

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-28650  
(P2019-28650A)

(43) 公開日 平成31年2月21日(2019.2.21)

(51) Int.Cl.	F I	テーマコード (参考)
<b>GO6T 7/00 (2017.01)</b>	GO6T 7/00 350C	5L096
<b>GO6N 3/08 (2006.01)</b>	GO6T 7/00 300F	
<b>GO6N 20/00 (2019.01)</b>	GO6N 3/08	
	GO6N 99/00 153	

審査請求 未請求 請求項の数 16 O L (全 20 頁)

(21) 出願番号 特願2017-146337 (P2017-146337)  
(22) 出願日 平成29年7月28日 (2017.7.28)

(71) 出願人 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 100124442  
弁理士 黒岩 創吾  
(72) 発明者 青葉 雅人  
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  
Fターム(参考) 5L096 CA02 DA02 EA39 FA46 GA55  
HA07 HA11 JA16 KA04 KA13  
KA15 MA07

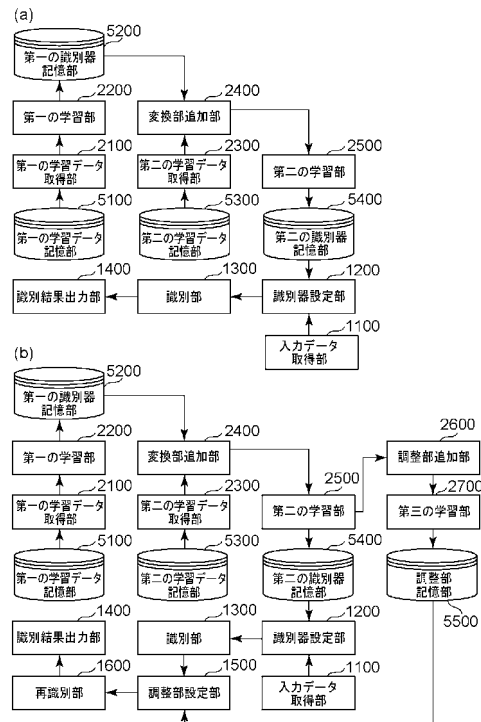
(54) 【発明の名称】 画像識別装置、学習装置、画像識別方法、学習方法及びプログラム

(57) 【要約】 (修正有)

【課題】直観に頼ることなく、高精度な画像認識を行える画像識別装置、学習装置、画像識別方法、学習方法及びプログラムを提供する。

【解決手段】画像処理装置は、撮像装置のセンサ値による入力画像を取得する取得手段1100と、変換部を有する識別器を用いて、取得されたセンサ値による入力画像を識別する識別手段1300と、を有する。識別器のうち少なくとも変換部は、撮像装置のセンサ値による学習画像と当該学習画像に付与された正解データとに基づいて学習していることを特徴とする。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

撮像装置のセンサ値による入力画像を取得する取得手段と、  
変換部を有する識別器を用いて、前記取得されたセンサ値による入力画像を識別する識別手段と、を有し、

前記識別器のうち少なくとも前記変換部は、撮像装置のセンサ値による学習画像と当該学習画像に付与された正解データとに基づいて学習されていることを特徴とする画像識別装置。

## 【請求項 2】

前記識別器は、現像された学習画像と当該学習画像に対して付与された正解データとを含む第 1 の学習データに基づいて学習された識別器に前記変換部を追加した識別器を、前記撮像装置のセンサ値による学習画像と当該学習画像に付与された正解データとを含む第 2 の学習データに基づいて学習することにより生成されることを特徴とする請求項 1 に記載の画像識別装置。

10

## 【請求項 3】

前記識別器は、ニューラルネットワークであることを特徴とする請求項 1 または 2 に記載の画像識別装置。

## 【請求項 4】

前記識別器は、多層の畳み込み演算によって画像特徴を抽出する特徴抽出部と、前記抽出された画像特徴からパターン分類を行う分類部と、を含む畳み込みニューラルネットワークであることを特徴とする請求項 3 に記載の画像識別装置。

20

## 【請求項 5】

前記識別器は、前記特徴抽出部の各層に対して、前記センサ値による入力画像の各画素における輝度値または距離情報の少なくとも一方を重み付き信号として入力することを特徴とする請求項 4 に記載の画像識別装置。

## 【請求項 6】

前記変換部は、

前記入力画像の各画素における輝度絶対量と、前記入力画像の全体から抽出したシーン特徴との、重み係数による重み付き線形和からなる入力信号を、非線形関数によって変換した出力信号を出力する第 1 層と、

30

前記シーン特徴と、前記第 1 層の出力信号との、重み係数による重み付き線形和からなる入力信号を、線形関数で変換した出力信号を出力する第 2 層と、

からなることを特徴とする請求項 3 から 5 のいずれか 1 項に記載の画像識別装置。

## 【請求項 7】

前記シーン特徴は、前記入力画像の全体から取得される画像特徴であることを特徴とする請求項 6 に記載の画像識別装置。

## 【請求項 8】

前記シーン特徴は、前記入力画像の全体から取得される画像特徴と、前記入力画像を撮像したときの撮像情報とであること特徴とする請求項 6 に記載の画像識別装置。

## 【請求項 9】

前記識別手段による識別結果に基づいて、前記変換部を調整する調整手段を更に有し、前記識別手段は、前記調整された変換部を有する識別器を用いて、前記入力画像を再識別することを特徴とする請求項 1 から 8 のいずれか 1 項に記載の画像識別装置。

40

## 【請求項 10】

前記撮像装置のセンサ値による入力画像は RAW 画像であることを特徴とする請求項 1 から 9 のいずれか 1 項に記載の画像識別装置。

## 【請求項 11】

前記識別手段は、前記入力画像を複数のクラスに分類する識別タスク、または前記入力画像を複数のクラスごとの領域に分割する識別タスクを実行することを特徴とする請求項 1 から 10 のいずれか 1 項に記載の画像識別装置。

50

**【請求項 1 2】**

撮像装置のセンサ値による入力画像を識別するために用いられる識別器を学習する学習装置であって、

センサ値による画像が現像された学習画像と当該学習画像に対して付与された正解データとを含む第 1 の学習データに基づいて第 1 の識別器を学習する第 1 の学習手段と、

前記学習された第 1 の識別器に前記変換部を追加して第 2 の識別器を生成する追加手段と、

センサ値による学習画像と当該学習画像に付与された正解データとを含む第 2 の学習データに基づいて前記第 2 の識別器を学習することにより、前記識別手段で用いられる識別器を生成する第 2 の学習手段と、

を有することを特徴とする学習装置。

10

**【請求項 1 3】**

撮像装置のセンサ値による入力画像を取得するステップと、

変換部を有する識別器を用いて、前記取得されたセンサ値による入力画像を識別するステップと、を有し、

前記識別器のうち少なくとも前記変換部は、撮像装置のセンサ値による学習画像と当該学習画像に付与された正解データとに基づいて学習されていることを特徴とする画像識別方法。

**【請求項 1 4】**

撮像装置のセンサ値による入力画像を識別するために用いられる識別器を学習する学習方法であって、

センサ値による画像が現像された学習画像と当該学習画像に対して付与された正解データとを含む第 1 の学習データに基づいて第 1 の識別器を学習するステップと、

前記学習された第 1 の識別器に前記変換部を追加して第 2 の識別器を生成するステップと、

センサ値による学習画像と当該学習画像に付与された正解データとを含む第 2 の学習データに基づいて前記第 2 の識別器を学習することにより、前記識別手段で用いられる識別器を生成するステップと、

を有することを特徴とする学習方法。

20

**【請求項 1 5】**

コンピュータを請求項 1 から 1 1 のいずれか 1 項に記載の画像識別装置として機能させるためのプログラム

30

**【請求項 1 6】**

コンピュータを請求項 1 2 に記載の学習装置として機能させるためのプログラム

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、画像を所定のクラスに分類する、画像を複数のクラスの領域に分割するなどの画像識別技術に関する。

**【背景技術】**

40

**【0002】**

画像を所定のクラスに分類する研究は、これまで広く行われてきており、近年では画像を非常に多くのクラスに分類するタスクも研究されてきている。非特許文献 1 には、深層学習による画像分類の技術が開示されている。

**【0003】**

また、画像を複数の領域に分割する領域分割の研究も多く行われており、画像から人物の領域、自動車の領域、道路の領域、建物の領域、空の領域などの、意味的な領域を切り出す課題が盛んに研究されている。このような課題は、意味的領域分割 (Semantic Segmentation) と呼ばれ、被写体の種類に対応した画像補正や、シーン解釈などに応用できると考えられている。

50

## 【0004】

意味的領域分割を行うにあたり、画像の各位置に関するクラスラベルの識別を小領域 (superpixel) 単位で行う手法が多く提案されている。小領域は主に類似した特徴を持つ小さな領域として画像から切り出されるもので、非特許文献2をはじめとして、さまざまな手法が提案されている。このようにして得られたそれぞれの小領域は、その内部の特徴量、あるいはその周辺のコンテキスト特徴量も一緒に用いてクラスラベルを識別することが行われる。通常はさまざまな学習画像を用いてこのような局所ベースの領域識別器を学習させることで領域識別を行う。

## 【0005】

近年では、深層学習を利用した領域分割の研究も行われてきている。非特許文献3では、CNN (Convolutional Neural Network) の中間層出力を特徴量として利用し、複数の中間層特徴による画素ごとのクラス判定結果を統合する。非特許文献3では、このようにして、画像の意味的領域分割を行っている。非特許文献3の手法では、前述のような小領域分割結果を利用することなく、画素ごとに直接クラス判定を行っている。

10

## 【0006】

このような画像分類や意味的領域分割などの画像識別を行うにあたり、識別器の入力データとして与えられるものは通常、カメラ内部処理もしくは撮影後にユーザの手により現像された画像である。本来、現像画像はユーザが目で見えて楽しむものであるため、画像の現像方法は見た目の美しさを基準にして決定される。しかしながら、このような通常の現像方法が画像識別のタスクにおいて適しているとは限らない。例えば、白い壁を美しく見せるために露出をややオーバー気味に飛ばした画像では、曇天のテクスチャレスな空と壁を区別することは困難になる。これに対して、暗めに撮影して壁のテクスチャが見えるような画像であるほうが、壁と空を分類するのに適していると考えられる。

20

## 【0007】

特許文献1では、撮像装置から得られたRAW画像の平均輝度値によって複数のガンマ補正関数の中から補正関数を選択することで、露出のアンダー/オーバーを抑えた画像を、表示用画像とは別に生成し、物体検出処理に利用することを提案している。

## 【先行技術文献】

## 【特許文献】

30

## 【0008】

【特許文献1】特開2014-11767号公報

## 【非特許文献】

## 【0009】

【非特許文献1】"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton, Proc. Neural Information Processing Systems, 2012.

【非特許文献2】"SLIC Superpixels", R. Achanta, A. S. Hajji, K. Smith, A. Lucchi, EPFL Technical Report, 2010.

40

【非特許文献3】"Fully Convolutional Networks for Semantic Segmentation", Long, Shelhamer, and Darrell, CVPR2015.

【非特許文献4】Hypercolumns for Object Segmentation and Fine-Grained Localization, B. Hariharan, P. Arbelaez, R. Girshick and J. Malik, CVPR2015.

## 【発明の概要】

## 【発明が解決しようとする課題】

50

## 【 0 0 1 0 】

しかしながら、特許文献 1 で、用意される補正関数は人が直観的にパラメータを設定したものであって、補正值の良し悪しはその直観に頼るものであり、高精度な画像識別を行えない場合があった。そこで、本発明は、高精度な画像識別を行えるようにすることを目的とする。

## 【課題を解決するための手段】

## 【 0 0 1 1 】

本発明は、撮像装置のセンサ値による入力画像を取得する取得手段と、変換部を有する識別器を用いて、前記取得されたセンサ値による入力画像を識別する識別手段と、を有し、前記識別器のうち少なくとも前記変換部は、撮像装置のセンサ値による学習画像と当該学習画像に付与された正解データとに基づいて学習されていることを特徴とする。

10

## 【発明の効果】

## 【 0 0 1 2 】

本発明によれば、高精度な画像識別を行うことができるようになる。

## 【図面の簡単な説明】

## 【 0 0 1 3 】

【図 1】各実施形態における画像処理装置の機能構成を示す概略ブロック図。

【図 2】第 1 の実施形態における学習処理、識別処理を示すフローチャート。

【図 3】第 1 の実施形態における意味的領域分割におけるクラスラベルを説明する図。

【図 4】第 1 の実施形態における識別器の構成を説明する図。

20

【図 5】第 1 の実施形態においてガンマ補正関数とその近似関数を説明する図。

【図 6】第 1 の実施形態における変換部の構成を説明する図。

【図 7】各実施形態における識別器の構成を説明する図。

【図 8】第 1 の実施形態における調整部の構成を説明する図。

【図 9】各実施形態における識別器の構成を説明する図。

## 【発明を実施するための形態】

## 【 0 0 1 4 】

## [ 第 1 の実施形態 ]

以下、本発明の第 1 の実施形態の詳細について図面を参照しつつ説明する。図 1 は、各実施形態に係る画像処理装置の機能構成を示す概略ブロック図であり、図 1 ( a ) が本実施形態に係る概略ブロック図である。画像処理装置は、学習時における学習装置、および識別時における画像識別装置として機能するものである。

30

## 【 0 0 1 5 】

このような画像処理装置は、CPU、ROM、RAM、HDD等のハードウェア構成を備え、CPUがROMやHD等に格納されたプログラムを実行することにより、例えば、後述する各機能構成やフローチャートの処理が実現される。RAMは、CPUがプログラムを展開して実行するワークエリアとして機能する記憶領域を有する。ROMは、CPUが実行するプログラム等を格納する記憶領域を有する。HDDは、CPUが処理を実行する際に要する各種のプログラム、閾値に関するデータ等を含む各種のデータを格納する記憶領域を有する。

40

## 【 0 0 1 6 】

まず、本実施形態における学習時の装置構成の概要について説明する。ここで学習とは、後述する識別時の処理を行うために利用される識別器を、事前に用意された学習画像から生成することである。処理内容の詳細については後述する。

## 【 0 0 1 7 】

第 1 の学習データ記憶部 5 1 0 0 には、あらかじめ第 1 の学習データが用意されている。第 1 の学習データは、複数の現像後の学習画像と、各学習画像に対して付与されたクラスラベルから構成される。第 1 の学習データ取得部 2 1 0 0 では、第 1 の学習データ記憶部 5 1 0 0 から、学習画像、クラスラベルを読み込む。第 1 の学習部 2 2 0 0 では、学習画像を識別器に入力することによって得られる識別結果とクラスラベルとの誤差から、第

50

1の識別器を学習する。学習して得られた第1の識別器は、第1の識別器記憶部5200に記憶される。

【0018】

第2の学習データ記憶部5300には、あらかじめ第2の学習データが用意されている。第2の学習データは、デジタルカメラ等の撮像装置で得られた、現像される前のセンサ値による学習画像と、各学習画像に対して付与されたクラスラベルで構成される。第2の学習データ取得部2300では、第2の学習データ記憶部5300から、学習画像、クラスラベルを読み込む。変換部追加部2400では、第1の識別器記憶部5200から、第1の識別器を読み込み、その入力側に変換部を追加することで、第2の識別器を生成する。第2の学習部2500では、第2の学習データにおける学習データを、第2の識別器に

10

【0019】

次に、識別時の装置構成の概要に関して説明する。ここで識別とは、未知の入力画像に対して画像識別を行うことである。処理内容の詳細は後述する。

【0020】

入力データ取得部1100では、撮像装置で得られた、現像される前のセンサ値による入力画像と、その入力画像に対応する撮影情報が読み込まれる。識別器設定部1200では、あらかじめ学習によって得られている第2の識別器を、第2の識別器記憶部5400から読み込んで設定する。識別部1300では、設定された第2の識別器の変換部に取得された入力画像を入力し、識別結果を得る。得られた識別結果は識別結果出力部1400に送られ、ユーザもしくは別機器に結果が提示される。

20

【0021】

本実施形態では、学習時の機能構成も、識別時の機能構成も同じ装置（画像処理装置）で実現されるものとして説明したが、それぞれ別の装置によって実現するようにしてもよい。また、第1の学習データ取得部2100、第1の学習部2200、第2の学習データ取得部2300、変換部追加部2400、および第2の学習部2500は、すべて同じ装置上で実現されるものとして説明したが、それぞれ独立したモジュールとしてもよい。また、装置上で実装されるプログラムとして実現してもよい。第1の学習データ記憶部5100、第1の識別器記憶部5200、第2の学習データ記憶部5300、および第2の識別器記憶部5400は、装置の内部もしくは外部のストレージとして実現される。

30

【0022】

入力データ取得部1100、識別器設定部1200、識別部1300、および識別結果出力部1400は、すべて同じ装置上で実現されるものでもよいし、それぞれ独立したモジュールとしてもよい。また、装置上で実装されるプログラムとして実現してもよいし、カメラ等の撮影装置内部において回路もしくはプログラムとして実装してもよい。第2の識別器記憶部5400は、学習時と識別時で別々の装置で実現される場合には、それぞれ異なるストレージであってもよい。その場合には、学習時に得られた識別器を、識別用の装置におけるストレージにコピーもしくは移動して用いればよい。

【0023】

次に、本実施形態に係る処理の詳細について説明する。ここでは、画像識別として、画像の意味的な領域分割を例にして説明を進める。まず、本実施形態の学習時の処理について説明する。図2(a)は、本実施形態における学習時の処理を示すフローチャートである。

40

【0024】

第1の学習データ取得ステップS2100では、第1の学習データ取得部2100が、第1の学習データ記憶部5100から、学習画像とクラスラベルを、学習データとして読み込む。第1の学習データ記憶部5100には、あらかじめ複数の現像済み学習画像とクラスラベルが用意されている。学習画像は、具体的にはデジタルカメラ等で撮影され、カメラ内部もしくはカメラ外部の現像プログラムによって現像された画像データである。通

50

常はJPEGやPNG、BMPなどの形式で与えられるが、本実施形態は学習画像のフォーマットに限定されるものではない。用意されている第1の学習画像の枚数を $N_1$ 枚とし、 $n$ 番目の第1の学習画像を $I_n$  ( $n = 1, \dots, N_1$ )と書くこととする。意味的領域分割におけるクラスラベルとは、学習画像の各領域に対して識別クラスがラベルとして割り振られているものである。

#### 【0025】

図3に意味的領域分割におけるクラスラベルの例を示す。図3(a)の500は学習画像の例で、図3(b)の540は学習画像500に対応するクラスラベルの例である。このように、画像に対応する正解クラスラベルが与えられた正解データを、画像識別では一般的にGT (Ground Truth)と呼ぶ。各領域に対して、空541、頭髮542、顔543、服544、花545、葉茎546といったクラスラベルが与えられている。ここでは意味的なクラスラベルを例に上げたが、光沢面やマット面、高周波領域といった領域の属性によるクラスラベルが与えられていてもよい。また、空と木の枝のような、複数種類の物体が混在して写っているクラスを定義してもよい。領域クラスは全部で $M$ 種類あるとする。学習画像 $I_n$ の全画素に対応するクラスラベル集合を $GT_n$ とする。

10

#### 【0026】

第1の学習ステップS2200では、第1の学習部2200が、第1の学習データによって、第1の識別器を学習する。ここでは、第1の識別器としてCNN (Convolutional Neural Network)を利用することとして説明する。CNNは、畳み込み層とプーリング層が何層も繰り返されることによって入力信号の局所的な特徴が次第にまとめられ、変形や位置ずれに対してロバストな特徴に変換されることで、識別タスクを行うニューラルネットワークである。

20

#### 【0027】

図4は、本実施形態で用いられるCNNの構造を説明する図である。CNNは、図4(a)の610と620によって構成されており、それぞれ畳み込み層、完全結合層と呼ばれている。これらは、それぞれの役割として特徴抽出とパターン分類を行っている部分であり、以降では一般性を失わないように、それぞれ特徴抽出部610、分類部620と記述することとする。611は入力層であり、入力画像630の各位置における畳み込み演算結果を信号として受け取る。612、613は中間層であり、複数の中間層を介して最終層615へと信号が送られる。各層では畳み込み層とプーリング層が逐次配置され、畳み込み演算と、プーリングによる信号の選択が繰り返される。特徴抽出部610の最終層615における出力信号は、分類部620へと送られる。分類部620では、各層の素子が前後の層と全結合しており、重み係数による積和演算によって信号が出力層640へと送られる。出力層640ではクラス数 $M$ と同数の出力素子があり、それぞれの素子の信号強度を比較して、最も大きな信号を出力する素子に対応するクラスが、その画素の出力クラスラベルとなる。

30

#### 【0028】

学習画像 $I_n$ をCNNに入力した際に出力層で得られる出力信号の値を教師信号と比較することで、学習が行われる。ここで、学習画像 $I_n$ の画素 $(i, j)$ における、クラス $m$ に対応する出力素子の教師信号 $t(n, i, j, m)$ は、 $GT_n$ の画素 $(i, j)$ のクラスラベルが $C$  ( $C = 1, \dots, M$ )である場合に、下記の数式1のように定義される。

40

#### 【0029】

##### 【数1】

$$t(n, i, j, m) = \begin{cases} 1 & \text{if } m = C \\ 0 & \text{otherwise} \end{cases} \quad (\text{数式1})$$

#### 【0030】

入力画像 $I_n$ を識別器に入力した結果得られた出力層における位置 $(i, j)$ のクラス $m$ に対応する出力素子 $(i, j, m)$ の出力信号を $y(n, i, j, m)$ とすると、出力

50

素子 ( i , j , m ) における誤差は下記の数式 2 のように計算される。

【 0 0 3 1 】

$$E ( n , i , j , m ) = ( y ( n , i , j , m ) - t ( n , i , j , m ) ) ^ 2 \quad ( \text{数式 2} )$$

誤差逆伝搬法によって出力層から入力層へと誤差が順次逆伝搬されることで、確率的勾配降下法などによって CNN における各層の重み係数が更新される。

【 0 0 3 2 】

学習における CNN の重み係数は、ランダムな初期値からスタートさせてもよいが、何かしらのタスクに対して学習済みの重み係数を初期値として与えてもよい。例えば、領域分割タスクのクラスラベルは画素ごとに与える必要があるが、画像分類タスクのクラスラベルは画像 1 枚に対して 1 つのラベルを付与すればよい。そのため、事前に人が GT としてクラスラベルを入力する手間は数十倍から数百倍の差がある。そのため、大量の画像分類タスクの学習画像が一般に公開されており、簡単に入手することができる。例えば、ILSVRC ( ImageNet Large-scale Visual Recognition Challenge ) では 120 万枚の画像分類用学習用画像が公開されている。よって、CNN の重み係数の初期値を、このような画像分類タスクで一度初期学習しておき、その学習結果による重み係数を初期値として、領域分割タスクの学習を始めてもよい。このようにして、第 1 の学習ステップにて学習された CNN の重み係数を、識別器のパラメータとして第 1 の識別器記憶部に記憶する。

10

【 0 0 3 3 】

次に、第 2 の学習データ取得ステップ S 2 3 0 0 では、第 2 の学習データ取得部 2 3 0 0 が、第 2 の学習データ記憶部 5 2 0 0 から、現像されていない学習画像、撮影情報、およびクラスラベルを、学習データとして読み込む。

20

【 0 0 3 4 】

第 2 の学習データ記憶部 5 2 0 0 には、あらかじめ複数の現像されていない学習画像とクラスラベルが用意されている。また、学習画像に関する撮影情報も得られている。学習画像は、具体的にはデジタルカメラ等で撮影され、現像処理を行う前の状態における、CCD や CMOS のような画像素子におけるセンサ値の羅列であり、一般的に RAW 画像と呼ばれる。RAW 画像における輝度値は、撮影情報を使って以下のように各画素における各色チャンネルの輝度の絶対量を求めることができる。

30

【 0 0 3 5 】

撮影情報として、この画像全体における輝度  $B_v$  値、センサの適正レベル値  $opt$  が得られているとする。ある画像における画素位置 ( i , j ) に対応する RAW 画像上のベイヤー配列における R チャンネルのセンサ測定値が  $R_{RAW}$  であったとき、画素 ( i , j ) の R チャンネルにおける輝度の絶対量である  $R_{B_v}$  の値は、下記の数式 3 で求めることができる。

【 0 0 3 6 】

【 数 2 】

$$R_{B_v} = B_v + \log_2 \left( \frac{R_{RAW}}{opt} \right) \quad ( \text{数式 3} )$$

40

【 0 0 3 7 】

G チャンネルと B チャンネルとの輝度絶対量である  $G_{B_v}$  と  $B_{B_v}$  も、同様にして数式 4、数式 5 より求められる。

【 0 0 3 8 】



【数 3】

$$G_{Bv} = Bv + \log_2 \left( \frac{G_{RAW}}{opt} \right) \quad (\text{数式 4})$$

【0039】

【数 4】

$$B_{Bv} = Bv + \log_2 \left( \frac{B_{RAW}}{opt} \right) \quad (\text{数式 5})$$

10

【0040】

このような変換式を用いることで、RAW画像による第2の学習画像に関する輝度絶対量のマップを得ることができる。

【0041】

用意されている第2の学習画像の枚数を $N_2$ 枚とし、 $n$ 番目の第2の学習画像から変換して得られた輝度絶対量のマップによる学習画像を $J_n$  ( $n = 1, \dots, N_2$ )と書くこととする。クラスラベルは、第1の学習画像におけるクラスラベルと同じ定義のものとする。学習画像 $J_n$ の全画素に対応するクラスラベル集合を $GT_n$ とする。RAW画像を伴った画像は、それを伴わない画像の収集に比べて困難であるため、第2の学習画像の数 $N_2$ は、一般的には $N_1$ より少ないことが多い。実際、多くのアマチュア写真はRAW画像を公開しないため、ウェブなどで収集できる画像のほとんどはRAW画像を伴わないものである。また、第2の学習データで用いた学習画像は、現像することによって第1の学習データに利用することも可能である。

20

【0042】

変換部追加ステップS2400では、変換部追加部2400が、第1の識別器記憶部5200から第1の識別器を読み込み、読み込まれた第1の識別器の入力層側に、変換部を追加する。

【0043】

まず、第1の識別器記憶部5200から、第1の学習ステップS2200にて学習されたCNNの重み係数を読み込む。読み込まれた重み係数をCNNに設定する。設定されたCNNに対して、図4(b)のようにして変換部651を追加する。変換部651の入力側には、第2の学習データ取得ステップS2300で用意した、輝度絶対量による学習画像651が入力される。変換部650を通過して変換された画像652は、現像後の画像と同様にしてCNNの入力層611に入力される。

30

【0044】

変換部は、CNNの新しい層として追加される。通常、RAW画像から現像画像への変換は、ガンマ補正とホワイトバランスによる修正が行われる。ガンマ補正関数は下記の数式6のように定義される。

【0045】

$$y = x \quad (\text{数式 6})$$

40

ここで、 $x$ は任意の画素におけるRAW画像の値、すなわち撮像装置におけるセンサ値であり、 $y$ はその画素の現像後の輝度値である。制御パラメータの値は、カメラやメーカ、撮影モードなどによって異なる。図5は、本実施形態におけるガンマ補正関数とその近似関数を説明する図であり、図5(a)にガンマ関数の例を示している。図5(a)の701は $\gamma = 1$ 、702は $\gamma = 0.5$ 、703は $\gamma = 2$ のときのガンマ補正関数のカーブである。ホワイトバランスは、これら補正された輝度値のチャンネルごとの重み付けにあたる。ここで、入力信号に対するガンマ補正関数を下記の数式7のように近似することを考える。

【0046】

50

$$y = w_2 \tanh(w_1 x + b_1 \cdot z_1) + b_2 \cdot z_2 \quad (\text{数式 7})$$

ここで、 $z_1$  および  $z_2$  は撮影環境によって変動する変数で、 $w_1$ 、 $w_2$ 、 $b_1$  および  $b_2$  は重み係数である。図 5 (b) に示すように、この関数は補正関数の近似とすることができる。図 5 (b) の 711、712、713 は、それぞれ下記の数式 8、数式 9、数式 10 のような関数である。

【0047】

$$y = 1.1 \tanh(x - 0.5) + 0.5 \quad (\text{数式 8})$$

$$y = 5 \tanh(x + 1) - 3.8 \quad (\text{数式 9})$$

$$y = 5 \tanh(x - 2) + 4.8 \quad (\text{数式 10})$$

図 6 は本実施形態における変換部の構成を示す図であり、数式 7 における形式は、図 6 (a) のような素子の組み合わせで表現することができる。輝度絶対量による学習画像 651 の任意の画素と、変換後の画像 652 における対応画素は、素子 653 および素子 654 によって結合される。学習画像 651 における輝度絶対値は、数式 7 では  $x$  に相当し、重み  $w_1$  で重みづけされて、入出力関数として非線形関数の  $\tanh$  を持つ素子 653 に入力される。素子 653 の出力信号は重み  $w_2$  で重みづけされ、単調増加の線形関数を入出力関数として持つ素子 654 に入力される。

【0048】

学習画像 651 からは、シーン特徴抽出器 655 を通して、画像のシーンを記述する特徴量 656 が抽出される。シーン記述特徴量 656 は、HOG や Fisher Vector、色ヒストグラムなどを想定することができるが、本実施形態はその特徴量の種類によって限定されるものではない。また、シーン記述特徴は上記のように画像特徴だけに限らない。例えば、撮像画像における付帯情報として、地軸に対する撮像装置の向き情報としてのジャイロセンサ値や、時計による時刻情報から特徴量を抽出してもよい。その例を図 6 (b) に示す。例えば、ジャイロセンサの値から地面方向を 3 軸の値で得ることができるため、これは正規化などすれば 3 次元のシーン記述特徴ベクトルとして利用することができる。また、時計による時刻情報は、1 時間を  $15 \text{ deg}$  として対応付けて  $\sin$ 、 $\cos$  による循環関数にすれば、1 日を 1 周期とした特徴量として利用できる。カレンダーに関しても同様に 1 カ月を  $15 \text{ deg}$  として 1 年を 1 周期とした特徴量として利用することができる。シーン記述特徴量 656 は、数式 7 では  $z_1$  および  $z_2$  に相当する。そして、重みベクトル  $b_1$  による積和演算により重みづけされて素子 653 へと入力され、重みベクトル  $b_2$  による積和演算により重みづけされて、素子 654 へと入力される。

【0049】

ここでは、 $z_1 = z_2$  として説明したが、シーン記述特徴は、例えば  $z_1$  を HOG、 $z_2$  を Fisher Vector といったように、別々のものとして分けてもよい。シーン記述特徴は、 $\tanh$  カーブのバイアスを調整するための特徴であって、重み係数ベクトル  $b_1$  および  $b_2$  で重み付けすることは、一種のシーン識別を行うことに相当する。例えば、晴れた日の屋外のシーンと、白い壁に囲まれた屋内のシーンでは、画像中に写っている物体の相違と、輝度絶対量の違いにより、異なるシーン記述特徴が得られるため、画像変換のバイアスとして異なる補正量をかけることになる。また、シーン記述特徴 656 を素子 653 および 654 に送る際に、重み係数  $b_1$  および  $b_2$  による線形和ではなく、多層構造のニューラルネットワークを加えてもよい。図 6 (c) はその例を示しており、657 および 658 はそれぞれ、入力層にシーン記述特徴 656 を入力し、1 つの出力信号  $f(z)$  および  $g(z)$  を出力する多層ニューラルネットワークである。この場合、数式 7 は以下の数式 11 のようになる。また、以降の説明では、各ニューラルネットワーク  $f$  および  $g$  の結合係数を  $b_1$  および  $b_2$  と置き換えて読めばよい。

【0050】

$$y = w_2 \tanh(w_1 x + f(z_1)) + b_2 \cdot g(z_2) \quad (\text{数式 11})$$

素子 654 の出力信号は、そのまま CNN の入力層へ渡す画像 652 の対応画素の値として扱われる。このような結合を  $R_{Bv}$ 、 $G_{Bv}$ 、 $B_{Bv}$  の各チャネルに対して持たせたとき、 $w_2$  および  $b_2$  の値は各チャネルのバランスを表現しており、これはホワイトバラ

10

20

30

40

50

ンスの値を近似するものである。このようにして、現像前の学習画像 6 5 1 は、輝度絶対量から現像処理と近似された変換により、画像変換されることになる。

【 0 0 5 1 】

第 2 の学習ステップ S 2 5 0 0 では、第 2 の学習部 2 5 0 0 が、変換部追加ステップ S 2 4 0 0 で追加された変換部とともに、識別器を学習する。変換部追加ステップ S 2 4 0 0 で設定された画像変換を定義する重み係数  $w_1$ 、 $w_2$ 、 $b_1$ 、 $b_2$  は、第 2 の学習データ取得部によって取得された学習画像とクラスラベルによって学習される。学習画像  $J_n$  が図 4 ( b ) の変換部 6 5 0 に入力され、特徴抽出 6 1 0 と分類部 6 2 0 を介して出力信号が得られたら、その値をクラスラベルと比較することにより、CNN 全体と変換部の重み係数を学習する。

10

【 0 0 5 2 】

特徴抽出部 6 1 0 と分類部 6 2 0 の結合係数は、第 1 の学習ステップ S 2 2 0 0 で得られた値を初期値とする。変換部 6 5 0 における重み係数は、ランダムな初期値から学習させてもよい。あるいは、変換部 6 5 0 だけ CNN とは独立に学習させ、その状態を初期値として CNN と一緒に学習させてもよい。変換部 6 5 0 だけを初期学習させるためには、変換部 6 5 0 を 3 層ニューラルネットワークとみなして、現像前の輝度絶対量マップによる学習画像 6 5 1 を入力とし、素子 6 5 3 の出力信号をネットワークの出力信号とみなす。教師信号として、適正露出による現像後画像の輝度値を与えることにより、誤差逆伝搬により回帰学習を行えばよい。変換部 6 5 0、特徴抽出部 6 1 0、分類部 6 2 0 の重み係数の初期値が決定されたら、全てを通して学習を行う。このようにして、変換部 6 5 0 と特徴抽出部 6 1 0、分類部 6 2 0 をすべてを通して学習画像で学習させることにより、変換部 6 5 0 のパラメータも、学習画像に対して識別誤差を軽減させる方向に修正することができる。これは、画像の現像方法を、見た目の良さではなく、識別し易いように修正していることに相当する。変換部 6 5 0 と畳み込み層 6 1 0、完全結合層 6 2 0 が学習されたら、得られた重み係数を第 2 の識別器記憶部 5 4 0 0 に記憶させる。

20

【 0 0 5 3 】

このようにして学習された識別器を用いて実際の入力画像を識別する工程を、以下に詳細説明する。図 2 ( b ) は、本実施形態に係る識別時の処理を示すフローチャートである。

【 0 0 5 4 】

まず、入力データ取得ステップ S 1 1 0 0 では、入力データ取得部 1 1 0 0 が、撮像装置から得られた現像前の画像データが取得される。入力データの方式は、第 2 の学習データ取得ステップ S 2 3 0 0 における現像前画像と同様であるとする。すなわち、撮像装置で得られたセンサ値による RAW 画像から、撮像情報を利用して、数式 3、数式 4、数式 5 を使って輝度絶対量のマップに変換したものである。

30

【 0 0 5 5 】

識別器設定ステップ S 1 2 0 0 では、識別器設定部 1 2 0 0 が、第 2 の識別器記憶部 5 4 0 0 から学習済みの識別器を読み込む。なお、ここでは入力データ取得ステップ S 1 1 0 0 の後に識別器設定ステップ S 1 2 0 0 を行うようにしているが、この 2 つのステップの手順は逆でもよい。識別器を常にメモリに確保して入力画像を次々に処理する場合には、識別器設定ステップ S 1 2 0 0 の後で入力データ取得ステップ S 1 1 0 0 以降の処理を繰り返し行うとしてもよい。識別器設定ステップ S 1 2 0 0 で設定される識別器は、図 4 ( b ) で表わされる変換部と CNN で構成される識別器である。

40

【 0 0 5 6 】

識別ステップ S 1 3 0 0 では、識別部 1 3 0 0 が、識別器設定ステップ S 1 2 0 0 で設定された識別器を用いて、入力データ取得ステップ S 1 1 0 0 で取得された入力画像の識別処理を行う。輝度絶対量のマップとして取得された入力画像は、図 4 ( b ) における識別器の変換部 6 5 0 に入力され、変換された画像は CNN の特徴抽出部 6 1 0 における入力層 6 1 1 へと入力される。畳み込み層 6 1 0 では入力画像の信号が各層に順伝搬され、変換された信号は全結合層 6 2 0 を介して、各識別クラスに割り当てられた出力素子 6 2

50

1 の出力信号になる。信号が最も大きい出力素子に対応するクラスラベルが、その画素のクラス識別結果となる。

【0057】

識別結果出力ステップ S 1 4 0 0 では、識別結果出力部 1 4 0 0 が、識別ステップ S 1 3 0 0 で得られた識別結果を出力する。識別結果出力ステップ S 1 4 0 0 で行われる処理は、識別結果を利用するアプリケーションに依存するものであって、本実施形態を限定するものではない。例えば、領域ごとに与える画像処理を、領域クラスによって変更するような画像補正アプリケーションであれば、各画素のクラスラベルを画像補正プログラムに出力すればよい。その際、各クラスの曖昧さによって処理の重み付けなどが必要であるなら、各クラスラベルに対応する出力素子 6 2 1 の出力信号値をクラス尤度としてそのまま出力してもよい。特定のクラスに関する識別結果だけが必要であるなら、他のクラスに関する結果を捨てて、必要なクラスの識別結果だけを出力すればよい。

10

【0058】

以上の説明では、画像識別処理として、画像の領域分割を例に説明したが、画像分類タスクに対しても、本実施形態は適用可能である。図 9 は各実施形態における識別器の構造を説明する図であり、画像分類タスクの場合は、図 9 ( a ) のように CNN の特徴抽出部 6 1 0 の最終層 6 1 5 の全画素における出力信号を、分類部 6 2 0 に入力すればよい。

【0059】

以上のように、本実施形態によれば、識別器への入力画像を変換する変換部を学習することにより、識別に適した画像が得られ、高精度な画像識別を実現することが可能になる。また、識別器の画像入力側に変換部を加えて現像前の学習画像を用いて追加学習することにより、現像処理も含めた識別器の学習を行うことができる。これにより、見た目重視の現像処理を介した画像で無理に識別処理を行うことなく、より高精度な識別が行えることが期待できる。また、変換部以外の部分を大量の現像後画像で事前学習し、それを初期値に変換部を含めた識別器の学習を行うプロセスにより、比較的少ない現像前画像による学習画像で、画像変換を含めた識別器を学習することができる。これは、現像後画像に比べて現像前画像による大量の学習画像を揃えることが困難な場合に有効である。

20

【0060】

[ 第 2 の実施形態 ]

次に、本発明の第 2 の実施形態について説明する。第 1 の実施形態では、識別器に CNN を用いた場合の例を示したが、識別器に CNN を用いた場合、畳み込み処理の繰り返しによって、順伝搬信号からはエッジ情報などが強く残ることになり、輝度値などの絶対値情報は徐々に情報が薄れていく傾向がある。色や明るさが有効な特徴であるような場合、そのような情報が失われることは識別精度低下の原因になる。例えば、パステルカラーの無地な家の壁や、太陽のランプのように光る物体などは、各色チャンネルの輝度があればより識別精度の向上が期待できる。

30

【0061】

また、撮像装置によっては、像面位相差 AF などによって距離情報が得られている場合もあり、距離情報との併用によって、被写体との距離によって発生するスケーリングに対するロバスト性が向上することも期待できる。本実施形態では、入力画像の各画素に対して与えられる絶対値情報を識別器に取り込む形態について説明する。

40

【0062】

なお、第 1 の実施形態で既に説明をした構成については同一の符号を付し、その説明を省略する。本実施形態における装置構成は、第 1 の実施形態と同じく図 1 ( a ) で表わされるため、詳細な説明は省略する。

【0063】

まず、本実施形態の学習処理について説明する。そのフローチャートは、図 2 ( a ) で示される第 1 の実施形態の学習処理のフローチャートと同じであるが、一部のステップにおける処理の内容が第 1 の実施形態とは異なる。

【0064】

50

第1の学習データ取得ステップS2100, 第1の学習ステップS2200、および第2の学習データ取得ステップS2300に関しては、第1の実施形態と同様であるため、説明は省略する。

【0065】

変換部追加ステップS2400では、まず、変換部追加部2400が、第1の識別器記憶部5200から、第1の学習ステップS2200にて学習されたCNNを読み込む。そして、変換部追加部2400は、このCNNに対して、図4(c)のように変換部650と、伝達部660を追加する。変換部650に関しては第1の実施形態と同様であるため、説明は省略する。伝達部660は、現像前の画像が持つ、各画素における絶対値情報を伝達するための多層ネットワークである。各層661、662、663、...、665は、CNNの各畳み込み層611、612、613、...、615に対応する層である。各層のチャンネル数は、画像の各画素における情報の数に相当する。例えば、絶対値情報として第1の実施形態で示した輝度絶対量( $R_{Bv}$ 、 $G_{Bv}$ 、 $B_{Bv}$ )を用いる場合には、各層は $R_{Bv}$ 、 $G_{Bv}$ 、 $B_{Bv}$ に対応する3つのチャンネルを持つことになる。層の各チャンネルにおける特徴面のサイズは、CNNの対応層の特徴面のサイズと等しいものとする。伝達部660の各層は単純にスケーリングの関係にあり、線形補間やバイキュービック補間、最近傍法などによる手法でリサイズされる。演算時間を重視するのであれば、間引きによるテーブル参照でリサイズを行ってもよい。これら伝達部における層間の結合部分には学習によって修正される結合係数は割り振られない。また、CNNの入力層611に対応する伝達部660の層661のサイズは入力画像サイズそのものであるため、入力画像の画素情報がそのまま直接設定されることになる。伝達部660の最も入力側の層611における、出力部の画素位置に対応する位置669の値は、分類部620の入力層にそのまま入力される。

10

20

【0066】

伝達部660の各層における値は、対応するCNNの畳み込み層に対して、バイアス係数とともにバイアス値として入力される。第1番目の畳み込み層におけるチャンネルmの位置(i, j)の素子に対する入力信号 $u_{1m}(i, j)$ は、以下の数式12のように表わされる。

【0067】

【数5】

$$u_{1m}(i, j) = \sum_{k=1}^K \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} z_{1-1}(i+p, j+q, k) h_{1pqkm} + b_{1m} + \sum_{r=1}^R J_1(i, j, r) b_{1r} \quad (\text{数式12})$$

30

【0068】

ここで右辺第1項はCNNにおける結合を表わしており、KはCNNの第1-1層におけるチャンネル数、Hは第1-1層と第1層の間における畳み込みフィルタの幅である。 $h_{1pqkm}$ は、第1層の第mチャンネルと第1-1層の第kチャンネルを結合する畳み込みフィルタの、フィルタ中心座標における位置(p, q)の値である。また、 $z_{1-1}(i, j, k)$ は、第1-1層における位置(i, j)の出力信号、 $b_{1m}$ は、第1層の第mチャンネルにおけるバイアス係数である。右辺第2項は伝達部660からの結合を表わしており、Rは画素情報のチャンネル数、 $J_1(i, j, r)$ は画素情報伝達部の第1層の第rチャンネルにおける値、 $b_{1r}$ は同バイアス係数である。これらの中で、学習によって修正されるパラメータは、 $h_{1pqkm}$ 、 $b_{1m}$ 、 $b_{1r}$ である。

40

【0069】

第2の学習ステップS2500では、第2の学習部2500が、第2の学習データ取得ステップS2300で取得した学習画像を用いて、CNNの内部結合係数を学習する。また、第2の学習部2500は、CNNの内部結合係数とともに、学習画像を用いて、変換部追加ステップS2400で追加された変換部650および伝達部660とCNNを結合する係数を学習する。上述した結合係数が学習されるパラメータとして追加されたことを

50

除けば、学習に関する基本的なアルゴリズムは第1の実施形態と同様であるため、その説明は省く。学習によって修正されたパラメータは、第2の識別器記憶部5400に記憶される。

#### 【0070】

次に、本実施形態の識別処理について説明する。そのフローチャートは、図2(b)で示される第1の実施形態の識別処理のフローチャートと同じであるが、一部のステップにおける処理の内容が第1の実施形態とは異なる。

#### 【0071】

入力データ取得ステップS1100、識別器設定ステップS1200の処理は、第1の実施形態と同様であるため、説明は省略する。

10

#### 【0072】

識別ステップS1300では、識別部1300が、現像前の輝度絶対値による入力画像を図4(c)に示すネットワークに入力することにより、識別結果を得る。識別時における順伝搬方向の信号の伝達に関しては、学習時と同じであるため、詳細な説明は省略する。

#### 【0073】

識別結果出力ステップS1400における処理は、第1の実施形態と同様であるため、説明は省略する。

#### 【0074】

絶対値情報としては、上記輝度絶対量以外にも、撮像系の像面位相差AFなどによって得られる距離情報を与えてもよい。距離情報は、対象物体の絶対的なサイズや立体形状に関する情報を与えるため、スケールによる類似物や、実物と写真や絵画などとの区別が付きやすくなる。例えば、看板に描かれた人物や巨大人物像と、実物の人間を区別する場合などで有効である。

20

#### 【0075】

距離情報用の伝達部670を追加した場合の構成を図4(d)に示す。653は、各画素の距離情報を持つ距離マップである。画像の画素密度に対して測距点が疎である場合には、線形補間やバイキュービック補間、あるいは最近傍法などによって、各画素の距離を補うことで、各画素に対する距離マップを算出すればよい。この場合、チャンネル数は、 $R_B$ 、 $v$ 、 $G_{Bv}$ 、 $B_{Bv}$ 、距離情報の4チャンネルになる。図4(d)の構成による学習処理や識別時の処理は、図4(c)の輝度絶対量による例と同様であるため、説明は省略する。さらに、輝度情報と距離情報を併用する場合には、図7(b)のように伝達部を2つ並列に用意すればよい。

30

#### 【0076】

なお、距離情報を利用する場合には、画像情報とは異なる勾配特徴を追加することも可能である。距離の勾配を特徴に組み込むことにより、写真と立体物の区別が容易にできるようになる。その場合は、図9(b)のように、特徴抽出部をもう1つ並列に並べる構成となる。この場合、学習時の変換部追加ステップS2400では、変換部650と伝達部660だけでなく、距離情報用の特徴抽出部710を追加することになる。距離情報用特徴抽出部710では、距離マップが入力層711に入力され、最終層715における出力信号は、分類部620への入力信号として与えられる。第2の学習ステップS2500では、学習画像に対する誤差信号が逆伝搬され、分類部620、画像用および距離情報用の特徴抽出部610と710、変換部650、画像用および距離情報用の伝達部660と670の重み係数が、学習によって更新される。

40

#### 【0077】

以上のように、本実施形態では、輝度値情報または距離情報の少なくとも一方をニューラルネットワークの中間層に入力して識別精度の向上を図ることができる。特に、CNNの出力層に向けて薄れがちな絶対値による情報が、特徴抽出の中間層に入れ込まれることによって、色や明るさが重要な情報である対象物体の識別に対して識別精度の向上が期待できる。また、像面位相差などによって得ることのできる距離情報も、同様な方法にて識

50

別器に利用することができ、さらに識別精度の向上が期待できる。

【0078】

[第3の実施形態]

次に、本発明の第3の実施形態について説明する。第1の実施形態では、入力画像のシーン特徴によって、入力画像全体に対して同じ画像変換を行う方法について説明をした。これは通常の現像方法を変換部で近似しつつ識別精度を向上させるための現像方法を学習によって得ることを意味する。ここで、さらに識別精度を向上させるために、領域によって異なる現像を行ってもよい。本実施形態では、領域ごとに変換部の変換パラメータを修正する形態について説明する。なお、第1、第2の実施形態で既に説明をした構成については同一の符号を付し、その説明は省略する。

10

【0079】

図1(b)は、本実施形態に係る画像処理装置の機能構成を示す概略ブロック図である。まず、学習時の学習装置として機能する際の装置構成について説明する。第1の学習データ取得部2100、第1の学習部2200、第2の学習データ取得部2300、変換部追加部2400、および第2の学習部2500は、第1の実施と同様であるため、説明を省略する。また、第1の学習データ記憶部5100、第1の識別器記憶部5200、第2の学習データ記憶部5300、および第2の識別器記憶部5400についても、第1の実施形態と同様であるため、説明は省略する。

【0080】

調整部追加部2600は、一次識別結果により変換部を調整する調整部を、識別器に追加する。第3の学習部2700は、第2の学習部2500によって学習された変換部と識別器を使って第2の学習データに対して識別処理を行い、そこで発生する誤差をもとに、調整部を学習する。学習された調整部のパラメータは、調整部記憶部5500に記憶される。

20

【0081】

続いて、識別時の画像識別装置として機能する際の装置構成について説明する。入力データ取得部1100、識別器設定部1200、および識別部1300は、第1の実施形態と同様であるため、説明を省略する。調整部設定部1500は、調整部記憶部5500から調整部のパラメータを読み込み、変換部を調整する調整部を設定する。再識別部1600は、識別部1300における識別結果と調整部を使って、識別器による識別処理を再度行う。得られた再識別結果は識別結果出力部1400に送られ、ユーザもしくは別機器に結果が提示される。

30

【0082】

次に、図2(c)を用いて、本実施形態に係る学習処理について説明する。なお、第1の学習データ取得ステップS2100から第2の学習ステップS2500までは、第1の実施形態と同様の処理のため、説明を省略する。

【0083】

調整部追加ステップS2600では、調整部追加部2600が、変換部による現像処理を、一次識別結果をもとに調整する調整部を追加する。図7は、各実施形態に係る識別器の構成を説明する図であり、本実施形態に係る構成例を図7(a)に示す。

40

【0084】

まず、本ステップでは、第2の学習ステップS2500までに学習された識別器に対して、学習画像を入力することで、クラス識別結果が素子621にて得られる。得られた識別結果から、第2の学習画像の中で、GTにおけるクラスラベルとの誤差が小さいもの、例えば誤差0.2以下の領域を学習データから除外する。残った学習データに関して、特徴抽出部610の最終層615における全チャンネルに関する出力信号ベクトル691を抽出する。なお、ここではベクトル691を最終層615における出力信号と説明したが、他の層の値を用いてもよい。例えば、入力層611の値を使ってもいいし、すべての層611から615の値を連結して用いてもよい。これらの値は調整部690に入力され、そこから出力される信号は変換部650へと与えられ、画像変換処理が調整される。

50

## 【 0 0 8 5 】

ここで、図 8 を用いて、本実施形態に係る調整部 6 9 0 の構成を説明する。特徴抽出部 6 1 0 における最終層 6 1 5 の出力信号と、出力層 6 4 0 における各クラスの出力信号は、連結されて特徴量 6 9 6 とされる。特徴量 6 9 6 は、重みベクトル  $b_3$  による重みづけがなされた上で、素子 6 9 3 に入力される。特徴量 6 9 6 は、また、重みベクトル  $b_4$  による重みづけがなされ、素子 6 5 4 に入力される。このような構造によって得られた調整値は、第 1 の実施形態と同様にしてガンマ補正関数の近似である。素子 6 5 4 への入力信号は、変換部による変換関数と調整部による変換関数の和となっていることから、2 つのガンマ補正関数の組み合わせによる現像処理を近似していることになる。変換部と調整部の双方によって補正された素子 6 5 4 の出力信号  $y$  は、下記の数式 1 3 のように記述される。

10

## 【 0 0 8 6 】

$$y = w_2 \tanh(w_1 x + f(z_1)) + b_2 \cdot g(z_2) + w_3 \tanh(b_3 \cdot z_3) + b_4 \cdot z_3 \quad (\text{数式 1 3})$$

ここで、 $z_3$  は、6 9 6 で示される畳み込み層における最終層の出力信号と、出力層における各クラスの出力信号を結合した特徴ベクトルであり、 $w_3$ 、 $b_3$ 、および  $b_4$  は重み係数である。右辺第 2 項は、調整部からの入力であり、ここでは調整項と呼ぶことにする。

## 【 0 0 8 7 】

次に、第 3 の学習ステップ S 2 7 0 0 では、第 3 の学習部 2 7 0 0 が、これら調整部 6 9 0 の重み係数を学習する。第 3 の学習部 2 7 0 0 は、調整部 6 9 0 に関する重み係数  $w_3$ 、 $b_3$ 、 $b_4$  以外のすべての重み係数の学習係数を 0 にして固定し、 $w_3$ 、 $b_3$ 、 $b_4$  についてのみ誤差逆伝搬して修正する。学習して得られた上記パラメータは、調整部記憶部 5 5 0 0 に記憶される。

20

## 【 0 0 8 8 】

ここで行われる学習は、特徴抽出部 6 1 0 における出力信号と、分類部におけるクラス識別信号を、入力特徴量として学習しているため、間違いパターンを学習していると解釈することができる。つまり、クラス識別結果と CNN の内部状態がどのようなときに間違いが発生し、そのときにどのような画像変換を行えば誤差が減少するかが学習されている。例えば、一次識別の時点で白い領域が飛び過ぎてしまい、識別結果が正しくなかった場合、本実施形態では、類似した間違いパターンのときに、その誤差を減らすために、輝度の明るい部分ではコントラストが強くなるように調整部の変換が修正される。

30

## 【 0 0 8 9 】

次に、図 2 ( d ) を用いて、本実施形態に係る識別処理について説明する。まず、第 1 の実施形態と同様の処理にて入力データ取得ステップ S 1 1 0 0 から識別ステップ S 1 3 0 0 を行い、一次識別結果を得る。

## 【 0 0 9 0 】

次に、調整部設定ステップ S 1 5 0 0 では、調整部設定部 1 5 0 0 が、調整部記憶部 5 5 0 0 から調整部の重み係数を読み込み、調整部 6 9 0 が設定される。

## 【 0 0 9 1 】

再識別ステップ S 1 6 0 0 では、再識別部 1 6 0 0 が、一次識別結果のクラス識別信号と、特徴抽出部 6 1 0 における出力信号を連結した特徴ベクトル 6 9 6 を調整部 6 9 0 に入力することで、調整項の追加された画像変換関数が領域ごとに得られる。入力画像は、調整項による修正を加えた画像変換を介して、領域ごとに調整された画像変換が行われ、その変換結果が特徴抽出 6 1 0 に入力される。これにより、分類部 6 2 0 を介して、出力層 6 4 0 にて再識別結果が得られる。なお、ここでは図示しないが、再識別ステップ S 1 6 0 0 の結果をさらに用いて、調整項を使って繰り返し再識別を行ってもよい。その場合、適当な繰り返し数で打ち切るか、もしくは調整項の信号の変化が小さくなった時点で計算を打ち切るなどすればよい。

40

## 【 0 0 9 2 】

50



識別結果出力ステップ S 1 4 0 0 は、第 1 の実施形態と同様の処理であるため、説明は省略する。

【 0 0 9 3 】

本実施形態では、このようにして、まず得られた一次識別結果をもとに、畳み込み層における出力信号とクラス識別信号を組み合わせた識別結果を反映した特徴量を用いて画像変換処理を調整させる。その調整の方法は、学習データによって誤差を縮小する方向に学習されているため、一次識別結果よりも精度よく識別されることが期待できる。

【 0 0 9 4 】

[ その他の実施形態 ]

上記の各実施形態では、CNNによる識別は、特徴抽出部 6 1 0 の最終層 6 1 5 と、分類部 6 2 0 接続した形式で説明を行った。しかし、被写体の細かいテクスチャが有効な特徴量である場合、最終層からの信号だけでは識別に不十分な場合もある。例えば、モルタルによる白壁と、曇天によるテクスチャのない空などを区別する場合などは、細かいテクスチャは重要な情報である。そのような場合には、特徴抽出部 6 1 0 のすべての層から信号を取り出すことで、それを分類部 6 2 0 に渡す方法もある。これはハイパーカラム構造と呼ばれ、非特許文献 4 などにも挙げられている公知の手法である。

10

【 0 0 9 5 】

ハイパーカラム構造を、これまで説明した構成に対して同様に採用しても、上記各実施形態と同様な処理を行うことができる。図 7 ( c ) に、ハイパーカラム構造の CNN を示す。6 8 1、6 8 2、6 8 3、...、6 8 5 は、出力層 6 4 0 の画素 6 2 1 の位置における、特徴抽出部 6 1 0 における各層 6 1 1、6 1 2、6 1 3、...、6 1 5 における出力信号である。これらの信号値は特徴ベクトルとして扱われ、分類部 6 2 0 へと入力される。

20

【 0 0 9 6 】

図 4 ( b ) ( c ) および図 5 ( d ) ( b ) に関しても、同様な構造を入れ込むことができる。構造が上記のようなハイパーカラム構造になっても、学習処理や識別処理に関しては同様のアルゴリズムで可能であるため、詳細な説明は省略する。

【 0 0 9 7 】

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 以上のプロセッサがプログラムを読み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。

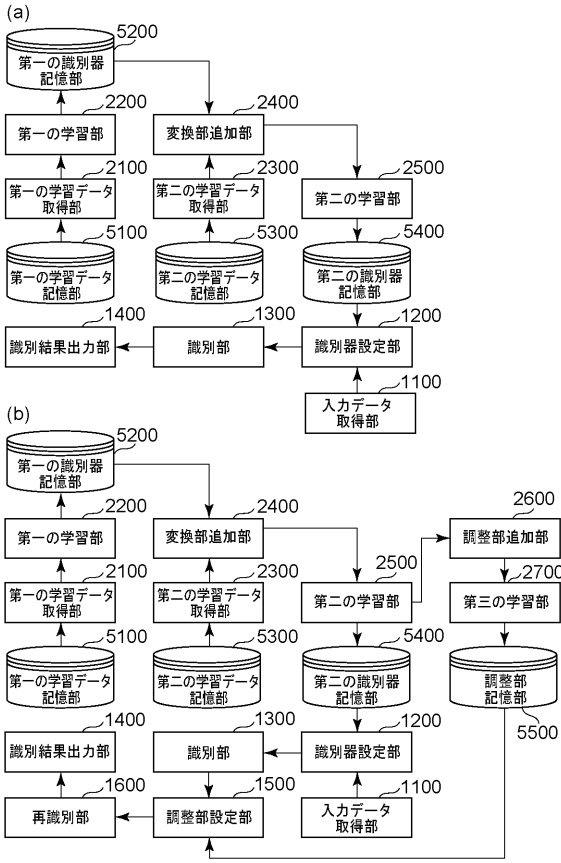
30

【 符号の説明 】

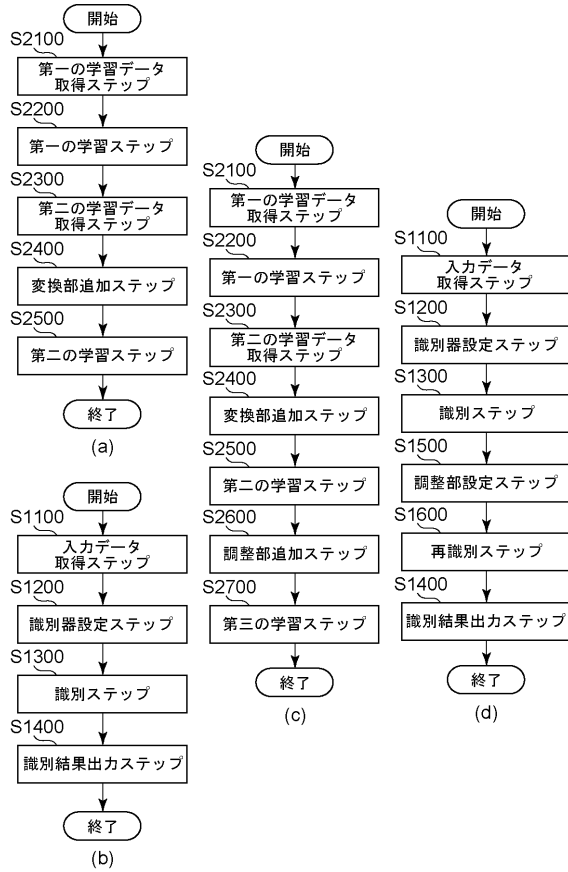
【 0 0 9 8 】

- 1 1 0 0 入力データ取得部
- 1 2 0 0 識別器設定部
- 1 3 0 0 識別部
- 1 4 0 0 識別結果出力部

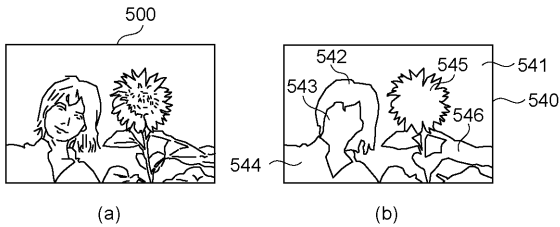
【図1】



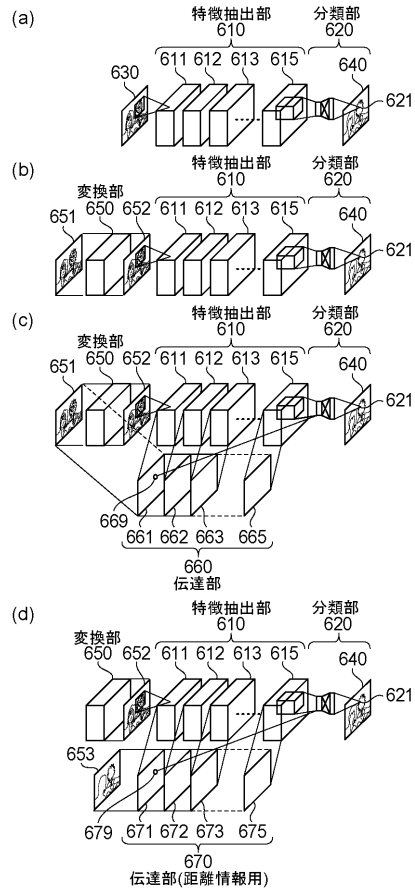
【図2】



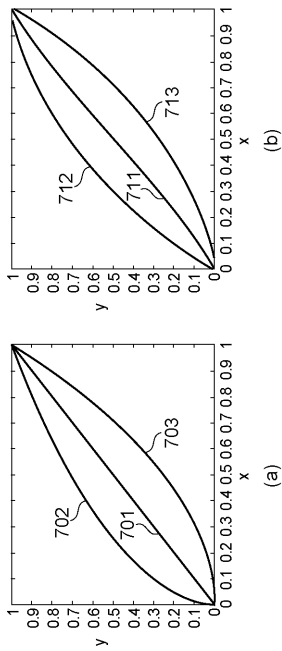
【図3】



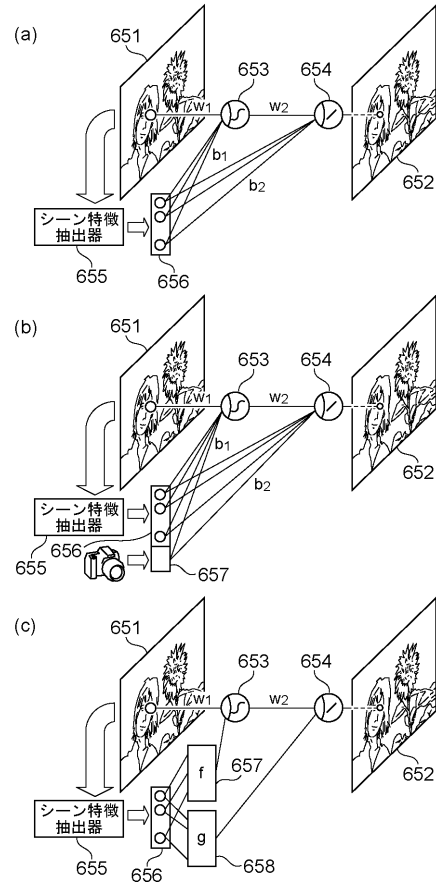
【図4】



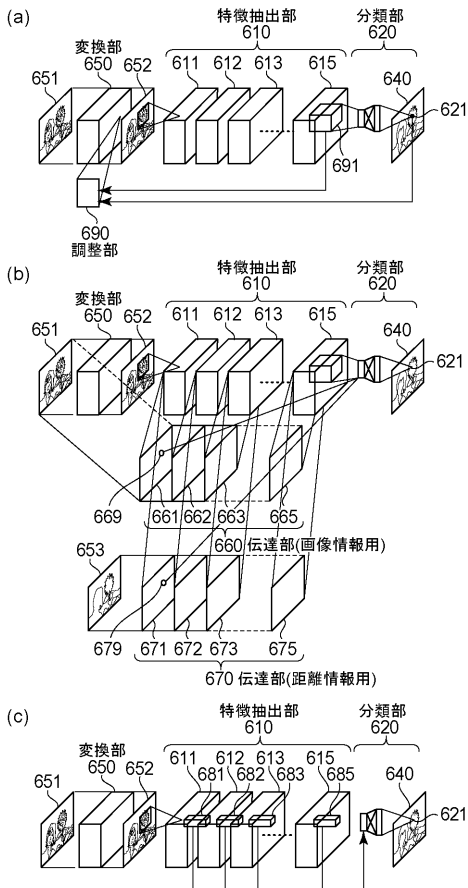
【 図 5 】



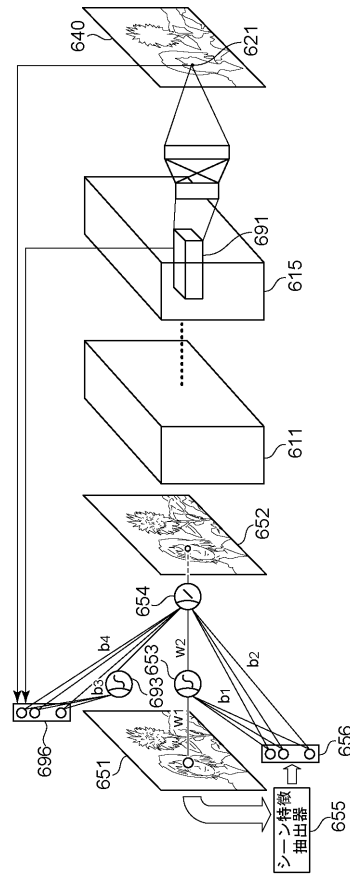
【 図 6 】



【 図 7 】



【 図 8 】



【 図 9 】

