(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7490359号
(P7490359)

(45)発行日　令和6年5月27日(2024.5.27)

(24)登録日　令和6年5月17日(2024.5.17)

(51)国際特許分類
Ｇ０６Ｔ　7/00　(2017.01)

ＦＩ
Ｇ０６Ｔ　7/00　３５０Ｂ

請求項の数　23　（全32頁）

(21)出願番号　特願2019-233229(P2019-233229)
(22)出願日　令和1年12月24日(2019.12.24)
(65)公開番号　特開2021-103347(P2021-103347A)
(43)公開日　令和3年7月15日(2021.7.15)
審査請求日　令和4年12月14日(2022.12.14)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100126240
弁理士　阿部 琢磨
(74)代理人　100223941
弁理士　高橋 佳子
(74)代理人　100159695
弁理士　中辻 七朗
(74)代理人　100172476
弁理士　冨田 一史
(74)代理人　100126974
弁理士　大朋 靖尚
(72)発明者　舘 俊太
東京都大田区下丸子３丁目３０番２号キ
最終頁に続く

(54)【発明の名称】　情報処理装置、情報処理方法及びプログラム

(57)【特許請求の範囲】
【請求項１】
入力画像に含まれる複数の物体を検出する情報処理装置であって、
前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に得るために用いる尤度マップを出力する学習モデルに、複数の異なる結合重みパラメータを設定することにより複数の前記尤度マップを生成する生成手段と、
前記入力画像に含まれる前記物体の正解としての位置の情報を取得する取得手段と、
前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差に基づいて、該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げるための第１の損失関数と、前記複数の尤度マップを用いて得られた尤度を領域毎に統合して得た尤度のうち、前記取得された位置を含む領域で得られた当該尤度を上げるための第２の損失関数と、に基づいて前記尤度マップごとに前記結合重みパラメータを更新する学習手段と、を有することを特徴とする情報処理装置。
【請求項２】
前記第１の損失関数は、前記第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差が小さいほど、該第１の注目領域の近傍の領域に対応付けられた前記尤度をより下げることを特徴とする請求項１に記載の情報処理装置。
【請求項３】
前記学習手段は、前記複数の尤度マップの前記取得された前記物体の位置を含む第２の

注目領域に対応付けられた前記尤度に基づいて、前記第２の注目領域に対応付けられた尤度を調整する第３の損失関数に更に基づいて、前記尤度マップごとに前記結合重みパラメータを更新することを特徴とする請求項１または２に記載の情報処理装置。

【請求項４】

前記第３の損失関数は、前記取得された物体の位置についての情報に基づいて前記入力画像に含まれる前記物体の位置が１つである場合は、前記複数の尤度マップのうちいずれか１つの尤度マップの前記第２の注目領域に対応付けられた尤度がより大きくなるように調整し、前記取得された物体の位置についての情報に基づいて前記入力画像に含まれる前記物体の位置が複数ある場合は、前記物体の数に応じて前記尤度マップの前記第２の注目領域に推定された尤度がより大きくなるように調整する損失値を決定することを特徴とする請求項３に記載の情報処理装置。

【請求項５】

前記第３の損失関数は、前記複数の尤度マップのうち２つの尤度マップ間において前記入力画像における前記物体の位置を含む前記第２の注目領域に対応付けられた尤度に基づいて、前記尤度マップの少なくとも一方における前記第２の注目領域に対応付けられた尤度を下げるように調整する損失値を決定することを特徴とする請求項３または４に記載の情報処理装置。

【請求項６】

前記学習手段により更新された前記結合重みづけパラメータを設定した前記学習モデルによって生成された前記尤度マップに基づいて、前記尤度が閾値より大きい領域に存在する前記物体を検出する検出手段と、を更に有することを特徴とする請求項１乃至５のいずれか１項に記載の情報処理装置。

【請求項７】

前記生成手段は、前記抽出された画像特徴のうち、前記結合重みパラメータが更新された学習済みモデル毎に異なる画像特徴の組み合わせを入力することによって前記尤度マップを生成することを特徴とする請求項１乃至６のいずれか１項に記載の情報処理装置。

【請求項８】

前記入力画像から画像特徴を複数の異なる組み合わせで抽出する抽出手段を更に有し、

前記生成手段は、前記抽出された異なる組み合わせの画像特徴を前記学習モデルに入力することで前記尤度マップを生成することを特徴とする請求項１乃至７のいずれか１項に記載の情報処理装置。

【請求項９】

前記物体とセンサとの距離を示す奥行き情報を取得する取得手段を更に有し、

前記生成手段は、前記奥行き情報をさらに前記学習モデルに入力することによって、複数の前記尤度マップを生成することを特徴とする請求項１乃至８のいずれか１項に記載の情報処理装置。

【請求項１０】

入力画像に含まれる複数の物体を検出する情報処理装置であって、

前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に得るための尤度マップを出力する学習済みモデルに、複数の異なる結合重みパラメータを設定することにより生成された複数の前記尤度マップを取得する第１の取得手段と、

前記取得された複数の尤度マップに基づいて、前記尤度が閾値より大きい注目領域を前記入力画像に含まれる前記物体として検出する検出手段と、を有し、

前記学習済みモデルは、前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差に基づいて、該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げるための第１の損失関数と、前記複数の尤度マップのうち所定の２つの尤度マップ間で共通である第２の注目領域に対応付けられた尤度に基づいて、前記尤度マップの少なくとも一方における前記第２の注目領域に対応付けられた尤度を調整するための第３の損失関数と、に基づいて、

前記第１と第３の損失関数によって出力された損失値を小さくするように前記尤度マップごとに前記結合重みパラメータを学習させた学習モデルであることを特徴とする情報処理装置。

【請求項１１】

前記第３の損失関数は、前記複数の尤度マップのうち２つの尤度マップ間において前記入力画像における前記物体の位置を含む前記第２の注目領域に対応付けられた尤度に基づいて、前記尤度マップの少なくとも一方における前記第２の注目領域に対応付けられた尤度を下げるための損失値を決定する関数であることを特徴とする請求項１０に記載の情報処理装置。

【請求項１２】

前記入力画像に含まれる前記物体の位置を取得する第２の取得手段を更に有し、

前記学習済みモデルは、前記複数の尤度マップを用いて得られた尤度を領域毎に統合して得た尤度のうち、前記取得された位置を含む領域で得られた当該尤度を上げるための第２の損失関数に更に基づいて、前記第１と第２と第３の損失関数によって出力された損失値を小さくするように前記尤度マップごとに前記結合重みパラメータを学習させた学習モデルであることを特徴とする請求項１０または１１に記載の情報処理装置。

【請求項１３】

前記第２の損失関数は、前記複数の尤度マップを統合した結果に対して、前記取得された位置に対応付けられた前記尤度が閾値より小さい場合は大きい損失値を出力することを特徴とする請求項１２に記載の情報処理装置。

【請求項１４】

前記第２の取得手段は、前記物体の位置から前記物体の数を取得し、

前記第２の損失関数は、前記取得された前記物体の数に基づいて、前記複数の尤度マップを統合した結果に対して、前記尤度が閾値より大きい領域が前記取得された数と一致しない場合は大きい損失値を出力することを特徴とする請求項１２または１３に記載の情報処理装置。

【請求項１５】

前記損失関数の出力する損失値が収束するように前記学習済みモデルのパラメータを更新することによって学習する学習手段を更に有することを特徴とする請求項１０乃至１４のいずれか１項に記載の情報処理装置。

【請求項１６】

前記入力画像から抽出された画像特徴に基づいて、前記物体の数を特定する特定手段を更に有し、

前記第１の取得手段は、前記特定された前記物体の数に応じて取得する前記尤度マップの数を調整することを特徴とする請求項１０乃至１５のいずれか１項に記載の情報処理装置。

【請求項１７】

入力画像に含まれる複数の物体を検出する情報処理装置であって、

前記物体を撮像した画像における該物体それぞれの領域を取得する取得手段と、

前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に得るための尤度マップを出力する学習モデルに、複数の異なる結合重みパラメータを設定することにより複数の前記尤度マップを生成する生成手段と、

前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差に基づいて、該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げるための第１の損失関数と、前記複数の尤度マップのうち所定の２つの尤度マップ間で共通である第２の注目領域に対応付けられた尤度に基づいて、前記尤度マップの少なくとも一方における前記第２の注目領域に対応付けられた尤度を調整するための第３の損失関数と、に基づいて前記尤度マップごとに前記結合重みパラメータを更新する学習手段と、

前記学習手段により更新された前記結合重みづけパラメータを設定した前記学習モデル

によって生成された前記尤度マップに基づいて、前記入力画像において物体毎に対応する領域を推定する推定手段と、を有することを特徴とする情報処理装置。

【請求項１８】

入力画像に含まれる複数の物体を検出する情報処理装置であって、

前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に対応付けた複数の尤度マップと、前記入力画像に含まれる前記物体の位置についての情報と、を取得する取得手段と、

前記複数の尤度マップを用いて得られた尤度を領域毎に統合して得た尤度のうち、前記取得された位置を含む領域で得られた当該尤度を上げるための第２の損失関数に基づいて、前記尤度マップごとに該尤度マップの領域毎の尤度を決定するためのパラメータを更新する学習手段と、有することを特徴とする情報処理装置。

【請求項１９】

入力画像に含まれる複数の物体を検出する情報処理装置であって、

前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に得るための複数の尤度マップを生成する生成手段と、

前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差に基づいて、該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げる第１の調整と、前記複数の尤度マップのうち所定の２つの尤度マップ間で共通である第２の注目領域に対応付けられた尤度に基づいて、前記尤度マップの少なくとも一方における前記第２の注目領域に対応付けられた尤度を調整する第２の調整と、を行うことによって前記尤度マップごとに尤度を調整する調整手段と、

前記調整された複数の尤度マップに基づいて、前記尤度が閾値より大きい領域を前記入力画像に含まれる前記物体として検出する検出手段と、

を有することを特徴とする情報処理装置。

【請求項２０】

入力画像に含まれる複数の物体を検出する情報処理装置であって、

前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に得るために用いる尤度マップを出力する学習モデルに、複数の異なる結合重みパラメータを設定することにより複数の前記尤度マップを生成する生成手段と、

前記入力画像に含まれる前記物体の正解としての位置の情報を取得する取得手段と、

前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域に対応付けられた尤度が閾値より大きい場合、該第１の注目領域の近傍の領域に対応付けられた前記尤度をより下げるための第１の損失関数と、前記複数の尤度マップを用いて得られた尤度を領域毎に統合して得た尤度のうち、前記取得された位置を含む領域で得られた当該尤度を上げるための第２の損失関数と、に基づいて前記尤度マップごとに前記結合重みパラメータを更新する学習手段と、を有することを特徴とする情報処理装置。

【請求項２１】

前記第１の損失関数は、前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域に対応付けられた尤度が閾値より大きい場合は該第１の注目領域の近傍の領域に対応付けられた前記尤度をより下げるための損失値を大きくなるように、前記第１の注目領域に対応付けられた尤度が閾値より小さい場合は該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げるための損失値をより小さくなるように損失値を出力する損失関数であることを特徴とする請求項２０に記載の情報処理装置。

【請求項２２】

コンピュータを、請求項１乃至２１のいずれか１項に記載の情報処理装置が有する各手段として機能させるためのプログラム。

【請求項２３】

入力画像に含まれる複数の物体を検出する情報処理方法であって、

前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤

度を領域毎に得るための尤度マップを出力する学習モデルに、複数の異なる結合重みパラメータを設定することにより複数の前記尤度マップを生成する生成工程と、
前記入力画像に含まれる前記物体の位置についての情報を取得する取得工程と、
前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差に基づいて、該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げるための第１の損失関数と、前記複数の尤度マップを用いて得られた尤度を領域毎に統合して得た尤度のうち、前記取得された位置を含む領域で得られた当該尤度を上げるための第２の損失関数と、に基づいて前記尤度マップごとに前記結合重みパラメータを更新する学習工程と、を有することを特徴とする情報処理方法。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は，画像中の複数の対象物体を認識する技術に関する。
【背景技術】
【０００２】
従来、画像中の特定の被写体を検出する方法が知られている。例えば非特許文献１ではニューラルネットワークを用いて被写体の位置・サイズ・カテゴリー等の認識を行う。非特許文献１に類する手法は入力画像を多層ニューラルネットワーク（深層ニューラルネットとも呼ばれ、以下ＤＮＮと略する）で処理する。これらの手法の一般的な形態は下記のようなものである。まず入力画像をＷ×Ｈのサイズに変換しＤＮＮに入力する。次にこの入力画像に対して畳み込み等の演算を複数回行い、画像を中間的にｗ×ｈ（ｗ≦Ｗ，ｈ≦Ｈ）の解像度の画像特徴へと変換する。ＤＮＮの後段の層ではこの特徴のｗ×ｈの各位置に対象物体が存在するか、しないかを判定する。存在すると判定した場合はさらに別途ＤＮＮで推定したサイズや精確な位置推定結果等、と合わせて検出結果を出力する。
【先行技術文献】
【非特許文献】
【０００３】
【文献】Ｊ．Ｒｅｄｍｏｎ，Ａ．Ｆａｒｈａｄｉ，ＹＯＬＯ９０００：Ｂｅｔｔｅｒ，Ｆａｓｔｅｒ，Ｓｔｒｏｎｇｅｒ，ＣＶＰＲ，２０１７
【発明の概要】
【発明が解決しようとする課題】
【０００４】
従来、物体の有無を判定する単位である１つのブロックの中に同種の複数の物体が隣接して存在する場合、「物体が１つ存在する」という誤検出を起こすことがあった。つまり、近接した同一種類の物体同士を分離して認識することは難しかった。
【０００５】
この課題は入力画像あるいは前述の中間的な画像特徴の解像度を上げることで回避できるが、その場合ＤＮＮの処理量および後処理（多重検出抑制等）の演算量が大きく増大する。
【０００６】
本発明は、上記の課題に鑑みてなされたものであり、複数の物体が隣接して存在する場合でも、そのそれぞれを精度よく検出することを目的とする。
【課題を解決するための手段】
【０００７】
上記の目的を達成する本発明に係る情報処理装置は、入力画像に含まれる複数の物体を検出する情報処理装置であって前記入力画像から抽出された画像特徴を入力すると前記物体が存在する可能性を示す尤度を領域毎に得るために用いる尤度マップを出力する学習モデルに、複数の異なる結合重みパラメータを設定することにより複数の前記尤度マップを生成する生成手段と、前記入力画像に含まれる前記物体の正解としての位置の情報を取得

する取得手段と、前記複数の尤度マップのうち同一の尤度マップ内の、前記複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域の尤度と、該第１の注目領域の近傍の領域に対応付けられた前記尤度との差に基づいて、該第１の注目領域の近傍の領域に対応付けられた前記尤度を下げるための第１の損失関数と、前記複数の尤度マップを用いて得られた尤度を領域毎に統合して得た尤度のうち、前記取得された位置を含む領域で得られた当該尤度を上げるための第２の損失関数と、に基づいて前記尤度マップごとに前記結合重みパラメータを更新する学習手段と、を有することを特徴とする。

【発明の効果】

【０００８】

本発明によれば、複数の物体が隣接して存在する場合でも、そのそれぞれを精度よく検出できる。

【図面の簡単な説明】

【０００９】

【図１】情報処理装置の機能構成例を示すブロック図
【図２】情報処理装置が実行する処理を説明するフローチャート
【図３】特徴抽出部が実行する処理を説明するフローチャート
【図４】画像特徴の模式図
【図５】尤度マップの更新と統合の模式図
【図６】尤度マップ生成部の概念図
【図７】尤度マップの結合重みの概念図
【図８】情報処理装置の機能構成例を示すブロック図
【図９】情報処理装置が実行する処理を説明するフローチャート
【図１０】損失関数の模式図
【図１１】情報処理装置の機能構成例を示すブロック図
【図１２】特徴の集計の模式図
【図１３】特徴の集計の詳細
【図１４】情報処理装置が実行する処理を説明するフローチャート
【図１５】情報処理装置の機能構成例を示すブロック図
【図１６】情報処理装置が実行する処理を説明するフローチャート
【図１７】ＧＵＩの一例を示す図
【図１８】情報処理装置の機能構成例を示すブロック図
【図１９】情報処理装置が実行する処理を説明するフローチャート
【図２０】尤度マップの一例を示す図
【図２１】情報処理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】

＜実施形態１＞

本実施形態は物体検出において複数の物体が近接・重畳している場合であっても、頑健に検出する一手法について説明する。ここでは物体の顔の検出を行う情報処理装置の例を説明する。ただし本発明は物体の顔に限定することなく各種物体検出に対して適用可能である。

【００１１】

複数の物体が含まれる画像から物体それぞれの検出する場合、物体同士が隣接していると１つの物体として検出することや一方の物体を検出しないことがある。そのようなケースに対し、本実施形態では、物体が存在する可能性を示す尤度を入力画像の各領域に対応づけられた尤度マップを２枚以上用意し、複数の尤度マップから物体の位置を検出する。複数の尤度マップは、それぞれの異なる位置にある物体を検出するように、損失関数あるいはニューロン間の結合を用いて、尤度マップ内または尤度マップ間の重みを学習する、あるいは尤度マップを用いて得られる尤度を更新する。

【００１２】

本発明の基本的な機能構成図である図１を用いて説明する。なお、これ以降図面を参照しながら説明する際に、図面間で同一種の符号は同一の構成モジュールを意味する。特筆すべき差異がない場合は重ねての説明を省く。

【００１３】

図２１は、情報処理装置のハードウェア構成例を示すブロック図である。中央処理ユニット（ＣＰＵ）２１１は、ＲＡＭ２１３をワークメモリとして、ＲＯＭ２１２や記憶装置２１４に格納されたＯＳやその他プログラムを読みだして実行し、システムバス２１９に接続された各構成を制御して、各種処理の演算や論理判断などを行う。ＣＰＵ２１１が実行する処理には、実施形態の情報処理が含まれる。記憶装置２１４は、ハードディスクドライブや外部記憶装置などであり、実施形態の情報処理にかかるプログラムや各種データを記憶する。入力部２１５は、カメラなどの撮像装置、ユーザー指示を入力するためのボタン、キーボード、タッチパネルなどの入力デバイスである。なお、記憶装置２１４は例えばＳＡＴＡなどのインターフェースを介して、入力部２１５は例えばＵＳＢなどのシリアルバスを介して、それぞれシステムバス２１９に接続されるが、それらの詳細は省略する。通信Ｉ／Ｆ２１６は無線通信で外部の機器と通信を行う。表示部２１７はディスプレイである。センサ２１８は画像センサや距離センサである。センサで計測した結果を画像として記憶装置２１４に記憶する。尚、ＣＰＵはプログラムを実行することで各種の手段として機能することが可能である。なお、ＣＰＵと協調して動作するＡＳＩＣなどの制御回路がこれらの手段として機能しても良い。また、ＣＰＵと画像処理装置の動作を制御する制御回路との協調によってこれらの手段が実現されても良い。また、ＣＰＵは単一のものである必要はなく、複数であっても良い。この場合、複数のＣＰＵは分散して処理を実行することが可能である。また、複数のＣＰＵは単一のコンピュータに配置されていても良いし、物理的に異なる複数のコンピュータに配置されていても良い。なお、ＣＰＵがプログラムを実行することで実現する手段が専用の回路によって実現されても良い。

【００１４】

図１は、情報処理装置の機能構成例を示すブロック図である。図１を用いて各機能構成について説明する。情報処理装置１は、画像入力部１０１、特徴抽出部１０２、尤度マップ生成部１０４、統合部１０７、出力部１０８、記憶部１０９から構成される。特徴抽出部１０２は、入力画像を処理して画像に含まれる様々な物体の特徴を示す画像特徴１０３を生成する。尤度マップ生成部１０４は、画像特徴１０３を入力すると、特定の物体が存在する可能性を示す尤度を領域毎に示す尤度マップを出力する学習モデルに基づいて、複数の尤度マップを生成する。ここでは、学習モデルの層間の結合重みづけパラメータのセットを異なる組み合わせで用意する。このパラメータセットは、後述する損失関数を用いて、隣接する同種類または同じくらいの大きさの物体が隣接していてもそれぞれを異なるマップで検出できるように学習させたものである。統合部１０７は、生成された複数の前記マップを統合して、特定の物体が存在する位置を示すマップを出力する。出力部１０８は、統合結果から、入力画像において認識対象となる物体が存在する位置を出力する。記憶部１０９は、情報処理装置がパラメータを学習する際に出力の目標値である教師値を記憶する。教師値は、予め用意され記憶されたデータを取得する構成以外にも、ユーザーや外部装置によって画像において特定の物体が存在する位置を示した教師値を入力できる構成でもよい。

【００１５】

＜認識処理の動作＞

次に情報処理装置が実行する処理を説明するフローチャートである図２を用いて、処理手順を説明する。以下の説明では、各工程（ステップ）について先頭にＳを付けて表記することで、工程（ステップ）の表記を省略する。ただし、情報処理装置１はこのフローチャートで説明するすべてのステップを必ずしも行わなくても良い。

【００１６】

本実施形態における、情報処理装置は、入力画像に含まれる複数の物体を検出する。例えば、物体が複数人映っている画像から、物体の存在する位置を検出する。そのために、

入力画像から物体を示す画像特徴を抽出する。次に、抽出された画像特徴に基づいて、物体が存在する可能性を示す尤度を出力するマップを少なくとも２つ以上生成する。ここで生成されるマップは、それぞれ異なる位置に存在する物体を検出できるように、それぞれ異なる検出結果（尤度）を出力するマップになるよう更新する。更新の方法は後述する。生成された複数のマップにおいて所定の値より大きい尤度が出力された注目領域について、同一のマップまたは異なるマップにおける注目領域の近傍の領域における尤度の値に基づいて損失値を出力する損失関数に基づいてマップを更新する。

【００１７】

まずＳ１で、画像入力部１０１が認識対象の物体が映った入力画像を入力する。入力画像は、ＲＧＢカラー画像以外でも、白黒画像やグレースケールの濃淡画像でも良い。また、カメラで撮像する画像以外でも、赤外線カメラによる赤外線写真や、ＬｉｄａｒやＴｏＦを代表とするアクティブ距離センサで得た距離画像でも良い。次にステップＳ２で特徴抽出部１０２が、入力画像から画像特徴を抽出する。画像特徴としては画素の色やテクスチャなどを集計した特徴ベクトルなど、公知の様々な方法が考えられる。マップ状の特徴であれば何れの方法でも適応でき、特定方法に限定されない。本実施形態の特徴抽出部１０２は多層ニューラルネット１０２ａを備えることとする。具体的には以下のような手順を行って画像特徴を抽出する。

【００１８】

＜画像特徴の抽出手順＞

Ｓ２において、多層ニューラルネット１０２ａによってマップ状の高次元画像特徴を抽出する方法について説明する。特徴抽出部が実行する処理を説明するフローチャートを図３に示す。まずステップＳ２０１においてニューラルネットが画像特徴を格納するための３次元配列Ｆ（ｘ，ｙ）を初期化する（ｘ，ｙは特徴の画素に関する添え字である）。次にステップＳ２０２からＳ２０７でニューラルネットの各層が入力画像に対して演算処理を行う。

【００１９】

Ｓ２０３において、本実施形態のニューラルネット１０２ａは図４に示すような各層の演算処理を行う。ここでいう演算処理とは、ニューラルネットワークの各層で、後段の検出処理において物体を検出するための画像特徴を入力画像から抽出する処理である。ニューラルネット１０２ａは入力画像４０１に対して複数回の畳み込みフィルタ処理を行う畳み込み処理４０２ａ、４０２ｂ、４０２ｃを備える。さらに各畳み込みの後に行う活性化関数処理を備える（図では略している）。さらにプール処理４０４ａ，４０４ｂを備える。Ｌ番目の層の畳み込みフィルタ処理および活性化関数処理を数式で表すと下記のようになる。

数式１

$$f_L(x, y, CHOUT) = \theta\left(\Sigma_{CHIN} \Sigma_{\Delta x, \Delta y = -K \sim +K} w^L(\Delta x, \Delta y, CHIN, CHOUT) \times f_{L-1}(x+\Delta x, y+\Delta x, CHIN) + B^L{}_{CHOUT} \right),$$

$$\theta(x) := 0 \quad if \quad x < 0$$

$$\theta(x) := x \quad Otherwise$$

【００２０】

ここで$f_L(x, y, z)$はＬ番目の層が出力する特徴マップの結果で、ｚ枚のマップからなる。（図４中に単位ｃｈとして付した数字は特徴マップの枚数である。）θ（・）は半波整流よりなる活性化関数、$w^L(\Delta x, \Delta y, CHIN, CHOUT)$（ただし$\Delta x, \Delta y \in \{-K, ..., 0, ..., K\}$）はＬ番目の層の畳み込みの重みパラメータ，$B^L$はＬ番目の層のバイアス項である。ＣＨＩＮはＬ－１番目の層が出力する特徴マップの番号、ＣＨＯＵＴはＬ番目の層が出力する特徴マップの番号を表す。なお上式ではＲＧＢ３チャンネルからなる入力画像Ｉ（ｘ，ｙ）は特徴マップ$f_0(x, y, z)$として扱うものとする。

【００２１】

なおここでは畳み込みの前後で特徴マップのｘ，ｙ方向のサイズが変化しないように、畳み込み処理の前に特徴マップf_{L-1}の周囲の画素に０値を充填してから畳み込むものとする（パディング処理と呼ばれる）。

【００２２】

プール処理４０４ａ、４０４ｂは特徴マップを所定の局所領域ごとに代表値で代表させることでマップのサイズを縮小する処理である。プール処理はＣＮＮの認識性能をロバストにする効果がある半面、結果の解像度が落ちるという性質がある。図６に示す例ではプール処理４０４ａ、４０４ｂはそれぞれ特徴マップを２×２画素ごとに統合して２分の１の解像度の特徴マップに縮小する処理である。

【００２３】

ここまでの演算処理により、ＣＮＮの各層の段階でそれぞれ特徴マップ４０３ａ、４０３ｂ、４０３ｃが生成される。以上はすべてＣＮＮに関する一般的な技術であり、上記の非特許文献１，また下記の非特許文献２，非特許文献３等で広く公知であるため、これ以上の詳細な説明は略す。必要に応じて先行文献を参照されたい。（非特許文献２：Ａ．Ｋｒｉｚｈｅｖｓｋｙ ｅｔ ａｌ．，ＩｍａｇｅＮｅｔ Ｃｌａｓｓｉｆｉｃａｔｉｏｎ ｗｉｔｈ Ｄｅｅｐ Ｃｏｎｖｏｌｕｔｉｏｎａｌ Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋｓ，ＮＩＰＳ ２０１２）。（非特許文献３：Ｍ．Ｄ． Ｚｅｉｌｅｒ，Ｒ． Ｆｅｒｇｕｓ，Ｖｉｓｕａｌｉｚｉｎｇ ａｎｄ Ｕｎｄｅｒｓｔａｎｄｉｎｇ Ｃｏｎｖｏｌｕｔｉｏｎａｌ Ｎｅｔｗｏｒｋｓ，ＥＣＣＶ ２０１４）。

【００２４】

なおＣＮＮは非特許文献２で行っているような大規模データベースを用いた分類タスクで重みパラメータをあらかじめ学習しておくものとする。この結果ＣＮＮの入力画像に近い低層の特徴マップ４０３ａは、画像中の線分の傾きのような単純な模様によく反応するマップが生成される。そして後段の高層の特徴マップ４０３ｃではより広い範囲の画像パターンを集約し、複雑なパターン形状に反応するようなマップが生成される。ＣＮＮの上記の性質は非特許文献３等で広く公知である。

【００２５】

次にＳ２０４では、ニューラルネット１０２ａが現在処理中の層が所定の層か否かを判定し、所定の層であればＳ２０５に進み、解像度を揃えてから（Ｓ２０５）、特徴マップ４０３を特徴の配列に連結して追加する（Ｓ２０６）。所定の層とは、設計時点であらかじめ取ってくると決めている層であって、一部でも良いし、全部の層を連結しても良い。所定の層ではない層に対しては、連結処理を行わない。これを繰り返して最終的に特徴マップが複数個連結された画像特徴１０３が得られる。上記の処理は数式では下記のように表される。すなわち、画素ごとに画像特徴を抽出し、特徴と画素とを対応づけた特徴マップを生成する。

数式２

$$F(x,y)=[f_1(x,y)^T, f_2(x,y)^T, \cdots, f_M(x,y)^T]^T$$

ここでf_1，f_2，・・・，f_Mは抽出された所定の特徴マップであり、特徴Ｆは上記特徴マップをＺ次元方向に連結したものである。ｘ，ｙは特徴マップの画素の座標である。

【００２６】

なお一般的にＣＮＮはプール処理を行うため、特徴マップの解像度は層によって異なっている。そのため前記の連結の前に連結特徴生成部２０４がアップサンプル処理４０５ａ、４０５ｂ（あるいはダウンサンプル処理）を行って解像度を変更し、各特徴マップを所定の解像度に揃える（Ｓ２０５）。ここでの解像度変更は値のコピーや線形補間といった一般的な方法を行えばよい。図６では各特徴マップ４０３ｂと４０３ｃをそれぞれ２倍と４倍にアップサンプルすることで、特徴マップf_1，f_2，f_3を入力画像Ｉ（ｘ，ｙ）と同じ解像度に揃えている。以上の結果、画像特徴１０３として特徴Ｆ（ｘ，ｙ）が得られた。

【００２７】

なお本形態では、入力画像Ｉ（ｘ，ｙ）と特徴Ｆ（ｘ，ｙ）の解像度は同一である。し

かし仮にアップサンプル（ダウンサンプル）の処理の倍率を変更すれば、入力画像Ｉ（ｘ，ｙ）よりも細かい（粗い）解像度の特徴を抽出することも可能である。なお本発明の説明においては特徴Ｆ（ｘ，ｙ）の個々の画素を一般化した名称として以降は「領域ブロック」と呼ぶこととする。以上が画像特徴を生成するＳ２の説明になる。

【００２８】

＜尤度マップの生成＞

Ｓ３では、尤度マップ生成部１０４が、入力画像から抽出された画像特徴を入力すると物体が存在する可能性を示す尤度を領域毎に対応付けた尤度マップを出力する学習モデルに、異なる結合重みパラメータを設定することにより複数の尤度マップを生成する。ここでは、尤度マップ生成部１～Ｎが、前段で得られた画像特徴に基づき、Ｎ枚の顔の尤度マップを生成する。尤度マップを複数生成するのは近接・重畳した複数の被写体でもなるべく漏れなく検出するためであり、基本的に各尤度マップの機能は同質である。なお、尤度マップ生成部１～Ｎは、それぞれが出力する尤度マップが、複数のマップで同じ物体を重複して検出しないように、かつ同じマップで隣接する物体が検出されないように、それぞれの異なるパラメータセットを学習済みであるものとする。学習方法については、学習処理の部分で後述する。

【００２９】

具体的には各尤度マップ生成部が、領域ごとに１×１サイズの畳み込み演算を行い、特定の物体（ここでは人物を検出したいので顔の特徴）があるか否かを示すスコア値（以降これを尤度スコアと呼ぶ）を算出する（下式）。ここではあらかじめ学習によって決定された重みパラメータｗ，ｂを用いて各領域の特徴の畳み込み演算処理を行う。これによりその領域に物体が存在する可能性を示す尤度を取得する。

数式３

$$v_k(x,y)=g\left(\Sigma_j\ w_{kj}\cdot f_{xyj}+b_k\right),$$

【００３０】

ここでv_kはｋ番目の尤度マップの尤度スコア（ｋ＝１，２，…，Ｎ）、f_{xyj}は高次元の特徴Ｆ（ｘ，ｙ）のｊ番目の次元の要素、ｗとｂは縦横１×１サイズの畳み込みのパラメータである。ｗとｂは画像特徴を元に、それぞれが独自に顔の尤度スコアのマップを生成するよう後述の学習処理によってあらかじめ調整されている。

【００３１】

関数ｇ（・）はスコア値を規格化する関数であり、

数式４

$$Sigmoid(x):=1/\{1+exp(-x)\}$$

等によって定義される。ここでは関数ｇは尤度スコアを０～１の範囲に規格化するものである。関数ｇとしては上記の形態に限らず規格化のためのその他の適当な非線形関数を用いてもよい。ここでは、尤度が高い（１に近い）ほど顔が存在する可能性が高い。

【００３２】

＜尤度マップの更新処理＞

Ｓ４～Ｓ７は尤度マップを更新する処理である。なお、尤度マップは、入力画像から各領域の尤度（対象物体が存在する可能性を示すスコア）を出力する過程で用いるマップであり、各領域には尤度を出力するための内部パラメータが設定されている。前述の学習の結果、尤度マップ内の内部パラメータは、尤度マップ毎に異なっており、同一の入力画像における共通する位置の領域に異なる尤度が出力される仕組みになっているものとする。

【００３３】

前段で得られた複数の尤度マップは、尤度マップ間で重複して同一の被写体を検出していたり、一方でどのマップでも尤度スコアが低い被写体があったり、等があり得る。そのため本処理ステップでは尤度マップの出力の調整を行う。

【００３４】

ただし尤度マップの結果が良好な場合必ずしも本ステップは必要ではない。また後述するように、本ステップと同様の機能が前段のＤＮＮの中で一体的に実現されるケースもあ

り得る。ただしここでは説明の便宜上、本処理を１つの独立した処理モジュールとして扱い、尤度マップ生成部１０４が各尤度マップの出力を調整する形態として説明する。

【００３５】

まず図１に示すように各尤度マップの間および各尤度マップの内部にはマップ間の結合１０５およびマップ内の結合１０６が備わっている。この結合を通じて、各マップの各領域ブロックは周囲のマップや領域の状態に応じて自身の状態を更新する。この結合の具体的な形態例の１つを図５に示す。図に例示するように、マップ間および同一マップ内のブロックの間にマップ間抑制結合２３およびマップ内抑制結合２４が備わっている。なお、図５の入力画像は、画像の左側にいる人物２人が重なっている。この２人の人物が同じ領域（ブロック）内で検出される例を示したのが尤度マップ群２２の尤度マップ２と３の出力である。この場合、検出結果の見え方としては、画像座標（５，３）に２つ分のスコア（閾値より大きい尤度）が検出されていることがわかる。異なる尤度マップでは、異なる人物に対して尤度が高くなるように学習済みの状態で、同じ画像座標に閾値より大きい尤度が検出されているのは、同じ人物のダブルカウントではなく、同じまたは隣接領域に存在する物体２つ分を検出したことになる。従来の方法では、入力画像に対して１枚の尤度マップを出力するため、隣接する物体がある領域には１人分の検出結果しか得られなかった。しかし、本実施形態では、同じ領域（ブロック）に隣接する物体がある場合でも、別々の尤度マップがそれぞれの尤度を算出するため、隣接する物体でも検出が可能になる。

【００３６】

この結合の入力信号の総和を下記式に基づいて算出し（Ｓ５）、各尤度マップv_kをそれぞれ更新する（Ｓ６）。なお、ωを含む重みは後述の学習処理の際に、一緒に学習される。色々な箇所で損失値の計算がなされ、すべての損失の総和を最小化するように学習が行われる。ωを含む全重みは、上記のような損失総和を最小にするように適宜調整される。z^{INTRA}とz^{INTER}はそれぞれ尤度マップ内、尤度マップ間、の反応の出方を調べ、それらの影響を加味して反応を増強するか減少するかを決める調整値になる。

数式５

$$z^{INTRA}{}_k(x,y)=\Sigma_{\Delta x,\Delta y\in R}\ \omega^{INTRA}{}_{\Delta x\Delta y}\cdot v_k(x+\Delta x,y+\Delta y),$$

$$z^{INTER}{}_k(x,y)=\Sigma_{k'\neq k}\ \omega^{INTER}{}_{kk'}\cdot v_{k'}(x,y)$$

数式６

$$v_k{}^{NEW}(x,y)=g(\alpha_k v_k+z_k{}^{INTRA}+z_k{}^{INTER}+\beta_k).$$

【００３７】

ここで$z_k{}^{INTRA}$と$z_k{}^{INTER}$はマップ内・マップ間の入力信号の総和，ω^{INTRA}とω^{INTER}は結合重みの強度である。$z_k{}^{INTRA}$と$z_k{}^{INTER}$は、それぞれ尤度マップ内、尤度マップ間、の反応の出方を調べ、それらの影響を加味して反応を増強するか減少するかを決める調整値になる。調整前の値$v_k{}^{NEW}(x,y)$にＺの二つの値を足し、発散しないようｇ（）でゲイン調整する。Ｒは近傍のブロックの集合であり、同一マップ内で相互に結合する範囲を示している。αおよびβは尤度マップの出力の調整値である。パラメータα，β，ωは後述の学習処理によってあらかじめ適切に調整されているとする。

【００３８】

上記の更新処理は複数回繰り返してもよいし、一度のみでもよい。図６にこの更新処理についてのブロック図を２つの形態例で示す。図６（Ａ）は再帰的な結合処理を示したものである。図６（Ｂ）は図６（Ａ）の再帰結合の繰り返しを３回に限定し、結合を全て等価なフィードフォワード結合に置き換えたものである。図６（Ｂ）のような形態は特に学習時やハードウェア実装時に好適な場合がある。

【００３９】

ここまでのＳ４～Ｓ７の処理は、複数の尤度マップを参照して、１つの推定結果を得るための統合処理の一例を示すものである。複数のマップで示された尤度を、統合前に互いに参照することによって、学習モデルの学習が十分に進んでいない状態でも、ルールベー

スで物体の位置が検出可能である。＜尤度マップの注目領域の尤度＞を数式５に基づいて、＜相互に入力＞して出力を調整する。１つの物体が２か所以上に検出されることや、２つ以上の物体が融合してしまうことといった状態を防ぐことを目的とした処理である。

【００４０】

上記に相当する機能が実現されるのであれば、別の形態例として図７のような形態も本発明の適用例の１つである。ここではマップ内・マップ間の結合重みに相当する処理としてニューラルネットの畳み込み処理２４を用いており、再帰的な結合は用いない。畳み込み処理２４は３チャンネルの尤度マップを入力とし、３チャンネルの尤度マップを出力としている。これにより図５のマップ内・マップ間の結合による出力調整と等価な出力調整処理を実現することが可能である。なお図中では例として黒丸が畳み込みの負の係数の重みを、白丸が正の係数の重みを示している。

【００４１】

＜統合処理＞

Ｓ８は、統合部１０７が、各尤度マップに分散している検出結果を統合し、統一した結果を生成するステップである。本ステップではまず統合部１０７が各尤度マップｋの各領域ブロックの尤度スコアを調べ、スコアが所定の閾値θ_kよりも大きい場合に顔が検出されたと判断して変数d_kに１の値をセットする（下式）。この処理によって、この処理によって、尤度が所定の値より大きい領域を物体の存在する候補領域とする。

数式７

$$d_k(x,y)=1 \quad if \quad v_k(x,y) \geqq \theta_k$$
$$d_k(x,y)=0 \quad Otherwise$$

【００４２】

上記のｄを要素とする３次元の行列（ｘ、ｙ、およびｋの３次元）をここでは尤度マップＤとする。さらにこの時、非最大値抑制処理も併せて行う。具体的にはｋ番目の検出結果d_kにおいて、所定の距離以内に複数の物体が近接して検出された場合は多重検出であると判断する。そして、尤度スコアが最も高い物体のみを真の検出として残し、低い方を偽として削除する（非最大値抑制処理は非特許文献１等で一般的な公知の方法であるのでここでは詳細を省略する）。

【００４３】

なお各尤度マップの中では非最大値抑制処理を行う一方で、各尤度マップ１～Ｎの間については非最大値抑制処理を行わない。複数の尤度マップの同一位置に複数の反応が生じた場合は、近接した位置に複数の物体が存在すると判断していずれの検出結果も残す。

【００４４】

なおさらにこのとき、被写体のより詳細な位置を推定してもよい。具体的な例の１つとしては例えば尤度マップの尤度スコア値からサブピクセル推定の方法で行う（サブピクセル推定は各尤度マップでそれぞれ独立に行う）。サブピクセル推定の詳細については非特許文献５等で広く公知であるためそちらを参照されたい。（非特許文献５；Ｐｓａｒａｋｉｓ ＆ Ｅｖａｎｇｅｌｉｄｉｓ，Ａｎ Ｅｎｈａｎｃｅｄ Ｃｏｒｒｅｌａｔｉｏｎ－Ｂａｓｅｄ Ｍｅｔｈｏｄ ｆｏｒ Ｓｔｅｒｅｏ Ｃｏｒｒｅｓｐｏｎｄｅｎｃｅ ｗｉｔｈ Ｓｕｂ－Ｐｉｘｅｌ Ａｃｃｕｒａｃｙ，２００５）。

【００４５】

また他の詳細な推定の方法としては位置やサイズを回帰推定するマップを別途用意し、マップの値に基づいて物体の位置・サイズを微調整してもよい（位置およびサイズ推定のための推定マップは各尤度マップｋで用意する必要がある）。本発明においてこれらの工夫の有無は発明の本質に関わらないため詳細を省略する。非特許文献１等で公知なため必要に応じて参照されたい。

【００４６】

以上のようにして尤度マップを統合した結果をまとめ、例えば図５の統合結果２３のような検出結果のリストとして出力する。図には検出した物体の位置と尤度スコアからなるリストの例を示している。

【００４７】
最後にステップＳ９で結出力部１０８が上記の統合結果に基づき顔の枠等を表示デバイス等に出力する。以上で情報処理装置の認識動作が終了する。
【００４８】
なお統合処理の他の派生の形態としては、１枚ずつ尤度マップを調べるのでなく、一旦全マップを重み付き和等して１枚のマップにしてから尤度スコアを調べる、等の形態も考えられる。また非最大値抑制の有無やその方法についてもさまざまな選択肢がある。また尤度スコア値の閾値θ_kについても１段階のみでなく２段階の閾値を用いる等も考えられる。このように尤度マップの統合の形態については複数考えられ、特定の形態に限定されない。
【００４９】
＜学習処理の動作＞
次に本認識装置の学習動作について説明する。学習動作の際の情報処理装置の機能構成例を図８に示す。ここでは学習に必要な損失値算出部２１０が追加されている。
【００５０】
学習の処理を説明するフローチャートは図９（Ａ）である。まず、図９（Ａ）のステップＳ２１で、画像入力部１０１が、学習画像のセット（バッチデータ）を選択して画像を入力する。また同時に、記憶部１０９が各学習画像に対応する教師値を統合部１０７および損失算出部２１０へ入力する（ステップＳ２２）。教師値は、各画像中に存在する物体の中心位置を示したものである。なお、教師値は検出する物体やタスクによって異なる。例えば、人物を検出する場合は、人の顔の中心位置にＧＴを与える。物体を検出する場合は重心位置等にＧＴを与える。具体的には、図１０（Ａ）の入力画像１００１に対する教師値は１００２であって、人物の顔の中心位置がある領域にラベルを付けた教師値になっている。教師データは、入力画像に対して正解の位置の座標のみを対応付けたデータでもよいし、人物の顔の中心位置には１、それ以外の位置には０を入れたマップ形式のデータでもよい。
【００５１】
次にＳ２３で特徴生成部１０２が、入力画像についての特徴を生成し、ついで尤度マップ生成部２０４ａ～２０４ｃが尤度マップを生成し、統合部１０７がそれらを統合した結果を生成する。尤度マップの統合は、予め決められた重みで統合してもよいし、統合方法を学習してもよい。
【００５２】
次に学習に必要となる認識結果の損失値の計算を行う。本実施形態の学習においては最終の統合結果についての損失値と、中間の生成物である尤度マップについての損失値の両方を学習計算に用いることとする。
【００５３】
＜統合した尤度マップに対する損失関数＞
まず統合した尤度マップに対する損失関数の方法は以下である。
【００５４】
Ｓ２４では、統合部１０７が、統合した尤度マップに対する損失関数と予め与えた教師値とを比較して物体の検出位置についての損失関数（第２の損失関数）を用いた損失値を出力する。前記複数の尤度マップを統合した結果に対して、前記取得された位置に対応付けられた前記尤度が閾値より小さい場合は前記尤度を上げるための損失関数である。なお、損失値はさまざまな形態の利用が可能である。ここでは例として下式のように二乗誤差を用いて、領域ブロック（ｘ，ｙ）ごとに損失値を計算し、総和する。まず、教師データが示す物体の正解位置と、学習モデルに入力画像を入力することで推定された物体の位置と検出された物体の数との差を求める。
数式８

$$Loss^{-}(D, T^{NUM}) = \Sigma_{xy}(\Sigma_{k} d_{k}(x, y) - T^{NUM}(x, y))^{2}$$

【００５５】

ただしＤは統合した尤度マップ、T^{NUM}（ｘ，ｙ）は教師値であり、領域ブロック（ｘ，ｙ）に顔の中心が位置する物体の総数を与える。上式の損失値を使って教師あり学習を行うことで、各尤度マップの反応結果が真の物体の数となるべく一致して反応するよう、ネットワークの重みパラメータが調整される（学習における重みパラメータの更新の方法については後述する）。つまり、数式８の損失関数は、正解位置に高い尤度が検出されなかった場合、または間違った位置に高い尤度が検出された場合は、すべての尤度マップに対して同程度の損失値を出力する。

【００５６】

また損失関数の別の形態として下式のように交差エントロピーを用いることも可能である。数式９によれば、ＧＴで示された正解位置に対応付けた各マップの尤度を比較する際に、いずれかの尤度マップを用いて物体の位置が検出できた場合がある。その場合は、その正解位置に対応づけられた尤度が低い他の尤度マップについては、その正解位置に物体が存在する可能性を示す尤度を小さくするように、各尤度マップ生成部（学習モデル）の層間の重みパラメータを学習する。ある領域に複数の物体が含まれる場合は、その物体の数に応じて大きな尤度を出力する尤度マップを準備する必要がある。そのため、ＧＴから物体の数だけ大きい尤度（１．０等）を示すＧＴマップを生成し、もしＧＴの数よりも大きな尤度を示す尤度マップの数が少ない場合は、より物体を積極するように周囲の画像特徴をがくしゅうする。数式９の損失関数を用いることで、数式８よりもより細かい学習ができる。

数式９

$$Loss(V,T)=\Sigma_{xy}\{\Sigma_k -t_k(x,y)\ log(v'_k(x,y)) -(1-t_k(x,y))\ log(1-v'_k(x,y))\}$$

【００５７】

ただしＴは教師値であり、０か１かを要素ｔの値として持つ３次元の行列である。Ｔの各要素は各領域・各マップに物体（の中心）が存在するか否かを示している。

【００５８】

なお、ここでは物体が同一ブロック（ｘ，ｙ）内に複数存在する場合にも適切に尤度マップの損失値を計算するため下記のような工夫を設ける。まず、あるブロック（ｘ，ｙ）にｎ個の物体が存在する場合、教師値Ｔ（ｘ，ｙ）の値として、先頭にｎ個の１の値、残りをＮ－ｎ個の０の値からなるＮ要素のベクトルｔを与える。次に、尤度マップv_k（ｘ，ｙ）の尤度スコア値を降順にソートし、この値をv'_k（ｘ，ｙ）とする（以降ｔおよびｖ’を＜ソート教師値＞および＜ソートスコア値＞と呼ぶ）。このようにしてから、数式９を用いて損失値を計算する。

【００５９】

具体例を１つ示す。仮にいまＮ＝４枚の尤度マップがあり、真値としてブロック（ｘ，ｙ）に２つの物体が存在するとする。さらに認識結果として当該ブロック（ｘ，ｙ）の尤度マップの尤度スコア値が

数式１０

$$V(x,y)=[v_1(x,y),v_2(x,y),v_3(x,y),v_4(x,y)]^T$$
$$=[0.1,\ 0.7,\ 0.5,\ 0.9]^T$$

と得られているとする。このとき、尤度スコア値をソートしたソートスコア値、およびソート教師値はそれぞれ

数式１１

$$V'(x,y)=[0.9,0.7,0.5,0.1]^T,$$
$$T(x,y)=[1,1,0,0]^T$$

となる。当該ブロックの損失値は

数式１２

$$Loss = \Sigma_k -t_k log(v_k') -(1-t_k) log(1-v_k')$$
$$= 0.105+0.358+0.693+0.105$$
$$= 1.261$$

と算出される（添え字を一部省略している）。各尤度マップが全体として全認識対象を過不足なく検出していれば、上記の損失値はゼロとなる。その際どの尤度マップがどの物体を検出したかは問わないことに注意されたい。この損失値は特定の対応関係（例えば尤度マップ１が前側の物体、尤度マップ２が後側の物体に反応するといった関係）を特に設けず、全体として検出精度が上がるよう各尤度マップ生成部のパラメータセットを学習させることを意味する。以上が尤度マップの＜統合結果＞についての損失値となる。

【００６０】

なおここで示したようにＤＮＮにおける損失値は、さまざまな形態の損失関数を採用することが可能である。本発明の適用対象は特定の形態の損失値に限定されない（なお学習の計算の都合上、損失関数は解析的に微分計算できる形が好適である）。

【００６１】

＜尤度マップの損失値算出＞

次に、それぞれの尤度マップについて、尤度マップ内の各領域に対応づけられたスコアに基づいた損失値の算出の方法について説明する。まずＳ２５では尤度マップ内の各領域に対応づけられたスコアに基づいた損失値を計算し、Ｓ２６では複数の尤度マップ間の対応する注目領域に対応づけられたスコアに基づいて損失値を計算する。模式図を図１０に示す。この二種類の損失値を適切に設計すれば、

（１）近接した複数の物体に対して複数の尤度マップが分担して反応する

（２）１つの物体に対して１つ以上の尤度マップが反応しない

の二つの性質を持った尤度マップ生成部１０６のパラメータセットを複数パターン学習で得ることが可能である。

【００６２】

まず単一の尤度マップについての損失関数（第１の損失関数）を下式のように定義する。複数の尤度マップのうち同一の尤度マップ内の第１の注目領域について、該注目領域の近傍の領域に対応付けられた尤度と該注目領域に対応づけられた尤度との差に基づいて、近傍領域の尤度を下げるための損失関数を用いる。第１の損失関数は、注目領域の尤度と近傍領域の尤度との差が小さいほどより大きい損失値を出力し、近傍領域の尤度を小さくする方向の調整をする。また、第１の損失関数は、注目領域の尤度と、近傍の領域の尤度との差が大きいほど、より小さい損失値を出力するか、または損失値を与えない。このような損失関数を用いることで、同一マップ内の隣接する領域で得られる尤度のコントラストがよりはっきりし、１つの領域において１つの物体を検出するための尤度マップを得られる。言い換えれば、１つのマップ内の特定領域において複数の物体を検出することを抑制できる。

数式１３

$$Loss^{INTRA} = -\Sigma_{xy} \Sigma_{k} \Sigma_{\Delta x, \Delta y \in R} \xi(\Delta x, \Delta y) v_k(x+\Delta x, y+\Delta y)$$

ただし関数ξは

数式１４

$$\xi(x, y) := \psi_1 \exp(-(x^2+y^2)/\rho_1) - \psi_2 \exp(-(x^2+y^2)/\rho_2)$$

である。図１０（Ａ）に示すような、正のピークとピークの近傍に負の窪みを持つメキシカンハット型の関数１００３である（ψ_1，ψ_2，ρ_1およびρ_2は関数形状を決定する定数のパラメータである）。本損失値は尤度マップに対してξを畳み込みカーネルとして畳み込んだ結果の総和の値である。Ｒは畳み込みを行う領域の範囲である。

【００６３】

複数の反応が近接して同時に一枚の尤度マップ上に生じると、損失値$Loss^{INTRA}$は大きな値を取る。例えば仮に図１０（Ａ）のように物体２人が近接して写っている入力画像１００１が入力されたとする。尤度マップ生成部が生成した尤度マップ群１００４では、尤度マップ１が両方の物体に対して反応している（反応の強さをグレースケールの濃淡で表す）。一方の尤度マップ群１００５では尤度マップ１と尤度マップ２に分散して反

応が生じている。この場合メキシカンハット型の関数の性質のため、前者では大きな、後者では小さな損失値が算出される。

【００６４】

上記は「近接した複数の物体が一枚の尤度マップで同時に検出される」ことを抑制する損失関数の一形態であるが、本発明が適用可能な形態としてはこの他の形態も考えられる。例えば下式のような損失値の定義もあり得る。

数式１５

$$Loss^{INTRA} = \Sigma_k \theta(\Sigma_{xy} V_k(x,y) - \theta_{th})$$

【００６５】

ここでθ（・）は半波整流の関数である。同損失値は、各尤度マップの尤度スコアの総和が所定閾値θ_{th}を越えるとペナルティを与える。すなわち、複数の尤度マップのうち同一の尤度マップ内の、第１の注目領域について、該注目領域に対応付けられた尤度が閾値より大きい場合は該注目領域の近傍の領域に対応付けられた尤度を下げるための損失関数を用いる。また、該注目領域に対応付けられた尤度が閾値より小さい場合は、第１の注目領域の近傍の領域に対応付けられた尤度を下げるための損失値をより小さくなるように損失値を出力する。このためこの損失値を用いて学習を行うと、どれか１つの尤度マップだけが突出して物体に反応するような動作が抑制される。以上が単一の尤度マップに対して定義される損失値の説明である。

【００６６】

次に複数の尤度マップ間の対応する注目領域に対応づけられたスコアに基づいて損失値を出力する損失関数（第３の損失関数）の例を示す。第３の損失関数は、複数の尤度マップのうち所定の２つの尤度マップにおいて尤度を比較する。入力画像の物体の位置（第２の注目領域）に大きな尤度が対応づけられた場合に、２つの尤度マップで競合する（対応領域の）尤度を下げるように学習モデルの重みパラメータを学習するための損失関数である。なお、第２の注目領域は、取得された教師値によって示された物体が存在する位置を含む各尤度マップの対応する領域である。ただし、注目領域に２つ以上の物体が存在する場合は、物体の数と尤度マップの数に応じて損失値が異なる。注目領域に物体が１つだけ存在する場合は、第２の損失関数は、２つの尤度マップにおいて共通する位置の注目領域に対応づけられた尤度に基づいて、２つの尤度マップにおける注目領域の少なくとも一方に対して大きな損失値を出力する。注目領域に物体が２つ以上存在する場合は、物体の数ｍと同じ数の尤度マップに対しては注目領域に推定された尤度が低いときはその尤度を大きくするような損失値が決定される。物体の数ｍより尤度マップが多い場合は、注目領域を尤度の大きい順にソートしたときのｍ＋１番目以降の尤度マップの注目領域に対し、尤度が小さくなるような損失値を決定する。また、入力画像の物体が存在しない位置に対応する尤度マップの領域に対して大きい尤度が推定された場合は、その尤度を小さくするために大きな損失値が決定される。つまり、第３の損失関数は、取得された物体の位置に基づいて入力画像に含まれる物体の位置が１つである場合は、複数の尤度マップのうちいずれか１つの尤度マップの第２の注目領域に対応付けられた尤度がより大きくなるようにする。また、取得された物体の位置についての情報に基づいて入力画像に含まれる前記物体の位置が複数ある場合は、物体の数に応じて尤度マップの第２の注目領域に推定された尤度が所定の閾値より大きくなるように調整するような損失値を決定する。損失値が小さいほど学習が進んでいると判断できるため、後の学習処理では算出された損失値を小さくする（または収束させる）ためにパラメータを調整する。この第２の損失関数によって、異なる尤度マップで同一物体を検出しないようにし、異なる尤度マップで異なる物体を検出できるように検出する対象を役割分担させることを目指す。なお、所定の２つの尤度マップとは、生成された複数の尤度マップからすべての組み合わせを指す。

数式１６

$$Loss^{INTER} = -\Sigma_{xy} \Sigma_k \Phi_k(T(x,y)) \log\{\Phi_k(V'(x,y))\}$$

と定義する。ただし、Φはソフトマックス関数

数式１７

$$\Phi_i(X) := \exp(x_i) / \Sigma_{j=1\,to\,N} \exp(x_j), \quad X = [x_1, \cdots, x_N]^T$$

である。Ｔ（ｘ，ｙ）およびＶ’（ｘ，ｙ）はＮ個の要素からなるベクトルであり、先掲の＜ソート教師値＞および＜ソートスコア値＞と同じものである。

【００６７】

上記の損失関数の結果の例を図１０（Ｂ）に示す。ここでは入力画像１００６のように単一の物体が写った画像が入力されている。これに対して反応結果１００８のようにマップ１とマップ２の両方が同時に反応した場合、損失値$Loss^{INTER}$は大きな値をとる。対して反応結果１００９のように正しくいずれか１つの尤度マップのみが反応している場合は、同損失値は小さな値をとる。

【００６８】

上記ではソフトマックス関数およびソートされた尤度スコアを用いたが、これは実現例の１つを示すのみである。「１つの物体に対してなるべく１つのマップの領域しか反応しない」ことを促進するような損失関数の設計であればさまざまな形態の採用が可能である。以上が尤度マップに関する損失値の定義である。

【００６９】

＜学習パラメータの更新＞

次にこのようにして得られた各種の損失関数が出力した損失値を使って、各マップ生成部のパラメータを更新する。本形態で学習更新の対象となるパラメータの１つは図８の尤度マップ生成手段２０４ａ，２０４ａ，２０４ｃ，の、それぞれの重みパラメータ w_k，b_kである。さらに各尤度マップ生成手段のマップ内結合とマップ間結合の重みパラメータ $\omega^{INTRA}\Delta x\Delta y$，　$\omega^{INTER}kk'$、および結合調整パラメータ α_k，β_kである（ここでｋはｋ番目の尤度マップ生成手段を表す添え字である）。図８では右上方向の矢印を付して学習対象のパラメータを示す。

【００７０】

上記各パラメータはそれぞれ乱数で初期化してから学習を開始する。特に各尤度マップはそれぞれ同質のマップであり、マップ間に競合的な損失値を与えて学習させることで、対象に対する反応が各マップに分散するように誘導することを企図している。もし各マップの重みパラメータｗｋ，　ｂｋの初期値が同一であると、常に同じ反応となって競合し、適切に学習が進まない。そのため各尤度マップは必ず異なる値で初期化する必要がある。

【００７１】

前述の方法で算出した損失値の総和の値をＥとし、Ｅを入力画像のバッチセットごとに算出し、これを減らすような勾配の方向に各パラメータを微小に更新すればよい。具体的には下式のようにパラメータの値を各々更新する（Ｓ２７）。例えば、Ｅが所定の値より小さい値に収束するまで、パラメータセットを更新する。

数式１８

$$E = Loss + \lambda_1 Loss^{INTRA} + \lambda_2 Loss^{INTER},$$
$$w^{t+1} = \eta_1 \partial E / \partial w^t + w^t,$$
$$b^{t+1} = \eta_1 \partial E / \partial b^t + b^t,$$
$$\omega^{t+1} = \eta_2 \partial E / \partial \omega^t + \omega^t,$$
$$\alpha^{t+1} = \eta_2 \partial E / \partial \alpha^t + \alpha^t,$$
$$\beta^{t+1} = \eta_2 \partial E / \partial \beta^t + \beta^t.$$

【００７２】

ただしＬｏｓｓは尤度マップの統合結果に対して算出された損失値、λ_1，λ_2は各種の損失値のバランスを調整するための定数、η_1，η_2は適当に設定された１以下の微小な係数（学習係数）である。なおここで各偏微分$\partial E / \partial x$の値はニューラルネットワークの一般的な方法である誤差逆伝搬法を用いて求める（誤差逆伝搬法については非特許文献４などに広く公知のためここでは省略する。また上式は見易さのため添え字を一部省略している）（非特許文献４：Ｙ．　Ｌｅ　Ｃｕｎ　ｅｔ　ａｌ．　Ｈａｎｄｗｒｉｔｔｅｎ

ｄｉｇｉｔ ｒｅｃｏｇｎｉｔｉｏｎ ｗｉｔｈ ａ ｂａｃｋ－ｐｒｏｐａｇａｔｉｏｎ ｎｅｔｗｏｒｋ． １９９０．）。

【００７３】

なお尤度マップ生成部の結合重みωは再帰的な結合を含むが、ωの学習には再帰的ネットワークの学習において一般的な方法を併せて用いるものとする（たとえば図６（Ｂ）のように有限繰り返し数の処理ブロックに展開した上で、誤差逆伝播法で更新する）。

【００７４】

なおここで学習対象としなかった特徴生成部１０２のニューラルネット１０２ａの重みについても、同様に誤差逆伝播法で学習してもよい（これは入力から出力まで一貫して重みパラメータを学習する形態でありＥｎｄ－ｔｏ－ｅｎｄ学習と呼ばれる）。

【００７５】

なおさらに、統合部１０７の検出閾値パラメータであるθ_k等を学習対象パラメータに加えるような形態なども考えられる。

【００７６】

＜派生の形態＞

ここまで各処理ブロックの機能モジュールやその学習形態について順を追って説明を行ってきた。ここでは考えられるその他の派生の形態についていくつかの例を加える。

【００７７】

例えば損失値の計算において、統合結果の損失値と尤度マップの損失値の算出方法についてそれぞれ述べたが、他の形態として、どちらか片方のみを用いたり、部分的に用いたり、学習の進み具合に応じてこれらを切り替えたりといった形態も考えられる。

【００７８】

また例えば、ここまでは物体を検出する際には、物体の中心位置を基準位置として学習し、検出したが、この基準位置を変えることもできる。例えば（尤度マップの数は増えるが）、物体の上下左右端を基準位置としてそれぞれを推定するようなマップを学習し、検出するような形態でもよい。

【００７９】

また、本実施形態ではＮ個の尤度マップ生成部、およびその結果としてのＮ枚の尤度マップを用いて対象を認識したが、この数Ｎを認識時に動的に変更するような方法も考えられる。例えば画像内に対象が多数重畳しているときは尤度マップ数が多いほうが、検出精度が高くなると考えられるが、物体の数に対してマップ数が多すぎるとノイズ状の反応が却って増えることや、余計な演算量が増えることがある。そのため尤度マップ数を適切に増減するような形態も考え得る。

【００８０】

これを説明するために図８に追加的にマップ数決定部２１１と重みパラメータ提供部２１２を示す。マップ数決定部２１１は画像特徴１０３に基づいて画像シーンを考慮して最も良好な結果が得られるようにマップの数ｎを決定する（例として対象物体で混雑した画像に対しては大きなｎを与える等）。次にその結果を受けた重みパラメータ提供部２１２は１～ｎの尤度マップ生成部に対してｎ個の重みパラメータの提供を行う。

【００８１】

マップ数決定部２１１がマップ数を決定するやり方として例えば以下の形態が考えられる。まず１個、２個、…、Ｎ個の尤度マップ生成部からなる、Ｎ通りの異なる設定の情報処理装置を用意し、各個に学習を済ませておく。次に入力画像ｘが与えられたときの検出結果の精度を各Ｎ通りの設定について調べ、精度の良し悪しの値を記憶しておく（検出精度の良さを測る基準として例えば先に挙げた統合部１０７の統合結果の損失値などを使えばよい）。

【００８２】

マップ数決定部２１１は画像ｘの画像特徴１０３を説明変量とし、検出結果の精度を目標変量として、各Ｎ通りの情報処理装置の検出精度を推定する回帰学習を行う（ニューラルネットやサポートベクトル回帰等の一般的な方法を用いる）。認識時にはこの回帰器の

推定結果を用いて、マップ数決定部２１１が各Ｎ通りの設定の検出精度の期待値を調べ、最も期待値の高かったマップ数ｎを採用する。なおこのときに、推定検出精度に使用マップの少なさ（＝総計算量）、も考慮するような合成指標を使って精度と計算量から使用マップ数を決定してもよい。以上のようにすることで、動的に尤度マップの構成を変更することが可能である。

【００８３】

またさらに別の派生の形態は以下のようである。これまで、本実施形態では説明の便宜上、各処理モジュールを明確に区別できるものとして説明してきた。ここで考えられる他の形態としては、ニューラルネットの各部が本実施形態の機能モジュールと同等機能を持ち、それらが境目なく結合した形態である。

【００８４】

例えば、特徴の生成部１０２、複数の尤度マップ生成部１０４、尤度マップ間の結合１０５や、マップ内の結合１０６、および統合部１０７、の各機能が、ＤＮＮの各層の上に分散して実現するような形態が考えられる。このような機能を実現するためには、どの層にどの機能的役割を実現させるかをある程度決めた上で、各機能が十分実現できる程度の層数および入出力チャンネル数、および層間の結合関係、を持ったＤＮＮを用意する。次に実現させたい機能的な役割に応じて中間層や最終層に対する損失関数を適切に設計し、学習データと教師値を与えて学習させる。学習がうまくいかない場合は中間層の出力を調べ、機能実現のために特徴表現や識別力が足りていないケース、出力調整に必要な繰り返し数（層数）が足りていないケース、等を同定して適宜必要な箇所のニューロンの層数やチャンネル数を増やす。１つの形態例として以上のような方法が考えられる。

【００８５】

またさらに別の派生の形態は以下のようである。同一種類の複数の物体に対して複数の尤度マップが反応することが本発明の要諦の１つであるが、ここでの「同一種類」とは物体カテゴリーの部分集合であってもよい。例えば、物体を見えのサイズ、アスペクト比や姿勢で分けたものをそれぞれ異なるカテゴリーとしてもよい。またさらに、任意の複数のカテゴリーをまとめた上位集合を作り、これを１つの種類であると見なす形態であってもよい。例えば、犬と猫を１つのカテゴリーとする等がこれに当たる。

【００８６】

例を示すと、種類の異なる複数のカテゴリーＡ，Ｂ，Ｃの物体があり、本情報処理装置に複数の尤度マップＸ，Ｙがあるとする。もし尤度マップＸ，ＹどちらもカテゴリーＡ，Ｂ，Ｃの物体全てを検出するように学習を行うのであれば、これは本発明の一形態である。さらに、尤度マップＸはカテゴリーＡ，Ｂに対して、尤度マップＹはカテゴリーＢ，Ｃに対して、それぞれ反応するように学習するような形態も、カテゴリーＢに対しては複数の尤度マップが反応するという点において本発明の形態の１つである。また尤度マップＸ，ＹともにカテゴリーＡ，Ｂ，Ｃの被写体すべてに反応するが、尤度マップＸは特にカテゴリーＡに優先的に反応し、尤度マップＹはカテゴリーＢに優先的に反応する、というように尤度マップを部分的に特性付けるような形態も考えられる。

【００８７】

以上ここまで本発明を適用して得られる派生的な形態について説明を加えてきた。本発明をなす根幹を述べると、＜同一種類の複数の物体が存在している＞ときに、それらが＜複数の同質の尤度マップ上に分散して検出される＞よう設計された各機能モジュールあるいはそれらの重みパラメータ、とまとめられる。上記目的に適う機能モジュール、重みパラメータ、（あるいは適切にパラメータを学習するための損失関数）は本発明の実施形態に含まれ、本発明の実現形態は特定の形態のみに限定されない。以上で実施形態１の説明を終える。

【００８８】

＜実施形態２＞

第二の実施形態では、第一の実施形態と同様に物体の検出を目的とする。情報処理装置の基本的な構成例のブロック図は図１１である。実施形態１と異なる点は各マップ３０４

ａ～３０４ｃに新たに複数の特徴集計部３０３ａ～３０３ｃが加わっている点である。本実施形態は各尤度マップ生成部に与える特徴にバリエーションを加えた形態となっている。これにより各尤度マップの出方に変化が生じ、近接した物体の分離度が向上する（理由は後述する）。

【００８９】

なお本実施形態では各尤度マップ生成部３０４の間および内部の結合は除いており、実施形態１の処理フロー中で行った尤度マップの更新は行わない。ただし派生的な形態として実施形態１と同様に上記結合を構成に含めて尤度マップの更新を行うことも考えられる。また、ハードウェア構成は実施形態１と同様に図２１のような構成を用いる。

【００９０】

画像特徴の抽出処理の説明のための模式図を図１２に示す。図１２（Ａ）はこれまでの画像特徴の形態図である。ニューラルネットの各階層の出力結果を連結して一種類の階層特徴

$$F(x,y)=[f_1(x,y)^T, f_2(x,y)^T, f_3(x,y)^T]^T$$

を生成して用いている。

【００９１】

図１２（Ｂ）は本実施形態で開示する画像特徴の抽出処理の形態である。階層特徴の生成時の集計の方法をＮ通りに変更することで、

$$F_k(x,y)=[f_{k1}(x,y)^T, f_{k2}(x,y)^T, f_{k3}(x,y)^T]^T$$

（ｋ＝１，２，…，Ｎ）と、Ｎ通りの特徴を生成している。

【００９２】

集計方法の具体例を図１３に示す。同図は特徴集計部３０３がＣＮＮの第ｊ層の出力の特徴$f^{raw}{}_j$に４通りのサブサンプルを行って４通りの画像特徴f_{1j}～f_{4j}を生成している。すなわち、モデル毎に異なる画像特徴が入力されるようになっている。ここではサブサンプルにより特徴マップの縦横それぞれの解像度を２分の１にしているが、２×２の領域ブロック範囲（Ｒａｎｇｅ（ｋ，ｊ）と記号を付して示す）の位相を都度変更しながらサブサンプルを行う。これにより各特徴f_{kj}がそれぞれ微妙に異なるバリエーションを持った画像特徴となっている。

【００９３】

本実施形態２の処理のフローのうち、特に画像特徴抽出部分（実施形態１でのステップＳ２に相当する）について詳細化したフローを図１４に示す。本フローでは特徴集計部３０３が、ステップＳ４２～Ｓ４９のループでＮ通りの画像特徴の集計および生成を行う。ステップＳ４３～ステップＳ４６では、尤度マップの番号ｋに応じて異なる集計範囲Ｒａｎｇｅ（ｋ，ｊ）を設定する（ステップＳ４４）。同範囲でサブサンプルを行って特徴f_{kj}を生成する（ステップＳ４５）。さらにステップＳ４７では$f_{kj}(x,y)$を連結して階層特徴$F_k(x,y)$を生成する。そして、尤度マップ生成部ｋへと出力する（ステップＳ４８）。

【００９４】

このようにサブサンプルの集計パターンを様々に変更することで、位相等の微妙に異なる特徴を複数の尤度マップ生成部に提供することができる。異なる特徴に基づいて物体の尤度スコアをそれぞれ判定するため、単一の特徴に基づいて判定するのに比較し、近接した物体パターンを分離・検出できる可能性が高い。なお同様の方法として、ニューラルネット１０２ａで高解像の入力画像を処理して解像度の高い画像特徴を得て用いることも考えられるが、ニューラルネット１０２ａの計算量の増大を伴う。本実施形態の形態は特徴の集計方法を変えるだけのため、計算量を特段に増やすことなく同種の効果が得られる。

【００９５】

なお上記は形態の一例であり、集計の方法の変化の付け方は他にもあり得る。（２×２ではなく１×２と２×１といった非正方形の範囲を用いる。サブサンプルの他に最大値プーリングや平均値サンプルを行う。一部分岐したＤＮＮを用いて特徴ごとに一部だけ異なる層の特徴マップを連結させる、等）。またベースとなるニューラルネットの特徴はここ

では階層型の画像特徴を用いているが、適用可能な形態はこれに限定されない。

【００９６】

なお学習時は実施形態１と同様に損失を計算し、誤差逆伝搬の方法で各重みを更新すればよい。

【００９７】

＜実施形態３＞

本実施形態では、本情報処理装置への情報入力部の１つとして画像の奥行き情報（以下２．５次元情報）を加え、これを利用する形態について説明する。さらにユーザーの指示を受ける表示切替指示部を設け、ユーザーの介在に基づいて認識結果の提示の仕方を切り替える形態について説明する。図１５に機能構成図を示す。また、ハードウェア構成は実施形態１と同様に図２１のような構成を用いる。

【００９８】

２．５次元情報の利用方法としては学習時に用いる場合と、認識時に用いる場合の２種類が考えられる。

【００９９】

＜学習時の２．５次元情報の利用＞

学習時の２．５次元情報の利用の仕方の１つとして、以下に例を挙げる。まず処理フローの図９（Ｂ）に示すように、観測した２．５次元情報のマップを真値の一種として与える（ステップＳ３３）。ここでは２．５次元情報のマップを変数Ｄｉｓｔ（ｘ，ｙ，ｄ）として表す。２．５次元情報のマップＤｉｓｔ（ｘ，ｙ，ｄ）は簡単のために尤度マップと同じ画像解像度を持った３次元の行列であるとする（なお奥行き方向ｄはｄ＝０，１，…，d_Nとあらかじめ離散化している。d_Nは最大の距離に対応する適当な値である）。行列Ｄｉｓｔの要素のうち、物体が存在する位置・奥行きに当たる要素には１、それ以外の要素には０が入っているとする。次に下式のようにマップ内の損失値の計算において２．５次元の値を利用する（ステップＳ３６）。

数式１９

$$\mathrm{Loss}^{\mathrm{INTRA}}(x,y) = -\sum_k \sum_{\Delta x,\Delta y,\Delta d \in R} \xi'(\Delta x,\Delta y,\Delta d) \times v'_k(x+\Delta x, y+\Delta y, d+\Delta d) \times \mathrm{Dist}(x+\Delta x, y+\Delta y, d+\Delta d)$$

ただし、ここでξ’はメキシカンハット関数を奥行き方向に拡張した下記の関数である。

数式２０

$$\xi'(x,y) := \psi_1 \exp(-(x^2+y^2+\kappa d^2)/\rho_1) - \psi_2 \exp(-(x^2+y^2+\kappa d^2)/\rho_2)$$

（κはスケール調整の定数）

$v'_k(x,y,d)$は尤度マップの尤度スコア$v_k(x,y)$を奥行き方向に複製して便宜的に３次元に拡張した変数である。（$v'_k(x,y,d) := v_k(x,y), d = 0,1,\cdots,d_N$）

上記式は、画像の像面上の距離、および奥行き上の距離、が両方共に近い物体を１つの尤度マップで検出することに対してペナルティを与えることを意味する。この損失値を使って学習することにより、奥行きおよび画像面上の距離の近い物体はなるべく各尤度マップに分散して反応が生じるように誘導される。

【０１００】

なお同様に距離情報を利用した派生の形態として、手前側の物体を大きな番号の尤度マップで検出し、遠い側の物体を小さな番号の尤度マップで検出したときにペナルティを与えるような損失値、といった形態も考えられる。ただしマップの数を大きく超える多数の物体が１列棒状に並んでいるような場合（集合写真等では頻繁に起こりえる）、このように距離と尤度マップを密接に対応させて学習させると、物体の検出分離度は逆に悪くなるので注意が必要である。本実施形態の数式１９の形態のように、奥行き情報を相対的にのみ用いる方法が本発明においてはより好適であると考えられる。

【０１０１】
＜認識時の２．５次元情報の利用＞
認識時に奥行き情報を利用する形態の１つとしては、２．５次元情報のマップを画像特徴の１つとして連結して認識に用いることである。
【０１０２】
また他の形態の１つは、認識時に、ユーザーの指示部である表示切替指示部４００１を用いて、ユーザーの指示と奥行き情報に基づいて表示を切り替える形態である。
【０１０３】
この後者の形態について詳細を述べる。入力画像を示す図１７（Ａ）および結果の表示例を示す図１７（Ｂ）～（Ｇ）を利用して説明する。まず認識処理が開始されると図１７（Ｂ）のような画像が入力される（ステップＳ６１）。次に尤度マップが生成される（ステップＳ６２）。次に尤度マップを統合することで、例として図１７（Ｃ）のような検出枠の結果が得られる（ステップＳ６３）。この結果をこのまま表示すると視認性が低い。そのため、考えられるユーザーインターフェースとしては２．５次元情報入力部４００２から入力された奥行き情報を用いて、検出物体のうちもっとも距離の近い物体を判定してその検出枠を表示することが考えられる（ステップＳ６４）。同時に拡大表示窓３０２ａにその拡大結果も表示する（ステップＳ６５）。結果例を図１７（Ｄ）に示す。次に表示切替指示部４００１の１つの形態例である左右矢印状のボタン３０１ａによってユーザーが指示を与え、表示の切り替えを行う（ステップＳ６６、ステップＳ６７）。このとき、ユーザーの左右ボタンの押し下げに応じ、被写体の左右位置順に結果を切り替えて表示する形態が考えられる。このように表示を切り替えた例を図１６（Ｅ）に示す。また物体の遠近順に切り替えて表示する、といった形態もあり得る。また、ユーザー指示部を用いずに、表示を一定時間ごとに自動的に切り替えるような形態も考えられる。
【０１０４】
なお派生の形態として、表示するのは上述のような検出枠等でなく、尤度マップの尤度スコア値を濃淡マップとして表示するようなことも考えられる。このようにマップの値をユーザーの指示に応じて表示することで、ニューラルネットの認識結果の内部状態の情報を提示するインターフェースとして利用することが考えられる。図１７（Ｆ）にその結果の例を示す。ここでは尤度マップの尤度スコア値を値の大きさに応じた灰色～黒色の矩形で示している。さらにユーザーが左右矢印状のボタン３０１ｂを押し下げて指示を与え、表示する尤度マップの結果を切り替えて表示した例を図１７（Ｇ）に示す。図１７（Ｆ）と図１７（Ｇ）は２つの異なる尤度マップの尤度スコアを切り替えて示している。そのため尤度スコア値の濃淡の一部には重なった領域が存在し、それぞれの値が異なっている点に注意されたい（例えば記号３０３を付した矩形）。ここでは尤度マップの尤度スコア値をそのまま全て表示しているが、検出結果と２．５次元情報を併せて用いることで、各検出された物体ごとに尤度スコア値を分けて表示する、等も考えられる。この機能によって、ユーザは学習済みモデルの学習が十分に進み、物体を正確に検出できていることを確認できる。
【０１０５】
＜実施形態４＞
実施形態４では画像の意味的領域分割（セマンティック・セグメンテーション）を行う情報処理装置について説明する。本発明が物体毎の物体検出タスクのみならず、領域ラベリングのタスク等種々のタスクに対しても広く適用可能であることを示す。
【０１０６】
重畳・近接した物体の領域を正しく同定・分離するのは一般に困難な課題である。非特許文献６などはこのためにまず画像中の物体の検出を行い、同定された物体領域を入力として再度ＤＮＮの処理を行い、各画素が前景か否かを判別して物体の前景領域を生成している。この方法は領域分割の精度が高いが、物体ごとにＤＮＮの処理を行うため演算量は多い（非特許文献６：Ｋ．Ｈｅ，Ｇ．Ｇｋｉｏｘａｒｉ，Ｐ．Ｄｏｌｌａｒ，Ｒ．Ｇｉｒｓｈｉｃｋ，Ｍａｓｋ Ｒ－ＣＮＮ，ＩＣＣＶ，２０１７）。

【０１０７】
本実施形態の情報処理装置は物体ごとにＤＮＮの処理を行わない。非特許文献６が物体ごとに前景領域を判別するマップを用意するのに対して、本発明の実施形態はＮ個の尤度マップのみを用いる（マップの数Ｎは画像中に同時に出現する対象物体の数より小さいことを想定している）。同マップを以降領域尤度マップと呼ぶ。本実施形態では、物体領域として物体の領域を同定することを目的とする。また複数の物体については、物体ごとに領域を分離・同定することを目的とする。
【０１０８】
＜学習動作＞
学習時には、教師値として図２０（Ｂ）のように領域のラベルの真値、ｌ（ｘ，ｙ）∈｛０，１，…，Ｌ｝を用意する。（図では物体領域の色の違いでラベルの値の違いを表現している）。ラベルｌ（ｘ，ｙ）が０の領域ブロックは物体が存在しない領域である。ラベルが１，２，…，Ｌの領域は物体領域であり、異なる数値でそれぞれ画像中の異なる物体を意味している。物体毎の領域を示した領域情報が本実施形態の教師データである。すなわち、教師値とは、各画素にどの物体が存在するかを示すラベル（例えば、左の人物は１、中央の人物は２、左の人物は３、人物がいない領域は０）を持った画像情報である。
【０１０９】
各領域尤度マップは、物体領域（ｌ（ｘ，ｙ）＞０）に対しては大きな尤度スコア値が、それ以外の領域（ｌ（ｘ，ｙ）＝０）に対しては小さな尤度スコアが出るように重みパラメータを学習する。用いる損失関数としては実施形態１の物体検出タスクと同様の交差エントロピー等であればよい。具体的には例えば実施形態１の数式８や数式９を用いる。また実施形態１の物体検出タスクでは物体の中心の領域ブロックに正の教師値を与えて学習したのに対して、領域判別タスクでは物体の領域に対応するブロックすべてに正の教師値を与えて学習する。
【０１１０】
ここで本発明の特性である＜複数の尤度マップが分散協調して認識を行う＞ことを実現する形態として、さらに以下のような二つの特性を実現する損失関数の項を加える。
（１）１枚の領域尤度マップは、近接や重畳している複数の物体領域に同時に反応しない
（２）複数の領域尤度マップは、同一物体の領域に対して同時に反応しない
まず上記（１）を実現する損失関数を説明する。形態としては種々あり得るが、例えば下式のようなものである。
数式２１

$$Loss^{INTRA} = \Sigma_k \Sigma_{xy} \Sigma_{\Delta x, \Delta y \in R} \{1-\delta(l(x,y)-l(x+\Delta x, y+\Delta y))\} \times v_k(x,y) \times v_k(x+\Delta x, y+\Delta y)$$

上記の損失関数は、１枚の領域尤度マップが、異なる複数の物体の領域に対して反応した場合に損失値のペナルティを与える。ただしδはディラックのデルタ関数であり、２つの領域のラベルが同じ値の時に１、異なる時に０を返す。v_kはｋ番目の領域尤度マップの物体領域の尤度スコア値である。またここでＲは同時反応を抑制する所定の近傍のブロックの範囲であり、この範囲の外であれば異なる物体の領域に反応してもペナルティを与えない。
【０１１１】
次に先述の（２）の特性を実現する損失関数についてであるが、これは実施形態１の数式１６等を使えば実現できる。
【０１１２】
以上に述べた損失関数を用いて、損失値の総和を下げるように学習対象の各パラメータを学習更新する。学習が進めば、領域尤度マップが正しく物体の領域に反応し、且つ近接・重畳した複数の被写体の領域はなるべく異なる複数の領域尤度マップに分散して検出されるようになる。
【０１１３】

＜認識動作＞

図１８は領域尤度マップを生成する情報処理装置の機能構成例のブロック図である。基本的な構成は実施形態１のものとほぼ同一であり、同一の処理を行うモジュールには同じ番号を付している。実施形態１と異なる点の１つは物体位置推定部５００が新たに加わっている点である。また他の異なる点の１つとしては尤度マップ生成部１０４および統合部１０７が検出する対象が物体の中心位置（実施形態１）か、物体の前景領域か（本実施形態）の違いがある。処理を説明するフローチャートは図１９、処理の過程と結果の一例は図２０に示す。また、ハードウェア構成は実施形態１と同様に図２１のような構成を用いる。

【０１１４】

ここで、領域尤度マップを用いた認識処理の流れを簡単に説明する。これまでの実施形態と同様に、まず、情報処理装置の画像入力部１０１が認識対象となる入力画像を入力する（ステップＳ７１）。次に、特徴抽出部１０２が、入力画像から画像特徴１０３を抽出する（ステップＳ７２）。複数の領域尤度マップからなる尤度マップ生成部１０４が対象被写体の領域ブロックか否かを示す尤度スコアのマップを生成する（ステップＳ７３）。領域尤度マップ１の結果例を図２０（Ｃ）、同領域尤度マップ２の結果例を図２０（Ｄ）に示す。図では尤度スコア値の大きさをグレースケールの矩形で表示している（黒いほどスコアが高い）。なお各領域尤度マップは近接した物体が異なるマップ上に分散して検出されるように、尤度マップ生成部１０４が用いる学習済みモデルはあらかじめ学習が施されているものとする（その方法については後述する）。

【０１１５】

ステップＳ７４～ステップＳ７６は領域尤度マップの統合処理になる。まずは統合部１０７が所定閾値以上の尤度スコア値を含む領域ブロックを物体の領域とする（ステップＳ７４）。図２０（Ｅ）に物体が存在する領域と判定された領域の例を示す。ここでは領域尤度マップ１で物体が存在する領域と判定した領域は黒の矩形で、同じく領域尤度マップ２で物体を検出した領域は灰色で示している（なお領域尤度マップ１と２の両方が物体領域とした箇所は、尤度スコア値がより高い方のマップの色で示している）。つまり、図２０（Ｅ）の物体領域マップは前述した尤度の大きさに応じた色分けではなく、領域尤度マップ毎に閾値以上の尤度を検出した位置を示すマップである。

【０１１６】

次にステップＳ７５では、物体位置推定部５００が物体の位置検出を行い、物体の位置の情報を統合部１０７に提供する。物体の検出の方法はこれまで実施形態１や非特許文献１等に開示されるような方法を別途行うものとする。ここでは実施形態１の方法を用いて、画像特徴１０３に基づいて検出したとする。図２０（Ｆ）に記号５０１ｆ～５０３ｆを付して検出された物体の検出枠の例を示す。

【０１１７】

次にステップＳ７６では、統合部１０７が、物体が存在する領域を個々の物体の領域に分割する。方法としてはまず検出枠と領域尤度マップとを対応づける。ここでは各枠内の領域のうち物体と判定した領域の数が最も多かった領域尤度マップを各枠に対応させる。（例として図２０（Ｆ）の検出枠５０３ｆの場合、領域尤度マップ１を対応させる。次に各検出枠に、対応する領域尤度マップの物体領域を各物体の領域として決定する（例として図２０（Ｇ）の領域５０３ｇ）。

【０１１８】

最後に、ステップＳ７７で、出力部１０８が、各物体が存在する領域を示す結果を出力する。例えば図２０（Ｇ）のように一体ずつ物体領域５０３ｇを表示してもよい。その際、複数の物体が重畳している領域（例えば図２０（Ｈ）に符号５０４ｈを付して示した灰色の矩形の領域）は各領域尤度マップの尤度スコア値の大小や２．５次元情報を用いて前側か後側かを推定してもよい。この場合被遮蔽領域に矩形５０４ｈのような色の変化をつけて表示してもよい。また図２０（Ｉ）のように全ての物体領域を重ね、物体ごとに領域の色を変える等して表示してもよい。実施形態３と同様にこれらの表示を切り替えるユー

ザー指示部等を備えてもよい。

【０１１９】

なお領域尤度マップの統合にはここで述べた以外に様々に細かな派生的工夫が考えられる。例えば物体領域の数ではなく、尤度スコア値の大小を対応づけの判定に用いる、２．５次元情報を用いて前側の物体の検出枠から領域の対応づけを行っていく。対応づけ済みの領域は取り除いて重畳領域を正確に対応づける。領域尤度マップの純度の高い（他のマップとの混在が少ない）枠の領域から対応づけしていく、等である。また物体の検出枠を使わずに、領域尤度マップのうち孤立した物体領域をそれぞれ別個の物体の領域とする簡便な方法等もあり得る（遮蔽された物体が複数の領域に過剰分割される可能性もあるため注意を要する）。このように様々な形態の方法が考え得るが、本発明の根幹に関わりのない表層的な相違のため、ここでは詳細を略す。

【０１２０】

以上、本発明が物体検出タスクのみならず領域判別タスクにも適用可能であることを示した。特に、複数の領域尤度マップを同時に使うことで、同種の物体が近接・重畳した場合に適することを示した。またこれは非特許文献６のような、物体の検出を行ってから物体ごとにＤＮＮで領域判定を行う演算量の大きな方法と異なる形態であることを示した。

【０１２１】

本発明は、以下の処理を実行することによっても実現される。即ち、上述した実施形態の機能を実現するソフトウェア（プログラム）を、データ通信用のネットワーク又は各種記憶媒体を介してシステム或いは装置に供給する。そして、そのシステム或いは装置のコンピュータ（またはＣＰＵやＭＰＵ等）がプログラムを読み出して実行する処理である。また、そのプログラムをコンピュータが読み取り可能な記録媒体に記録して提供してもよい。

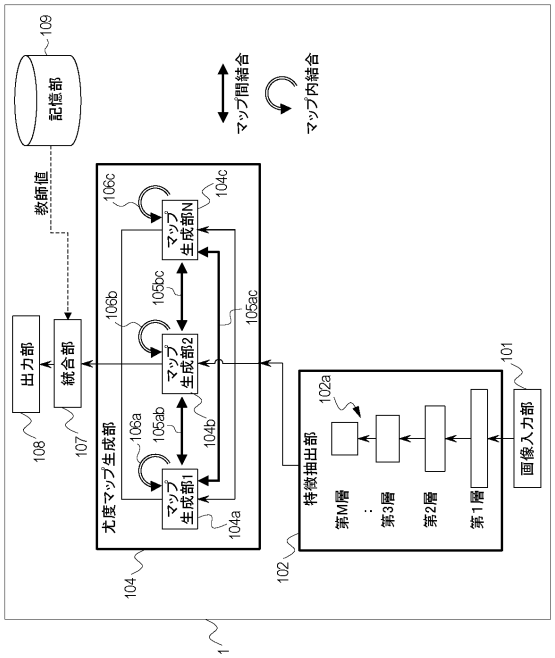
【符号の説明】

【０１２２】

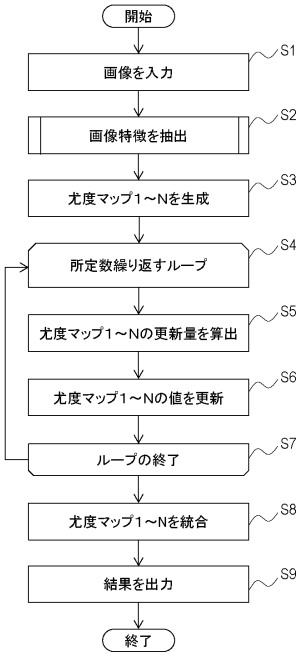
１０１　画像入力部
１０２　特徴抽出部
１０３　画像特徴
１０４　尤度マップ生成部
１０５　マップ間結合経路
１０６　マップ内結合経路
１０７　統合部
１０８　出力部
１０９　記憶部

【図面】

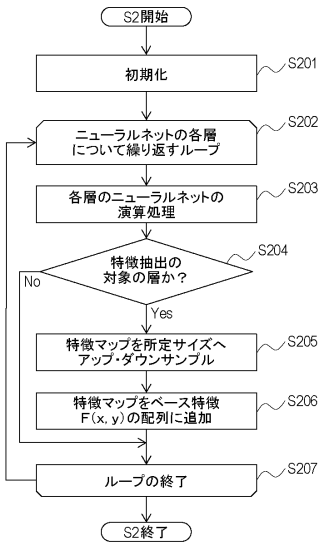
【図１】


109
記憶部
教師値
マップ間結合
マップ内結合
106c
104c
マップ生成部N
105bc
106b
105ac
出力部
統合部
マップ生成部2
101
画像入力部
102a
特徴抽出部
第M層
第3層
第2層
第1層
108
107
106a
105ab
104b
尤度マップ生成部
マップ生成部1
104a
104
102
1

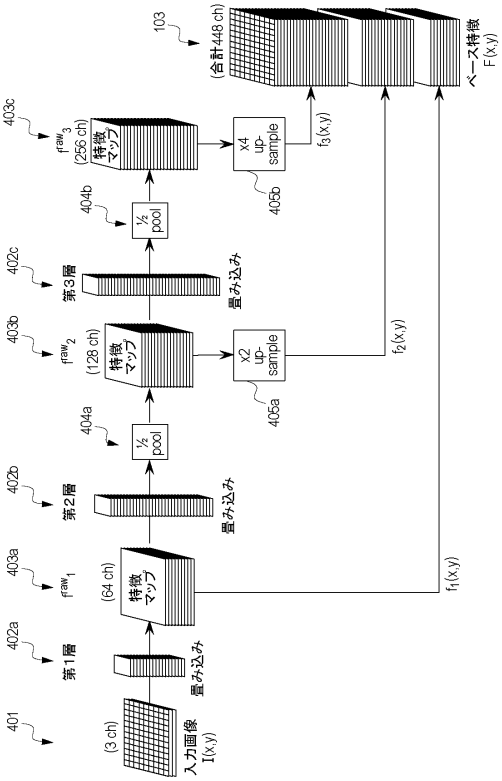
【図２】


開始
画像を入力
S1
画像特徴を抽出
S2
尤度マップ1～Nを生成
S3
所定数繰り返すループ
S4
尤度マップ1～Nの更新量を算出
S5
尤度マップ1～Nの値を更新
S6
ループの終了
S7
尤度マップ1～Nを統合
S8
結果を出力
S9
終了

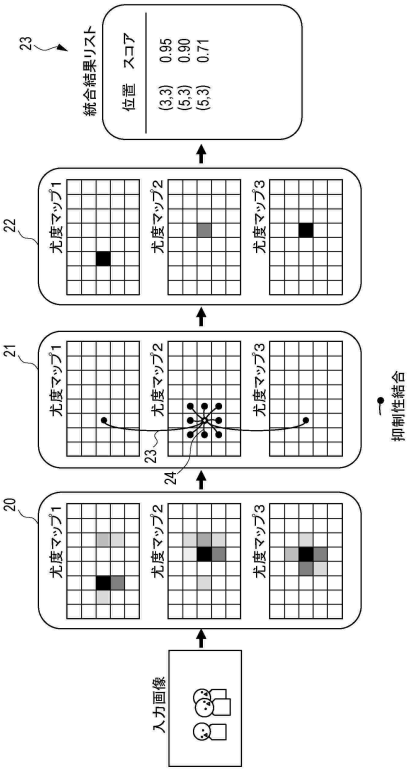
【図３】


S2開始
初期化
S201
ニューラルネットの各層
について繰り返すループ
S202
各層のニューラルネットの
演算処理
S203
特徴抽出の
対象の層か？
S204
No
Yes
特徴マップを所定サイズへ
アップ・ダウンサンプル
S205
特徴マップをベース特徴
F(x, y)の配列に追加
S206
ループの終了
S207
S2終了

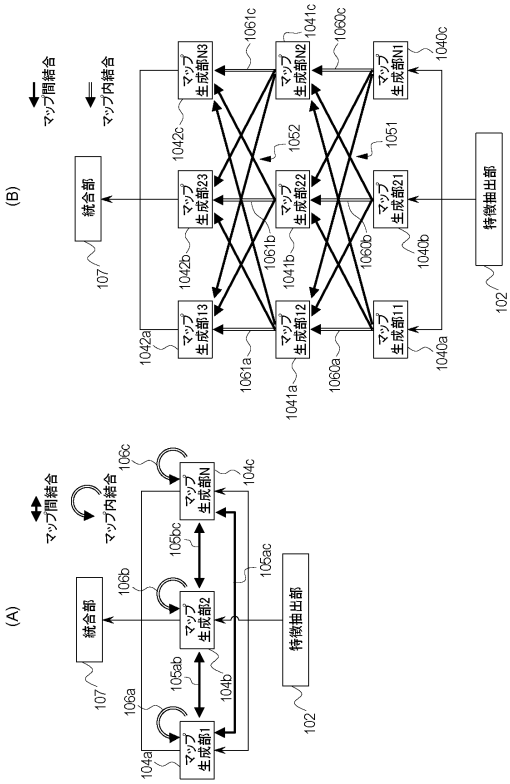
【図４】


103
(合計448 ch)
ベース特徴
F(x,y)
403c
(256 ch)
特徴マップ
x4 up-sample
f3(x,y)
405b
404b
½ pool
402c
第3層
畳み込み
403b
(128 ch)
x2 up-sample
f2(x,y)
405a
404a
402b
第2層
403a
(64 ch)
f1(x,y)
402a
第1層
401
(3 ch)
入力画像
I(x,y)

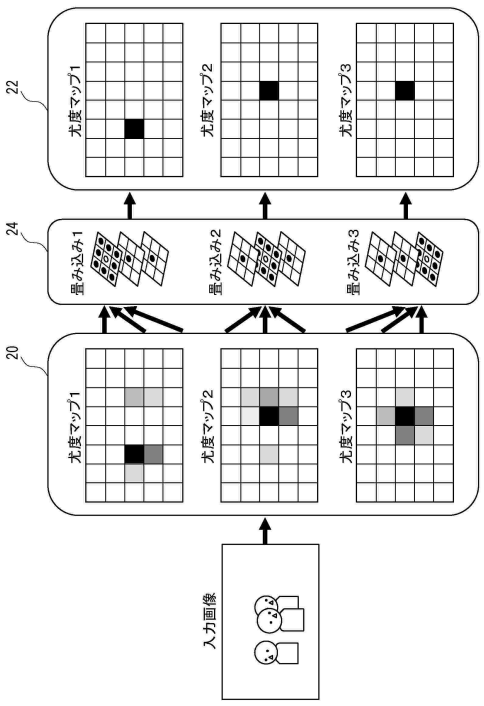
【図５】


23
統合結果リスト
位置 スコア
(3,3) 0.95
(5,3) 0.90
(5,3) 0.71
22
尤度マップ1
尤度マップ2
尤度マップ3
21
尤度マップ1
尤度マップ2
尤度マップ3
抑制性結合
23
24
20
尤度マップ1
尤度マップ2
尤度マップ3
入力画像

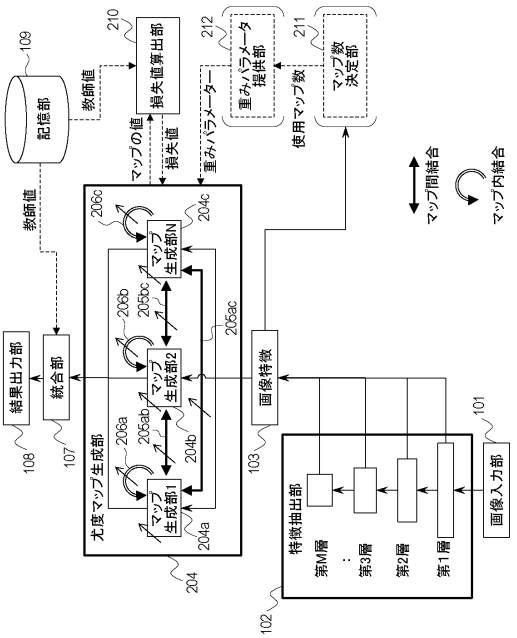
【図６】


マップ間結合
マップ内結合
(B)
統合部
107
マップ生成部N3
マップ生成部N2
マップ生成部N1
マップ生成部23
マップ生成部22
マップ生成部21
マップ生成部I3
マップ生成部I2
マップ生成部I1
特徴抽出部
102
1061c
1041c
1060c
1042c
1052
1051
1042b
1061b
1041b
1060b
1040b
1042a
1061a
1041a
1060a
1040a
1040c
(A)
106c
106b
106a
104c
105bc
105ac
105ab
104b
104a

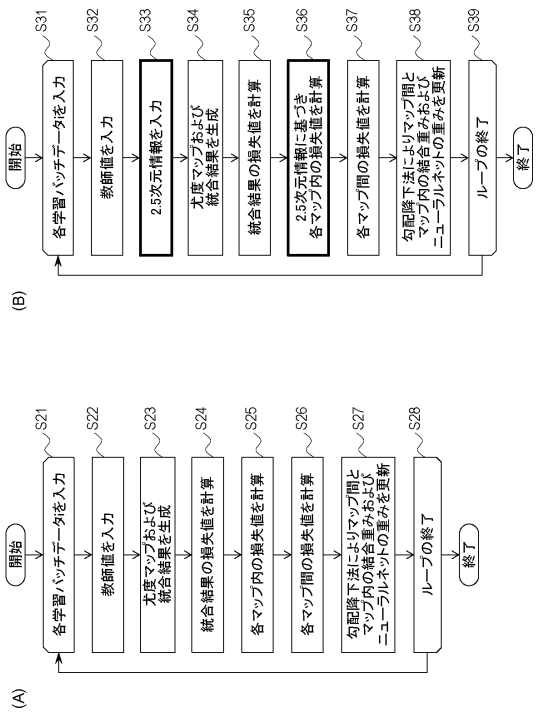
【図７】


22
尤度マップ1
尤度マップ2
尤度マップ3
24
畳み込み1
畳み込み2
畳み込み3
20
尤度マップ1
尤度マップ2
尤度マップ3
入力画像

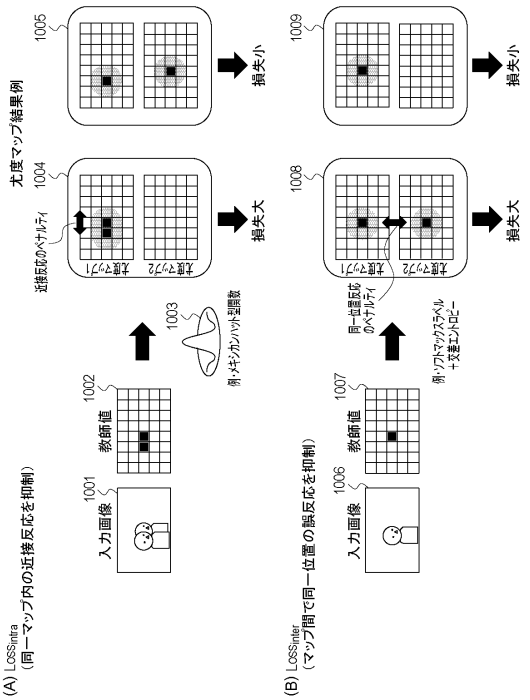
【図８】


109
配信部
教師値
210
損失値算出部
マップの値
損失値
212
重みパラメータ提供部
重みパラメータ
211
マップ数決定部
使用マップ数
マップ間結合
マップ内結合
結果出力部
108
統合部
107
尤度マップ生成部
204
マップ生成部1
マップ生成部2
マップ生成部N
206a
206b
206c
205ab
205bc
205ac
204a
204b
204c
画像待機
103
特徴抽出部
102
第M層
第3層
第2層
第1層
画像入力部
101

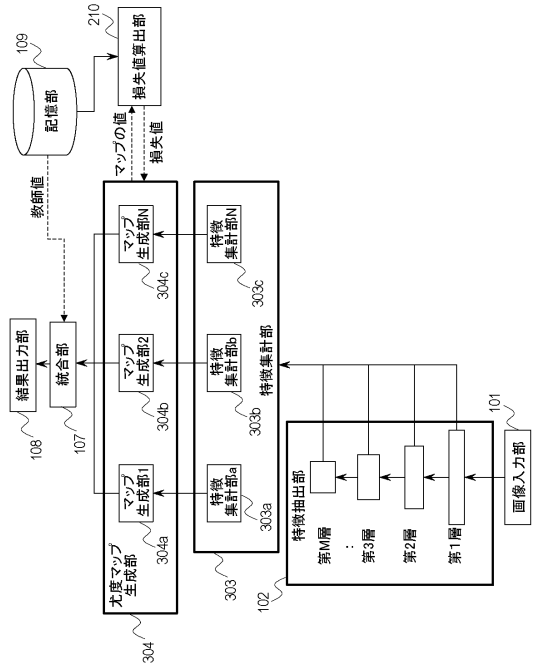
【図９】



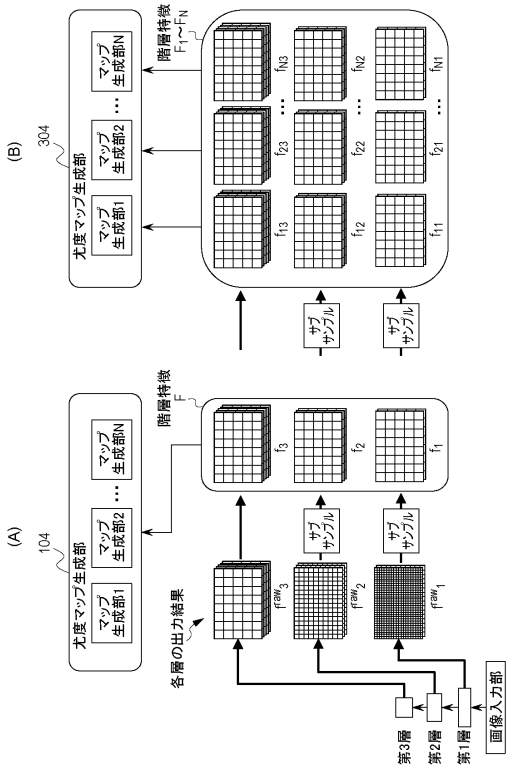
【図１０】



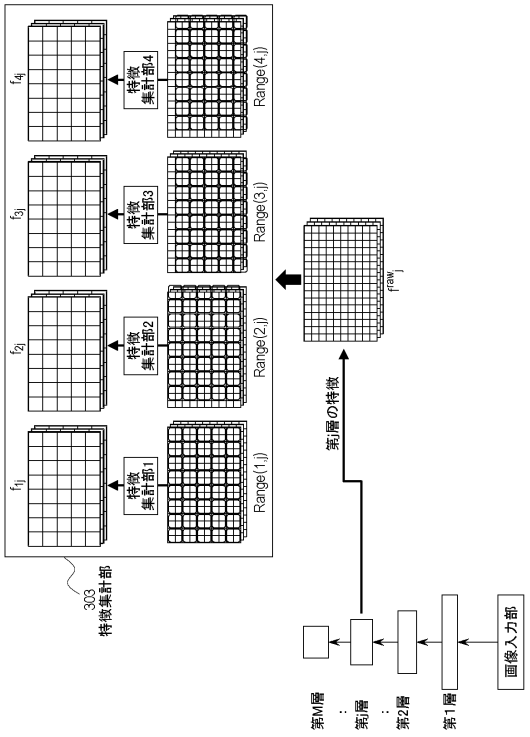
【図１１】



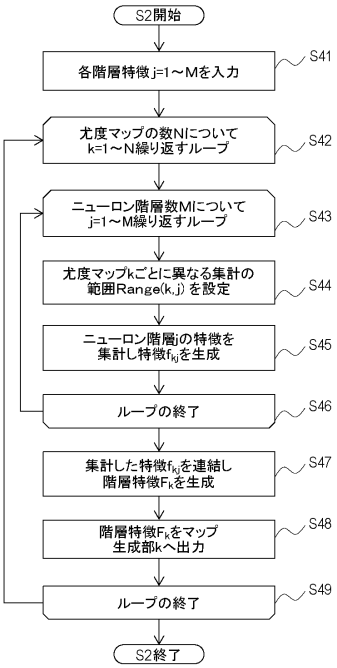
【図１２】



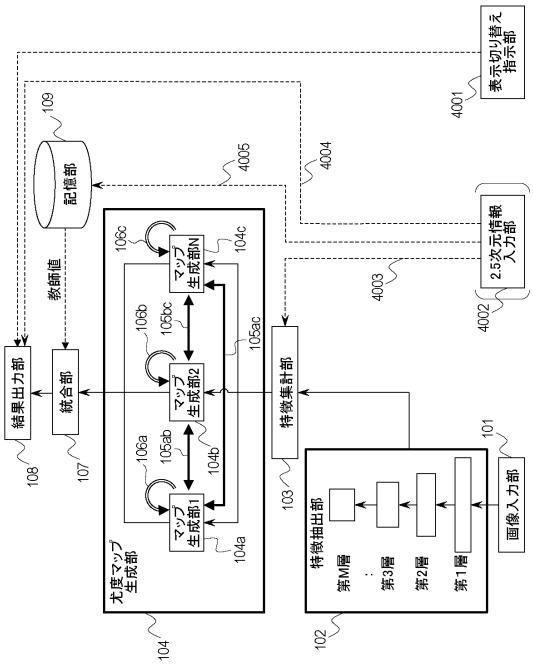
【図１３】



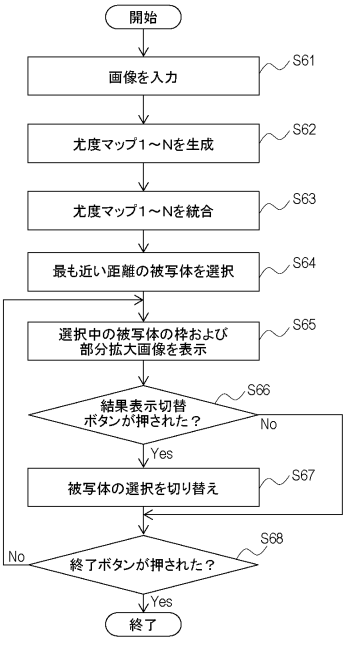
【図１４】



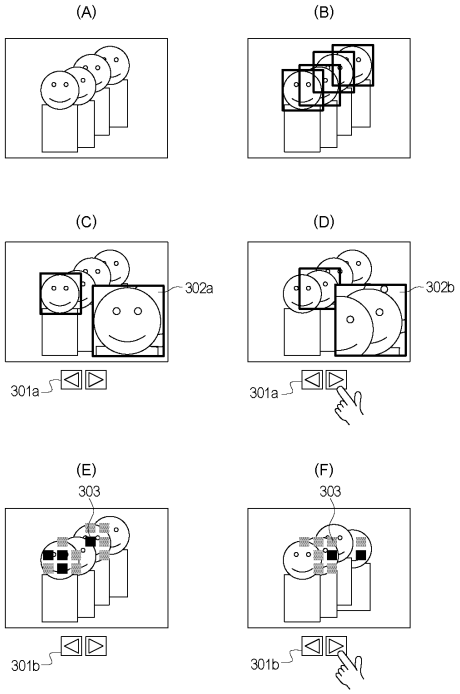
【図１５】



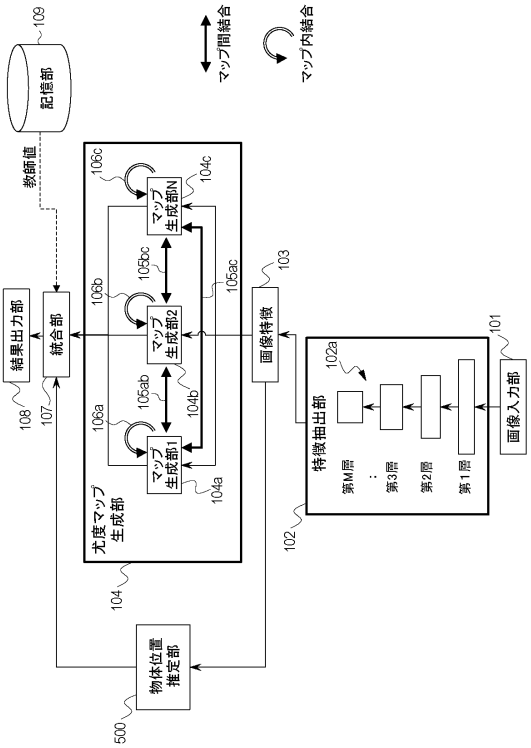
【図１６】



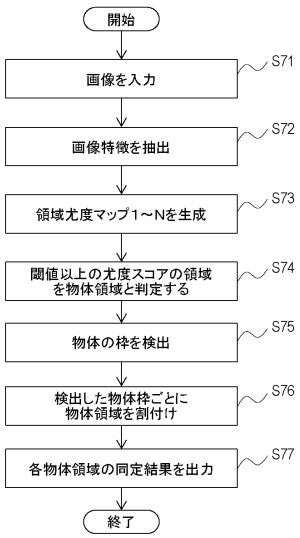
【図１７】



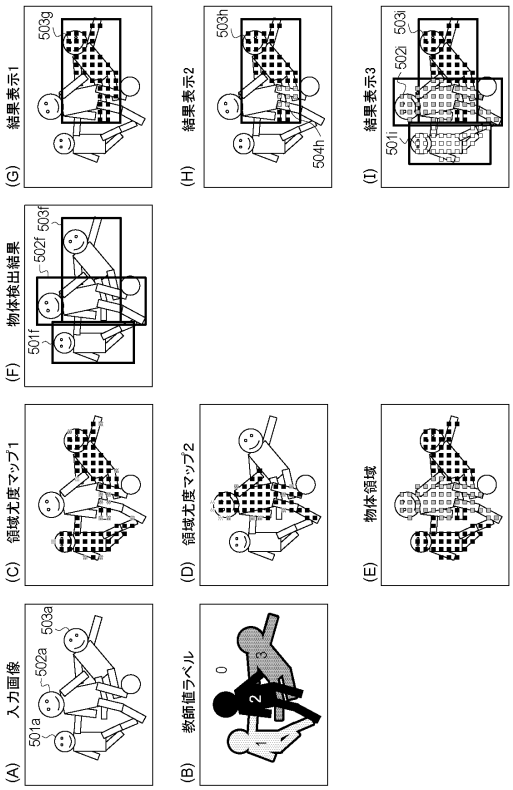
【図１８】



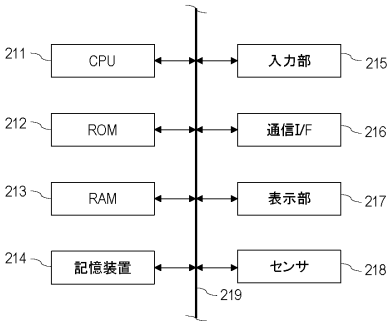
【図１９】



【図２０】



【図２１】


211
CPU
212
ROM
213
RAM
214
記憶装置
219
入力部
215
通信I/F
216
表示部
217
センサ
218

フロントページの続き

ヤノン株式会社内
審査官　吉川 康男

(56)参考文献　特開２０１９－０３２７７３（ＪＰ，Ａ）
ＡＩによる三次元モデル自動構成に基づくＶＲ地震シミュレーションシステム，第２４回画像センシングシンポジウム ＳＳＩＩ２０１８ IS1-28，2018年06月13日
Joseph Redmon;Santosh Divvala;Ross Girshick;Ali Farhadi，You Only Look Once: Unified, Real-Time Object Detection，2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，IEEE，2016年06月27日，779-788，https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7780460
畳み込みニューラルネットワークを用いた歩行者検出の高速化，電子情報通信学会技術研究報告 Ｖｏｌ．１１７ Ｎｏ．５０５ ＮＬＰ２０１７－１０３，2018年03月06日
Burak Uzkent;Aneesh Rangnekar;Matthew J. Hoffman，Aerial Vehicle Tracking by Adaptive Fusion of Hyperspectral Likelihood Maps，2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)，IEEE，2017年07月21日，233-242，https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8014769

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　７／００