

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7655789号
(P7655789)

(45)発行日 令和7年4月2日(2025.4.2)

(24)登録日 令和7年3月25日(2025.3.25)

(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 C

H 0 4 N 7/18 (2006.01) H 0 4 N 7/18 G

請求項の数 10 (全18頁)

(21)出願番号	特願2021-93766(P2021-93766)	(73)特許権者	000001007
(22)出願日	令和3年6月3日(2021.6.3)		キャノン株式会社
(65)公開番号	特開2022-185872(P2022-185872 A)	(74)代理人	東京都大田区下丸子3丁目30番2号 110003281
(43)公開日	令和4年12月15日(2022.12.15)		弁理士法人大塚国際特許事務所
審査請求日	令和6年5月28日(2024.5.28)	(72)発明者	北澤 源基
			東京都大田区下丸子3丁目30番2号
			キャノン株式会社内
		審査官	小池 正彦

最終頁に続く

(54)【発明の名称】 画像処理装置、画像処理方法、撮像装置

(57)【特許請求の範囲】

【請求項1】

それぞれの画像における追尾対象物体の特徴量の画像間における距離と、画像中の追尾対象物体および非追尾対象物体のそれぞれの特徴量間の距離と、1以上の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域との重複率と、に基づく損失を求める計算手段と、

前記損失に基づいて、画像から物体の特徴量を抽出するためのモデルの学習を行う学習手段と、

前記学習手段により学習済みのモデルに基づき、画像中の追尾対象物体を追尾する追尾処理を行う追尾手段と

を備えることを特徴とする画像処理装置。

【請求項2】

前記計算手段は、

第一の画像中の追尾対象物体および非追尾対象物体の特徴量を、該第一の画像を入力した前記モデルの演算の結果である特徴量マップから取得し、該第一の画像に後続する第二の画像中の追尾対象物体の特徴量を、該第二の画像を入力した前記モデルの演算の結果である特徴量マップから取得する

ことを特徴とする請求項1に記載の画像処理装置。

【請求項3】

前記計算手段は、

第一の画像中の物体の画像領域をリサイズし、該リサイズした画像領域を入力した前記モデルの演算の結果を、該第一の画像中の物体の特徴量として取得し、該第一の画像に後続する第二の画像中の物体の画像領域をリサイズし、該リサイズした画像領域を入力した前記モデルの演算の結果を、該第二の画像中の物体の特徴量として取得する

ことを特徴とする請求項 1 に記載の画像処理装置。

【請求項 4】

前記計算手段は、それぞれの画像における追尾対象物体の特徴量の画像間における距離と、画像中の追尾対象物体および非追尾対象物体のそれぞれの特徴量間の距離と、第一の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域との重複率と、該第一の画像に後続する第二の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域との重複率と、に基づく損失を求めることを特徴とする請求項 1 ないし 3 のいずれか 1 項に記載の画像処理装置。

10

【請求項 5】

前記計算手段は、画像のセットごとに前記損失を求め、画像のセットごとに求めた損失を、前記非追尾対象物体の検出において得られる、物体らしさを示すスコアに応じて重み付けした総和を、最終的な損失として求め、

前記学習手段は、前記最終的な損失に基づいて前記モデルの学習を行う

ことを特徴とする請求項 1 ないし 4 のいずれか 1 項に記載の画像処理装置。

【請求項 6】

前記計算手段は、画像のセットごとに前記損失を求め、画像のセットごとに求めた損失の平均を最終的な損失として求め、

前記学習手段は、前記最終的な損失に基づいて前記モデルの学習を行う

ことを特徴とする請求項 1 ないし 4 のいずれか 1 項に記載の画像処理装置。

20

【請求項 7】

前記第一の画像における非追尾対象物体は、該第一の画像中で追尾対象物体に最も近接している非追尾対象物体であることを特徴とする請求項 2 または 3 に記載の画像処理装置。

【請求項 8】

画像を撮像する撮像部と、

請求項 1 ないし 7 のいずれか 1 項に記載の画像処理装置と

を備えることを特徴とする撮像装置。

30

【請求項 9】

画像処理装置が行う画像処理方法であって、

前記画像処理装置の計算手段が、それぞれの画像における追尾対象物体の特徴量の画像間における距離と、画像中の追尾対象物体および非追尾対象物体のそれぞれの特徴量間の距離と、1 以上の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域との重複率と、に基づく損失を求める計算工程と、

前記画像処理装置の学習手段が、前記損失に基づいて、画像から物体の特徴量を抽出するためのモデルの学習を行う学習工程と、

前記画像処理装置の追尾手段が、前記学習工程で学習済みのモデルに基づき、画像中の追尾対象物体を追尾する追尾処理を行う追尾工程と

を備えることを特徴とする画像処理方法。

40

【請求項 10】

コンピュータを、請求項 1 ないし 7 のいずれか 1 項に記載の画像処理装置の各手段として機能させるための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、画像中の物体を追尾するための技術に関するものである。

【背景技術】

【0002】

50

画像中の物体を追尾するための技術として、輝度や色情報を利用するものやテンプレートマッチングを利用するもの、Deep Neural Networkを利用するもの等が存在する。Deep Neural Networkを利用する方法の一つにTracking by Detectionと呼ばれる方法がある。この方法では、画像中の物体を検出し、該物体を過去の物体と対応付けることによって追尾を行う。非特許文献1では、この対応付けをMetric Learningによって学習されたNeural Networkから得られた特徴量を用いて行う。Metric Learningとは、類似度が高いデータほど空間上での距離が近く分布するような空間への変換を学習する方法である。

【先行技術文献】

10

【特許文献】

【0003】

【文献】特開2018-026108号公報

【非特許文献】

【0004】

【文献】Nicolai Wojke, Alex Bewley, Dietrich Paulus: Simple online and realtime tracking with a deep association metric (ICIP2017)

【発明の概要】

【発明が解決しようとする課題】

20

【0005】

Metric Learningによって学習されたNeural Networkにより、各物体に対応する特徴量を抽出し、それらを過去の特徴量と対応付けることによって追尾を行う方法を考える。そのような方法では、追尾対象となる物体の姿勢が変化したり、遮蔽が発生したりするような場合、追尾対象となる物体と見た目の類似したその他の物体を誤って対応付けしてしまい、追尾が失敗することがある。

【0006】

本発明では、追尾対象となる物体と見た目の類似した物体が存在するような場合であっても、追尾対象となる物体とその他の物体を正しく識別して追尾を行うための技術を提供する。

30

【課題を解決するための手段】

【0007】

本発明の一様態は、それぞれの画像における追尾対象物体の特徴量の画像間における距離と、画像中の追尾対象物体および非追尾対象物体のそれぞれの特徴量間の距離と、1以上の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域との重複率と、に基づく損失を求める計算手段と、前記損失に基づいて、画像から物体の特徴量を抽出するためのモデルの学習を行う学習手段と、前記学習手段により学習済みのモデルに基づき、画像中の追尾対象物体を追尾する追尾処理を行う追尾手段とを備えることを特徴とする。

【発明の効果】

【0008】

40

本発明の構成によれば、追尾対象となる物体と見た目の類似した物体が存在するような場合であっても、追尾対象となる物体とその他の物体を正しく識別して追尾を行うことができる。

【図面の簡単な説明】

【0009】

【図1】学習装置100の機能構成例を示すブロック図。

【図2】学習装置100による学習モデルの学習処理のフローチャート。

【図3】物体検出用CNNと特徴量抽出用CNNとの関係を示す図。

【図4】追尾対象物体の特徴量を取得するための処理を説明する図。

【図5】 d_1 、 d_2 、 r 、 m を求めるための処理を説明する図。

50

【図 6】特徴量間距離を二次元空間上に示したイメージ図。

【図 7】追尾装置 200 の機能構成例を示すブロック図。

【図 8】追尾装置 200 による追尾対象物体の追尾処理のフローチャート。

【図 9】変形例 1 に係る正解データの構成例を示す図。

【図 10】変形例 2 に係る、追尾対象物体および非追尾対象物体の特徴量の取得方法を説明する図。

【図 11】 $I \circ U_1$ 、 $I \circ U_2$ の計算方法を示す図。

【図 12】コンピュータ装置のハードウェア構成例を示すブロック図。

【発明を実施するための形態】

【0010】

以下、添付図面を参照して実施形態を詳しく説明する。尚、以下の実施形態は特許請求の範囲に係る発明を限定するものではない。実施形態には複数の特徴が記載されているが、これらの複数の特徴の全てが発明に必須のものとは限らず、また、複数の特徴は任意に組み合わせられてもよい。さらに、添付図面においては、同一若しくは同様の構成に同一の参照番号を付し、重複した説明は省略する。

【0011】

[第 1 の実施形態]

まず、連続して撮像された画像における追尾対象（追尾対象物体）とその他の物体（非追尾対象物体）とを正しく識別して該追尾対象物体の追尾を行うために用いる学習モデルの学習を行う学習装置としての画像処理装置について説明する。

【0012】

本実施形態に係る学習装置 100 の機能構成例を図 1 のブロック図に示す。図 1 に示す機能構成例を有する学習装置 100 による学習モデルの学習処理について、図 2 のフローチャートに従って説明する。

【0013】

ステップ S101 では、取得部 110 は、第一の画像と、該第一の画像の正解データ（第一の正解データ）と、を取得する。第一の正解データは、第一の画像中の追尾対象物体について予め作成されたものであり（既知）、該第一の画像における該追尾対象物体の画像領域を示す領域情報と、該追尾対象物体に固有のラベルと、を含む。追尾対象物体の画像領域を示す領域情報は、例えば、画像領域が矩形領域である場合には、該矩形領域の中心位置（追尾対象物体の位置）、該矩形領域の幅および高さ、を含む情報である。なお、ここでは、追尾対象物体の画像領域が矩形領域である場合には、該矩形領域の中心位置を追尾対象物体の位置としているが、該矩形領域のどの位置を追尾対象物体の位置としても良い。

【0014】

また取得部 110 は、第一の画像に後続して撮像された第二の画像と、該第二の画像の正解データ（第二の正解データ）と、を取得する。第二の正解データは、第一の正解データと同様、第二の画像中の追尾対象物体について予め作成されたものであり（既知）、該第二の画像における該追尾対象物体の画像領域を示す領域情報と、該追尾対象物体に固有のラベルと、を含む。第一の画像と第二の画像とで同じ追尾対象物体には同じラベルが付与されている。

【0015】

第一の画像および第二の画像のそれぞれは、例えば、動画像における第 1 フレームの画像、該第 1 フレームに後続する第 2 フレームの画像、である。また、例えば、第一の画像および第二の画像のそれぞれは、定期的若しくは不定期的に撮像された複数の静止画像における着目静止画像、該着目静止画像よりも後に撮像された静止画像、である。また、必要に応じて取得した画像の一部分を切り出す等の処理を行ってもよい。

【0016】

ステップ S102 では、物体検出部 120 は、第一の画像を物体検出用の学習モデルである物体検出用 CNN (Convolutional Neural Network) に

10

20

30

40

50

入力して該物体検出用CNNの演算処理を実行する。これにより物体検出部120は、第一の画像中におけるそれぞれの物体について、該物体の画像領域を示す領域情報と、該物体のスコアと、を該物体の検出結果として取得する。物体の画像領域を示す領域情報は、例えば、画像領域が矩形領域である場合には、該矩形領域の中心位置、該矩形領域の幅および高さ、を含む情報である。また、物体のスコアは、値域を0～1とするスコアであり、大きいほど物体らしさが高いことを意味する。画像中から物体を検出する技術としては、“Liu, SSD: Single Shot Multibox Detector: ECCV 2016”等が挙げられる。

【0017】

ステップS103では、特徴量抽出部130は、第一の画像を特徴量抽出用の学習モデルである特徴量抽出用CNNに入力して該特徴量抽出用CNNの演算処理を実行する。これにより特徴量抽出部130は、該第一の画像を縦横等分割した場合における各領域の特徴量を表す特徴量マップ（第一の特徴量マップ）を取得する。そして特徴量抽出部130は、第一の正解データが示す追尾対象物体の画像領域の中心位置に対応する第一の特徴量マップ上の対応位置を特定し、該特定した対応位置における特徴量を「追尾対象物体の特徴量」として取得する。また、特徴量抽出部130は、ステップS102で取得した領域情報が示す画像領域の中心位置に対応する第一の特徴量マップ上の対応位置を特定し、該特定した対応位置における特徴量を「非追尾対象物体の特徴量」として取得する。

【0018】

同様に、特徴量抽出部130は、第二の画像を特徴量抽出用の学習モデルである特徴量抽出用CNNに入力して該特徴量抽出用CNNの演算処理を実行する。これにより特徴量抽出部130は、該第二の画像を縦横等分割した場合における各領域の特徴量を表す特徴量マップ（第二の特徴量マップ）を取得する。そして特徴量抽出部130は、第二の正解データが示す追尾対象物体の画像領域の中心位置に対応する第二の特徴量マップ上の対応位置を特定し、該特定した対応位置における特徴量を「追尾対象物体の特徴量」として取得する。

【0019】

ここで、追尾対象物体の特徴量を取得するための処理について、図4を用いて説明する。図4では、横640画素×縦480画素のサイズを有する第一の画像に、該第一の画像を横24個×縦16個の分割領域に分割した場合におけるそれぞれの分割領域の特徴量が登録されているマップである第一の特徴量マップを重ねた状態を示している。第一の特徴量マップは、チャンネル数がC（任意の自然数）の特徴量が（24×16）個登録されたマップである。第一の画像の左上隅の位置を原点（0，0）とするので、第一の画像上の位置（x，y）は0 ≤ x ≤ 639，0 ≤ y ≤ 479の範囲をとり、第一の特徴量マップ上の位置（p，q）（分割領域単位的位置）は0 ≤ p ≤ 23，0 ≤ q ≤ 15の範囲をとる。

【0020】

位置402は、第一の正解データが示す追尾対象物体の位置（x，y）＝（312，276）である。このとき、特徴量抽出部130は、 $(312 \times 24 / 640) = 11.7$ を計算し、11.7の小数部を削除した11をpとし、 $(276 \times 16 / 480) = 9.2$ を計算し、9.2の小数部を削除した9をqとする。これにより特徴量抽出部130は、第一の画像上の追尾対象物体の位置（312，276）に対応する第一の特徴量マップ上の位置（11，9）を求める。矩形403は、第一の画像上の追尾対象物体の位置（312，276）に対応する第一の特徴量マップ上の位置（11，9）を示す。そして特徴量抽出部130は、位置403における特徴量（位置403における1×1×Cの配列要素）を、第一の画像における追尾対象物体の特徴量として取得する。

【0021】

また特徴量抽出部130は、第二の画像についても同様の処理を行うことで、第二の画像における追尾対象物体の特徴量を取得する。つまり特徴量抽出部130は、第二の正解データが示す追尾対象物体の位置に対応する第二の特徴量マップ上の位置を求め、該求めた位置における特徴量（該求めた位置における1×1×Cの配列要素）を第二の画像にお

10

20

30

40

50

ける追尾対象物体の特徴量として取得する。

【 0 0 2 2 】

なお、ステップ S 1 0 2 で取得した領域情報が示す画像領域の中心位置に対応する第二の特徴量マップ上の対応位置を特定する場合も、(x 、 y) から (p 、 q) を求める上記の方法を用いて特定する。

【 0 0 2 3 】

ここで、物体検出用 C N N と、特徴量抽出用 C N N と、の関係について図 3 を用いて説明する。物体検出用 C N N および特徴量抽出用 C N N はどのような構成で実装しても構わない。例えば、図 3 (a) に示す如く、物体検出用 C N N と特徴量抽出用 C N N をそれぞれ別個に設け、それぞれが入力画像 (第一の画像、第二の画像) に対して独立して動作するように構成しても良い。また、図 3 (b) に示す如く、物体検出用 C N N と特徴量抽出用 C N N とで共通の構成を共有 C N N として設け、物体検出用 C N N において共有 C N N に含まれない構成を改めて物体検出用 C N N とし、特徴量抽出用 C N N において共有 C N N に含まれない構成を改めて特徴量抽出用 C N N としても良い。この場合、入力画像は共有 C N N に入力され、共有 C N N が入力画像に対する処理結果を物体検出用 C N N および特徴量抽出用 C N N に入力する。物体検出用 C N N は共有 C N N から出力された処理結果を入力として動作し、特徴量抽出用 C N N は共有 C N N から出力された処理結果を入力として動作する。

10

【 0 0 2 4 】

次に、ステップ S 1 0 4 では、損失計算部 1 4 0 は、ステップ S 1 0 3 で求めた特徴量を用いて、第一の画像における追尾対象物体の特徴量と、第二の画像における該追尾対象物体の特徴量と、の間のユークリッド距離である特徴量間距離 d_1 を求める。また損失計算部 1 4 0 は、ステップ S 1 0 3 で求めた特徴量を用いて、第一の画像における追尾対象物体の特徴量と、該第一の画像における非追尾対象物体の特徴量と、の間のユークリッド距離である特徴量間距離 d_2 を求める。特徴量間の距離は、それぞれの特徴量が属する特徴量空間におけるそれぞれの特徴量間の距離であり、特徴量間の距離を求める方法については周知であるため、これについての説明は省略する。

20

【 0 0 2 5 】

次に、ステップ S 1 0 5 では、損失計算部 1 4 0 は、損失関数のパラメータを求める。本実施形態では、以下の (式 1) で与えられる T r i p l e t l o s s を損失関数として用いる。

30

【 0 0 2 6 】

$$l o s s = \max (r d_1 - d_2 + m , 0) \quad \dots \quad (式 1)$$

この損失関数では、 d_1 が大きく、 d_2 が小さくなることで損失が発生するため、同一物体に対する特徴量間距離が、異なる物体に対する特徴量間距離よりも小さくなるような特徴量を獲得することができる。

【 0 0 2 7 】

この (式 1) において、 r と m が損失関数のパラメータであり、 r は距離パラメータ、 m はマージンパラメータである。 r を大きく設定すると同一物体同士の特徴量間距離が小さくなるように学習され、 m を大きくすると異なる物体同士の特徴量間距離が遠くなるように学習される。本実施形態では、これらのパラメータを正解データとステップ S 1 0 2 で検出した物体の領域を用いて動的に設定することによって、より物体の識別性能が高い特徴量を学習する。

40

【 0 0 2 8 】

なお、使用する損失関数については、距離学習が行えるものであれば、特定の関数に限らない。例えば、損失関数として C o n t r a s t i v e l o s s や S o f t m a x l o s s を用いても良い。

【 0 0 2 9 】

ここで、 d_1 、 d_2 、 r 、 m を求めるための処理について、図 5 に示す具体例を挙げて説明する。図 5 (a) は第一の画像を示しており、図 5 (b) は第二の画像を示している。

50

【 0 0 3 0 】

図 5 (a) の第一の画像は、追尾対象物体 5 0 1 と、非追尾対象物体 5 0 2 , 5 0 3 と、を含む。矩形 5 0 6 は、第一の正解データによって規定される該追尾対象物体 5 0 1 の画像領域である。矩形 5 0 7 は、ステップ S 1 0 2 で検出された非追尾対象物体 5 0 2 の画像領域である。矩形 5 0 8 は、ステップ S 1 0 2 で検出された非追尾対象物体 5 0 3 の画像領域である。

【 0 0 3 1 】

図 5 (b) の第二の画像は、追尾対象物体 5 0 4 と、非追尾対象物体 5 0 5 と、を含む。矩形 5 0 9 は第二の正解データによって規定される該追尾対象物体 5 0 4 の画像領域であり、矩形 5 1 0 はステップ S 1 0 2 で検出された非追尾対象物体 5 0 5 の画像領域である。

10

【 0 0 3 2 】

このような場合、特徴量抽出部 1 3 0 は、第一の正解データが示す「第一の画像における追尾対象物体 5 0 1 の位置」に対応する第一の特徴量マップ上の対応位置における特徴量と、第二の正解データが示す「第二の画像における追尾対象物体 5 0 4 の位置」に対応する第二の特徴量マップ上の対応位置における特徴量と、の間の特徴量間距離 d_1 を求める。

【 0 0 3 3 】

また特徴量抽出部 1 3 0 は、非追尾対象物体 5 0 2 , 5 0 3 のうち一方 (図 5 (a) では非追尾対象物体 5 0 2) を選択物体として選択し、該選択物体についてステップ S 1 0 2 で検出された位置に対応する第一の特徴量マップ上の対応位置における特徴量と、第一の正解データが示す「第一の画像における追尾対象物体 5 0 1 の位置」に対応する第一の特徴量マップ上の対応位置における特徴量と、の間の特徴量間距離 d_2 を求める。第一の画像中に複数の非追尾対象物体が含まれている場合に、該複数の非追尾対象物体からの選択物体の選択方法については特定の選択方法に限らない。例えば、追尾対象物体の位置に最も近い位置の (最も近接している) 非追尾対象物体を選択物体として選択する。

20

【 0 0 3 4 】

また特徴量抽出部 1 3 0 は、以下の (式 2) を用いて距離パラメータ r を求め、以下の (式 3) を用いてマージンパラメータ m を求める。

【 0 0 3 5 】

$$r = 1 \quad \dots \quad (\text{式 2})$$

$$m = m_0 + I o U \quad \dots \quad (\text{式 3})$$

ここで、 m_0 は任意の実数である。I o U (I n t e r s e c t i o n o v e r U n i o n) は、2 つの画像領域の面積の重複率である。図 5 の場合、追尾対象物体 5 0 1 の画像領域 5 0 6 と、選択物体として選択された非追尾対象物体 5 0 2 の画像領域 5 0 7 と、の面積の重複率を I o U として求める。

30

【 0 0 3 6 】

ステップ S 1 0 6 では、損失計算部 1 4 0 は、上記の処理で得られた d_1 , d_2 , r , m を用いて (式 1) を計算することで、損失 $l o s s$ を求める。

【 0 0 3 7 】

図 6 は、本実施形態に係る上記の方法で損失を計算した際の特徴量間距離を二次元空間上に示したイメージ図である。特徴量 6 0 1 は第一の画像における追尾対象物体の特徴量、特徴量 6 0 2 は第二の画像における追尾対象物体の特徴量である。 d_1 は特徴量 6 0 1 と特徴量 6 0 2 との間の特徴量間距離である。特徴量 6 0 3 および特徴量 6 0 4 は、第一の画像における非追尾対象物体の特徴量である。 d_2 は特徴量 6 0 1 と特徴量 6 0 4 との間の特徴量間距離、 d_2' は特徴量 6 0 1 と特徴量 6 0 3 との間の特徴量間距離である。また、 m 、 m_0 、I o U は、(式 3) に示したマージンパラメータである。

40

【 0 0 3 8 】

この方法では、第一の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域とで重複が発生していない場合は、特徴量間距離 d_2' が、特徴量間距離 d_1 とマージン

50

パラメータ m_0 の和より大きくなるように学習される。一方、第一の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域とで重複が発生している場合は、特徴量間距離 d_2 が、特徴量間距離 d_1 とマージンパラメータ m の和より大きくなるように学習される。このため、追尾対象物体の画像領域とのIoUが高い非追尾対象物体、すなわち、追尾対象物体との距離が近く、誤った追尾の切り替わりが発生する危険性が高い非追尾対象物体ほど、特徴量間距離が大きくなるように学習を行うことができる。

【0039】

なお、損失関数は、追尾対象物体の画像領域とのIoUに基づいてパラメータを変更するものであれば、上記の形式に限定されるものではない。また、損失を計算する際は、第一の画像と第二の画像のセットを複数セット用意し、それぞれのセットについて上記の処理を行ってセットごとの損失を求め、セットごとに求めた損失の平均を最終的な損失としても構わない。

【0040】

ステップS107では、学習部150は、ステップS106で求めた損失 $loss$ を極小化するように特徴量抽出用CNNにおけるパラメータ（重み値など）を更新することで、該特徴量抽出用CNNの学習を行う。なお、学習部150は、特徴量抽出用CNNの学習に加えて、物体検出用CNNの学習を同時に行うようにしても良い。重みなどのパラメータの更新は、例えば誤差逆伝搬法を用いて行うことが考えられる。

【0041】

次に、上記の学習したCNNを用いて画像における追尾対象物体を追尾するための処理を行う追尾装置としての画像処理装置について説明する。本実施形態に係る追尾装置200の機能構成例を図7のブロック図に示す。図7に示す機能構成例を有する追尾装置200による追尾対象物体の追尾処理について、図8のフローチャートに従って説明する。

【0042】

ステップS201では、画像取得部210は、1枚の撮像画像を取得する。該1枚の撮像画像は、連続して撮像される撮像画像のうち最近に撮像された撮像画像であり、例えば、動画画像における最新のフレームの撮像画像であっても良いし、定期的若しくは不定期的に撮像された静止画像において最近に撮像された静止画像であっても良い。

【0043】

ステップS202では、物体検出部220は、ステップS201で取得した撮像画像を物体検出用CNNに入力して該物体検出用CNNの演算処理を実行する。これにより物体検出部220は、ステップS201で取得した撮像画像中におけるそれぞれの物体について、該物体の画像領域を示す領域情報と、該物体のスコアと、を該物体の検出結果として取得する。

【0044】

ステップS203では、特徴量抽出部230は、ステップS201で取得した撮像画像を特徴量抽出用CNNに入力して該特徴量抽出用CNNの演算処理を実行することで、該撮像画像の特徴量マップを取得する。そして特徴量抽出部230は、ステップS202で取得したそれぞれの物体について、該物体の領域情報が示す画像領域の中心位置に対応する特徴量マップ上の対応位置を特定し、該特定した対応位置における特徴量を「該物体の特徴量」として取得する。

【0045】

そして、ステップS201で取得した撮像画像が、追尾処理を開始してから最初に取得した撮像画像（初期時刻の撮像画像）であれば、処理はステップS204に進む。一方、ステップS201で取得した撮像画像が、追尾処理を開始してから2枚目以降の撮像画像である（初期時刻の撮像画像ではない）場合には、処理はステップS207に進む。

【0046】

ステップS204では、特徴量保存部240は、ステップS203で取得したそれぞれの物体の特徴量に対して、該物体に固有のID（物体ID）を付与する。例えば、スコアが高い物体から順に、物体ID = 1, 2, 3, ...を付与する。

10

20

30

40

50

【 0 0 4 7 】

そしてステップ S 2 0 5 では（ステップ S 2 0 4 からステップ S 2 0 5 に進んだ場合）、特徴量保存部 2 4 0 は、ステップ S 2 0 4 で物体 I D が付与されたそれぞれの物体の特徴量を追尾装置 2 0 0 に保存する。

【 0 0 4 8 】

ステップ S 2 0 7 では、特徴量照合部 2 5 0 は、ステップ S 2 0 3 で取得したそれぞれの物体の特徴量（現時刻の特徴量）のうち未選択の 1 つを選択特徴量として選択する。そして特徴量照合部 2 5 0 は、追尾装置 2 0 0 に保存済みの特徴量から、物体ごとに、現時刻に近い順に n （ n は自然数）個の特徴量（保存特徴量）を取得する。そして特徴量照合部 2 5 0 は、該取得したそれぞれの保存特徴量と、選択特徴量と、の間の特徴量間距離（特徴量間距離の求め方については上記と同様）を求める。なお、特徴量照合部 2 5 0 は、物体ごとに、該物体について取得した n 個の保存特徴量から 1 つの保存特徴量（例えば、 n 個の保存特徴量の平均値）を求め、物体ごとの該求めた 1 つの保存特徴量と、選択特徴量と、の間の特徴量間距離を求めるようにしても良い。

10

【 0 0 4 9 】

ステップ S 2 0 8 では、特徴量照合部 2 5 0 は、ステップ S 2 0 7 で求めた特徴量間距離のうち最小の特徴量間距離を特定する。そして特徴量照合部 2 5 0 は、該特定した最小の特徴量間距離に対応する保存特徴量の物体に付与されている物体 I D を、選択特徴量の物体に付与する。また特徴量照合部 2 5 0 は、「ステップ S 2 0 7 で求めた特徴量間距離のうち最小の特徴量間距離に対応する保存特徴量の物体に付与されている物体 I D」が先に他の物体に付与されている場合には、新たな物体 I D を発行して選択特徴量の物体に付与する。物体 I D は、スコアが高い物体から順に降順とする方法が考えられるが、これに限らない。

20

【 0 0 5 0 】

このようなステップ S 2 0 7 の処理およびステップ S 2 0 8 の処理を、ステップ S 2 0 3 で取得した全ての物体の特徴量について行うことで、ステップ S 2 0 2 で検出した全ての物体に対して物体 I D を付与することができる。なお、現時刻の特徴量に物体 I D を付与するための上記の処理は一例であり、上記の処理に限定することを意図したものではない。例えば、ハンガリアンアルゴリズムを用いても良い。また、ステップ S 2 0 7 およびステップ S 2 0 8 では、次のような処理を行うようにしても良い。

30

【 0 0 5 1 】

ステップ S 2 0 7 では、特徴量照合部 2 5 0 は、ステップ S 2 0 3 で取得したそれぞれの物体の特徴量（現時刻の特徴量）のうち未選択の 1 つを選択特徴量として選択する。そして特徴量照合部 2 5 0 は、過去に追尾装置 2 0 0 に保存された物体の特徴量（保存特徴量）と、選択特徴量と、の間の特徴量間距離（特徴量間距離の求め方については上記と同様）を求める。

【 0 0 5 2 】

ステップ S 2 0 8 では、特徴量照合部 2 5 0 は、ステップ S 2 0 7 で求めた特徴量間距離のうち閾値未満である特徴量間距離を特定する。なお、特徴量照合部 2 5 0 は、閾値未満である特徴量間距離として複数の特徴量間距離を特定した場合には、該特定した複数の特徴量間距離のうち最小の特徴量間距離を特定する。そして特徴量照合部 2 5 0 は、該特定した特徴量間距離に対応する保存特徴量のうち保存タイミングが最近の保存特徴量の物体に付与されている物体 I D を、選択特徴量の物体に付与する。また特徴量照合部 2 5 0 は、ステップ S 2 0 7 で求めた特徴量間距離のうち閾値未満の特徴量間距離がなかった場合には、新たな物体 I D を発行して選択特徴量の物体に付与する。

40

【 0 0 5 3 】

つまり、現時刻の特徴量において、過去に追尾装置 2 0 0 に保存された着目特徴量との特徴量間距離が閾値未満であれば、現時刻の特徴量の物体と、着目特徴量の物体と、を対応づけるべく、現時刻の特徴量に、着目特徴量の物体の物体 I D を付与する。

【 0 0 5 4 】

50

そしてステップS205では(ステップS208からステップS205に進んだ場合)、特徴量保存部240は、ステップS208で物体IDが付与されたそれぞれの物体の特徴量を追尾装置200に保存する。

【0055】

また特徴量照合部250は、ステップS205で追尾装置200に保存した物体IDと、該物体IDを付与した物体の画像領域と、を追尾結果として出力する。追尾結果は、追尾装置200若しくはその他の装置が、追尾結果に含まれる物体の画像領域を示す枠を撮像画像に重ねて該物体の物体IDと共に表示しても良いし、外部の装置に対して送信しても良い。なお、上記の図8のフローチャートに従った処理は、追尾処理を行う撮像画像ごとに行われる。

10

【0056】

このように、本実施形態では、撮像画像中における追尾対象物体との距離が近く、誤った追尾の切り替わりが発生する危険性の高い非追尾対象物体ほど、特徴量間距離が大きくなるように学習を行う。これにより、より物体の識別性能が高い特徴量を獲得することができ、見た目の類似した物体が存在する場合でも、高い精度で追尾を行うことができる。

【0057】

<変形例1>

本変形例では、画像中のそれぞれの物体(追尾対象物体および/または非追尾対象物体)に対して正解データが付与されているケースについて説明する。本変形例に係る正解データについて図9を用いて説明する

20

図9(a)は、図9(b)の画像中の物体901, 902, 903のそれぞれの画像領域904, 905, 906の中心位置(水平位置座標、垂直位置座標)、幅、高さ、ラベルが登録されたテーブルとしての正解データの構成例を示している。なお、第一の画像、第二の画像にはそれぞれ同一のラベルが付与された物体が少なくとも1つ存在するものとする。

【0058】

損失計算部140において損失を計算する際は、あるラベルが付与された物体を追尾対象物体、それ以外のラベルが付与された物体を非追尾対象物体とみなす。また、上記の計算を別のラベルが付与された物体についても行い、それらの平均を取ることで最終的な損失を決定してもよい。

30

【0059】

<変形例2>

本変形例では、特徴量抽出部130による追尾対象物体および非追尾対象物体の特徴量の取得方法について、図10を例にとり説明する。画像1001において領域1002は、正解データが示す追尾対象物体の画像領域であり、領域1003および領域1004は、物体検出部120によって検出された非追尾対象物体の画像領域である。画像1005は領域1002内の画像であり、画像1006は領域1003内の画像であり、画像1007は領域1004内の画像である。

【0060】

特徴量抽出部130は、画像1005~1007を規定の画像サイズ(ここでは32画素×32画素)の画像に変換する(リサイズなど)。つまり特徴量抽出部130は、画像1005を32画素×32画素の画像1008に変換し、画像1006を32画素×32画素の画像1009に変換し、画像1007を32画素×32画素の画像1010に変換する。

40

【0061】

そして特徴量抽出部130は、画像1008を特徴量抽出用CNNに入力して該特徴量抽出用CNNの演算処理を実行することで、該画像1008に対応する物体の特徴量を取得する。また特徴量抽出部130は、画像1009を特徴量抽出用CNNに入力して該特徴量抽出用CNNの演算処理を実行することで、該画像1009に対応する物体の特徴量を取得する。また特徴量抽出部130は、画像1010を特徴量抽出用CNNに入力して

50

該特徴量抽出用CNNの演算処理を実行することで、該画像1010に対応する物体の特徴量を取得する。物体の特徴量は、任意の次元数のベクトルとして取得することが考えられる。

【0062】

このような処理を第一の画像および第二の画像について行うことでそれぞれの画像における物体の特徴量を取得することができるので、以降は第1の実施形態と同様の処理を行えば良い。

【0063】

<変形例3>

本変形例では、上記の距離パラメータ r およびマージンパラメータ m をそれぞれ(式4)、(式5)に従って求める。

【0064】

$$r = 1 + I o U_1 \cdot I o U_2 \quad (\text{式4})$$

$$m = m_0 \quad (\text{式5})$$

ここで、 m_0 は任意の正の実数である。 $I o U_1$ 、 $I o U_2$ の計算方法について、図11を用いて説明する。図11(a)は第一の画像を示しており、(b)は第二の画像を示している。

【0065】

図11(a)に示す如く、第一の画像には、追尾対象物体1101および非追尾対象物体1102(第一の画像に複数の非追尾対象物体が含まれている場合には、追尾対象物体1101に最も近接している非追尾対象物体)が含まれている。 $I o U_1$ は、追尾対象物体1101の画像領域1105と、非追尾対象物体1102の画像領域1106との $I o U$ である。

【0066】

また、図11(b)に示す如く、第二の画像には、追尾対象物体1103および非追尾対象物体1104(第二の画像に複数の非追尾対象物体が含まれている場合には、追尾対象物体1103に最も近接している非追尾対象物体)が含まれている。 $I o U_2$ は、追尾対象物体1103の画像領域1107と、非追尾対象物体1104の画像領域1108との $I o U$ である。

【0067】

つまり、第一の画像および第二の画像の両方で追尾対象物体の画像領域と重複部分を有する非追尾対象物体が存在する場合、 r が1より大きな値を取る。この方法で損失を計算することにより、追尾対象物体に近接する非追尾対象物体が存在する場合に、特徴量間距離が小さくなるように学習を行うことができる。これにより、追尾対象物体が他の物体に遮蔽された、または他の物体を遮蔽した場合にも、頑健に追尾対象物体を識別する特徴量を獲得することができる。

【0068】

なお、損失関数は、第一の画像および第二の画像の両方で追尾対象物体の画像領域と重複部分を有する非追尾対象物体が存在する場合にパラメータを変更するものであれば、上記の形式に限定されるものではない。

【0069】

<変形例4>

本変形例では、損失を計算する際、第一の画像と第二の画像のセットを複数セット用意し、それぞれのセットについて上記の処理を行ってセットごとの損失を求め、セットごとに求めた損失を該セットに対応する非追尾対象物体のスコアに応じて重み付けしたものの総和を、最終的な損失とする。本変形例では損失 $loss$ を以下の(式6)に従って求める。

【0070】

【数1】

10

20

30

40

50

$$loss = \frac{1}{N} \sum_{i=1}^N p_i \cdot loss_i \quad (式6)$$

【0071】

10

ここで、Nは上記のセットの総数、 $loss_i$ はi番目のセットについて求めた損失である。また、 p_i は、i番目のセットについて(式1)の d_2 を求める際に選択した非追尾対象物体のスコアである。この方法で損失を計算すると、スコアの高い非追尾対象物体、すなわち誤った追尾の切り替えが発生する危険性の高い物体の損失に対する寄与が大きくなる。これにより、追尾の性能に重要な事例を重視して学習することができ、効率的な学習を行うことが可能となる。尚、損失の重み付けは、スコアによって変更されるものであれば、上記の形式に限定されるものではない。

【0072】

<変形例5>

本変形例では、損失関数としてContrastive lossを用いる。Contrastive lossは以下の(式7)で表される。

20

【0073】

【数2】

$$loss = \frac{1}{2} [y(rd)^2 + (1-y)max(m-d, 0)^2] \quad (式7)$$

30

【0074】

ここで、dは2つの物体に対応する特徴量の特徴量間距離であり、yは2つの物体が同一物体である場合に1、異なる物体である場合に0となる。この損失関数では、2つの物体が同一物体である場合にはdが小さく、異なる物体である場合にはdが大きくなるように学習される。また、rとmはそれぞれ距離パラメータとマージンパラメータであり、第1の実施形態と同様に、rを大きく設定すると同一物体同士の特徴量間距離が小さくなるように学習され、mを大きくすると異なる物体同士の特徴量間距離が遠くなるように学習される。パラメータを設定する際は、例えば、第1の実施形態と同様に(式2)および(式3)を用いることができる。

40

【0075】

<変形例6>

第1の実施形態では、学習モデルとしてCNNを用いた。しかし、学習モデルに適用可能なモデルはCNNに限らず、他のモデル(例えば機械学習モデル)を用いても構わない。

【0076】

第1の実施形態やその変形例では、それぞれの画像における追尾対象物体の特徴量の画像間における距離と、画像中の追尾対象物体および非追尾対象物体のそれぞれの特徴量間の距離と、1以上の画像における追尾対象物体の画像領域と非追尾対象物体の画像領域と

50

の重複率と、に基づく損失の求め方の一例を説明した。しかし、これらの情報を用いて損失を求める方法は上記の方法に限らない。

【 0 0 7 7 】

また、第 1 の実施形態では、学習装置 1 0 0 と追尾装置 2 0 0 とを別個の画像処理装置として説明したが、1 つの画像処理装置に納めても良い。つまり、学習装置 1 0 0 が行うものとして説明した処理と、追尾装置 2 0 0 が行うものとして説明した処理と、の両方を 1 つの画像処理装置が行うようにしても良い。このような場合、例えば、動画像を撮像する、もしくは定期的若しくは不定期的に静止画像を撮像する撮像部と、学習装置 1 0 0 が行うものとして説明した処理と追尾装置 2 0 0 が行うものとして説明した処理との両方を実行可能な画像処理装置と、を搭載した撮像装置を構成することができる。このような撮像装置では、自身が撮像した撮像画像を用いて学習装置 1 0 0 の動作（物体検出用 C N N や特徴量抽出用 C N N の学習）を実行することができる。そして撮像装置は、該動作後に自身が撮像した撮像画像における物体の追跡処理を、学習済みの物体検出用 C N N や特徴量抽出用 C N N を用いて実行することができる。

10

【 0 0 7 8 】

また、第 1 の実施形態では、特徴量間距離としてユークリッド距離を用いたが、特徴量間距離をどのような計算で求めても良く、特定の計算方法に限らない（どのような種類の距離を特徴量間距離として採用しても良い）。

【 0 0 7 9 】

[第 2 の実施形態]

20

図 1 や図 7 に示した各機能部はハードウェアで実装しても良いし、ソフトウェア（コンピュータプログラム）で実装しても良い。後者の場合、このようなコンピュータプログラムを実行可能なコンピュータ装置は、上記の学習装置 1 0 0 や追尾装置 2 0 0 に適用可能である。学習装置 1 0 0 や追尾装置 2 0 0 に適用可能なコンピュータ装置のハードウェア構成例について、図 1 2 のブロック図を用いて説明する。

【 0 0 8 0 】

C P U 1 2 0 1 は、R A M 1 2 0 2 や R O M 1 2 0 3 に格納されているコンピュータプログラムやデータを用いて各種の処理を実行する。これにより C P U 1 2 0 1 は、コンピュータ装置全体の動作制御を行うと共に、学習装置 1 0 0 や追尾装置 2 0 0 が行うものとして説明した各種の処理を実行もしくは制御する。

30

【 0 0 8 1 】

R A M 1 2 0 2 は、R O M 1 2 0 3 や外部記憶装置 1 2 0 6 からロードされたコンピュータプログラムやデータを格納するためのエリアを有する。また、R A M 1 2 0 2 は、I / F 1 2 0 7 を介して外部から受信したデータを格納するためのエリアを有する。また R A M 1 2 0 2 は、C P U 1 2 0 1 が各種の処理を実行する際に用いるワークエリアを有する。このように R A M 1 2 0 2 は、各種のエリアを適宜提供することができる。

【 0 0 8 2 】

R O M 1 2 0 3 には、コンピュータ装置の設定データ、コンピュータ装置の起動に係るコンピュータプログラムやデータ、コンピュータ装置の基本動作に係るコンピュータプログラムやデータ、などが格納されている。

40

【 0 0 8 3 】

操作部 1 2 0 4 は、キーボード、マウス、タッチパネルなどのユーザインターフェースであり、ユーザが操作することで各種の指示を C P U 1 2 0 1 に対して入力することができる。

【 0 0 8 4 】

表示部 1 2 0 5 は、液晶画面やタッチパネル画面を有し、C P U 1 2 0 1 による処理結果を画像や文字などでもって表示することができる。例えば、表示部 1 2 0 5 には、上記の追尾処理の結果を表示することができる。また、表示部 1 2 0 5 は、画像や文字を投影するプロジェクタなどの投影装置であっても良い。

【 0 0 8 5 】

50

外部記憶装置 1206 は、ハードディスクドライブなどの大容量情報記憶装置である。外部記憶装置 1206 には、OS（オペレーティングシステム）や、学習装置 100 や追尾装置 200 が行うものとして説明した各種の処理を CPU 1201 に実行もしくは制御させるためのコンピュータプログラムやデータが保存されている。外部記憶装置 1206 に保存されているデータには、学習装置 100 や追尾装置 200 に保存されているものとして説明したデータ、上記の説明において既知のデータとして取り扱ったデータ、CNN に係るデータ、などが含まれている。外部記憶装置 1206 に保存されているコンピュータプログラムやデータは、CPU 1201 による制御に従って適宜 RAM 1202 にロードされ、CPU 1201 による処理対象となる。

【0086】

I/F 1207 は、外部の装置との間のデータ通信を行うための通信インターフェースである。例えば、I/F 1207 は、動画像や静止画像を撮像可能な撮像装置、動画像や静止画像を保持しているサーバ装置、またはこれらの装置が接続されているネットワーク、などに接続することができる。この場合、撮像装置が撮像した動画像や静止画像、サーバ装置から供給された動画像や静止画像は、I/F 1207 を介してコンピュータ装置に供給され、該供給された動画像や静止画像は、RAM 1202 や外部記憶装置 1206 に格納される。

【0087】

CPU 1201、RAM 1202、ROM 1203、操作部 1204、表示部 1205、外部記憶装置 1206、I/F 1207 はいずれも、システムバス 1208 に接続されている。

【0088】

なお、図 12 に示したハードウェア構成は、学習装置 100 や追尾装置 200 に適用可能なコンピュータ装置のハードウェア構成の一例であり、適宜変形/変更が可能である。

【0089】

また、上記の各実施形態や各変形例で使用した数値、処理タイミング、処理順、処理の主体、データ（情報）の送信先/送信元/格納場所などは、具体的な説明を行うために一例として挙げたもので、このような一例に限定することを意図したものではない。

【0090】

また、以上説明した各実施形態や各変形例の一部若しくは全部を適宜組み合わせて使用しても構わない。また、以上説明した各実施形態や各変形例の一部若しくは全部を選択的に使用しても構わない。

【0091】

（その他の実施形態）

本発明は、上述の実施形態の 1 以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける 1 つ以上のプロセッサがプログラムを読出し実行する処理でも実現可能である。また、1 以上の機能を実現する回路（例えば、ASIC）によっても実現可能である。

【0092】

発明は上記実施形態に制限されるものではなく、発明の精神及び範囲から離脱することなく、様々な変更及び変形が可能である。従って、発明の範囲を公にするために請求項を添付する。

【符号の説明】

【0093】

100：学習装置 110：取得部 120：物体検出部 130：特徴量抽出部 140：損失計算部 150：学習部

10

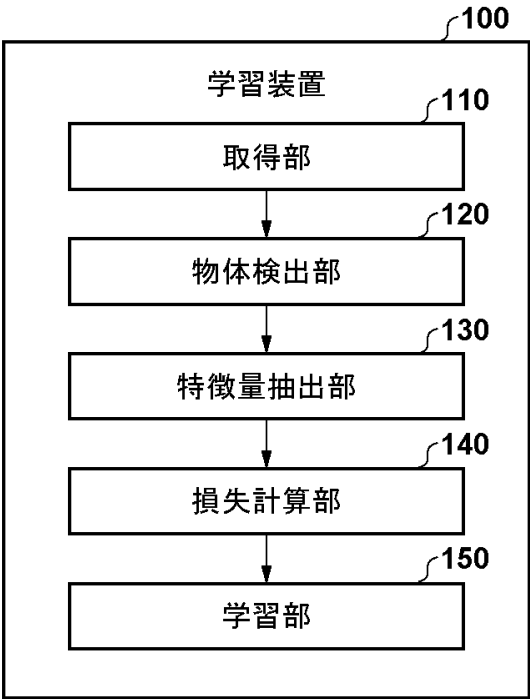
20

30

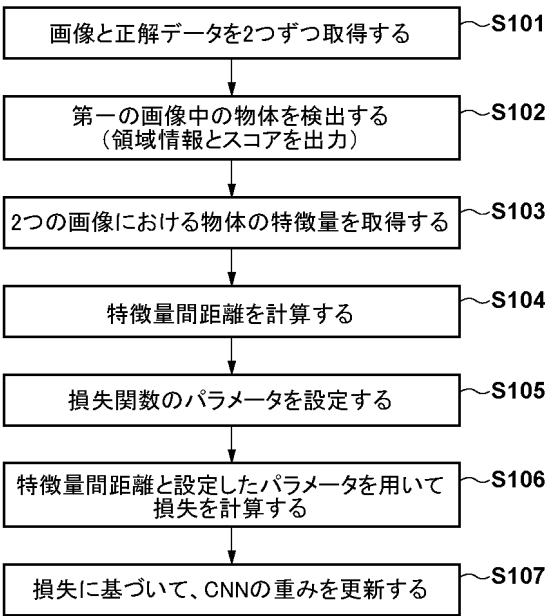
40

50

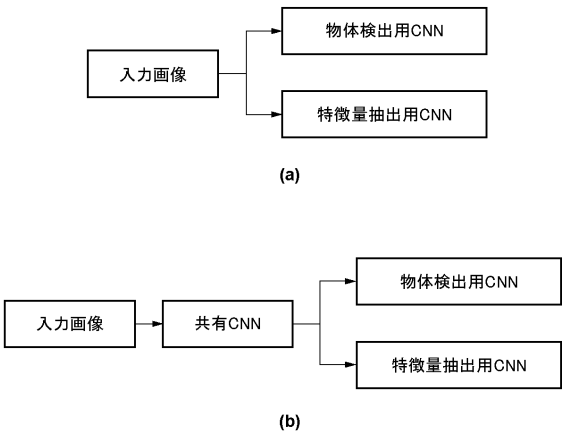
【図面】
【図 1】



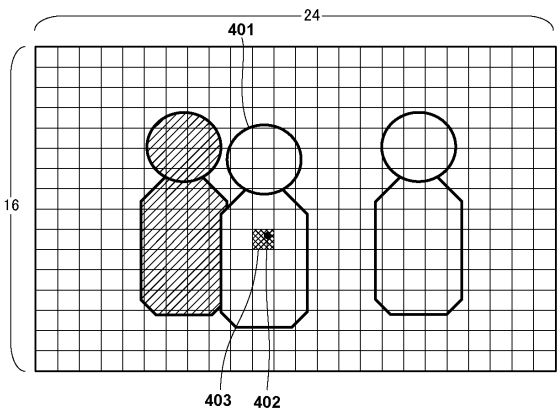
【図 2】



【図 3】



【図 4】



10

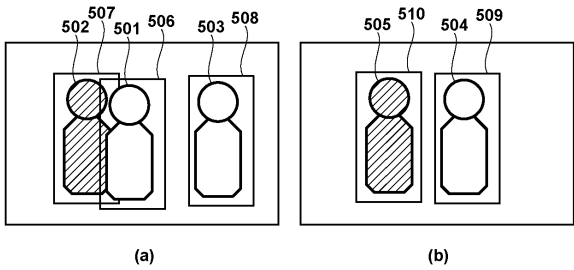
20

30

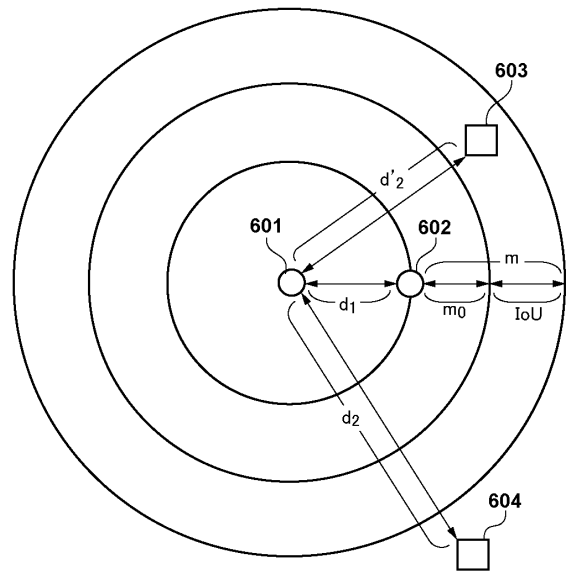
40

50

【図 5】

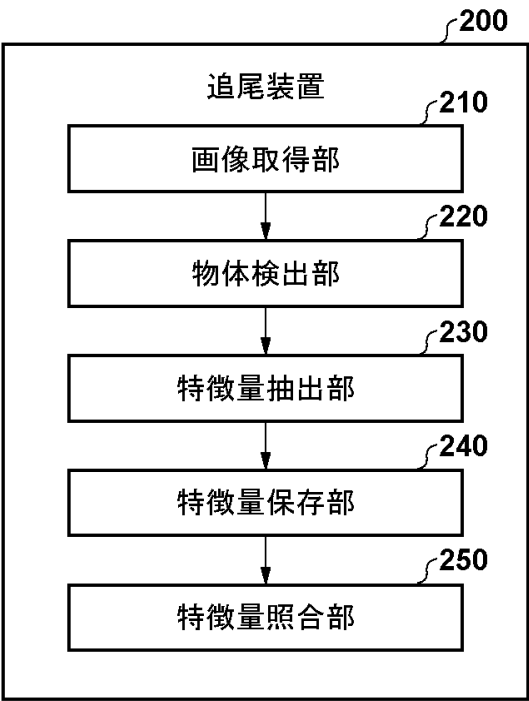


【図 6】

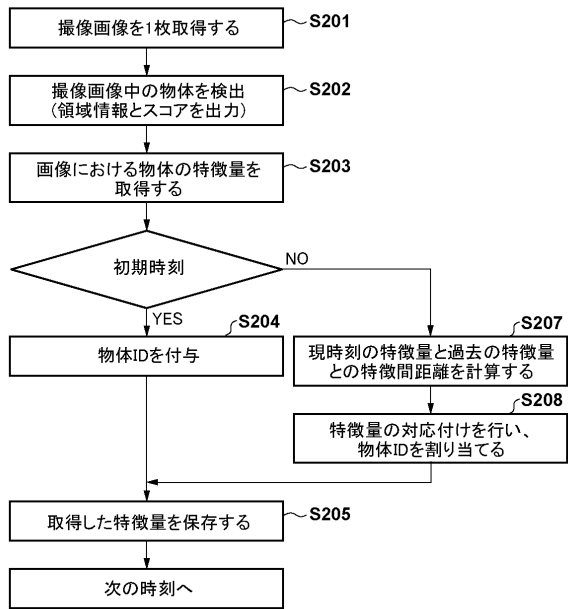


10

【図 7】



【図 8】



20

30

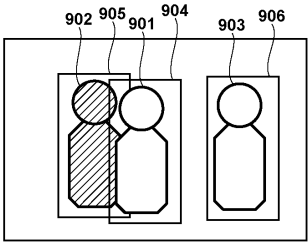
40

50

【図 9】

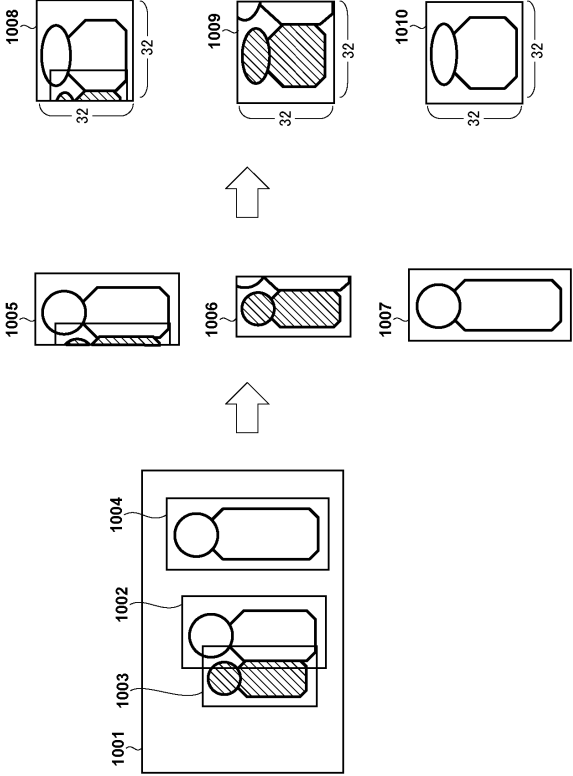
	中心位置座標 (水平方向)	中心位置座標 (垂直方向)	幅	高さ	ラベル
物体901	312	276	126	331	A
物体902	129	263	125	335	B
物体903	496	271	130	338	C

(a)



(b)

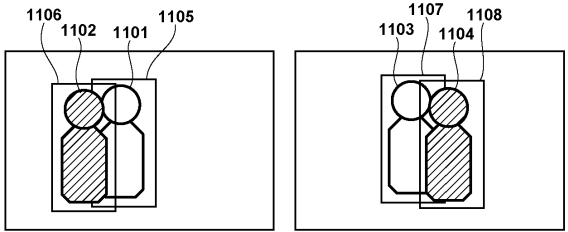
【図 10】



10

20

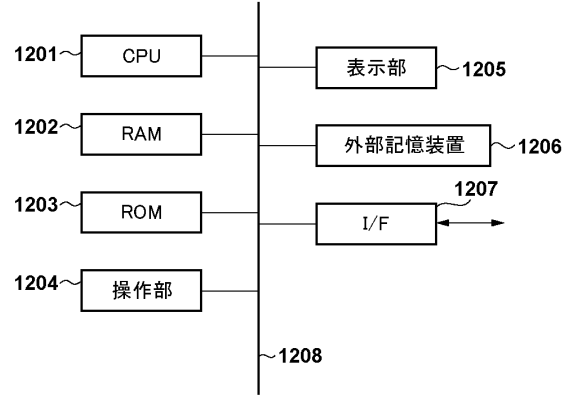
【図 11】



(a)

(b)

【図 12】



30

40

50

フロントページの続き

- (56)参考文献 米国特許出願公開第 2 0 2 0 / 0 1 9 3 1 4 1 (U S , A 1)
特開 2 0 1 8 - 0 2 6 1 0 8 (J P , A)
Ente Guo, et al. , Real-time Detection and Tracking Network with Feature Sharing , 2020 IE
EE International Conference on Visual Communications and Image Processing (VCIP) , 米国
 , IEEE , 2020年 , p.519-522 , <https://ieeexplore.ieee.org/document/9301779> , DOI: 10.1
109/VCIP49819.2020.9301779
石川晴也,外5名, オンライン複数物体追跡において分割モデルの検証, 電気学会知覚情報 /
次世代産業システム研究会 PI-20-17, 日本, 電気学会, 2020年03月23日, p.73-78
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 T 7 / 2 4 6
G 0 6 T 7 / 0 0
H 0 4 N 7 / 1 8