

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号

特許第7077046号

(P7077046)

(45)発行日 令和4年5月30日(2022.5.30)

(24)登録日 令和4年5月20日(2022.5.20)

(51)国際特許分類

F I

G 0 6 T 7/00 (2017.01)

G 0 6 T

7/00

3 5 0 C

請求項の数 18 (全23頁)

(21)出願番号	特願2018-23792(P2018-23792)	(73)特許権者	000001007
(22)出願日	平成30年2月14日(2018.2.14)		キヤノン株式会社
(65)公開番号	特開2019-139618(P2019-139618 A)	(74)代理人	東京都大田区下丸子3丁目30番2号 100099324
(43)公開日	令和1年8月22日(2019.8.22)		弁理士 鈴木 正剛
審査請求日	令和3年2月10日(2021.2.10)	(72)発明者	舘 俊太
			東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
		(72)発明者	真継 優和
			東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
		(72)発明者	森 克彦
			東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
		審査官	千葉 久博

最終頁に続く

(54)【発明の名称】 情報処理装置、被写体の判別方法及びコンピュータプログラム

## (57)【特許請求の範囲】

## 【請求項1】

画像の特徴抽出のベースとなるベース特徴を生成するベース特徴生成手段と、

前記画像の局所領域に関する第1解像度の第1特徴を前記ベース特徴に基づいて生成する第1特徴生成手段と、

前記画像の、前記局所領域よりも広い広域領域に関する第2特徴であって、前記第1解像度よりも粗い第2解像度の第2特徴を前記ベース特徴に基づいて生成する第2特徴生成手段と、

前記第2特徴に基づいて、前記画像に存在する被写体を判別するための判別パラメータを前記画像の部分領域毎に設定する設定手段と、

前記第1特徴と前記判別パラメータとに基づいて前記被写体を判別する判別手段と、  
を備えることを特徴とする、

情報処理装置。

## 【請求項2】

前記判別手段は、前記部分領域毎に前記被写体が含まれることの尤度を前記第1特徴に基づいて生成し、当該尤度と前記判別パラメータとに基づいて前記被写体を判別す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】

画像の局所領域に関する第1特徴を生成する第1特徴生成手段と、

前記画像の、前記局所領域よりも広い広域領域に関する第2特徴を生成する第2特徴生成手段と、

前記第2特徴に基づいて、前記画像に存在する複数種類の被写体を判別するための、前記複数種類それぞれに対応する複数の判別パラメータを前記画像の部分領域毎に設定する設定手段と、

前記部分領域毎に前記複数種類の被写体それぞれに対応する複数の尤度を前記第1特徴に基づいて生成し、前記複数の尤度と前記複数の判別パラメータとに基づいて前記複数種類の被写体を判別する判別手段と、を備えることを特徴とする、  
情報処理装置。

【請求項4】

前記設定手段は、複数のサイズそれぞれに対応する複数の判別パラメータを、前記部分領域毎に設定し、

前記判別手段は、前記部分領域毎に前記複数のサイズそれぞれに対応する複数の尤度を前記第1特徴に基づいて生成し、前記複数の尤度と、前記複数の判別パラメータとに基づいて前記被写体を判別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5】

前記ベース特徴生成手段は、前記画像に対してニューラルネットワークで複数回の畳み込み演算を行うことにより、複数層の特徴マップを生成するとともに、これらの層の特徴マップを統合することにより前記ベース特徴を生成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6】

前記ベース特徴生成手段は、前記複数層の特徴マップの解像度を合致させ、前記複数層の特徴マップを統合することを特徴とする、

請求項5に記載の情報処理装置。

【請求項7】

前記第1特徴生成手段は、前記複数層の特徴マップのうち前記第1解像度の層の特徴マップを変換することにより前記第1特徴を生成し、

前記第2特徴生成手段は、前記第2解像度の層の特徴マップを変換することにより前記第2特徴を生成することを特徴とする、

請求項5又は6に記載の情報処理装置。

【請求項8】

前記画像の撮像時の条件を表すカメラ情報を取得する取得手段を更に備え、

前記第2特徴生成手段は、前記カメラ情報に基づいて、前記第2特徴を生成することを特徴とする、

請求項1から7のいずれか1項に記載の情報処理装置。

【請求項9】

前記第2特徴生成手段は、予め分類された画像特徴である画像分類特徴と前記カメラ情報とを連結することにより前記第2特徴を生成することを特徴とする、

請求項8に記載の情報処理装置。

【請求項10】

前記カメラ情報は、合焦している領域に関する合焦情報、水平線推定結果に関する水平線情報または測光値に関する測光情報であることを特徴とする、

請求項8または9に記載の情報処理装置。

【請求項11】

前記設定手段は、前記画像分類特徴を用いて複数のサイズごとに前記判別パラメータを設定することを特徴とする、

請求項9に記載の情報処理装置。

【請求項12】

前記判別パラメータは、予め学習により決定された閾値であることを特徴とする、

10

20

30

40

50

請求項 1 から 11 のいずれか一項に記載の情報処理装置。

【請求項 13】

前記部分領域毎に領域の属性を判定する判定手段を備えており、

前記第 2 特徴生成手段は、前記判定手段による判定結果に基づいて前記第 2 特徴を生成することを特徴とする、

請求項 1 から 3 のいずれか一項に記載の情報処理装置。

【請求項 14】

前記設定手段は、前記判定手段による判定結果に基づいて前記判別パラメータを設定することを特徴とする、

請求項 13 に記載の情報処理装置。

10

【請求項 15】

前記第 1 特徴に基づいて、前記画像を構成する複数の領域ブロック各々の属性の尤度を判定する判定手段と、

関心領域である前記部分領域を特定する特定手段と、

前記特定手段により特定された前記部分領域各々の位置及びサイズを検出するとともに、前記部分領域毎の特徴を抽出する特徴抽出手段と、を更に備え、

前記設定手段は、前記部分領域毎の位置及びサイズと前記判定手段による判定結果とに基づいて前記判別パラメータを前記部分領域毎に設定す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 16】

20

画像の特徴抽出のベースとなるベース特徴を生成するステップと、

前記画像の局所領域に関する第 1 解像度の第 1 特徴を前記ベース特徴に基づいて生成するステップと、

前記画像の、前記局所領域よりも広い広域領域に関する第 2 特徴であって、前記第 1 解像度よりも粗い第 2 解像度の第 2 特徴を前記ベース特徴に基づいて生成するステップと、

前記第 2 特徴に基づいて、前記画像に存在する被写体を判別するための判別パラメータを前記画像の部分領域毎に設定するステップと、

前記第 1 特徴と前記判別パラメータとに基づいて前記被写体を判別するステップと、を有することを特徴とする、

被写体の判別方法。

30

【請求項 17】

画像の局所領域に関する第 1 特徴を生成するステップと、

前記画像の、前記局所領域よりも広い広域領域に関する第 2 特徴を生成するステップと、

前記第 2 特徴に基づいて、前記画像に存在する複数種類の被写体を判別するための、前記複数種類それぞれに対応する複数の判別パラメータを前記画像の部分領域毎に設定するステップと、

前記部分領域毎に前記複数種類の被写体それぞれに対応する複数の尤度を前記第 1 特徴に基づいて生成し、前記複数の尤度と前記複数の判別パラメータとに基づいて前記複数種類の被写体を判別するステップと、を有することを特徴とする、

被写体の判別方法。

40

【請求項 18】

コンピュータを請求項 1 から 15 のいずれか一項に記載された情報処理装置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、画像に映った被写体を判別するための情報処理技術に関する。

【背景技術】

【0002】

画像に映った被写体（人物、物体等）を認識する装置では、認識精度を高めるためのさま

50

ざまな工夫が行われている。例えば、特許文献 1 に開示された技術では、物体を認識する際に画像が撮影されたシーンの分類を行う。すなわち物体の画像上の出現位置をシーンごとに集計し、分類する。そして、分類の結果に応じて物体の事前知識を表すスコアを調整する。これにより、例えば「街角」のシーンの画面上方には、「車」が出現する頻度は少ないといった事前知識に基づいた物体検出を行うことにより誤検出を低減させている。

【先行技術文献】

【特許文献】

【0003】

【文献】WO2012/046426

【発明の概要】

【発明が解決しようとする課題】

【0004】

特許文献 1 に示されるシーンの分類を前提とする手法では、分類精度が適切でない場合は、逆に認識精度が低下してしまう。また、カメラの方位や仰角が異なると物体の出現位置の分布が変化するため、認識精度を高めるには多数のシーンの分類結果を用意する必要がある。そのため、効果が期待できるのは一部の画像に限られるという問題がある。

【0005】

本発明は、特段の制約なしに取得した画像中の被写体を精度よく判別することができる技術を提供することを目的とする。

【課題を解決するための手段】

【0006】

本発明の一つの側面における情報処理装置は、画像の特徴抽出のベースとなるベース特徴を生成するベース特徴生成手段と、前記画像の局所領域に関する第 1 解像度の第 1 特徴を前記ベース特徴に基づいて生成する第 1 特徴生成手段と、前記画像の、前記局所領域よりも広い広域領域に関する第 2 特徴であって、前記第 1 解像度よりも粗い第 2 解像度の第 2 特徴を前記ベース特徴に基づいて生成する第 2 特徴生成手段と、前記第 2 特徴に基づいて、前記画像に存在する被写体を判別するための判別パラメータを前記画像の部分領域毎に設定する設定手段と、前記第 1 特徴と前記判別パラメータとに基づいて前記被写体を判別する判別手段と、を備えることを特徴とする。

【発明の効果】

【0007】

本発明によれば、特段の制約なしに撮影された画像中の被写体を精度よく判別することができる情報処理装置を提供することができる。

【図面の簡単な説明】

【0008】

【図 1】第 1 実施形態に係る情報処理装置のハードウェア構成図。

【図 2】第 1 実施形態に係る情報処理装置の機能ブロック構成図。

【図 3】第 1 実施形態における被写体の判別方法の手順説明図。

【図 4】ベース特徴の生成処理の詳細手順説明図。

【図 5】(A) は局所特徴、(B) は広域特徴の生成処理の手順説明図。

【図 6】判別パラメータを推定する処理の詳細手順説明図。

【図 7】被写体判別動作の詳細処理手順説明図。

【図 8】特徴マップの生成過程を示す説明図。

【図 9】カメラ情報と連結して広域特徴の生成過程を示す説明図。

【図 10】(A) ~ (C) は入力画像と各種カメラ情報との関係を示す説明図。

【図 11】学習処理時の機能構成を示すブロック図。

【図 12】判別パラメータ設定部における学習処理の手順説明図。

【図 13】判別パラメータ設定部における回帰器学習処理の手順説明図。

【図 14】(A), (B) は学習処理に用いる学習データの例を示す説明図。

【図 15】判別パラメータ設定学習の流れを示す説明図。

10

20

30

40

50

【図 16】第 2 実施形態の情報処理装置の機能構成図。

【図 17】(A), (B) は第 2 実施形態における被写体の判別方法の手順説明図。

【図 18】第 3 実施形態の情報処理装置の機能構成を示すブロック図。

【図 19】第 3 実施形態における被写体の判別方法の手順説明図。

【図 20】(A) ~ (D) は判別パラメータ学習の模式図。

【発明を実施するための形態】

【0009】

以下、本発明を適用した情報処理装置の実施の形態例を説明する。

情報処理装置は、ストレージを有するコンピュータとコンピュータプログラムとを用いて実施することができる。

10

[ 第 1 実施形態 ]

図 1 は第 1 実施形態に係る情報処理装置のハードウェア構成図である。この情報処理装置は、コンピュータ 10 とその周辺デバイスとを有する。コンピュータ 10 は、システムバス 19 に接続された CPU 11、GPU 12、ROM 13、RAM 14、外部記憶装置 15 を有する。システムバス 19 には、また、入力デバイスインタフェース 16、出力デバイスインタフェース 17、映像機器インタフェース 18 も接続される。

【0010】

CPU (Central Processing Unit) 11 は、コンピュータ 10 全体を制御する。GPU (Graphics Processing Unit) 12 は、画像処理などの高負荷の演算を行う演算装置として機能する。ROM (Read Only Memory) 13 は、変更を必要としない制御プログラムやパラメータなどを格納する。RAM (Random Access Memory) 14 は、CPU 11 のワークメモリであり、プログラムやデータなどを一時記憶する。外部記憶装置 15 は、ストレージの一例として、半導体メモリ、ハードディスク、光磁気ディスク、着脱可能なメモリカード等の記憶メディアを有するものである。外部記憶装置 15 には、本発明のコンピュータプログラム、画像 (映像を含む)、後述する各種特徴マップ、判別パラメータ、特徴変換に用いる所定の規則、カメラ情報などが記憶される。また、特徴生成の学習に用いる学習画像、被写体判別に用いられる教師値、学習済ニューラルネットワークモデルなども記憶される。

20

【0011】

入力デバイスインタフェース 16 は、ポインティングデバイスやキーボードなどの入力デバイス 21 とのインタフェースである。出力デバイスインタフェース 17 は、データ表示用のモニタ 22 とのインタフェースである。映像機器インタフェース 18 は、カメラ 23 等の撮像機器とのインタフェースである。

30

【0012】

コンピュータ 10 は、CPU 11 が本発明のコンピュータプログラムを読み込んで実行することにより、被写体の判別方法の実施に適した情報処理装置として動作する。なお、ニューラルネットワーク等による学習、パラメータ設定、判別、画像処理等については、GPU 12 が処理する。もちろん、CPU 11 が十分高性能である場合には、GPU 12 は省略してよい。

【0013】

コンピュータ 10 が情報処理装置として動作するときの機能ブロック構成例を図 2 に示す。また、本実施形態の情報処理装置が実行する被写体の判別方法の処理手順例を図 3 に示す。以後の説明では処理の各ステップを、以下「S」と略記する。

40

画像入力部 101 は、画像を取得する (S10)。本実施形態では、カメラで撮影した撮影画像を取り込むものとする。取り込んだ画像を「入力画像」と呼ぶ。カメラ情報入力部 110 は撮影時の条件などを表すカメラ情報の入力処理を行う。カメラ情報の内容については、後で詳しく説明する。

【0014】

ベース特徴生成部 108 は、入力画像における特徴抽出のベースとなるベース特徴を生成し、これを記憶する (S11)。ベース情報の内容及びその生成過程については、後で詳

50

しく説明する。

#### 【 0 0 1 5 】

第 1 特徴生成部 1 0 2 は、ベース特徴を所定の規則に従って変換することにより入力画像の局所領域の特徴を含む第 1 特徴を生成し、生成した第 1 特徴を記憶する ( S 1 2 )。局所領域の特徴を、以後、「局所特徴」と呼ぶ。第 2 特徴生成部 1 0 3 は、ベース特徴を所定の規則に従って変換するとともに、必要に応じてカメラ情報をも考慮することにより、入力画像の広域領域の特徴を含む第 2 特徴を生成し、生成した第 2 特徴を記憶する ( S 1 3 )。「広域領域」とは局所領域よりも広い範囲にわたる特徴をいう。この広域領域の特徴を、以後「広域特徴」と呼ぶ。

#### 【 0 0 1 6 】

被写体のうち例えば人物の顔を検出対象とし、その顔の検出をタスクとする場合、入力画像に人物と人工物や自然物とが乱雑に映っていると誤検出や未検出が多数生じてしまう場合がある。本実施形態では、誤検出や未検出を抑制するため、判別パラメータ設定部 1 0 4 で、入力画像に応じて、被写体を判別する際に用いる判別パラメータを推定する。そして、これを入力画像を構成する領域ブロックごとに設定する ( S 1 4 )。「領域ブロック」は座標で特定される画素群の領域をいう。判別パラメータには、例えば被写体の候補の尤度がそれ以上であればその候補が被写体であると判別するための閾値 ( ) を用いる。この閾値 ( ) は、第 2 特徴に基づいて決定される推定値であり、判別の誤りが所定値以下となるように予め学習により決定された閾値である。

#### 【 0 0 1 7 】

被写体判別部 1 0 6 は、第 1 特徴と判別パラメータとに基づいて入力画像中の被写体を判別する ( S 1 5 )。本実施形態では、検出対象である人物 (あるいは人物の顔) を判別するために、被写体判別部 1 0 6 が、3 種類の尤度生成部 1 0 6 a、1 0 6 b、1 0 6 c と閾値処理部 1 0 6 d の機能を備えるものとする。尤度生成部 1 0 6 a、1 0 6 b、1 0 6 c は、第 1 特徴生成部 1 0 2 で生成された第 1 特徴に基づいて尤度スコアを生成する。この尤度スコアは、入力画像の領域ブロックごとに、顔が小さい人物、顔が中程度の人物、顔が大きい人物がそれぞれ存在することの確からしさ (尤度) を表す尤度スコアを並べたものである。被写体判別部 1 0 6 は、これらの尤度マップに基づいて被写体の候補を決定する。そして、閾値処理部 1 0 6 d で、入力画像に応じて推定された閾値 ( ) と被写体の候補とを閾値処理することにより、当該候補が、検出対象である人物 (あるいは人物の顔) かどうかを判別する。結果出力部 1 0 7 は、判別結果をモニタ 2 2 等に出力する ( S 1 6 )。

#### 【 0 0 1 8 】

< ベース特徴の生成 >

図 3 の S 1 1 の処理の内容を図 4 及び図 8 を参照して詳しく説明する。図 4 はベース特徴生成部 1 0 8 が実行する処理の詳細手順説明図であり、図 8 は処理の概念説明図である。ベース特徴は、例えば CNN ( Convolutional Neural Network : 畳み込みニューラルネットワーク、以下 CNN と略す ) を用いて生成することができる。ただし、S 1 1 の処理では、CNN の全結合層と呼ばれる層は用いず、畳み込み層と呼ばれる種類の層のみを用いる。図 8 中、「c h」が付された数字は、特徴マップの枚数である。

#### 【 0 0 1 9 】

ベース特徴生成部 1 0 8 は、まず、ベース特徴  $F(x, y)$  を格納する空の配列を用意する ( S 1 2 0 1 )。つまり、上記の配列を初期化する。(  $x, y$  ) は画素の座標 (  $X, Y$  座標系 ) を表す添え字である。初期化後、ベース特徴生成部 1 0 8 は、CNN により、図 8 に示された複数回の畳み込み演算を繰り返すことで複数層の特徴マップを生成する ( 図 4 の S 1 2 0 2 ~ S 1 2 0 7 )。図 8 の例で言えば、ベース特徴生成部 1 0 8 は、RGB ( 赤、緑、青 ) の 3 c h の入力画像  $I(x, y)$  4 0 1 に対して、畳み込み演算 4 0 2 a を行い、6 4 c h の特徴マップ 4 0 3 a を生成する。また、この特徴マップ 4 0 3 a について、再度畳み込み演算 4 0 2 b を行い、その演算結果について 1 / 2 プーリング 4 0 4 a を行うことで 1 2 8 c h の特徴マップ 4 0 3 b を生成する。同様にして、特徴マップ 4

10

20

30

40

50

0 3 bについて再度畳み込み演算 4 0 2 cを行い、その演算結果について1 / 2 プーリング 4 0 4 bを行うことで2 5 6 c hの特徴マップ 4 0 3 cを生成する。1 / 2 プーリング 4 0 4 a、4 0 4 bは、特徴マップ 4 0 3 a、4 0 3 bをそれぞれ所定の局所領域ごとに代表値で代表させることにより、マップサイズを縮小する処理である。この処理には、C N Nの認識性能をロバストにする効果がある（ただし、演算が進むに従って解像度が低くなる）。本実施形態の1 / 2 プーリング 4 0 4 a、4 0 4 bでは、それぞれ特徴マップを（2 × 2 画素）ごとに統合して、1 / 2 の解像度の特徴マップに縮小する。L 番目の層の畳み込み演算及び1 / 2 プーリングの処理を数式で表すと下記ようになる。

【 0 0 2 0 】

[ 数 1 ]

$$f^L(x, y, z) = \left( CHIN(x, y) - K \sim + K w^L(x, y, CHIN, CHOUT) \right) \times f^{L-1}(x, y, CHIN) + B^L CHOUT$$

【 0 0 2 1 】

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出力結果であり、z 枚の特徴マップを表す。x, y は画素の位置（座標）を表す。また、 $(\cdot)$ は、半波整流よりなる活性化関数（ReLU (Rectified Linear Unit) function）であり、入力した値が0 以下のとき0 となり、1 より大きいとき入力をそのまま出力される。また、 $w^L(x, y, CHIN, CHOUT)$ （ただし  $x, y \in \{-K, \dots, 0, \dots, K\}$ ）は、L 番目の層の畳み込みの重みパラメータである。B<sup>L</sup>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識別番号、CHOUTは、L 番目の層が出力する特徴マップの識別番号を表す。

なお、上式では入力画像  $I(x, y)$  は、特徴マップ  $f^0(x, y, z)$  として扱うものとする。ここでは、畳み込み演算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、畳み込み演算の前に特徴マップ  $f^{L-1}$  の周囲の画素に0 値を充填してから畳み込む。このようにして、C N Nの各層において、それぞれ複数層となる特徴マップ 4 0 3 a、4 0 3 b、4 0 3 c が生成される。

【 0 0 2 2 】

C N Nでは、大規模データベースを用いた画像の分類タスクにより予め重みパラメータを学習しておく。これにより、C N Nの入力画像 4 0 1 (=  $I(x, y)$ ) に近い所定数の層（低層）1 0 9 l の特徴マップ 4 0 3 a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特徴マップとなり得る。また、C N Nの入力画像 4 0 1 から最も遠い所定数の層（高層）1 0 9 h の特徴マップ 4 0 3 c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特徴マップとなり得る。中間の層（中層）1 0 9 m の特徴マップ 4 0 3 bは、特徴マップ 4 0 3 aと特徴マップ 4 0 3 c との中間のパターン形状に反応する特徴マップとなり得る。

【 0 0 2 3 】

ベース特徴生成部 1 0 8 は、C N Nにおいて処理中の層が、特徴抽出の対象となる所定の層か否かを判定する（S 1 2 0 4）。所定の層は、本実施形態の例では予め定められた枚数（ch数）を表す低層、中層、高層のいずれかである。所定の層でない場合は（S 1 2 0 4 : No）、S 1 2 0 7 へ移行する。所定の層であれば（S 1 2 0 4 : Yes）、特徴マップを所定サイズへアップサンプリングし（S 1 2 0 5）、特徴マップをベース特徴  $F(x, y)$  の配列に追加する（S 1 2 0 6）。このような処理をループの条件（特徴マップがn（自然数）枚に達したかどうか）を満たすまで繰り返し（S 1 2 0 7）、最終的にn枚の特徴マップが連結されたベース特徴  $F(x, y)$  が生成される。上記のベース特徴生成の処理は、数式では下記のように表される。

[ 数 2 ]

$$F(x, y) = [f^1(x, y)^T, f^2(x, y)^T, \dots, f^n(x, y)^T]^T$$

【 0 0 2 4 】

10

20

30

40

50

ここで、 $f_1, f_2, \dots, f_n$ は抽出された特徴マップであり、ベース特徴 $F(x, y)$ は、上記のように抽出された特徴マップを $z$ 次元方向に統合したものとなる。なお、CNNは1/2プーリング404a, 404bを行うため、特徴マップの解像度は、層によって異なるものとなる。そのため、ベース特徴生成部108は、上記の統合前に、各特徴マップを所定の解像度、例えば入力画像 $I(x, y)$ の解像度に合致させる処理を行う(51205)。具体的には、特徴マップ403bについて2倍のアップサンプリング(X2 up-sample)405aを行う。また、特徴マップ403cについて4倍のアップサンプリング(X4 up-sample)405bを行う。解像度を合致させるための変更は、画素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。

【0025】

10

以上のようにして、CNNの低層109l、中層109m及び高層109hを統合した448chのベース特徴 $F(x, y)$ が生成される。これにより、第1の性質として、様々な被写体の、様々なスケール及びバリエーションの情報が、ベース特徴 $F(x, y)$ の中に含まれることとなる。第2の性質として、2次元の画素群の領域 $(x, y)$ ごとに3次元以上の高次元の特徴ベクトルが対応したものとなる。

このような二つの性質により、ベース特徴 $F(x, y)$ は、様々な被写体の検出ないし認識のタスクに有効に用いることができるようになる。

【0026】

なお、本実施形態では、入力画像 $I(x, y)$ とベース特徴 $F(x, y)$ の解像度を合致させるようにしたが、この限りでない。アップサンプリング処理の倍率を検出対象に応じて変更することで、入力画像 $I(x, y)$ よりも粗い解像度、または、細かい解像度のベース特徴 $F'(x', y')$ を生成することもできる。そのため、本明細書では、上述した「領域ブロック」を、ベース特徴 $F(x, y)$ の画素群の領域を表す一般化した名称として用いる。

20

【0027】

<第1特徴の生成>

次に、第1特徴生成部102による第1特徴を生成する処理(512)の詳細手順について、図5(A)を参照して説明する。この処理は、51208から51211のループとなる。第1特徴生成部102は、所定の規則の例として、ベース特徴 $F(x, y)$ から第1解像度となる低層の特徴109lを抽出し、低層の特徴109lから領域ブロック $x, y$ (座標 $(x, y)$ の位置の画素群、以下同じ)ごとの局所特徴 $F_{xy}$ を生成し、これを記憶する。局所特徴 $F_{xy}$ は、入力画像401に判別の検出対象、例えば人物が映っている場合に、後段で人物の顔などの局所的な部分の特徴が各領域ブロック $x, y$ に存在するかどうかの判別に用いる。その特徴量は、後段において判別に必要なデータをコンパクトに表現したベクトルとなる。

30

【0028】

第1特徴生成部102は、多次元の特徴マップであるベース特徴 $F(x, y)$ 及び8近傍の特徴を取得する(51209)。具体的には、ベース特徴 $F(x, y)$ のうち領域ブロック $x, y$ に対応する特徴ベクトル、及び、各領域ブロックの周囲の8つの領域ブロックの特徴ベクトルを抽出する。そして、これらを1次元に連結して局所特徴 $F_{xy}$ として記憶する(51210)。数式で表すと以下のとおりである。次式において、 $T$ はベクトルの転置である。

40

[数3]

$$F_{xy} = [F(x-1, y-1)^T, F(x, y-1)^T, F(x+1, y-1)^T, \\ F(x-1, y)^T, F(x, y)^T, F(x+1, y)^T, \\ F(x-1, y+1)^T, F(x, y+1)^T, F(x+1, y+1)^T, ]^T$$

【0029】

<第2特徴の生成>

次に、第2特徴生成部103による第2特徴を生成する処理(513)の詳細について図5(B)及び図9を参照して説明する。この処理は、51301から51307の順に行

50



われるが、S 1 0 3 から 1 3 0 5 はループとなる。第 2 特徴は、入力画像の広域特徴を含む特徴であり、判別パラメータである閾値 ( ) を推定する際の手掛かりに用いられる。第 2 特徴も第 1 特徴と同じくベース特徴  $F(x, y)$  に基づいて生成する。

#### 【 0 0 3 0 】

第 2 特徴生成部 1 0 3 は、まず、ベース特徴  $F(x, y)$  から第 1 解像度よりも粗い第 2 解像度となる高層の特徴 1 0 9 h を抽出する ( S 1 3 0 1 )。そして、抽出した高層の特徴 1 0 9 h を 1 次元に整列させる ( S 1 3 0 2 )。つまり、並べ直す。これは、図 8 の例でいえば特徴マップ 4 0 3 c を整列させたものに相当する。整列に際しては、特徴マップ 4 0 3 c を R A M 1 4 に記憶しておき、随時読みだして使用できるようにする。

#### 【 0 0 3 1 】

次に、第 2 特徴生成部 1 0 3 は、S 1 3 0 3 から S 1 3 0 5 のループで、高層の特徴 1 0 9 h について、ニューラルネットワークの全結合層の演算処理によって特徴変換を行う。全結合層の演算処理では、1 次元に整列された高層の特徴 1 0 9 h について重み付けを行う。全結合層の重み付けは、本実施形態では、図 9 に画像分類カテゴリ 2 0 8 として示す出力層を用いて、予め画像分類タスクが判定できるように学習しておく。このような学習形態を「マルチタスク学習」と呼ぶ。ここでは、例えば 1 0 0 0 クラス程度の分類タスクの学習を行う。このマルチタスク学習を行った後に、出力層の前の中間層を用いて特徴変換を行うことで、人物やその人物の目の大きさ、髪色など、検出対象の大まかな画像の特徴を表す画像分類特徴 1 1 5 を得ることができる。

#### 【 0 0 3 2 】

第 2 特徴生成部 1 0 3 は、次に、画像分類特徴 1 1 5 とカメラ情報入力部 1 1 0 を通じて入力されたカメラ情報とを連結して広域特徴 G を生成する ( S 1 3 0 6 )。カメラ情報は入力画像の撮像時の条件を表す情報である。カメラ情報の一例として、本実施形態では図 1 0 ( A ) に示されるように、カメラの焦点がどの領域ブロックに合っているかを表す合焦情報 1 5 0 1 を数値化した合焦情報値 1 5 0 2 を用いる。合焦情報値 1 5 0 2 は、焦点が当たっている領域ブロックでは 0、そうでない場合は焦点の深さに応じた + の数値となる。この合焦情報値 1 5 0 2 を領域ブロックごとの閾値 ( ) に加算することにより、例えば「焦点の合っていないエリアは被写体の誤検出が多いために閾値 ( ) を上げる」といった、状況に応じた閾値 ( ) の設定ができる。その後、第 2 特徴生成部 1 0 3 は、連結した特徴を広域特徴 G として記憶し、第 2 特徴の生成処理を終える ( S 1 3 0 7 )。

#### 【 0 0 3 3 】

なお、カメラ情報は、合焦情報値 1 5 0 2 のほか、図 1 0 ( B ) に示すような水平線推定結果 1 5 0 3 と各領域ブロックとの位置関係に関する水平線情報値 1 5 0 4 を使用することもできる。水平線は、カメラが重力センサを有する場合、その重力センサの検出情報から推定することができる。そして、領域ブロックごとの閾値 ( ) に加算する数値として、水平線より上の領域ブロックは - 1、それ以外の領域ブロックは、近くなるほどプラスの数値となるようにする。カメラ情報の他の種類としては、図 1 0 ( C ) に示す物理的な測光値に関する測光情報値 1 5 0 5 などを使用することもできる。なお、図 1 0 ( A ) ~ ( C ) はカメラ情報の例示であって、被写体の適切な判別パラメータ ( 閾値 ( ) ) の推定及び設定の際に手がかりとなる情報であれば、あらゆる種類のカメラ情報を用いることができる。カメラ情報は、外部記憶装置 1 5 に保持され、随時、カメラ情報入力部 1 1 0 から第 2 特徴生成部 1 0 3 に提供される。

#### 【 0 0 3 4 】

##### < 判別パラメータ設定 >

次に、図 3 の S 1 4 に示した判別パラメータを設定する処理について詳細に説明する。ここでは、判別パラメータ設定部 1 0 4 が、判別パラメータとして人物の顔を検出する際に用いる閾値 ( ) を算出し、領域ブロックごとに設定する場合の例を説明する。

判別パラメータ設定部 1 0 4 は、顔サイズ  $s$  ごとに、かつ、入力画像の領域ブロック (  $i, j$  ) ごとに適切な閾値 ( ) が設定されるように、図 9 に示される 3 つのパラメータ回帰器 1 0 4 a、1 0 4 b、1 0 4 c を備える。パラメータ回帰器 1 0 4 a は、小さい顔の

10

20

30

40

50

閾値（顔（小）閾値）用の回帰器である。パラメータ回帰器 104c は大きい顔の閾値（顔（大）閾値）用の回帰器である。パラメータ回帰器 104b は、中間のサイズの顔の閾値（顔（中）閾値）用の回帰器である。領域ブロックの添え字を（ $i, j$ ）とするのは、局所特徴の領域ブロック（ $x, y$ ）よりも解像度を低くするためである。

#### 【0035】

局所特徴では、顔の尤度スコアのほかに位置精度が重要となるため、より高い解像度の領域ブロック単位（ $x, y$ ）で判別を行う必要があるが、判別パラメータ設定部 104 による閾値（ $i, j$ ）の推定では、位置精度はそれほど重要でない。むしろ、より粗い解像度の領域ブロック単位（ $i, j$ ）で推定を行うのが好適である。その理由として、本実施形態では、閾値（ $i, j$ ）を入力画像の領域ブロックごとに異なるパラメータ回帰器でそれぞれ学習するため、領域ブロックの解像度が高すぎると処理量やメモリ消費の点で不利になる。また学習時の事例が少なくなるため、過学習が起こりやすい点も理由の一つに挙げられる。このように被写体の尤度判別と閾値（ $i, j$ ）の設定とが相補的な関係にある点が本実施形態における特長の一つである。

10

#### 【0036】

判別パラメータ設定部 104 による具体的な処理の手順は以下の通りである。判別パラメータ設定部 104 は、図 7 に示すように、顔サイズ  $s$ （ただし  $s \in \{\text{小}, \text{中}, \text{大}\}$ ）、及び入力画像上の各領域ブロック（ $i, j$ ）について処理を繰り返す（S1401～S1405）。判別パラメータ設定部 104 は、抽出された広域特徴  $G$  を用いて、各顔サイズ  $s$  及び各領域ブロック（ $i, j$ ）に閾値（ $ijs$ ）を算出する（S1403）。算出は、下記式に示される一般的なロジスティック回帰器の演算により行う。

20

[数 4]

$$ijs = 1 / (1 + \exp\{-W_{ijs}^T \cdot G + b_{ijs}\})$$

#### 【0037】

ただし、 $W$  は広域特徴  $G$  と同じ長さのベクトルからなる重みパラメータ、 $b$  はバイアス値である。また、重み  $W$  の値及びバイアス値  $b$  は予め学習して得ているものとする。以上の演算により、判別パラメータである閾値（ $ijs$ ）が、顔サイズ  $s$  及び入力画像の領域ブロック（ $i, j$ ）ごとに算出され、設定される。

#### 【0038】

判別パラメータ設定部 104 の処理により設定される閾値（ $ijs$ ）の例が図 9 に示されている。図 9 では、閾値（ $ijs$ ）が 0 に近い領域ブロックほど白く、閾値（ $ijs$ ）が 1 に近い領域ブロックほど黒く示されている。すなわち、閾値（ $ijs$ ）が 1 に近い黒い領域ブロックほど検出を抑制し、尤度スコアが高くない限り検出をしないという設定がなされる。

30

#### 【0039】

<被写体判別>

次に、図 3 の S15 に示された判別処理の詳細な手順例を図 7 及び図 9 を参照して説明する。図 7 において、被写体判別部 106 は、人物の顔のサイズ  $s$  ごとに S1501 から S1509 のループ処理を行う。すなわち、被写体判別部 106 は、顔（小）尤度スコアを生成する尤度生成部 106a、顔（中）尤度スコアを生成する尤度生成部 106b、顔（大）尤度スコアを生成する尤度生成部 106c として機能する。

40

各尤度スコアは、各領域ブロック（ $x, y$ ）について局所特徴  $F_{xy}$  を入力とし、サイズ  $s$  の顔が被写体の候補としてその領域ブロックに存在する確度を表す評価情報である（S1501～S1504）。具体的には、局所特徴  $F_{xy}$  を入力変数とし、例えばサポートベクトルマシン（SVM: support vector machine）を用いて、各領域ブロック（ $x, y$ ）についてのサイズ  $s$  の顔の尤度スコア  $L_s(x, y)$  を以下の式により算出する。

[数 5]

$$L_s(x, y) = R(k - s_k - v_{sk} \cdot F_{xy} + b_s)$$

#### 【0040】

$v_{sk}$  は、SVM がサイズ  $s$  の顔を判別するための  $k$  番目のサポートベクトル、 $s_k$  は同じくサポートベクトルの重み係数、 $b_s$  はバイアス項である。これらのパラメータは、顔サイ

50

ズ  $s$  ごとに予め後述する方法で学習を行い、記憶してあるものとする。 $R(\cdot)$  は、 $SV$   $M$  の出力を尤度に変換するための規格化関数である。ここでは簡易に下式のような関数で  $SV$   $M$  のスコアを  $0 \sim 1$  に規格化する。なお、 $\alpha$  は定数である。

【数 6】

規格化関数  $z' = R(z)$  の定義：

$$z' = 1 \quad \text{if } z \geq 0$$

$$z' = z / \alpha \quad \text{if } 0 < z < \alpha$$

$$z' = 0 \quad \text{if } z < 0$$

【0041】

$SV$   $M$  は、線形入力素子を利用した 2 クラスのパターン識別器の一つである。尤度生成部 106a、106b、106c は、全ての領域ブロック  $(x, y)$  と全ての顔サイズ  $s$  について条件判定の処理を行う。そのために、 $SV$   $M$  のように軽い処理の判別器であることが好適である。換言すれば処理が軽ければ決定木や多層パーセプトロン等、どのような判別器を用いることができる。

10

【0042】

尤度生成部 106a、106b、106c は、また、抽出した  $L_s(x, y)$  の値を閾値 ( $ijs$ ) と比較する。ただし、被写体の尤度スコアを算出した領域ブロック  $(x, y)$  と、閾値 ( $ijs$ ) を設定した領域ブロック  $(i, j)$  とは解像度が異なるため、下記の式に示すように座標を変換して得られる領域ブロックの閾値 ( $i'j's$ ) と比較する。下記の式において、 $\alpha$  と  $\beta$  は 2 つの領域ブロック間の座標変換のパラメータである。

20

【数 7】

$$i' = \alpha x + \beta$$

$$j' = \beta y + \alpha$$

【0043】

そして、次の数式に示す条件を判定する。

【数 8】

$$L_s(x, y) > i'j's$$

【0044】

このようにして尤度生成部 106a、106b、106c は、図 9 に示す通り、顔 (小) 尤度スコア、顔 (中) 尤度スコア、顔 (大) 尤度スコアを生成する。

30

各尤度スコアが生成されると、被写体判別部 106 は、閾値処理部 106d で各領域ブロック  $(x, y)$  について条件判定を行う (S1506)。すなわち、閾値処理部 106d は、尤度スコア  $L_s(x, y)$  が 8 近傍の尤度スコアの中で最大の値か否かを判定する。また、閾値 ( $i'j's$ ) 以上かどうかを判定する。肯定的の場合 (S1506: Yes)、条件を満たす領域ブロック ( $i$  番目の領域ブロック) の座標  $(x, y)$  を中心とする位置にサイズ  $s$  の顔が存在すると判定する (S1507)。つまり、そのサイズ  $s$  の顔を図 9 に示される被写体の候補と判定する。その後、S1505 の処理に戻る。

一方、S1506 の結果が否定的である場合 (S1506: No)、ただちに S1505 の処理に戻る。図 3 の S16 では、上記の被写体判別部 106 による判定の結果が、最終検出結果 121 として結果出力部 107 に出力される。

40

【0045】

< 学習処理 >

次に、情報処理装置が行う学習処理について、図 11 から図 14 を参照して説明する。図 11 は、学習処理を行うときの情報処理装置の機能ブロック構成図である。図 2 の機能ブロック構成図に、カメラ情報保持部 130、学習画像保持部 131、教師値保持部 132、尤度集計部 133、判別パラメータ教師値保持部 134 が追加されている。なお、ベース特徴生成部 108 については、図示を省略してある。

【0046】

図 12 に示す学習処理の手順説明図を参照すると、画像入力部 101 は、学習画像保持部 131 から学習画像を取得する。また、尤度生成部 106a、106b、106c は、各

50

画像に対応する顔の教師値を教師値保持部 132 から取得する。なお、ここでは、各画像に対応する人物の顔の教師値を用いるが、必ずしも顔に限られない。図 14 (A) は学習画像セット X、同 (B) は学習画像セット Y を示す。各学習画像セット X, Y は、それぞれの図の (1) 行に示される学習画像と、図の (2) 行に示される顔の教師値のセットからなる。顔の教師値は、領域ブロックが小サイズの顔、中サイズの顔、大サイズの顔、のいずれかを含むか、あるいはいずれも含まないかを表す値である。具体的には、顔の基準点を含む領域ブロックを正事例 (顔があった) “1”、それ以外の領域ブロックを負事例 (顔がなかった) “0” の教師値とする。各教師値は、簡便には、顔サイズごとに入力デバイス 21 を介して操作者が付与してもよいが、自動認識処理によってそれを行うようにしてもよい。顔の基準点は、顔領域の重心とするが、この限りでない。

10

#### 【0047】

第 1 特徴生成部 102 は、各学習画像から局所特徴を生成する (S112)。局所特徴を生成する処理の内容は、S12 の処理と同様となる。

まず、尤度生成部 106a, 106b, 106c は、局所特徴に基づき被写体 (顔) を正しく判別できるよう尤度判定用の SVM を次のように学習する (S113)。すなわち、尤度生成部 106a, 106b, 106c は、与えられた顔についての教師値の “0”、“1” の値を目標変数、対応する領域ブロックの連結層特徴量  $F_{xy}$  を説明変数とし、顔の正事例と負事例が判別できるように SVM を学習する。学習は、サイズ  $s$  ごとに行い、SVM のパラメータ ( $[v_{sk}, w_{sk}, b_s]$ ) を得る。以上が、尤度生成部 106a の学習手順である。

20

#### 【0048】

情報処理装置では、その後、判別パラメータ設定部 104 の回帰器 104a, 104b, 104c を学習する (S114)。回帰器 104a, 104b, 104c は、画像の領域ブロックごとに備えられるパラメータ回帰器  $ij$  である。パラメータ回帰器  $ij$  は、それぞれが、前述した数 2 式に従うロジスティック回帰器である。学習の目的は、これらのロジスティック回帰器が入力画像の広域特徴に応じて適切な閾値 ( $\theta_{ij}$ ) が推定できるように、回帰器  $ij$  の重みパラメータ ( $[W_{ij}, b_{ij}]$ ) を得ることである。ここでは、説明の簡単のために以降では顔のサイズ  $s$  については考慮せず、全て同じ顔として扱うものとする。サイズ  $s$  を考慮したときにも本質的に学習動作は同一であり、単純に以降に述べる学習手順をサイズ別に行えばよい。

30

#### 【0049】

図 13 はこの学習の手順説明図であり、S1141 から S1149 の処理のループを実行する。図 15 は、学習処理の概念説明図である。図 13 及び図 15 を参照すると、画像入力部 101 は、学習画像保持部 131 に保持された複数の画像からランダムに  $n$  個サンプリングしてバッチセット 501 を選択する (S1142)。

尤度生成部 106a は、このバッチセット 501 に対して SVM を用いて、各学習画像の被写体の尤度を並べた尤度マップを生成する (S1143)。

尤度集計部 133 は、尤度マップの局所的なピークの位置を抽出して記憶する (S1144)。そして、尤度集計部 133 は、ピークを集計して正事例 (ピークの位置に顔があった) か負事例 (ピークの位置に顔はなかった) の尤度のピークのスコアを領域ごとに集計して尤度スコアの分布を生成する (S1145)。このようにして集計された結果が図 15 に示す尤度スコア分布 503a, 503b である。尤度スコア分布 503a, 503b として示されるように、尤度スコアは画像中の位置ごと、すなわち領域ブロック ( $i, j$ ) ごとに集計する。なお、尤度スコア分布 503a, 503b の細線の曲線は、各画像のスコア分布、太線の曲線はバッチの平均のスコア分布である。

40

#### 【0050】

一般に、画像のパターンから特徴を判別する場合、画像中央部では比較的精度がよく、画像の端ではパターンの一部が隠れるために認識精度が下がる。すなわち、図 15 に示すように、正事例と負事例のスコア分布は、画像や領域ブロックの場所によって変化する。本実施形態では、各バッチセット 501 の画像の領域ブロック ( $i, j$ ) ごとに適切な目標

50

となる閾値 ( $\theta_{ij}$ ) を設定する。例えば、顔検出の誤検出率が一定未満に抑えられるように、バッチセット 501 中の負事例の 95% を棄却できるような閾値を目標閾値 ( $\hat{\theta}$ ) として設定する。例えば、画像中央付近のパラメータ回帰器  $\theta_{3,2}$  に対しては  $\hat{\theta}_{3,2} = 0.4$  と比較的低い閾値を、画像隅のパラメータ回帰器  $\theta_{4,3}$  に対しては  $\hat{\theta}_{4,3} = 0.7$  と高い閾値を回帰推定の目標値として与える ( $S1146$ )。

【0051】

次に、第2特徴生成部 103 は、バッチセット 501 の各バッチの画像の広域特徴 505 を生成する ( $S1147$ )。ついで、第2特徴生成部 103 は、広域特徴 505 を説明変数とし、目標の閾値 ( $\hat{\theta}_{ij}$ ) に近い閾値 ( $\theta_{ij}$ ) が、前述した数 4 の式で算出できるように重みパラメータ ( $[W_{ij}, b_{ij}]$ ) を求める。これは目標値  $\hat{\theta}_{ij}$  と推定値  $\theta_{ij}$  の誤差の総和の値  $E$  をバッチごとに算出し、これを減らすような勾配方向に重みパラメータを更新すればよい。具体的には機械学習で広く公知な勾配法を用いて下式のように重みの値を微小に更新する ( $S1148$ )。

[数9]

$$\begin{aligned} W_{ij}^{t+1} &= W_{ij}^t - \eta \frac{E}{W_{ij}^t} \\ b_{ij}^{t+1} &= b_{ij}^t - \eta \frac{E}{b_{ij}^t} \end{aligned}$$

【0052】

ただし、 $\eta$  は微小な係数である。上記の更新動作を所定の  $m$  回数繰り返し、得られた重みパラメータ ( $[W_{ijm}, b_{ijm}]$ ) をパラメータ回帰器  $\theta_{ij}$  のパラメータとする。

このようにして学習されたパラメータ回帰器  $\theta_{ij}$  は、画像の広域特徴に応じて、領域ブロックごとに適切な目標閾値が設定できるように鍛えられたものとなる。

【0053】

以上の学習処理は、前述のように、顔のサイズ  $s$  ごとにパラメータ回帰器  $\theta_{ijs}$  として学習を行う。その結果、例えば乱雑な背景があるような画像では、誤検出しやすい小サイズの顔の検出が抑制されるような閾値 ( $\theta$ ) が設定される。また、例えば、ポートレート写真のような画像特徴を持つ画像に対しては、中心付近の大サイズの顔の失敗事例が少ないと考えられるため、中心近辺の閾値が低く設定され、大サイズの顔が検出され易くなる。

以上が被写体判別部 106、判別パラメータ設定部 104 の学習処理の内容となる。

【0054】

<変形例>

本実施形態の別の派生的形態として、第1特徴はCNNの低層の特徴マップ 403 a のみを用いて生成し、第2特徴は高層の特徴マップ 403 c のみを使って生成するといったベース特徴の使い分けをしてもよい。また、ニューラルネットワーク以外の特徴の形態として、色のヒストグラム、SIFT特徴といった従来の画像特徴も適応可能である。また、本実施形態の特徴は、矩形のブロック領域を要素単位として算出したが、第1特徴の生成単位は矩形ブロック形状に限定されない。例えば、色味の近さ等で画素をグループ分けして得られるスーパーピクセルと呼ばれる領域を単位とし、各領域について被写体の尤度スコアの計算を行ってもよい。

また、本実施形態では、判別パラメータ設定部 104 は、領域ブロック別かつ顔サイズ別に適切な閾値設定を行ったが、閾値の学習・推定を領域ブロックごとに行わない形態も考えられる。すなわち、領域ブロックに対してではなく、画像全体に対して顔サイズごとに一つの閾値を得てもよい。さらに顔サイズにも分けない、といった派生の形態も考えられる。また、判別パラメータ設定部 104 に与えた閾値の目標値は誤検出率が所定条件を満たすようにという基準で決定したが、この限りでない。他の派生として適合率 (precision) が所定値以上であるという基準や、ベイズエラーを最小にする基準、といった形態を採用することも考えられる。

【0055】

[第2実施形態]

第2実施形態では、意味的領域分割などの他のタスクにも適応できることを説明する。この場合、判別パラメータとして、第1実施形態のような閾値 ( $\theta$ ) ではなく、認識結果を

10

20

30

40

50

統合する際のスコアバランスの重み係数 を設定する。第 2 実施形態の情報処理装置は、ハードウェア構成は第 1 実施形態と同じであるが、コンピュータプログラムにより実現される機能ブロック構成が第 1 実施形態と異なるものとなる。

#### 【 0 0 5 6 】

図 1 6 は、第 2 実施形態の情報処理装置の機能ブロック構成図である。第 2 実施形態の情報処理装置では、画像入力部 2 0 1 が、撮影画像の中から必要な入力画像を取り込む。また、第 1 特徴生成部 2 0 2 が、取り込んだ入力画像の局所領域の特徴を含む第 1 特徴を生成する。また、第 2 特徴生成部 2 0 3 が、取り込んだ入力画像の広域特徴を含む第 2 特徴を生成する。また、判別パラメータ設定部 2 0 4 が、第 2 特徴から判別パラメータを設定する。また、被写体判別部 2 0 6 が、第 1 特徴及び設定された判別パラメータに基づいて画像中の被写体を判別する。また、結果出力部 2 0 7 が、判別結果を出力する。また、ベース特徴生成部 2 0 8 が、第 1 特徴及び第 2 特徴を生成するベースとなる特徴情報を生成し、記憶する。また、カメラ情報入力部 2 1 0 が、カメラ情報をベース特徴生成部 2 0 8 に入力する処理を行う。この実施形態では、新たに、領域ブロックのカテゴリごとの判定を行う領域カテゴリ判定部 2 1 1 を有する。

#### 【 0 0 5 7 】

ベース特徴生成部 2 0 8 は、図 1 6 に示すように、RGBヒストグラム 2 0 8 a、SIFT (Scale-Invariant Feature Transform) 特徴 2 0 8 b をベース特徴として生成する。また、カメラ情報入力部 2 1 0 から、水平線情報値 2 0 8 c、測光情報値 2 0 8 d といった、領域ブロックのカテゴリを判別する際に有効となるカメラ情報もベース特徴として取り込む。ここでは図 1 0 ( A ) ~ ( C ) に示す合焦情報値 1 5 0 2、水平線情報値 1 5 0 4、測光情報値 1 5 0 5 を用いる。

#### 【 0 0 5 8 】

領域カテゴリ判定部 2 1 1 は、入力画像の各領域ブロックの領域カテゴリの尤度スコアを算出する。領域カテゴリは、例えば、空、地面、建物、それ以外の物体の各属性を表す。領域カテゴリの教師値は、図 1 4 ( A ) , ( B ) の ( 3 ) 行に示すようなものである。このような領域カテゴリの判別を行うことが、第 2 実施形態によるタスクの目的である。

#### 【 0 0 5 9 】

第 2 実施形態では、第 1 実施形態と異なり、ベース特徴生成部 2 0 8 がニューラルネットワークを用いていない。また、領域カテゴリ判定部 2 1 1 を、空領域判定部 2 1 1 a、地面領域判定部 2 0 1 1 b、建物領域判定部 2 1 1 c、及び、物体領域判定部 2 1 1 d で構成する。これらの判定部 2 1 1 a ~ 2 1 1 d は、第 1 特徴生成部 2 0 2 で生成される第 1 特徴に基づいて領域カテゴリをスコア化したマップ、すなわち領域カテゴリスコアマップを生成する。この領域カテゴリスコアマップは、被写体判別部 2 0 6 及び第 2 特徴生成部 2 0 3 で使用される。

#### 【 0 0 6 0 】

第 2 実施形態における被写体の判別方法の処理手順例を図 1 7 ( A ) , ( B ) を参照して説明する。まず、第 1 実施形態と同様に、画像入力部 2 0 1 で取り込んだ入力画像からベース特徴生成部 2 0 8 でベース特徴を生成する ( S 2 1、S 2 2 )。第 1 特徴生成部 2 0 2 は、ベース特徴から局所特徴及び広域特徴を抽出し、第 1 特徴を生成する ( S 2 3 )。領域カテゴリ判定部 2 1 1 は、領域カテゴリスコアマップを生成する ( S 2 4 )。この S 2 4 の詳細な処理の手順を図 1 7 ( B ) に示す。すなわち、領域カテゴリ判定部 2 1 1 は、例えば上述の SVM を用いて、空、地面、建物、その他の物体の 4 種類の属性を表す領域ブロックのカテゴリの尤度スコアを算出する。そしてこれらの尤度スコアを並べて領域カテゴリスコアマップを生成する ( S 2 4 0 1 ~ S 2 4 0 5 )。第 2 特徴生成部 2 0 3 は、ベース特徴と領域カテゴリスコアマップとから、第 2 特徴を生成する ( S 2 5 )。

#### 【 0 0 6 1 】

判別パラメータ設定部 2 0 4 は、領域カテゴリの重み係数のパラメータ  $= [ \quad 1, \quad 2, \quad 3, \quad 4 ]$  を推定する ( S 2 6 )。ここでは、領域カテゴリごとにロジスティック回帰器を用意し、説明変数を広域特徴 G として用いて、適切なカテゴリ間のバランスの係数 を

下式で推定する。

[ 数 1 0 ]

$$c = 1 / ( 1 + \exp \{ - W_c^T \cdot G + b_c \} ) \quad ( c = 1, \dots, 4 )$$

【 0 0 6 2 】

被写体判別部 2 0 6 では、尤度生成部 2 0 6 a が、推定された重み  $w = [ w_1, w_2, w_3, w_4 ]$  と各カテゴリの領域カテゴリスコアマップとを掛け算して調整後の尤度スコアを算出する ( S 2 7 )。

[ 数 1 1 ]

$$L'_c(x, y) = c \cdot L_c(x, y) \quad ( c = 1, \dots, 4 )$$

【 0 0 6 3 】

上記のロジスティック回帰器では、予め、適切な係数  $c$  が得られるように、重みパラメータ  $[ W_c, b_c ]$  を調整しておく。具体的には、調整後の尤度  $L'_c(x, y)$  が平均的に各カテゴリの教師値 (  $[ 0, 1 ]$  の二値 ) との誤差を縮小するように勾配法等で調整すればよい。このとき、第 1 実施形態と異なり、広域特徴に領域カテゴリのスコアマップを連結して用いることも考えられる。このような工夫を追加すると、各領域カテゴリの間の誤検出の起こし易さなどを考慮した学習が可能となる。例えば建物と物体は取り間違えやすい、といったカテゴリ間の共起性も一種の情報として考慮し、適切な重み係数  $c$  の学習を行うことができる。例えば、建物と物体の両方の尤度スコアが高い場合は、どちらの重み係数も下げるように学習されるようにする。

閾値処理部 2 0 6 d は、領域カテゴリを閾値処理する ( S 2 8 )。結果出力部 2 0 7 は、判別結果を、例えばモニタ 2 2 に出力して表示させる ( S 2 9 )。

【 0 0 6 4 】

< 変形例 >

第 2 実施形態の変形例として、重み係数を領域カテゴリごとに掛け算するのでなく、例えば  $4 \times 4$  のパラメータを持つ行列  $MAT$  を用いて、全カテゴリのスコアマップの結果を混合して調整後の尤度スコアを求めてもよい。この場合の領域ブロック (  $x, y$  ) の 4 つのカテゴリの尤度  $L(x, y)$  は、下式により算出することができる。この式は、領域ブロック (  $x, y$  ) の 4 つのカテゴリの尤度を  $4 \times 1$  サイズのベクトルとして並べたものに相当する。

[ 数 1 2 ]

$$L'(x, y) = MAT \cdot L(x, y)$$

【 0 0 6 5 】

[ 第 3 実施形態 ]

情報処理装置の第 3 実施形態について説明する。この実施形態では、これまで説明したような定形の領域ブロックに対してではなく、不定形の領域ブロックをベースとしたタスクに対しても適応可能であることを説明する。第 3 実施形態の情報処理装置のハードウェア構成は第 1 実施形態と同じであり、コンピュータプログラムによって実現される機能ブロック構成が異なる。第 3 実施形態の情報処理装置の機能ブロック構成図を図 1 8 に示す。第 1 及び第 2 実施形態と異なる点は、被写体の判別を領域ブロックではなく、人間が画像に注意を向ける関心領域 ( RO I : Region of Interest ) に着目して判別する点である。

【 0 0 6 6 】

第 3 実施形態の情報処理装置は、図 1 8 に示されるように、特徴生成部 3 0 2、RO I 選択部 3 0 4、RO I 特徴抽出部 3 0 6、被写体尤度生成部 3 0 7、領域尤度生成部 3 0 9、判別パラメータ設定部 3 1 1、閾値処理部 3 1 5 の機能を有する。これらは、撮像モードによって検出したい被写体の種類や適切な閾値が異なると考えられ、これについて対応するための機能である。

【 0 0 6 7 】

第 3 実施形態の情報処理装置は、また、風景モードや追尾撮像モードといったカメラ側のインタフェースであって、ユーザーが選択する撮像モードを指示する撮像モード指示部 3 1 2 を有する。さらに、判別パラメータ設定部 3 1 1 において用いるサポートベクトル回

10

20

30

40

50

帰器 (SVR) の係数を保持する SVR 係数保持部 313 を有する。

#### 【0068】

第3実施形態における被写体の判別方法を図19を参照して説明する。ここでは、検出対象となる被写体を特定の人物とする場合の例を示す。情報処理装置は、まず、入力画像の取り込み及びその入力画像の局所特徴の生成処理を行う (S31 ~ S32)。これらの処理の内容は、第1実施形態と同様となる。

ROI 選択部 304 は、「人物らしさ」の高い ROI を複数選択し、それぞれ対応する候補領域を生成する (S33)。ROI 特徴抽出部 306 は、ROI ごとにその位置 (座標 (x, y)) 及びサイズ s を検出するとともに、各 ROI の特徴量を算出し、集計する (S34、ROI プーリング処理と呼ばれる)。被写体尤度生成部 307 は、第1及び第2実施形態と同様の手法で被写体尤度 (人物の尤度) を算出し、算出結果を基に検出候補を生成する (S35)。

#### 【0069】

また、領域尤度生成部 309 は、特徴生成部 302 で生成された局所特徴を用いて、領域ブロックごとに、芝生、群衆、ボールといった領域カテゴリ (属性) の尤度を推定して領域カテゴリ尤度を生成する (S36)。判別パラメータ設定部 311 は、例えば情報処理装置の操作者が撮像モード指示部 312 で指示したモードに応じて、予めモードごとに学習済である重みパラメータである SVR 係数を SVR 係数保持部 313 から読み出す (S37)。指示されるモードには、例えば風景モード、マクロモード、追尾モード、ポートレートモードなどがある。

#### 【0070】

判別パラメータ設定部 311 は、ROI 特徴抽出部 306 が抽出した ROI のサイズ s、及び、ROI の座標 (x, y) と、領域尤度生成部 309 が生成した ROI 中の領域カテゴリの尤度 310 とに基づいて判別パラメータを生成する (S38)。本実施形態の判別パラメータは、各 ROI の判別に用いる閾値 (ROI) である (S38)。閾値処理部 315 は、閾値 (ROI) と S35 で生成した検出候補の尤度とを比較し、すなわち閾値処理し、検出候補が被写体 (人物) かどうかを判別し (S39)、その結果を出力する (S40)。情報処理装置は、このような形態の処理により、例えば「ボールの周辺にある小サイズの人物は検出対象の被写体である可能性が高い」などの情報を考慮して、人物の判別を行うことができる。

#### 【0071】

なお、SVR はサポートベクトルのマージン最大化を基準として回帰学習を行う手法である。SVR で最適な閾値 ( ) を推定するためには、まず、各学習事例について目標閾値 ( ^ ) を準備する。まず、図20 (A) に示すように、ROI の位置及びサイズを特徴次元とする特徴空間上の人物の正事例と負事例の分布を考える。各事例には、被写体尤度生成部 307 が生成した人物の尤度の値 (L) が予め付されている。

#### 【0072】

判別パラメータ設定部 311 は、以下のようにして、SVR の教師値となる最適な閾値 ( ^ ) を事例ごとに設定する。具体的には、各事例について周囲の k 近傍 (当該事例の領域に隣り合う k (所定の定数) の領域) の事例を判別する際の適切な閾値 ( ^ ) を設定する。ここでは、図20 (A) 中の正事例 801 について、周囲の事例の尤度スコアを参照し、近傍のいずれの事例についても閾値 ( ^ ) を越えるような負事例がないように、 ^ = 0.5 を目標閾値として設定している。同様にして、全ての事例について、閾値 ( ^ ) を設定する。設定した結果の例を図20 (B) に示す。

このようにして得られた閾値 ( ^ ) のセットが SVR の学習時の教師値となる。SVR には、図20 (C) に入力事例 803 として示すように、事例が存在しない位置に入力事例が入ってくる。そのため、ロバスト性を持たせて適切な閾値が推定できるように、マージン最大化の基準で SVR の係数の学習を行う。以上が SVR の学習方法になる。

#### 【0073】

また、図20 (D) のように説明変数の特徴量としてより高次元での特徴を推定するとき



にも、同様の方法を用いることができる。また、このような閾値の推定の手法はS V R以外に、カーネル密度推定等の方法を用いてもよく、特定の方法に限定されない。

また、操作者が選択する撮像モードに応じて学習データと教師値の複数のセットを用意し、それぞれ学習することもできる。これにより、追尾モードの時はそれ以外のモードの時よりも、大サイズの人物の検出が抑制される、といった学習を行うことも可能である。

【 0 0 7 4 】

第3実施形態では、第1実施形態と異なり、定形的な領域ブロックをベースとしない認識手法にも適用可能となる。特に、判別パラメータの設定についても、S V Rなどの手法を使うことによって、位置やサイズ等の連続値を考慮した判別パラメータの設定が行える。

【 0 0 7 5 】

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。すなわち、上述した各実施形態の機能を実現するコンピュータプログラムを、ネットワークまた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C P U等）がコンピュータプログラムを読み出して実行する処理によっても本発明の実施は可能である。この場合、そのコンピュータプログラム、および該コンピュータプログラムを記憶した記録媒体は本発明を構成することになる。

10

20

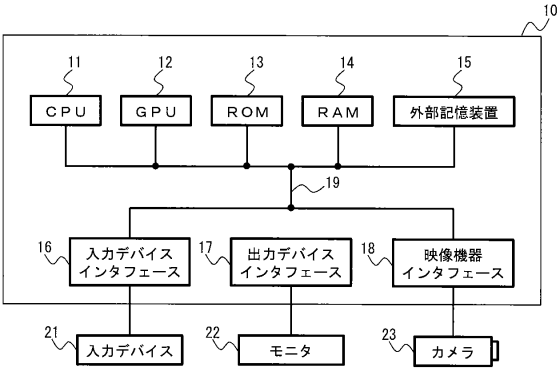
30

40

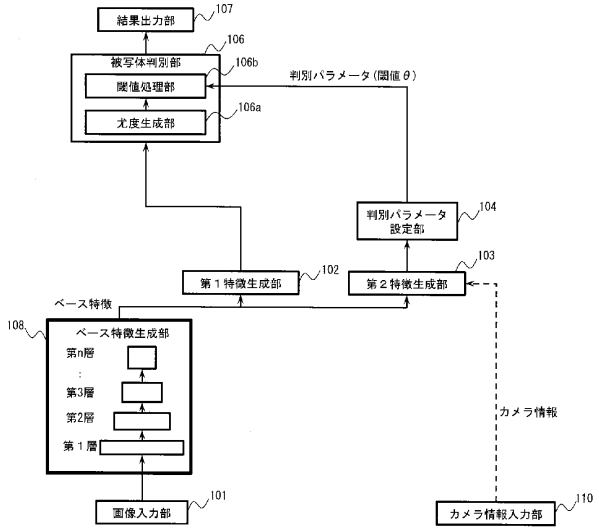
50

【図面】

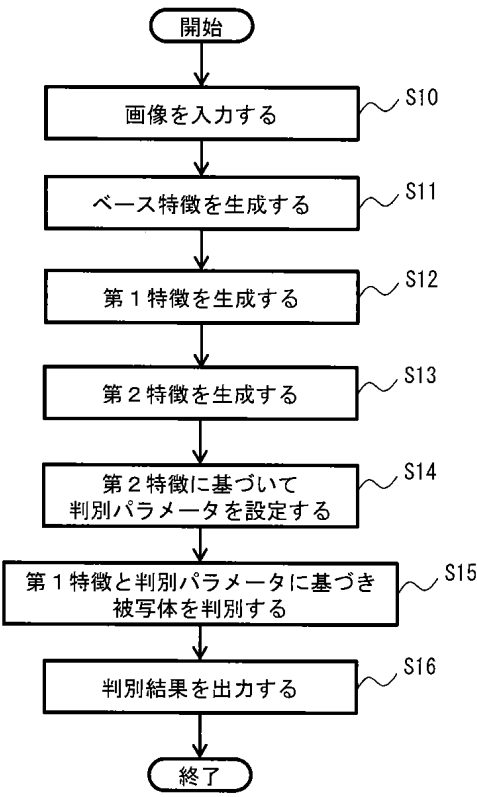
【図 1】



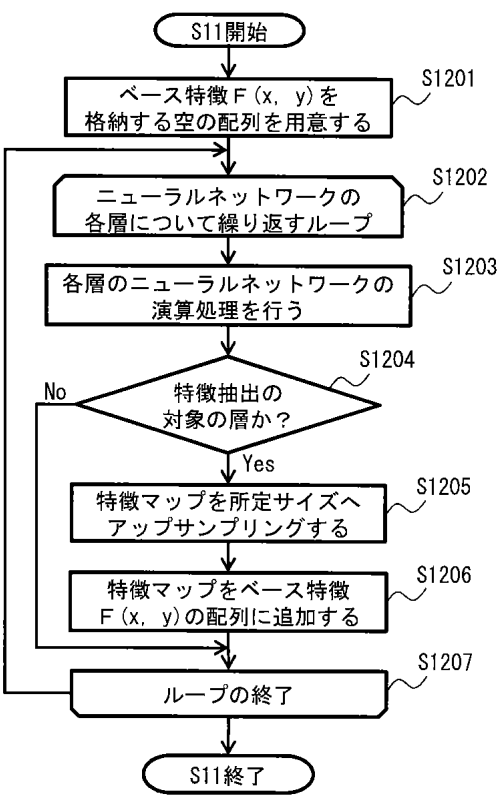
【図 2】



【図 3】



【図 4】



10

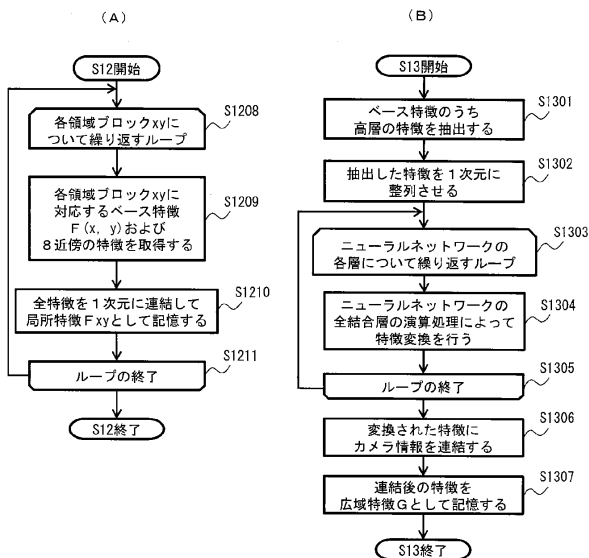
20

30

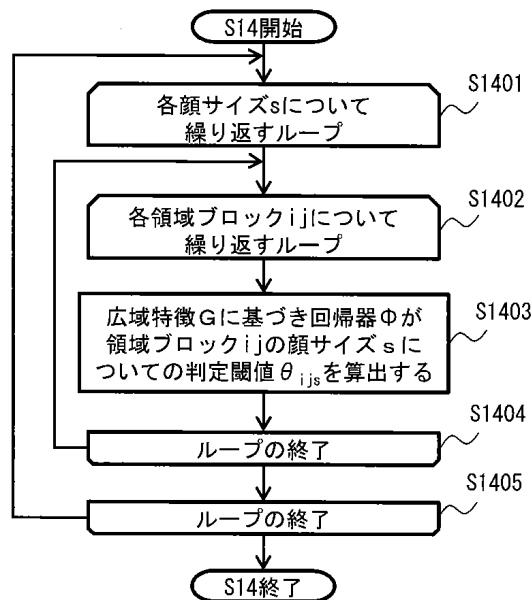
40

50

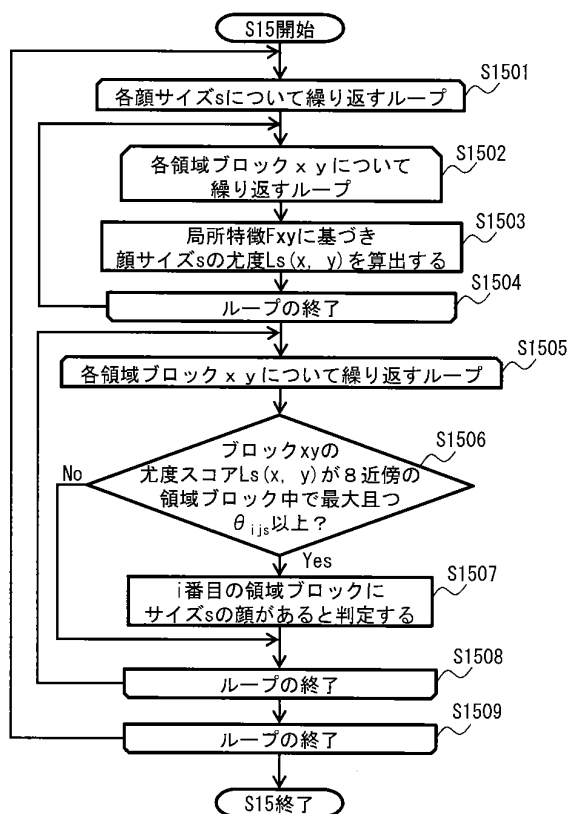
【図 5】



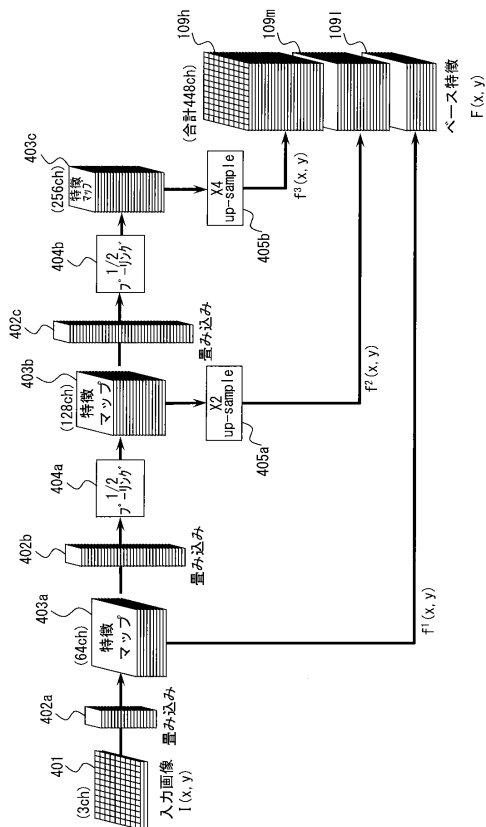
【図 6】



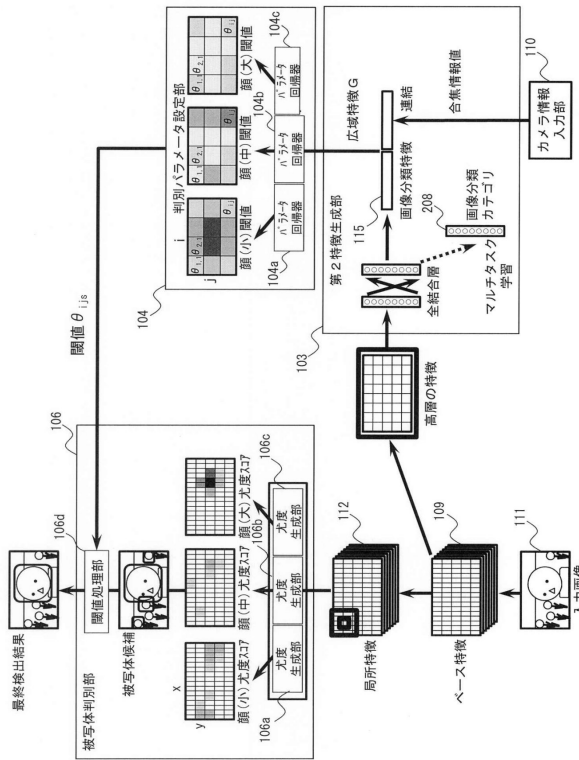
【図 7】



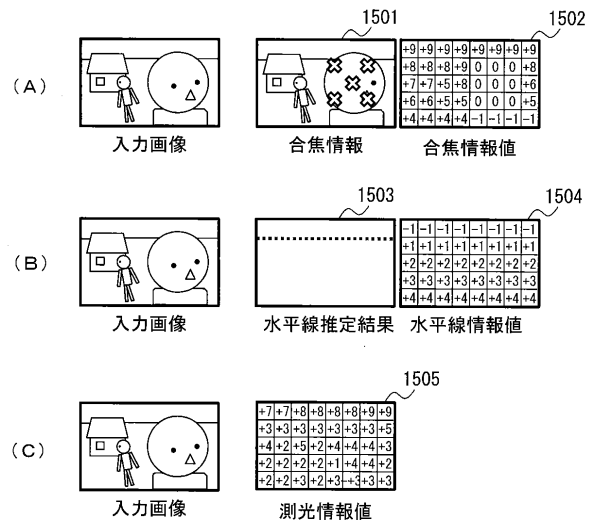
【図 8】



【図 9】



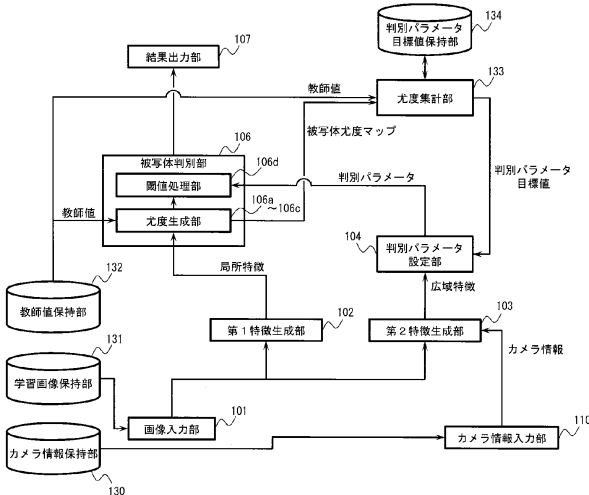
【図 10】



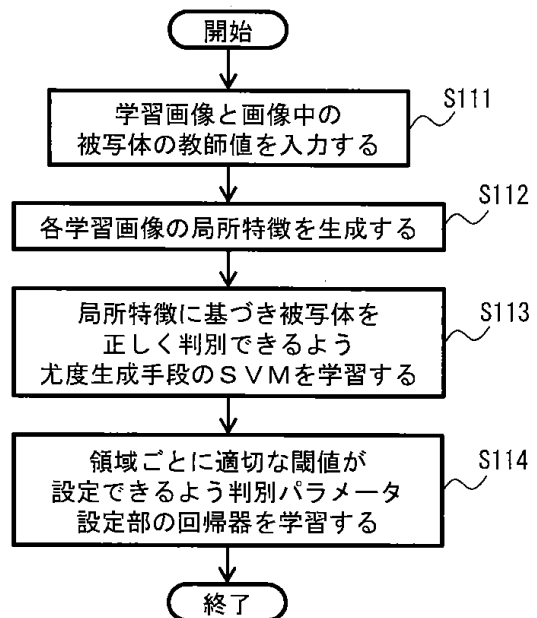
10

20

【図 11】



【図 12】

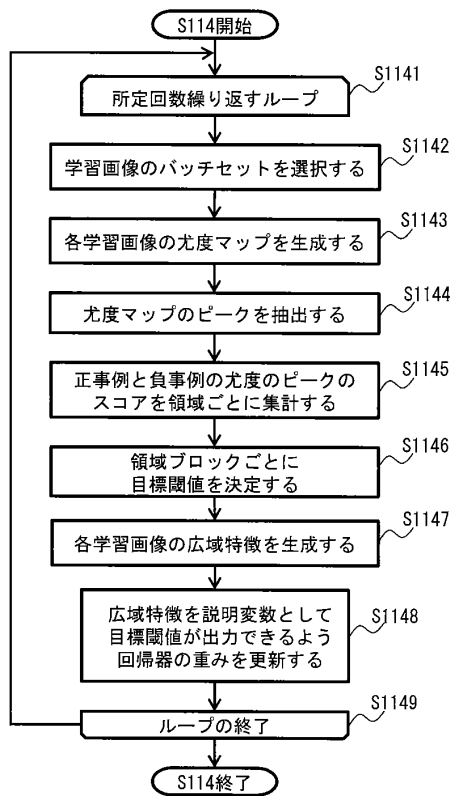


30

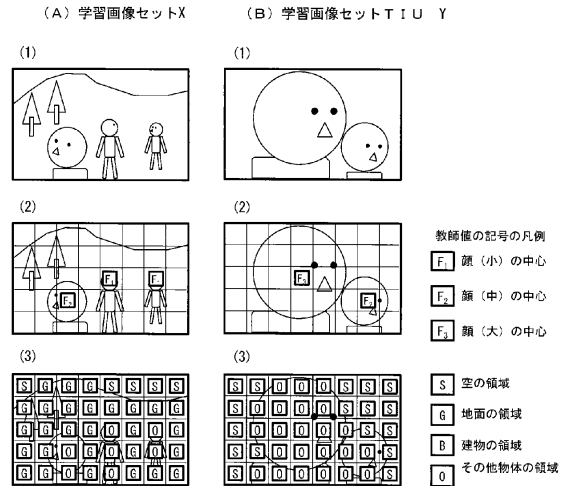
40

50

【 図 1 3 】



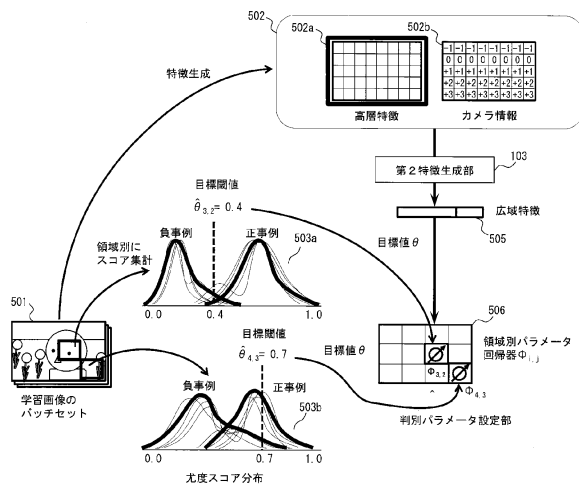
【 図 1 4 】



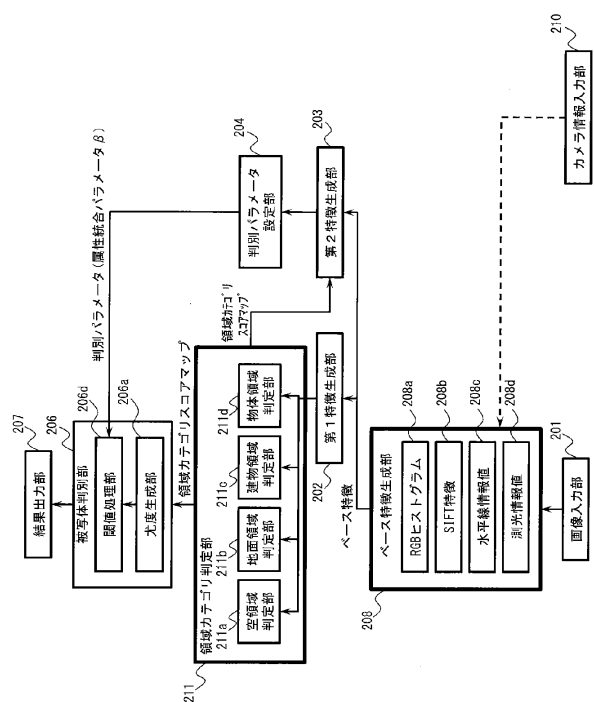
10

20

【 図 1 5 】



【 図 1 6 】

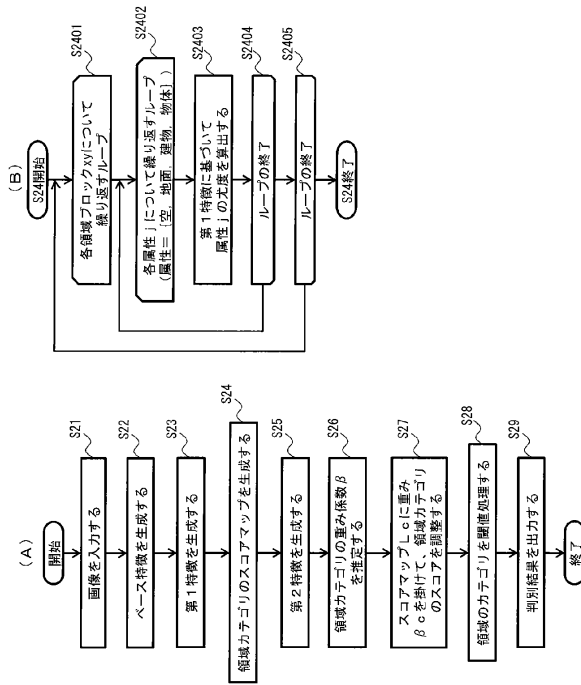


30

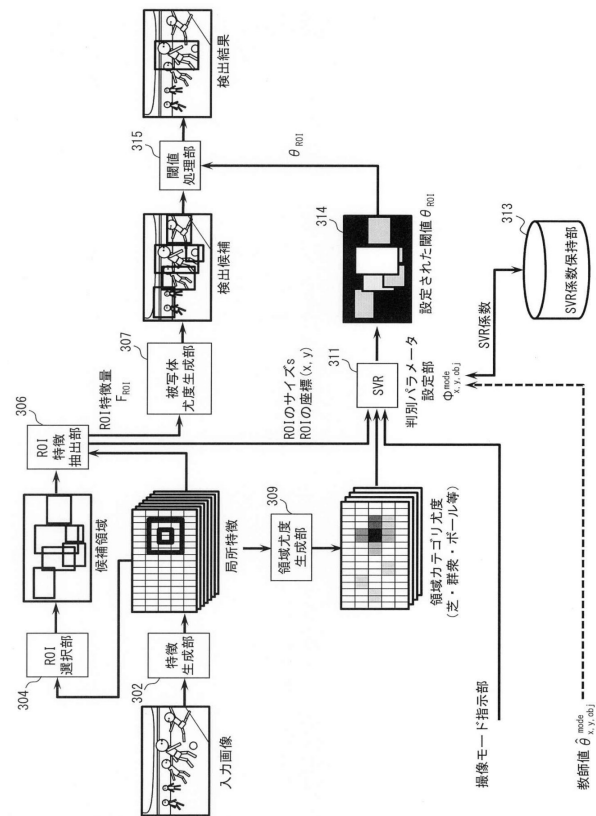
40

50

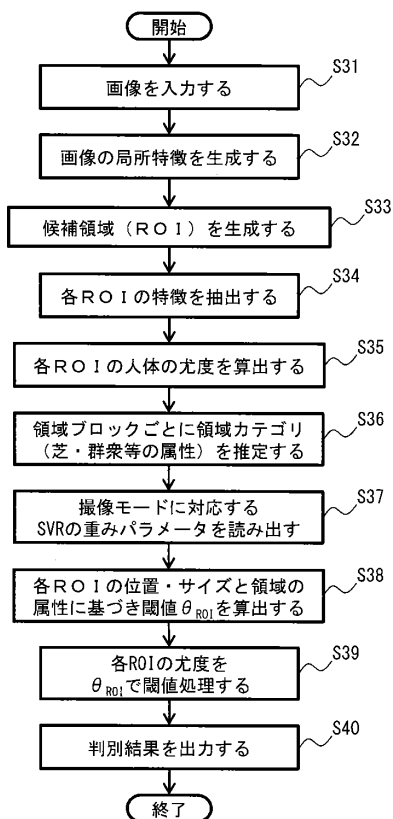
【図 17】



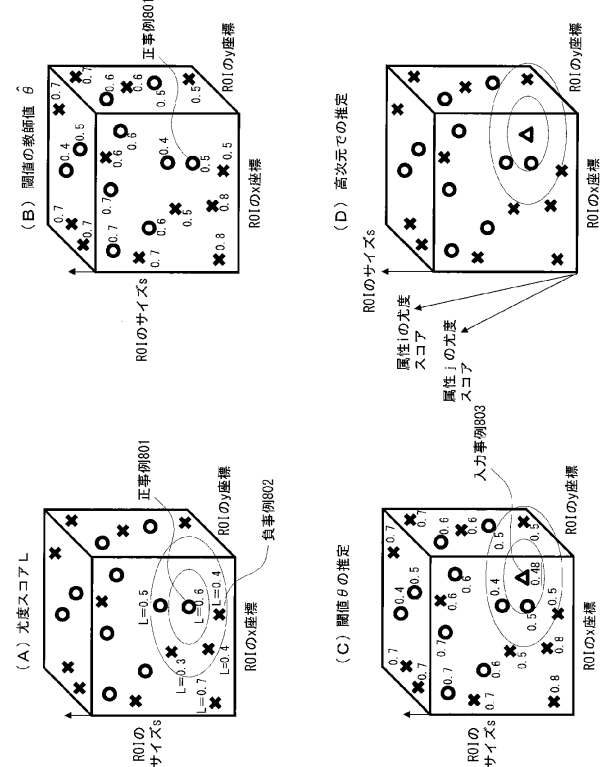
【図 18】



【図 19】



【図 20】



10

20

30

40

50

---

フロントページの続き

- (56)参考文献 特開 2 0 1 7 - 1 0 2 6 2 2 ( J P , A )  
特開 2 0 1 7 - 9 7 5 1 0 ( J P , A )  
特開 2 0 1 7 - 5 3 8 9 ( J P , A )  
特開 2 0 1 6 - 1 8 5 3 8 ( J P , A )  
特開 2 0 1 3 - 4 5 4 3 3 ( J P , A )  
特表 2 0 0 6 - 5 1 3 4 6 8 ( J P , A )
- (58)調査した分野 (Int.Cl. , D B 名)  
G 0 6 T 7 / 0 0