

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6332937号
(P6332937)

(45) 発行日 平成30年5月30日(2018.5.30)

(24) 登録日 平成30年5月11日(2018.5.11)

(51) Int. Cl.			F I		
G06T	7/00	(2017.01)	G06T	7/00	350B
G06T	7/60	(2017.01)	G06T	7/60	110
H04N	5/232	(2006.01)	H04N	5/232	

請求項の数 15 (全 24 頁)

(21) 出願番号	特願2013-220439 (P2013-220439)
(22) 出願日	平成25年10月23日(2013.10.23)
(65) 公開番号	特開2015-82245 (P2015-82245A)
(43) 公開日	平成27年4月27日(2015.4.27)
審査請求日	平成28年9月28日(2016.9.28)

(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	100090273 弁理士 園分 孝悦
(72) 発明者	猿田 貴之 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	佐藤 卓馬

最終頁に続く

(54) 【発明の名称】 画像処理装置、画像処理方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

被写体の撮影画像が撮影された際の撮影情報を、前記撮影画像に含まれる複数の領域それぞれで取得する取得手段と、

前記取得手段で取得された前記撮影情報に基づいて、前記被写体の撮影画像の所定の領域毎に、被写体の分類を示す複数のクラスのうち少なくとも一つのクラスについて、存在確率を推定する推定手段と、

前記推定手段で推定された前記存在確率と、前記撮影画像に係る画像情報とに基づいて、前記所定の領域毎に前記クラスを識別する識別手段と、
を有することを特徴とする画像処理装置。

【請求項2】

前記取得手段は、前記撮影情報と共に前記撮影画像を取得し、

前記推定手段は、前記取得手段で取得された前記撮影情報に基づいて、前記取得手段で取得された前記撮影画像の所定の領域毎に、前記存在確率を推定することを特徴とする請求項1に記載の画像処理装置。

【請求項3】

前記推定手段は、前記所定の領域を前記撮影画像の画素として、前記撮影情報に基づいて前記所定の領域毎に前記存在確率を推定することを特徴とする請求項1又は2に記載の画像処理装置。

【請求項4】

10

20

前記撮影画像を、少なくとも1つ以上の画素で構成される領域毎に分割する分割手段を更に有し、

前記推定手段は、前記所定の領域を前記分割手段で分割された領域として、前記撮影情報に基づいて前記所定の領域毎に前記存在確率を推定することを特徴とする請求項1から3までのいずれか1項に記載の画像処理装置。

【請求項5】

前記推定手段は、前記撮影情報と、前記存在確率との関係に関して予め記憶されている学習情報に基づいて、前記所定の領域毎に前記存在確率を推定することを特徴とする請求項1から4までのいずれか1項に記載の画像処理装置。

【請求項6】

前記所定の領域の間の特徴の類似度を推定する類似度推定手段を更に有し、

前記識別手段は、前記存在確率と、前記画像情報と、前記類似度推定手段で推定された前記類似度を示す類似度情報とに基づいて、前記所定の領域毎に前記クラスを識別することを特徴とする請求項1から5までのいずれか1項に記載の画像処理装置。

【請求項7】

前記所定の領域の間の特徴の類似度を推定する類似度推定手段と、

前記類似度推定手段で推定された類似度を示す類似度情報に基づいて、前記撮影画像における複数の前記所定の領域を統合する統合手段と、を更に有し、

前記推定手段は、前記撮影情報に基づいて、前記統合手段で統合された領域毎に前記存在確率を推定することを特徴とする請求項1から6までのいずれか1項に記載の画像処理装置。

【請求項8】

前記類似度推定手段は、前記撮影情報と、前記画像情報とのうち少なくとも何れか1つに基づいて、前記所定の領域の間の特徴の類似度を推定することを特徴とする請求項7に記載の画像処理装置。

【請求項9】

撮影に係るシーンを推定するシーン推定手段を更に有し、

前記推定手段は、前記撮影情報と、前記シーン推定手段で推定された前記シーンに係るシーン情報とに基づいて前記所定の領域毎に前記存在確率を推定することを特徴とする請求項1から8までのいずれか1項に記載の画像処理装置。

【請求項10】

前記シーン推定手段は、前記撮影情報と、前記画像情報とのうち少なくとも何れか1つに基づいて、前記シーンを推定することを特徴とする請求項9に記載の画像処理装置。

【請求項11】

前記撮影画像から予め設定された対象物を検出する検出手段を更に有し、

前記推定手段は、前記撮影情報と、前記検出手段で検出された前記対象物に係る検出情報とに基づいて前記所定の領域毎に前記存在確率を推定することを特徴とする請求項1から8までのいずれか1項に記載の画像処理装置。

【請求項12】

前記検出手段は、前記撮影画像から人物の人体及び顔を前記対象物として検出することを特徴とする請求項11に記載の画像処理装置。

【請求項13】

前記取得手段は、撮影部で撮影された前記撮影画像と、前記撮影部が前記被写体を撮影した際に取得した前記撮影情報とを取得することを特徴とする請求項1から12までのいずれか1項に記載の画像処理装置。

【請求項14】

被写体の撮影画像が撮影された際の撮影情報を、前記撮影画像に含まれる複数の領域それぞれで取得するステップと、

前記取得された前記撮影情報に基づいて、前記被写体の撮影画像の所定の領域毎に、被写体の分類を示す複数のクラスのうち少なくとも1つのクラスについて、存在確率を推定

10

20

30

40

50

するステップと、

前記推定された前記存在確率と、前記撮影画像に係る画像情報とに基づいて、前記所定の領域毎に前記クラスを識別するステップと、
を有することを特徴とする画像処理方法。

【請求項 15】

コンピュータを、請求項 1 から 13 までのいずれか 1 項に記載の画像処理装置の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

10

本発明は、画像処理装置、画像処理方法及びプログラムに関する。

【背景技術】

【0002】

画像認識手法のひとつとして、撮影して得られた画像を複数の領域に分割して、分割した領域毎に被写体の分類に関するクラスを識別する手法がある。この手法は、各領域から抽出される特徴量に基づいて、各領域のクラスを識別する。適切に画像を領域分割することは、被写体の認識やシーンの認識、被写体に応じた画質の補正等の多くの画像処理を容易にする。

非特許文献 1 に開示されている技術は、色情報、テクスチャ情報に基づいて入力画像をスーパーピクセルと呼ばれる小領域に分割する。そして、この技術は、分割した各小領域のクラスを Recursive - Neural - Networks (RNNs) と呼ばれる識別器を用いて識別する。

20

非特許文献 2 に開示されている技術は、条件付き確率場 CRF (Conditional - Random - Field) を用いて、領域分割及びクラス識別を同時に行う。この技術は、各画素から抽出される特徴に基づくだけでなく、隣接する画素間のクラス共起まで考慮した上で各画素のクラスを識別する。この技術は、特徴が曖昧なため単体では認識の難しい画素に対して、周辺の画素との関係を考慮して識別する。より具体的にいうと、この技術は、各画素をノードとして、ノードのエネルギー (Unary Potential) 及びノード間のエネルギー (Pairwise Potential) を定義して、それらの画像全体の総和を最小化する。そして、この技術は、エネルギーを最小化する各画素のクラスを識別する。

30

【0003】

上記の 2 つの技術では、領域分割及びクラス分類に使用する情報を画像 (特徴量) から得ているが、画像から得ることができる情報だけではなく、撮影した際に得ることができる画像以外の情報を用いて領域分割を行う技術もある。

特許文献 1 に開示されている技術は、GPS による場所情報を用い、地域に応じて画像中の領域 (SP) の被写体のクラスの推定スコアを変える。例えば、赤道直下なら "SNOW" はない」等である。また、この技術は、撮影時期、撮影した方角等も情報として使用したり、被写体のクラス間の空間的配置関係の共起のテーブルを場所情報によって変えたりしている。

40

特許文献 2 に開示されている技術は、画像を撮影する際に被写体距離情報も取得し、被写体距離情報に基づいて画像を領域分割する。そして、この技術は、分割領域毎にシーン判定を行い、その判定結果に基づいて分割領域毎に画像処理を実行する。

特許文献 3 に開示されている技術は、画像セグメンテーション等による任意のオブジェクト抽出結果と、撮影の際に得ることができる距離情報とに基づいて前景領域、背景領域を分割する。例えば、この技術は、所定のオブジェクトの距離よりも近い距離の画素を前景領域とし、それ以外を背景領域として分割する。

【先行技術文献】

【特許文献】

【0004】

50

【特許文献1】米国特許第7860320号明細書

【特許文献2】特開2012-4716号公報

【特許文献3】特開2011-253354号公報

【非特許文献】

【0005】

【非特許文献1】R. Socher, "Parsing Natural Scenes and Natural Language with Recursive Neural Networks", International Conference on Machine Learning 2011.

【非特許文献2】P. Krahenbuhl, "Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials", Neural Information Processing Systems 2011.

【非特許文献3】P. Felzenszwalb, "Efficient Graph-Based Image Segmentation", International Journal of Computer Vision 2004.

【非特許文献4】P. Felzenszwalb, "A Discriminatively Trained, Multiscale, Deformable Part Model", Computer Vision and Pattern Recognition 2008.

【発明の概要】

【発明が解決しようとする課題】

【0006】

上述した従来の技術は、各領域中から抽出された特徴量を入力とし、予め学習された識別器を用いて画像中の各領域のクラスを識別する。しかしながら、これらの技術では、各領域のクラスを精度よく識別することができなかった。

本発明は、画像の領域毎に被写体の分類に関するクラスを精度よく識別することを目的とする。

【課題を解決するための手段】

【0007】

そこで、本発明の画像処理装置は、被写体の撮影画像が撮影された際の撮影情報を、前記撮影画像に含まれる複数の領域それぞれで取得する取得手段と、前記取得手段で取得された前記撮影情報に基づいて、前記被写体の撮影画像の所定の領域毎に、被写体の分類を示す複数のクラスのうち少なくとも1つのクラスについて、存在確率を推定する推定手段と、前記推定手段で推定された前記存在確率と、前記撮影画像に係る画像情報とに基づいて、前記所定の領域毎に前記クラスを識別する識別手段と、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、画像の領域毎に被写体の分類に関するクラスを精度よく識別することができる。

【図面の簡単な説明】

【0009】

【図1】システム構成の一例を示す図である。

【図2】識別対象画像の一例を示す図(その1)である。

【図3】識別対象画像の一例を示す図(その2)である。

【図4】画像認識装置のハードウェア構成の一例を示す図である。

【図5】実施形態1の画像認識装置の機能構成等の一例を示す図である。

【図6】実施形態1の処理の一例を示すフローチャート(その1)である。

【図7】識別対象画像の一例を示す図(その3)である。

【図8】実施形態1における学習装置の機能構成等の一例を示す図である。

10

20

30

40

50

- 【図 9】実施形態 1 の処理の一例を示すフローチャート（その 2）である。
- 【図 10】識別対象画像の一例を示す図（その 4）である。
- 【図 11】識別対象画像の一例を示す図（その 5）である。
- 【図 12】実施形態 2 の画像認識装置の機能構成等の一例を示す図である。
- 【図 13】実施形態 2 の処理の一例を示すフローチャートである。
- 【図 14】実施形態 3 の画像認識装置の機能構成等の一例を示す図（その 1）である。
- 【図 15】実施形態 3 の処理の一例を示すフローチャート（その 1）である。
- 【図 16】ノード間の結合の一例を示す図である。
- 【図 17】実施形態 3 の画像認識装置の機能構成等の一例を示す図（その 2）である。
- 【図 18】実施形態 3 の処理の一例を示すフローチャート（その 2）である。
- 【図 19】実施形態 4 の画像認識装置の機能構成等の一例を示す図である。
- 【図 20】実施形態 4 の処理の一例を示すフローチャートである。
- 【図 21】識別対象画像の一例を示す図（その 6）である。
- 【図 22】実施形態 5 の画像認識装置の機能構成等の一例を示す図（その 1）である。
- 【図 23】実施形態 5 の処理の一例を示すフローチャート（その 1）である。
- 【図 24】実施形態 5 の画像認識装置の機能構成等の一例を示す図（その 2）である。
- 【図 25】実施形態 5 の処理の一例を示すフローチャート（その 2）である。
- 【図 26】物体検出の一例を示す図である。
- 【発明を実施するための形態】

【0010】

以下、本発明を実施するための最良の形態について図面を用いて説明する。

<実施形態 1>

図 1 は、システム構成等の一例を示す図である。

図 1 のシステムでは、カメラ 10 と、画像認識装置 20（画像処理装置）とが、ネットワークを介して接続されている。なお、画像認識装置 20 と、カメラ 10 とは、一体となってもよい。

カメラ 10 は、画像認識装置 20 による画像処理の対象となるシーン 30 を撮影する。

画像認識装置 20 は、カメラ 10 で撮像（撮影）されたシーン 30 の各領域におけるクラスを識別する。より具体的にいうと、画像認識装置 20 は、カメラ 10 で撮像された識別対象画像（撮影画像）と、カメラ 10（撮影部）で前記識別対象画像が撮影された際に取得された撮影情報とに基づいて、前記識別対象画像の領域分割及びクラス識別を行う。ここでいうクラスとは、被写体の分類に関するクラスカテゴリー名である。クラスの詳細については、図 2 等を用いて後述する。また、撮影情報とは、カメラ 10 で撮影した際に取得される画像以外の情報のことであり、画像を出力するまでにカメラ 10 で取得される全ての情報を意味する。例えば、撮影情報は、フォーカスを合わせる際等に取得される距離情報や、シャッタースピード等の撮影におけるカメラパラメータを決定するための色温度及び測光値に関する情報、それらによって決定されたカメラパラメータ等の情報である。その他にも、撮影情報には、撮影日時情報、GPS 情報、カメラ内の姿勢センサーによる天地判定に関する情報等がある。

【0011】

図 2 は、識別対象画像の一例を示す図である。

本実施形態では、図 2 の（a）に示されるように撮影された識別対象画像 100 の各画素のクラスを識別する例について説明する。

ここでいうクラスとは、図 2 の（b）に示されるような各画素に対応して振られる sky、tree、car 等といった被写体の分類に関するクラスカテゴリー名である。

図 3 は、識別対象画像における各画素の識別に関する概念の一例を示す図である。

各画素 103 にクラスが割り当てられることで、領域分割が実現される。図 3 の（b）は、図 3 の（a）の左上部分を拡大して示しており、各画素 103 が sky カテゴリーに割り当てられている様子を示している。

【0012】

図4は、画像認識装置20のハードウェア構成の一例を示す図である。

CPU401は、画像認識装置20全体を制御する。CPU401がROM403やHD404等に格納されたプログラムを実行することにより、後述する画像認識装置20の機能構成及び画像認識装置20に係るフローチャートの処理が実現される。

RAM402は、CPU401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である。

ROM403は、CPU401が実行するプログラム等を格納する記憶領域である。

HD404は、CPU401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である。

操作部405は、ユーザによる入力操作を受け付ける。

表示部406は、画像認識装置20の情報を表示する。

ネットワークI/F407は、画像認識装置20と、外部の機器とを接続する。

【0013】

図5は、本実施形態における画像認識装置20の機能構成等の一例を示す図である。

撮影部500は、カメラ10に相当し、識別対象画像を取得する。

本実施形態の画像認識装置20は、入力部501、存在確率分布推定部502、識別部504を有する。更に、画像認識装置20は、記憶部として存在確率分布推定用辞書保持部503、認識用辞書保持部505を有している。なお、存在確率分布推定用辞書保持部503及び認識用辞書保持部505は、不揮発性の記憶装置として計算機20と接続された構成としてもよい。画像認識装置20が有するこれらの各機能の詳細については、図6等を用いて後述する。

【0014】

図6は、本実施形態における処理の一例を示すフローチャートである。

S110で、入力部501は、撮影部500によって撮影された識別対象画像及びその際に得られた撮影情報を入力データとして受信する。

S120で、存在確率分布推定部502は、S110において入力された撮影情報に基づいて、存在確率分布推定用辞書保持部503に保持されている辞書(学習情報)を用いて識別対象画像の所定の位置毎に各クラスの存在確率分布を推定する。なお、存在確率分布は、各クラスの存在度合の分布に係る分布情報の一例である。存在確率分布推定部502は、推定した存在確率分布に関する情報(以下、単に存在確率分布という)を識別部504に送信する。辞書の詳細については、図8等を用いて後述する。

S130で、識別部504は、S110で入力された識別対象画像の各画素のクラスを識別する。識別部504は、前記識別の際に、認識用辞書保持部505に保持されている識別器と、S120で推定された存在確率分布とを用いて識別を行う。

【0015】

次に、図6に示したフローチャートに従って、各処理のより具体的な処理について述べる。

S110で、入力部501は、図1で示したような対象のシーン30を捉えた識別対象画像及びその際に得られた撮影情報を撮影部500から取得する。この識別対象画像及び撮影情報は、予め取得されて外部装置に記憶されていてもよい。この場合、入力部501は、前記識別対象画像及び前記撮影情報を前記外部装置から取得する。

ここで、撮影情報の詳細について説明する。本実施形態では、撮影情報を Ph_k とおく。kは撮影情報を表すインデックスである。撮影情報の一例として、撮影部500がフォーカスを合わせる際に得ることができる距離情報について説明する。

図7は、識別対象画像40に対して被写体距離を得る測距点42の一例を示す図である。

測距点42は、測距枠41の各分割領域の中心点となる。同一分割領域内の中心点以外の領域では距離情報を得ることができないので、前記中心点以外の領域における距離は、中心点の距離と同じ値であるものとする。また、前記中心点以外の領域における距離は、線形補間等の補間方法で補間した値としてもよい。このようにして、識別対象画像40の

10

20

30

40

50

各画素の位置に対応した被写体距離を得ることができる。ここでは、各画素の位置に対応した被写体距離を $Ph_1(x, y)$ とおく。撮影部 500 は、その他の撮影情報 $Ph_k(x, y)$ も同様に計算しておく。なお、撮影部 500 は、色温度、測光値等のシーンに対応して一意に決定される情報に関して、全ての画素の位置で同じ値となるように設定しておく。本実施形態では撮影部 500 がフォーカスを合わせる際に得ることができる距離情報について説明したが、撮像センサーの他に、測距用のセンサーを取り付けて画像撮影の際に距離情報を得るようにしてもよい。

【0016】

図6の説明に戻る。

S120で、存在確率分布推定部502は、S110において入力された撮影情報 $Ph_k(x, y)$ に基づいて、識別対象画像の所定の位置毎に各クラスの存在確率分布を推定する。本実施形態では、各画素の位置に対応した各クラスの存在確率分布を $P_c(x, y)$ とおく。ここで、 c はクラスカテゴリーを表すインデックスである。

本実施形態では、式1のように撮影情報を入力として、出力を各クラスの存在確率とする設定関数 g_c をクラス毎に予め学習しておき、記憶領域に記憶しておく。設定関数 g_c の学習方法に関しては後述する。なお、本実施形態では撮影情報を入力とする設定関数を用いる例について説明するが、GPSや撮影日時等の撮影情報毎に設定関数を記憶領域に記憶しておいてもよい。そして、存在確率分布推定部502は、それらの撮影情報に対応する設定関数を選択して存在確率分布を推定するようにしてもよい。

各画素の位置に対応した各クラスの存在確率 $P_c(x, y)$ は以下の式1によって推定される。

【数1】

$$P_c(x, y) = g_c(Ph_1(x, y), Ph_2(x, y), \dots, Ph_k(x, y)) \quad \text{式1}$$

また、識別対象画像から各画素において得ることができる輝度値や特徴量等の情報を $I(x, y)$ とおき、式2として識別対象画像から得ることができる情報を追加するようにしてもよい。

【数2】

$$P_c(x, y) = g_c(Ph_1(x, y), Ph_2(x, y), \dots, Ph_k(x, y), I(x, y)) \quad \text{式2}$$

【0017】

また、本実施形態で用いる撮影情報は式3のようにある撮影情報の差分等であってもよい。

【数3】

$$Ph_l(x, y) = Ph_{l1}(x + \alpha, y + \beta) - Ph_{l1}(x, y) \quad \text{式3}$$

式3において、 α 、 β は予め定められた定数である。これにより、 α 、 β で定義された距離離れた画素との撮影情報の類似性を見ることができる。例えば、式4のようにある定数 Y を定義するようにしてもよい。

【数4】

$$Ph_l(x, y) = Ph_{l1}(x, Y) - Ph_{l1}(x, y) \quad \text{式4}$$

例えば、撮影情報を距離情報とし、 $Y = 0$ とすれば、 $Y = 0$ の位置と、画像の上部との距離の差分となる。そのため、画像上部の画素のクラスカテゴリーが sky カテゴリーである場合、 sky カテゴリーでない場合と比べて、画像認識装置20が参照している画素が sky カテゴリーであれば距離の差分値が小さくなる確率が高くなる。一方、画像上部の画素のクラスカテゴリーが sky カテゴリーでない場合、 sky カテゴリーである場合と比べて、距離の差分値が大きくなる確率が高くなる。即ち、画像認識装置20は、クラ

10

20

30

40

50

スカテゴリによっては距離値の差分によって同一カテゴリか否かを推定することができるようになる。

画像認識装置 20 は、この設定関数 g_c を用いて存在確率を計算することによって、複数の撮影情報を反映したカテゴリの存在確率を推定することができる。例えば、色温度、測光値が他の画素より高く、被写体距離が他の画素より遠い画素の $s k y$ カテゴリの存在確率は高くなる。その他、他の領域に比べて著しく暗い領域は、逆光領域である可能性が高くなるため、 $s k y$ カテゴリではなく人体等の主被写体である確率が高くなる。

【0018】

S130 で、識別部 504 は、S110 において入力された識別対象画像の各画素のクラスを識別する。識別部 504 は、S120 で推定された存在確率分布と、認識用辞書保持部 505 に保持されている識別器とを用いて各画素のクラスを識別する。識別器は画像情報を用いて学習されている。より具体的な学習方法、画像情報については後述する。

本実施形態では、式 5 のように画像情報を入力として、出力を各クラスの事後確率とする識別器を用いる。ここで、各画素の位置に対応する画像情報、即ち、特徴量を $F(I, x, y)$ 、事後確率を $P_{c|F=F(I, x, y)}$ とおく。I は識別対象画像である。また、画像情報を入力とし、出力を尤度とする関数(識別器)を $L_c(F(I, x, y))$ とおけば、式 5 のようになる。

【数 5】

$$P_{c|F=F(I, x, y)}(x, y) = L_c(F(I, x, y)) \cdot P_c(x, y) \quad \text{式 5}$$

識別部 504 は、識別器によって求められた尤度と、存在確率とから各画素の事後確率を計算し、式 6 にあるように各画素で最も高い事後確率のクラスを割り当てることで、領域分割及びクラス識別を実現する。

【数 6】

$$\hat{c}(x, y) = \arg \max_{c \in S_c} P_{c|F=F(I, x, y)}(x, y) \quad \text{式 6}$$

式 6 において、 S_c はクラスの集合を示していて、クラス数を N とする。識別部 504 は、識別対象画像の全画素に対して同様の処理を行う。

【0019】

次に、本実施形態で使用する辞書を学習するフローについて説明する。

図 8 は、本実施形態における学習装置 300 の機能構成の一例を示す図である。なお、学習装置 300 のハードウェア構成は、図 4 と同様とする。ここでは、図 5 の画像認識装置 20 とは別に学習装置 300 が構成されているものとするが、画像認識装置 20 内に学習装置 300 が構成されていてもよい。即ち、CPU 401 が ROM 403 や HD 404 等に格納されたプログラムを実行することにより、学習装置 300 の機能構成及び学習装置 300 に係るフローチャートの処理が実現されるようにしてもよい。

学習装置 300 は、学習データ設定部 301、存在確率分布推定部 302、認識用辞書学習部 303 を有する。更に、学習装置 300 は、記憶部として学習用データ保持部 304、存在確率分布推定用辞書保持部 503、認識用辞書保持部 505 を有している。学習装置 300 が有するこれらの各機能の詳細については、図 9 等を用いて後述する。

【0020】

図 9 は、本実施形態における学習に関する処理の一例を示すフローチャートである。

S910 で、学習データ設定部 301 は、学習に用いる学習データを設定する。本実施形態では学習データとして、図 10 に示されるような画像 50 と、その各画素のクラス名が定義されているグラントゥルース 51 とがあればよい。また、学習データの各画素のカメラ情報が必要となる。S910 で、学習データ設定部 301 は、学習用データ保持部 304 に保持されている学習用データの中で学習に使用するデータを設定する。

S920 で、存在確率分布推定部 302 は、S910 において設定された学習データを用いて存在確率分布を推定する設定関数 g_c を学習する。存在確率分布推定部 302 は、設定関数 g_c を回帰学習によって得ることができる。回帰学習の手法としては、SVR (

10

20

30

40

50

Support Vector Regression)や回帰木(Regression Tree)等、様々な手法があるが、何れの手法を用いてもよい。また、存在確率分布推定部302は、設定関数 g_c をテーブルとして設定してもよい。

【0021】

回帰学習で学習する設定関数 g_c では、式1に示すように、入力情報はカメラ情報となり、出力情報は入力されたクラスに対する存在確率となる。本実施形態における学習は学習データの各画素の撮影情報 $Ph_k(x, y)$ に対して各クラスカテゴリーである確率 $P_c(x, y)$ を推定すればよく、学習装置300は、各画素の撮影情報と、その画素のクラスカテゴリーとを学習データから抽出する。

図7の画像を例により具体的に説明する。図7のように各測距点の距離値が得られている場合、距離値が無限遠の場合は距離値が無限遠でない場合と比べてskyカテゴリーの存在確率が高くなる。また、距離値が他の測距点に比べて近い場合は、遠い場合と比べてbodyカテゴリーの存在確率が高くなる。それ以外の値の場合、skyカテゴリー以外のカテゴリーの存在確率が同程度となる。学習装置300は、以上のような存在確率を撮影情報から推定する設定関数を学習する。

【0022】

S930で、認識用辞書学習部303は、入力情報を画像情報(特徴量)として、識別対象画像の各画素のクラス尤度を出力する識別器を学習する。学習データとしては、S920と同様に、図10に示されるような各画素に対して一意にクラス名を与えたデータが使用される。

識別器の学習手法としては、SVM(Support Vector Machine)やRandomized Trees等の手法を用いればよい。識別器に用いる画像情報(特徴量)としては、各画素の色特徴やテクスチャ特徴等を用いればよい。例えば、RGB、HSV、Lab、YCbCr色空間の各成分、Gabor filter、LOGのフィルタ応答を用いるとする。色特徴は、4(色空間) \times 3(成分)の12次元となる。また、フィルタ応答に関してはGabor filter、LOGフィルタの数に対応した次元数となる。識別器の学習の際、学習装置300は、クラス毎に学習データの中から正事例と、負事例とを選択して、その正事例、負事例を識別する識別器を学習する。

【0023】

以上、本実施形態によれば、画像認識装置20は、撮影情報に基づいて各クラスの存在確率分布を推定することにより、識別対象画像の各領域のクラスを精度よく識別することができるようになる。これにより、画像認識装置20による領域分割精度が向上する。

【0024】

<実施形態2>

本実施形態における画像認識装置20は、実施形態1とは異なり、上述したクラスの識別を画素毎に行うのではなく、識別対象画像を予め小領域に分割しておき、前記分割した小領域毎にクラスの識別を行う。画像認識装置20は、予め小領域に分割してから識別を行うことで、局所的なクラスカテゴリーの誤認識を防ぐことができる。そのために、画像認識装置20は、撮影情報又はその撮影情報を用いて推定する各クラスの存在確率分布を小領域毎に推定する。例えば、画像認識装置20は、撮影情報により距離情報が得られている場合、小領域毎の距離の変化を見ることができ、大きな変化を捉えることができる。より具体的にいうと、画像認識装置20は、図11に示すように、撮影された識別対象画像100の小領域101のクラスを識別する。図11は、識別対象画像の一例を示す図である。ここで、小領域とは、画像中における1画素以上であって所定値以下の画素で構成される領域を意味する。小領域の作成方法(分割方法)については、図13等を用いて後述する。

図12は、本実施形態における画像認識装置20の機能構成等の一例を示す図である。

本実施形態における画像認識装置20は、図5に示した機能構成に加えて領域分割部506を有している。領域分割部506の詳細な説明については、図13等を用いて後述する。その他の構成については、図5と同様であるため説明を省略する。

【 0 0 2 5 】

図 1 3 は、本実施形態における画像認識装置 2 0 が領域分割をして各領域のクラスを識別する処理の一例を示すフローチャートである。

S 2 1 0 の処理は、S 1 1 0 の処理と同様であるため説明を省略する。

S 2 2 0 で、領域分割部 5 0 6 は、S 1 1 0 において入力された識別対象画像を小領域に分割する。例えば、領域分割部 5 0 6 は、非特許文献 3 に記載されているような S P (スーパーピクセル) と呼ばれる小領域に分割する。そして、領域分割部 5 0 6 は、領域分割の結果を識別部 5 0 4 に送信する。

S 2 3 0 で、存在確率分布推定部 5 0 2 は、S 1 2 0 の処理とほぼ同様の処理を行うが、S 2 2 0 で分割された小領域毎にクラスを推定する。より具体的にいうと、存在確率分布推定部 5 0 2 は、領域内の各画素に対して、式 7 のように存在確率を算出して領域内の画素数で平均化すればよい。式 7 の $P_c(x, y)$ は、実施形態 1 と同様に式 1 又は式 2 から算出される。R は小領域を示しており、i はある 1 つの小領域を表すインデックスである。また、E は平均を示す。

【 数 7 】

$$P_c(R_i) = E \left(\sum_{(x,y) \in R_i} P_c(x,y) \right) \quad \text{式 7}$$

【 0 0 2 6 】

また、存在確率分布推定部 5 0 2 は、存在確率ではなく撮影情報を小領域毎に平均してもよい。その場合、存在確率分布推定部 5 0 2 は、式 8 のように各撮影情報を平均化する。

【 数 8 】

$$P_h(R_i) = E \left(\sum_{(x,y) \in R_i} P_h(x,y) \right) \quad \text{式 8}$$

そして、各クラスの存在確率は式 9 のように推定される。

【 数 9 】

$$P_c(R_i) = g_c(P_h_1(R_i), P_h_2(R_i), \dots, P_h_k(R_i)) \quad \text{式 9}$$

これにより、存在確率分布推定部 5 0 2 は、実施形態 1 と異なり撮影情報を小領域内で平均化するため、ノイズロバストとなり大きな変化を捉えることができる。

【 0 0 2 7 】

S 2 3 0 で、識別部 5 0 4 は、各小領域のクラスを識別する。実施形態 1 と同様に、識別器として、SVM や Randomized Trees 等を用いればよい。また、識別器に用いる特徴量は各領域から抽出される。例えば、実施形態 1 と同様に、RGB、HSV、Lab、YCbCr 色空間の各成分、Gabor filter、LoG のフィルタ応答を用いるとする。また、学習装置 3 0 0 は、領域毎に特徴づけを行うため、各小領域内の画素毎に得ることができる特徴量から統計量を求める。用いる統計量は、平均、標準偏差、歪度、尖度の 4 つとする。歪度は分布の非対称性の度合いを示す統計量であり、尖度は分布が平均の近くに密集している度合いを示す統計量である。したがって、色特徴は 4 (色空間) × 3 (成分) × 4 (統計量) の 48 次元となり、テクスチャ特徴の次元数は (フィルタ応答数) × 4 (統計量) となる。また、この他に小領域の重心座標や小領域の面積等を特徴量としてもよい。

識別部 5 0 4 は、識別器によって求められた尤度と、各小領域の存在確率とから、式 1 0 にあるように各小領域で最も高い事後確率のクラスを割り当てることで、各小領域のクラス識別を実現する。

10

20

30

40

【数 1 0】

$$\hat{c}(R_i) = \arg \max_{c \in S_c} P_{c|F=F(I,R_i)}(R_i) \quad \text{式 1 0}$$

【0 0 2 8】

識別部 5 0 4 は、識別対象画像の全小領域に対して同様の処理を行う。また、ここでは、識別部 5 0 4 が領域毎にクラス識別を行うようにしたが、実施形態 1 と同様に画素毎に識別を行って、小領域内で平均するようにしてもよい。その場合はまず、識別部 5 0 4 は、式 1 1 のように小領域内でクラス毎に事後確率の和を計算する。

【数 1 1】

$$P_c(R_i) = \sum_{(x,y) \in R_i} P_{c|F=F(I,x,y)}(x,y) \quad \text{式 1 1}$$

10

次に、識別部 5 0 4 は、式 1 2 のように各領域において最大の事後確率をもつクラスを割り当てる。

【数 1 2】

$$\hat{c}(R_i) = \arg \max_{c \in N} P_c(R_i) \quad \text{式 1 2}$$

【0 0 2 9】

以上、本実施形態によれば、画像認識装置 2 0 は、識別対象画像を予め小領域に分割してから撮影情報に基づいた存在確率分布を推定する。そして、画像認識装置 2 0 は、前記推定した存在確率分布を利用してクラス識別を行うことで、より精度よくクラス識別を行うことができる。これにより、画像認識装置 2 0 による領域分割精度が向上する。

20

【0 0 3 0】

<実施形態 3 >

本実施形態における画像認識装置 2 0 は、撮影情報を用いて各クラスの存在確率分布を推定するだけでなく、画像情報及び撮影情報の少なくとも 1 つを用いて各画素間の特徴量の類似度を求め、クラスの識別の際に前記類似度を示す類似度情報を利用する。これにより、画像認識装置 2 0 は、更に領域分割及びクラス識別精度を向上させることができる。

画像認識装置 2 0 は、類似度を利用することにより、対象とする画素に対して予め定められた近傍の画素の情報も考慮して識別を行うことができるため、領域分割及びクラス識別精度が向上する。本実施形態では実施形態 1 と同様に画素単位で識別を行う手法に関して説明を行うが、実施形態 2 と同様に予め領域分割された小領域毎に識別するようにしてもよい。

30

図 1 4 は、本実施形態における画像認識装置 2 0 の機能構成等の一例を示す図である。

本実施形態における画像認識装置 2 0 は、図 5 に示した機能構成に加えて類似度推定部 5 0 7 を有している。類似度推定部 5 0 7 の詳細については、図 1 5 等を用いて後述する。その他の構成については、図 5 と同様であるため説明は省略する。

【0 0 3 1】

図 1 5 は、本実施形態における画像認識装置 2 0 が、各画素間の類似度を利用して領域分割及び各領域のクラス識別をする際の処理の一例を示すフローチャートである。

40

S 3 1 0 の処理は、S 1 1 0 の処理と同様であるため説明を省略する。

S 3 2 0 の処理は、S 1 2 0 の処理と同様であるため説明を省略する。

S 3 3 0 で、類似度推定部 5 0 7 は、画像情報及び撮影情報の少なくとも 1 つを用いて各画素間の類似度を算出する。類似度推定部 5 0 7 は、推定した類似度に関する情報（以下、単に類似度という）を識別部 5 0 4 に送信する。S 3 3 0 の処理の詳細については後述する。S 3 3 0 で、類似度推定部 5 0 7 が画像情報のみを利用する場合には、入力部 5 0 1 から画像情報のみが入力され、撮影情報のみを利用する場合には撮影情報のみが入力され、両方を利用する場合には両方の情報が入力される。

【0 0 3 2】

50

S 3 4 0 で、識別部 5 0 4 は、S 3 1 0 で入力された識別対象画像を認識用辞書保持部 5 0 5 に保持されている識別器を用いて各画素のクラスを識別する。その際に、識別部 5 0 4 は、S 3 2 0 で推定された存在確率分布、入力画像から得ることができる画像情報及び S 3 3 0 で推定された画素間の類似度に基づいて、各画素のクラスを識別する。S 3 4 0 の処理の詳細については後述する。

次に、図 1 5 に示したフローチャートに従って、S 3 3 0 の処理及び S 3 4 0 の処理についてより具体的に説明する。

S 3 3 0 で、類似度推定部 5 0 7 は、画像情報及び撮影情報の少なくとも 1 つを用いて各画素間の類似度を算出するが、本実施形態では類似度を Gaussian カーネルで表現する。類似度推定部 5 0 7 が類似度を計算する画素の位置を x_i 、 x_j とおき、それぞれの画素の位置で得ることができる画像情報又は撮影情報（特徴量）を f_i 、 f_j とおき、画素間の類似度を $S(f_i, f_j)$ とおけば、式 1 3 のように定義される。

10

【数 1 3】

$$S(f_i, f_j) = \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|Color_i - Color_j|^2}{2\theta_\beta^2}\right) \quad \text{式 1 3}$$

【0 0 3 3】

ここでは類似度を定義する画像情報（特徴量）として、画素の位置 p 、色成分ベクトル $Color$ を用いたが、撮影情報を用いることもできる。その場合は、S 3 1 0 で入力された撮影情報 $Ph_k(x_i)$ 、 $Ph_k(x_j)$ を利用する。その場合、式 1 4 のように定義することができる。

20

【数 1 4】

$$S(f_i, f_j) = \exp\left(-\frac{|Ph_k(x_i) - Ph_k(x_j)|^2}{2\theta_r^2}\right) \quad \text{式 1 4}$$

また、画像情報及び撮影情報を合わせて式 1 5 のように表すこともできる。

【数 1 5】

$$S(f_i, f_j) = \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|Color_i - Color_j|^2}{2\theta_\beta^2} - \frac{|Ph_k(x_i) - Ph_k(x_j)|^2}{2\theta_r^2}\right) \quad \text{式 1 5}$$

30

なお、 θ_α 、 θ_β 、 θ_r は、ハイパーパラメータで後述する識別器の学習の際にクロスバリデーションによって求めてもいいし、ユーザが決定してもよい。ここでは撮影情報の中から、1 つを選択して類似度を推定する例について説明したが、更に他の撮影情報を利用する場合はハイパーパラメータを設定して追加すればよい。

【0 0 3 4】

S 3 4 0 で、識別部 5 0 4 は、識別器を用いて各画素のクラスを識別する。本実施形態では非特許文献 2 に開示されているような条件付き確率場（CRF）を用いて説明する。

40

条件付き確率場とは、構造化データのラベリング及びセグメンテーションを扱うことのできる確率的なフレームワークである。条件付き確率場は、要素（ノード）自体のポテンシャル（Unary ポテンシャル）と、要素（ノード）間の関係の度合を表すポテンシャル（Pairwise ポテンシャル）との和の最適化を行う。ここで、観測データを $X = \{x_1, x_2, \dots, x_M\}$ とする。M は全観測データ数であり、本実施形態の場合、画素数に対応する。関連するクラスラベルを $Y = \{y_1, y_2, \dots, y_M\}$ とする。Y_i の取り得る値を $C = \{c_1, c_2, \dots, c_N\}$ とする。N はクラスカテゴリ数を表している。条件付き確率場 (X, Y) は式 1 6 のようにギブス分布で表される。

【数16】

$$P(Y|X) = \frac{1}{Z(X)} \exp(-E(Y)) \quad \text{式16}$$

【0035】

ここで、 $Z(X)$ は分布を正規化するための分配関数を表し、 $E(Y)$ は以下の式17で表される Gibbs エネルギーである。

【数17】

$$E(Y) = \sum_i \psi_u(y_i) + \sum_{i < j} \psi_p(y_i, y_j) \quad \text{式17}$$

10

式17で、 $\psi_u(y_i)$ は、各ノードのポテンシャル (Unary ポテンシャル) を表し、 $\psi_p(y_i, y_j)$ は、ノード間の関係度のポテンシャル (Pairwise ポテンシャル) を表す。Unary ポテンシャルは各ノードに割り当てられ、本実施形態においては各画素に割り当てられる。Unary ポテンシャルは以下の式18で定義する。

【数18】

$$\psi_u(y_i) = -\log(\text{prob}(y_i)) \quad \text{式18}$$

$\text{prob}(y_i)$ は i 番目の画素がクラスカテゴリ y_i である確率を表す。確率値は実施形態1と同様に SVM を用いて識別した際の尤度に存在確率を乗算した値を、シグモイド関数を用いて変換することで近似的に求められる値である。

20

【0036】

次に、Pairwise ポテンシャルについて説明する。条件付き確率場では図16に示すようにノード間の結合を定義する必要がある。図16は、ノード間の結合の一例を示す図である。図16では、各ノード60と、ノード間の結合61とが示されている。図16の(b)では隣接ノードのみが結合しており、図16の(c)では全ノードが結合している。本実施形態では図16の(c)に示されるような全結合条件付き確率場を利用する。Pairwise ポテンシャルは以下の式19のように表される。

【数19】

$$\psi_p(y_i, y_j) = \mu(y_i, y_j) \cdot k(f_i, f_j) \quad \text{式19}$$

30

ここで、 $\mu(y_i, y_j)$ はラベル間の適合性を表す関数で、例えば式20の Potts モデルを用いる。

【数20】

$$\mu(y_i, y_j) = [y_i \neq y_j] \quad \text{式20}$$

【0037】

この場合は、類似度の高いノードに異なるラベルを与えると、ペナルティが与えられる。本実施形態においては、 $\mu(y_i, y_j)$ としてクラス共起行列を学習する。これは、あるクラスカテゴリ同士が一緒に存在しやすい又は存在しにくいという画像中の物体の共起という情報を利用する方法である。例えば、人と車、車と自転車、人と馬等と一緒に存在しやすいが、馬とバス、羊とバイクと一緒に存在しにくい等である。学習に際しては、例えば (クラス数) \times (クラス数) の行列を用意して、学習画像において同画像中に共起するクラスカテゴリのセルに投票していけばよい。クラス共起行列 μ のクラスカテゴリ C_i と、クラスカテゴリ C_j の共起確率 $\mu(C_i, C_j)$ とは、式21のようになる。

40

【数21】

$$\mu(C_i, C_j) = \frac{(\text{共起画像の数}) + \gamma}{(\text{全学習画像数}) + \delta} \quad \text{式21}$$

ここで、 γ は共起確率を正規化するための正の定数であり、学習の際に予め決定して

50

おく。

次に、 $k(f_i, f_j)$ は特徴量で定義される関数であり、式 1 3 から式 1 5 までで定義した類似度を用いて、式 2 2 のように表される。

【数 2 2】

$$k(f_i, f_j) = \omega_1 S(f_i, f_j) + \omega_2 \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\epsilon^2}\right) \quad \text{式 2 2}$$

【0038】

式 2 2 の第 2 項は Smoothness kernel である。式 2 2 では画素の位置のみを考慮したが、撮影情報を考慮してもよい。 10

撮影情報を考慮する場合の $k(f_i, f_j)$ は式 2 3 のように表せる。

【数 2 3】

$$k(f_i, f_j) = \omega_1 S(f_i, f_j) + \omega_2 \exp\left(-\frac{|Ph_k(x_i) - Ph_k(x_j)|^2}{2\theta_\epsilon^2}\right) \quad \text{式 2 3}$$

ω_1 、 ω_2 は各カーネルの重みであり、 θ_ϵ はハイパーパラメータである。これらは、先のと同様に学習の際にクロスバリデーションによって求めればよい。これらの式によって定義された条件付き確率場に対して式 2 4 で表される MAP 推定が行われ、各ノードのクラスが決定される。 20

【数 2 4】

$$Y = \arg \max_{Y \in L^N} P(Y | X) \quad \text{式 2 4}$$

上式において、 $P(Y | X)$ について厳密解を求めるためには、全てのクラスカテゴリーの組み合わせを L^N 通り計算する必要がある。しかし、非特許文献 2 に開示されている平均場近似と、高次元フィルタリングとを組み合わせた手法を用いれば、画素数 M に対して線形のオーダーまで計算コストをおさえることができる。

以上の処理により、識別部 5 0 4 は、各画素をノードとした条件付き確率場を定義して MAP 推定を行うことで識別対象画像の全画素のクラスを識別し、領域分割することができる。 30

【0039】

本実施形態の処理に関して条件付き確率場を用いて説明したが、非特許文献 1 にあるような Recursive-Neural-Networks (RNNs) 等の別の識別器を用いてもよい。その場合は推定された類似度を識別器の特徴量の 1 つとして利用すればよい。このように、本実施形態における画像認識装置 2 0 は、撮影情報に基づいて各クラスの存在確率分布を推定し、更に各画素間の類似度を利用することでより精度よく識別対象画像の各画素のクラスを識別することができ、領域分割精度を向上させることができる。

また、本実施形態ではノードを画素単位に設定して説明を行ったが、予め領域分割を行い、各領域をノードとして条件付き確率場を用いて各領域のクラスを識別するようにしてもよい。その場合、図 1 7 に示されているような機能構成を有する画像認識装置 2 0 を用いる。図 1 7 は、本実施形態における画像認識装置 2 0 の機能構成等の一例を示す図である。図 1 7 に示される画像認識装置 2 0 の機能構成は、図 1 4 に示される機能構成に加えて、更に領域分割部 5 0 6 を有している。 40

【0040】

図 1 8 は、図 1 7 の機能構成を有する画像認識装置 2 0 が実行する処理の一例を示すフローチャートである。

図 1 8 における各処理については上述したため詳細な説明は省略するが、S 4 4 0 で、類似度推定部 5 0 7 は、画素間の類似度を求めるのではなく、領域間の類似度を求める。 50

ここでは、各領域の撮影情報から得ることができる特徴量のヒストグラムの類似度を基に類似度推定部507が領域間の類似度を求める手法について説明する。

それぞれの領域を R_i 、 R_j 、領域間の類似度を $S(R_i, R_j)$ とおく。そして、領域 R_i の撮影情報 $Ph_k(x, y)$ (ここでの x, y は領域 R_i に含まれる画素を示している)のヒストグラムを $hist(R_i, l)$ とおき、ピン数を L 、あるピンを表すインデックスを l ($l = 1, 2, \dots, L$)とおく。その場合、領域間の類似度 $S(R_i, R_j)$ は以下の式25のように表される。

【数25】

$$S(R_i, R_j) = \frac{\sum_{l=1}^L \min(hist(R_i, l), hist(R_j, l))}{\sum_{l=1}^L hist(R_i, l)} \quad \text{式25}$$

10

また、類似度を算出する上述以外の方法として、類似度推定部507は、ヒストグラム間の距離を求めて類似度を算出する方法を用いてもよい。また、その他にも、類似度推定部507は、複数の撮影情報を用いてもよいし、画像から得られる特徴量を用いてもよい。

【0041】

以上、本実施形態によれば、画像認識装置20は、各画素間又は予め分割された領域間における特徴量の類似度を利用してクラスを識別することができる。これにより、画像認識装置20は、より精度よくクラスを識別することができ、領域分割制度を向上させることができる。

20

【0042】

<実施形態4>

本実施形態における画像認識装置20は、撮影情報を用いて各クラスの存在確率分布を推定するだけでなく、画像情報及び撮影情報の少なくとも1つを用いて各領域間の類似度を求め、類似度を利用して分割された小領域同士を統合する。そして、画像認識装置20は、統合した統合領域のクラスを、識別器を用いて識別する。

図19は、本実施形態における画像認識装置20の機能構成等の一例を示す図である。

図19に示される画像認識装置20の機能構成は、図5に示される機能構成に加えて、更に、領域分割部506、類似度推定部507、領域統合部508を有している。これらの機能の詳細な説明については、図20等を用いて後述する。画像認識装置20は、類似度を利用して小領域を統合してからクラス識別を行うことで、小領域のみで識別する場合に比べて、ロバスト性を向上させることができる。

30

図20は、画像認識装置20が識別対象画像を領域分割し、領域間の類似度を基に分割領域を統合し、その各統合領域のクラスを識別する処理の一例を示すフローチャートである。

S510の処理は、S210の処理と同様であるため説明を省略する。

S520の処理は、S220の処理と同様であるため説明を省略する。

S530で、類似度推定部507は、式21で定義されるような領域間の類似度を求める。類似度推定部507は、推定した類似度を領域統合部508に送信する。

【0043】

40

S540で、領域統合部508は、S530で推定された領域間の類似度に基づいて、S520で生成(分割)された隣接する小領域を統合する。領域統合部508による領域統合の一例を図21に示す。図21は、領域分割及び領域統合の一例を示す図である。

領域統合部508は、識別対象画像100の小領域101のうち隣接していて、かつ、類似度が予め定められた閾値よりも高い小領域同士を統合する。図21の(c)には、統合された統合領域102が示されている。より具体的にいうと、領域統合部508は、壁や空の領域が小領域に分割されてしまった場合に、小領域中の距離値の平均や分散を利用して統合することができる。領域統合部508は、小領域のうち隣接している領域同士の類似度を算出して、その類似度の値が予め定められた閾値より大きい場合に統合する。領域統合部508は、統合した統合領域の情報(以下、単に統合領域という)を識別部A2

50

30に送信する。

S550で、存在確率分布推定部502は、S540で統合された統合領域の存在確率を推定する。推定方法は実施形態2と同様であるため、説明を省略する。

S560で、識別部504は、S550で推定された存在確率分布を利用して、各統合領域のクラスを識別する。推定方法は実施形態2と同様であるため、説明を省略する。

【0044】

以上、本実施形態によれば、画像認識装置20は、撮影情報によって推定される存在確率分布の他に、類似度を利用して小領域を統合することで、領域分割精度をより向上させることができる。

【0045】

<実施形態5>

本実施形態における画像認識装置20は、撮影情報を用いて各クラスの存在確率分布を推定するだけでなく、シーン推定を行いそのシーン推定結果(シーン情報)も用いて各クラスの存在確率を推定する。ここで、シーンとは、屋外、夜景、風景、ポートレートといった撮影シーンでもよいし、オフィス、雪山、海といった撮影されている場面を表すシーンでもよい。

図22は、本実施形態における画像認識装置20の機能構成等の一例を示す図である。

図22に示される画像認識装置20の機能構成は、図5に示される機能構成に加えて、更に、シーン推定部509、記憶部としてシーン推定用辞書保持部510を有している。これらの機能の詳細な説明については、図23等を用いて後述する。また、画像認識装置20は、物体検出結果も用いて各クラスの存在確率を推定することもできる。画像認識装置20が物体検出、特に顔検出、人体検出の結果を用いて各クラスの存在確率を推定する方法についても後述する。

【0046】

図23は、本実施形態における画像認識装置20がシーン推定結果を利用して、領域分割及び各領域のクラス識別をする際の処理の一例を示すフローチャートである。

S610の処理は、S110の処理と同様であるため説明を省略する。

S620で、シーン推定部509は、画像情報及び撮影情報の少なくとも1つを用いて識別対象画像のシーンを推定する。ここで、識別器を用いたシーン推定方法について説明する。より具体的には、Bag-of-words手法を用いたSVMを利用した手法について説明する。Bag-of-words手法とは、特徴量をVisualWord化して、その各VisualWordが画像中にどれくらい表れるかという頻度ヒストグラムを識別器で学習して特徴量とする手法である。

【0047】

VisualWord化は、例えば、kmeans等によって行われる。VisualWord化する特徴量には、画像情報又は撮影情報を用いればよい。例えば、画像情報なら画像からグリッド状に抽出したSIFT特徴量や、撮影情報なら各測距点で得られた距離値等を利用する。両方の情報を利用する場合には、各情報から得ることができる特徴量を頻度ヒストグラム化して、そのヒストグラムを連結して特徴量として識別器で学習すればよい。また、シーン推定部509は、GPSや撮影日時等の撮影情報を用いる場合はまず、それらの撮影情報によって識別器を選択して、選択した識別器によってシーンを識別してもよい。それによって、シーン推定部509は、GPSや撮影日時に対応したシーンを識別することができる。

識別器の学習データは画像に対して、定義したシーンのうち1つのシーン名が付与されているデータを用意すればよい。シーン推定部509は、認識の際、シーン推定の際に1つのシーンに決定してもよいが、1つのシーンに決定する必要はなく、本実施形態では定義した各シーンの尤度を推定する。シーン数をNSceneとすると、シーン推定部509は、各シーンの尤度を要素とする $1 \times N_{Scene}$ の行列になるシーン尤度行列を出力する。シーン推定部509は、推定結果であるシーン尤度行列の情報(以下、単にシーン尤度行列という)を存在確率分布推定部502に送信する。

10

20

30

40

50

【 0 0 4 8 】

S 6 3 0 で、存在確率分布推定部 5 0 2 は、S 6 2 0 で推定されたシーン推定結果及び撮影情報を利用して各クラスの存在確率分布を推定する。本実施形態では、撮影情報を $P h_k$ とおく。k は撮影情報を表すインデックスである。存在確率分布推定部 5 0 2 は、S 6 1 0 で入力された撮影情報 $P h_k(x, y)$ に基づいて識別対象画像の所定の位置毎に各クラスの存在確率を推定する。本実施形態において、各画素の位置に対応した各クラスの存在確率を $P_c(x, y)$ とおく。ここで、c はクラスカテゴリーを表すインデックスである。

本実施形態では、実施形態 1 と同様に撮影情報を入力として、出力を各クラスの存在確率とする設定関数 g_c を予め学習しておく。そのため、各画素の位置に対応した各クラスの存在確率 $P_c(x, y)$ は、以下の式 2 6 によって推定される。

【数 2 6】

$$P_c(x, y) = g_c(P h_1(x, y), P h_2(x, y), \dots, P h_k(x, y)) * P_c(S_c) \quad \text{式 2 6}$$

【 0 0 4 9 】

ここで、 $P_c(S_c)$ とは、S 6 2 0 で推定されたシーン尤度行列から求められるクラス c の存在確率である。例えば、シーン推定結果において"屋外"の尤度が他に比べて高いとする。その場合、s k y カテゴリーの存在確率が他の存在確率よりも高くなる。そのため、予め学習の際に各シーンの学習画像を構成している各クラスカテゴリーの頻度を集計しておく必要がある。より具体的にいうと、各シーンの学習画像群の各クラスカテゴリーの画素数を集計して頻度を求めておけばよい。各クラスの存在確率 $P_c(x, y)$ は、以下の式 2 7 のように定義してもよい。

【数 2 7】

$$P_c(x, y) = g_c(P h_1(x, y), P h_2(x, y), \dots, P h_k(x, y), P_c(S_c)) \quad \text{式 2 7}$$

本実施形態において $P_c(S_c)$ は、学習された各シーンのクラスカテゴリー c の尤度と、各シーン確率との積和により求められる。学習データとしては実施形態 1 と同様に図 7 のような各画素のクラスカテゴリーが定義されているデータを用意すればよいが、加えて対象画像 5 0 のシーン名を定義しておく必要がある。

以上により、存在確率分布推定部 5 0 2 は、識別対象画像の所定の位置毎に各クラスの存在確率分布を推定することができる。

S 6 4 0 の処理は、S 1 3 0 の処理と同様であるため説明を省略する。

以上の処理により、画像認識装置 2 0 は、シーン推定結果も考慮して推定した各クラスの存在確率を利用することで領域分割精度を向上させることができる。

【 0 0 5 0 】

また、シーン推定部 5 0 9 の他に物体検出器を利用して各クラスの存在確率を推定してもよい。

その場合における画像認識装置 2 0 の機能構成は、図 2 4 のようになる。図 2 4 は、物体検出器を利用して各クラスの存在確率を推定する画像認識装置 2 0 の機能構成等の一例を示す図である。図 2 4 に示される画像認識装置 2 0 の機能構成は、図 2 2 に示される機能構成に加えて、更に、物体検出部 5 1 1、記憶部として物体検出用辞書保持部 5 1 2 を有している。

図 2 5 は、図 2 4 に示される機能構成を有する画像認識装置 2 0 の処理の一例を示すフローチャートである。

S 7 1 0 及び S 7 2 0 の処理は、S 6 1 0 及び S 6 2 0 の処理と同様であるため説明を省略する。また、S 7 5 0 の処理は、S 6 4 0 の処理と同様であるため説明を省略する。

S 7 3 0 で、物体検出部 5 1 1 は、識別対象画像中の物体検出を行う。本実施形態では、図 2 6 の (b)、図 2 6 の (c) に示すような人体検出及び顔検出を利用する例について説明するが、対象物を予め設定して識別器を学習することができるものであれば他のものであってもよい。図 2 6 は、物体検出の一例を示す図である。

10

20

30

40

50

各物体検出器の学習方法については説明を省略するが、例えば、非特許文献4に開示されているようなDeformablePartsModelというモデルを定義した物体検出器を用いればよい。

【0051】

S740で、存在確率分布推定部502は、S720及びS730において推定されたシーン推定結果、物体検出結果（検出情報）及び撮影情報を利用して各クラスの存在確率分布を推定する。存在確率分布推定部502は、設定関数 g_c を利用すれば、各画素の位置に対応した各クラスの存在確率 $P_c(x, y)$ を、以下の式28又は式29によって推定することができる。

【数28】

$$P_c(x, y) = g_c(Ph_1(x, y), Ph_2(x, y), \dots, Ph_k(x, y)) * Pc(Sc) * Pc(O_1) * Pc(O_2) \quad \text{式28}$$

10

【数29】

$$P_c(x, y) = g_c(Ph_1(x, y), Ph_2(x, y), \dots, Ph_k(x, y), Pc(Sc), Pc(O_1), Pc(O_2)) \quad \text{式29}$$

式28又は式29において、 $P_c(O_1)$ 、 $P_c(O_2)$ は、S730において推定された物体検出結果から求められるクラスカテゴリーの存在確率である。

【0052】

本実施形態のように人体検出結果、顔検出結果を利用する場合、例えば、人物クラスカテゴリーの存在確率が他のクラスカテゴリーの存在確率に比べて高くなる。人体検出結果による人物クラスカテゴリーの存在確率を $P_{c=body}(O_{body})$ とおく。また、人物クラスカテゴリー以外のクラスの存在確率を $P_{c \neq body}(O_{body})$ とおく。例えば、図26の(b)のように人体検出結果が矩形領域で抽出される場合は、矩形領域43内の $P_{c=body}(O_{body})$ 及び $P_{c \neq body}(O_{body})$ は、式30、式31のように表される。

20

【数30】

$$P_{c=body}(O_{body}) = \frac{1}{N} + \frac{N-1}{N} \Pr(body) \quad \text{式30}$$

【数31】

$$P_{c \neq body}(O_{body}) = \frac{1 - \Pr(body)}{N-1} \quad \text{式31}$$

30

ここで、 $\Pr(body)$ には、人体検出器によるスコアを用いる。また、矩形領域43以外の $P_{c=body}(O_{body})$ 及び $P_{c \neq body}(O_{body})$ は、式32のように表される。

【数32】

$$P_{c=body}(O_{body}) = P_{c \neq body}(O_{body}) = \frac{1}{N} \quad \text{式32}$$

【0053】

また、顔検出結果による人物クラスカテゴリーの存在確率を $P_{c=body}(O_{face})$ とおく。また、人物クラスカテゴリー以外のクラスの存在確率を $P_{c \neq body}(O_{face})$ とおく。例えば、図26の(c)のように人体検出結果が矩形領域で抽出される場合は、矩形領域44内の $P_{c=body}(O_{face})$ 及び $P_{c \neq body}(O_{face})$ は、式33、式34のように表される。

40

【数33】

$$P_{c=body}(O_{face}) = \frac{1}{N} + \frac{N-1}{N} \Pr(face) \quad \text{式33}$$

【数34】

$$P_{c \neq body}(O_{face}) = \frac{1 - \Pr(face)}{N-1} \quad \text{式34}$$

ここで、 $\Pr(face)$ には、顔検出器によるスコアが用いられる。また、矩形領域

50

4 4 以外の $P_{c=body}(O_{face})$ 及び $P_{c \neq body}(O_{face})$ は、式 3 5 のように表される。

【数 3 5】

$$P_{c=body}(O_{face}) = P_{c \neq body}(O_{face}) = \frac{1}{N} \quad \text{式 3 5}$$

他の物体検出器の結果も同様に計算される。

【0054】

以上、本実施形態によれば、画像認識装置 20 は、撮影情報によって推定される存在確率分布の他に、シーン推定結果及び物体検出結果も考慮して推定した各クラスの存在確率を利用することで領域分割精度を向上させることができる。

10

なお、実施形態 2 と同様に、画像認識装置 20 は、先に領域分割を行ってから、各小領域におけるクラス識別を行うようにしてもよい。また、実施形態 3、実施形態 4 と同様に、画像認識装置 20 は、類似度推定部を有していてもよく、類似度推定部によって推定された類似度を用いて小領域を統合したり、各領域におけるクラスを識別したりしてもよい。

【0055】

< その他の実施形態 >

また、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、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し、そのシステム或いは装置のコンピュータ（又は 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。

20

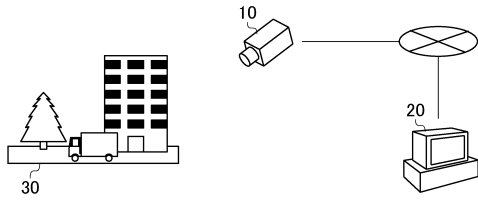
【0056】

以上、上述した各実施形態によれば、画像認識装置 20 は、画像の領域毎に被写体の分類に関するクラスを精度よく識別することができる。

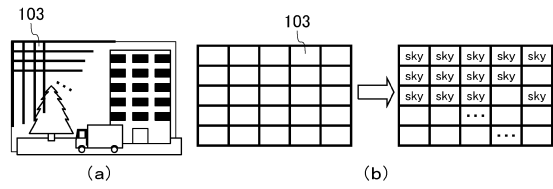
【0057】

以上、本発明の好ましい形態について詳述したが、本実施形態は係る特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

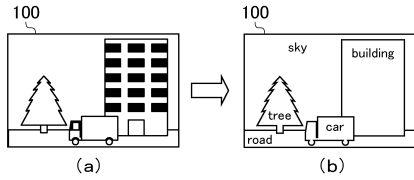
【図1】



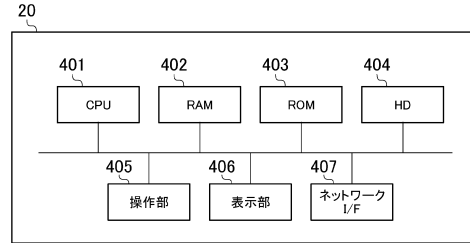
【図3】



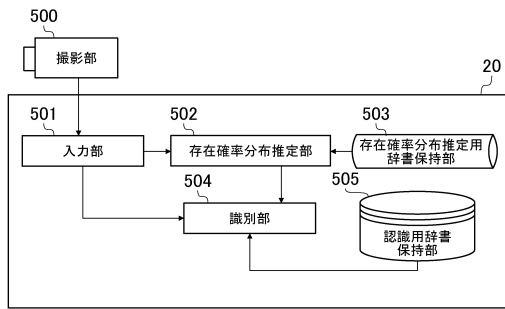
【図2】



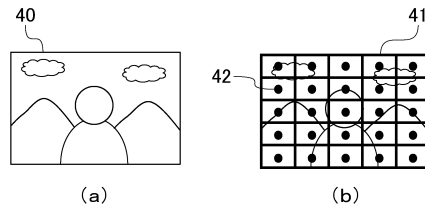
【図4】



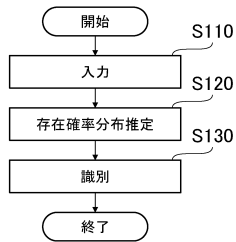
【図5】



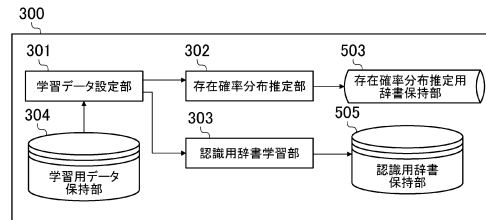
【図7】



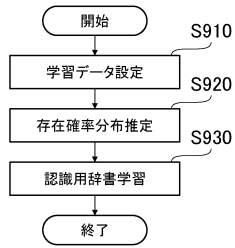
【図6】



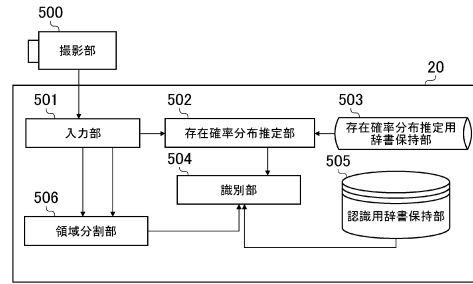
【図8】



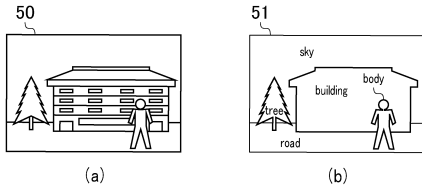
【図9】



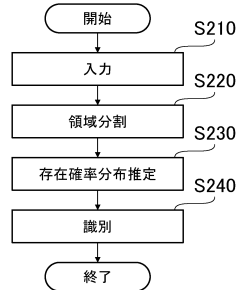
【図12】



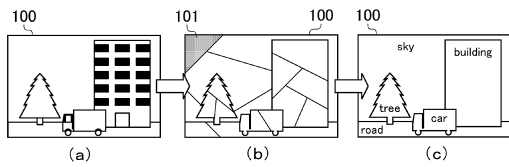
【図10】



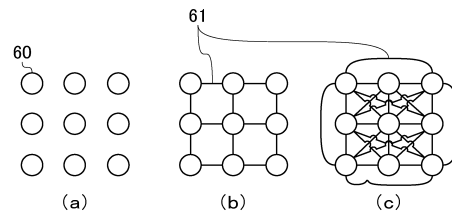
【図13】



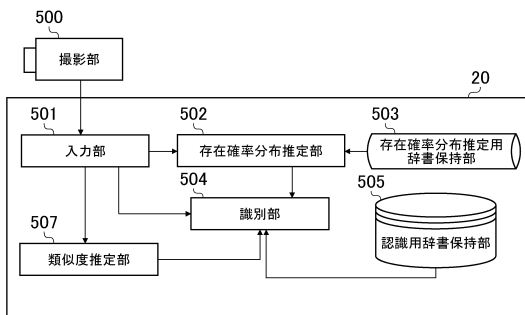
【図11】



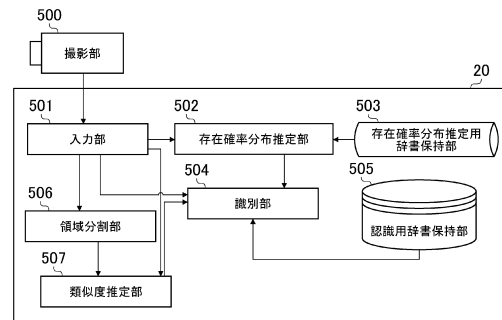
【図16】



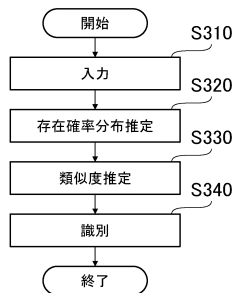
【図14】



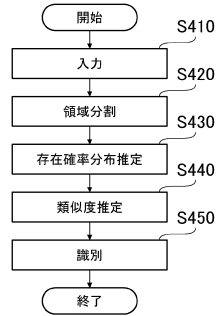
【図17】



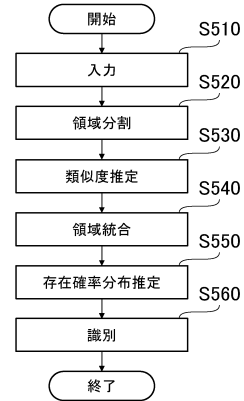
【図15】



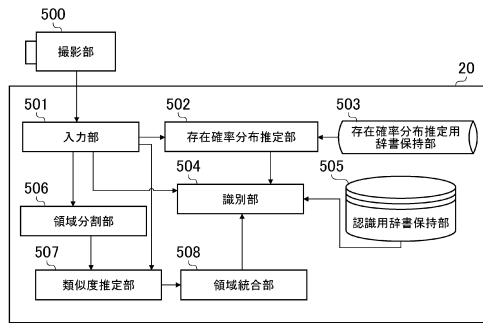
【図18】



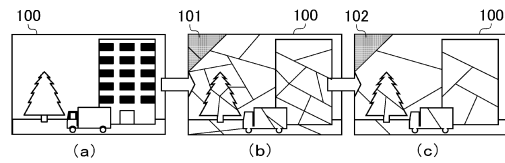
【図20】



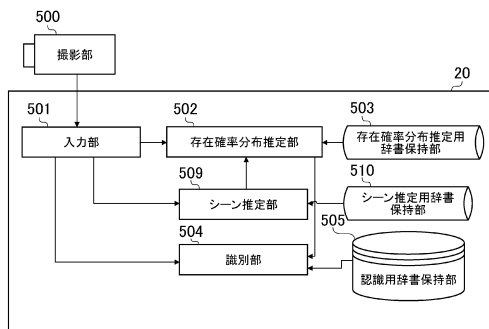
【図19】



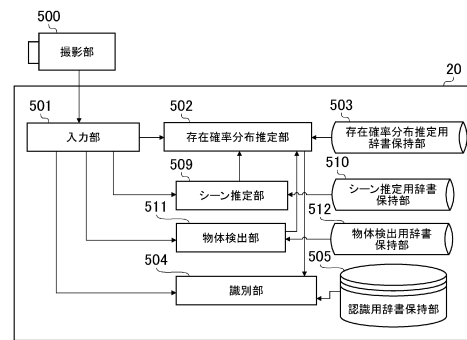
【図21】



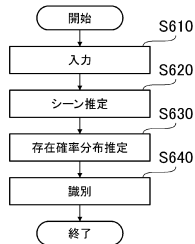
【図22】



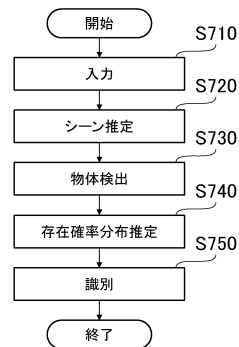
【図24】



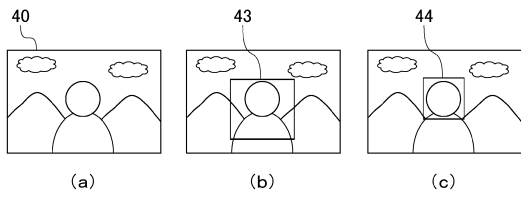
【図23】



【図25】



【 図 26 】



フロントページの続き

(56)参考文献 特表2009-541896(JP,A)
特表2012-507761(JP,A)
特開2002-298138(JP,A)
国際公開第2004/095374(WO,A1)
国際公開第2012/046426(WO,A1)
特開2012-043151(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06T	7/00
G06T	7/60
H04N	5/232