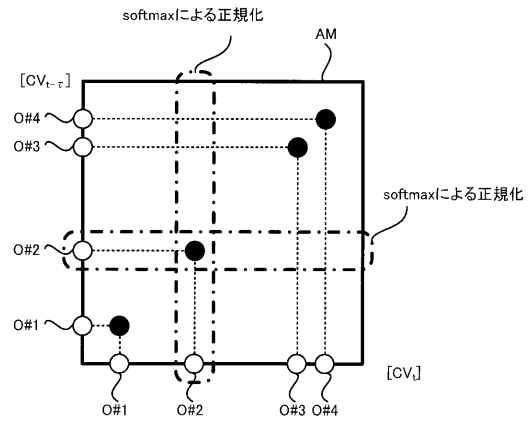
画像	位置 (正解ラベル)	
IM_s	O_s	PI_s_label
$IM_{s+\tau}$	$O_{s+\tau}$	$PI_{s+\tau_label}$
...

10

【図9】



【図10】



20

30

40

50

フロントページの続き

- (56)参考文献 国際公開第2018/030048(WO, A1)
国際公開第2019/008951(WO, A1)
特表2019-536154(JP, A)
FEICHTENHOFER et al., Detect to Track and Track to Detect, Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV 2017), 米国, IEEE, 2017年10月29日, pp.3057-3065, DOI:10.1109/ICCV.2017.330, ISBN 978-1-5386-1032-9
- (58)調査した分野 (Int.Cl., DB名)
G06T 7/20 - 7/292
G06T 7/00