

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6440303号
(P6440303)

(45) 発行日 平成30年12月19日(2018.12.19)

(24) 登録日 平成30年11月30日(2018.11.30)

(51) Int.Cl. F I
G06T 7/77 (2017.01) G O 6 T 7/77
G06T 7/00 (2017.01) G O 6 T 7/00 3 5 0 C

請求項の数 7 (全 16 頁)

(21) 出願番号	特願2014-244417 (P2014-244417)	(73) 特許権者	397065480
(22) 出願日	平成26年12月2日 (2014.12.2)		エヌ・ティ・ティ・コムウェア株式会社
(65) 公開番号	特開2016-110232 (P2016-110232A)		東京都港区港南一丁目9番1号
(43) 公開日	平成28年6月20日 (2016.6.20)	(74) 代理人	100064908
審査請求日	平成29年3月13日 (2017.3.13)		弁理士 志賀 正武
		(74) 代理人	100108578
			弁理士 高橋 詔男
		(74) 代理人	100108453
			弁理士 村山 靖彦
		(72) 発明者	長森 藤江
			東京都港区港南一丁目9番1号 エヌ・テ
			ィ・ティ・コムウェア株式会社内
		(72) 発明者	中江 俊博
			東京都港区港南一丁目9番1号 エヌ・テ
			ィ・ティ・コムウェア株式会社内
			最終頁に続く

(54) 【発明の名称】 対象認識装置、対象認識方法、およびプログラム

(57) 【特許請求の範囲】

【請求項1】

認識対象画像に任意面積の候補領域を複数設定する画像変換部と、
 前記画像変換部により設定された各候補領域に含まれる画像情報に対して認識処理を行
 い、前記各候補領域に対象候補ごとの確信度を計算する認識計算部と、

前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前
 記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度を合計した評
 価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値
 の分布をマップとして生成するマップ生成部と、

前記マップ生成部により作成されたマップに基づいて対象の位置を特定する位置算出部
 と、

を有する、対象認識装置。

【請求項2】

認識対象画像に任意面積の候補領域を複数設定する画像変換部と、
 前記画像変換部により設定された各候補領域に含まれる画像情報に対して認識処理を行
 い、前記各候補領域に対象候補ごとの確信度を計算する認識計算部と、

前記認識対象画像の各位置において、当該位置の候補領域について、認識対象画像に含
 まれる対象候補を表す各クラスIDごとに、前記確信度と、当該候補領域の面積に応じた
 係数とをパラメータとし、前記面積が大きいほど値が大きくなる評価値を算出し、前記ク
 ラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生

10

20

成するマップ生成部と、

前記マップ生成部により作成されたマップに基づいて対象の位置を特定する位置算出部と、

を有する、対象認識装置。

【請求項 3】

認識対象画像に任意面積の候補領域を複数設定する画像変換部と、

前記画像変換部により設定された各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算する認識計算部と、

前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度の合計値と、前記重複する複数の候補領域の面積に応じた係数とをパラメータとし、前記面積が大きいほど値が大きくなる評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成するマップ生成部と、

前記マップ生成部により作成されたマップに基づいて対象の位置を特定する位置算出部と、

を有する、対象認識装置。

【請求項 4】

前記画像変換部は、前記認識対象画像の特徴量を抽出し、当該抽出した特徴量に基づいて、前記認識対象画像内に含まれる特徴の大きさに応じた候補領域を設定し、前記認識対象画像に設定された複数の候補領域のうち、所定の領域面積より小さい領域面積の候補領域を削除する、

請求項 1 から 3 のうち何れか 1 項に記載の対象認識装置。

【請求項 5】

学習画像に基づいて対象を学習する学習処理部を有し、

前記学習処理部は、外部から供給された学習画像の全体を用いて学習結果を生成し、

前記認識計算部は、前記画像変換部により設定された候補領域の画像と、前記学習結果とを用いて対象候補の確信度を計算する、

請求項 1 から 4 のうち何れか 1 項に記載の対象認識装置。

【請求項 6】

認識対象画像に任意面積の候補領域を複数設定するステップと、

前記各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算するステップと、

前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度を合計した評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成するステップと、

前記マップに基づいて対象の位置を特定するステップと、

を有する、対象認識方法。

【請求項 7】

コンピュータに、

認識対象画像に任意面積の候補領域を複数設定させ、

前記各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算させ、

前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度を合計した評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成させ、

前記マップに基づいて対象の位置を特定させる、ためのプログラム。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【0001】

本発明は、対象認識装置、対象認識方法、およびプログラムに関する。

【背景技術】

【0002】

従来、撮像画像に対して画像認識を行って画像内に含まれる対象を認識する技術がある。この対象認識技術においては、近年、ニューラルネットワークを多層化したディープラーニング技術を利用して高い精度で対象を認識することが行われている。このディープラーニング技術は、一般的にはニューラルネットワークの入力層と出力層との間の中間層において複数段階に亘って層を重ねることにより高い精度で対象を認識する。このディープラーニング技術において、特に、畳み込みニューラルネットワーク (Convolutional Neural Network ; CNN) が、従来の画像特徴量に基づいて対象を認識するよりも高い性能を有することで注目されている。

10

【0003】

畳み込みニューラルネットワークによれば、ラベルが付与された認識対象画像を学習し、認識対象画像に含まれる主な対象を認識する。このため、畳み込みニューラルネットワークは、学習画像内に主な対象が複数存在する場合には、手動で主な対象が含まれる領域矩形で指定し、当該指定された領域に対応した画像にラベルを付与して学習を行っていた。

【0004】

この学習処理において、指定された領域に対してラベルを付与する必要があった。その後の学習処理において、指定された領域に対応した画像を畳み込みニューラルネットワークに入力し、畳み込みニューラルネットワークから出力された特徴ベクトルを学習している。

20

また、従来の対象の認識プロセスは、認識対象画像に対してエッジ抽出処理等を実施することにより局所的な特徴に基づいて候補領域を抽出すると共に、候補領域を畳み込みニューラルネットワークに入力して特徴ベクトルを抽出した上で分類を行い、分類された最も確信度が高い候補領域を認識結果として得ていた。確信度とは、ある画像領域とラベルとともに学習された画像の主体の類似度が、他のクラスの類似度より相対的にどの程度高いかを示す量である。

【先行技術文献】

30

【非特許文献】

【0005】

【非特許文献1】 “ Rich feature hierarchies for accurate object detection and semantic segmentation ” , Tech report, Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, UC Berkeley (2014)

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、従来の畳み込みニューラルネットワークを利用した対象認識技術は、学習画像に複数の対象が存在する場合には自動的に所望する一つの対象に絞り込むことができないため、上述したように画像に含まれる対象が含まれる複数の領域を一つずつ手動により指定する必要があった。

40

【0007】

すなわち、従来の畳み込みニューラルネットワークを利用した対象認識技術における学習プロセスは、1つの認識対象画像に対して1つのラベルを付与して学習データを作成する必要があるため、領域指定のための負担が大きい。また、対象の認識精度を高めるための学習プロセスの処理対象となる画像 (以下、教師画像) を畳み込みニューラルネットワークの処理専用で作成して記憶させる必要がある。

【0008】

また、従来の畳み込みニューラルネットワークを利用した対象認識技術における認識プ

50

ロセスは、認識対象が物体などの輪郭が明確なものが多く、教師画像も物体の輪郭等が明確なものであることが多い。このことから、特定の物体の判断が困難な画像に対しては良好な認識精度を発揮することができなかつた。例えば、秋葉原、浅草などの街全体の風景のように、多数の物体が含まれている画像や、特定の物体を判別しにくい画像や、画像内に近距離物体が遠距離物体とが混在している画像や、サイズが大きく異なる物体が混在している画像などを高い認識精度で処理することができなかつた。

【0009】

さらに、従来の畳み込みニューラルネットワークを用いた対象認識技術に限らず、既存の他の対象認識技術においても同様の問題があつた。

【0010】

そこで本発明は、上記の課題に鑑みてなされたものであって、より高い認識精度で画像内の主な対象および当該対象の位置を判定することができる対象認識装置、対象認識方法、およびプログラムを提供することを目的としている。

【課題を解決するための手段】

【0011】

(1) 本発明の一態様は、認識対象画像に任意面積の候補領域を複数設定する画像変換部と、前記画像変換部により設定された各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算する認識計算部と、前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度を合計した評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成するマップ生成部と、前記マップ生成部により作成されたマップに基づいて対象の位置を特定する位置算出部と、を有する、対象認識装置である。

【0012】

(2) 本発明の一態様は、認識対象画像に任意面積の候補領域を複数設定する画像変換部と、前記画像変換部により設定された各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算する認識計算部と、前記認識対象画像の各位置において、当該位置の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度と、当該候補領域の面積に応じた係数とをパラメータとし、前記面積が大きいほど値が大きくなる評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成するマップ生成部と、前記マップ生成部により作成されたマップに基づいて対象の位置を特定する位置算出部と、を有する、対象認識装置である。

【0013】

(3) 本発明の一態様は、認識対象画像に任意面積の候補領域を複数設定する画像変換部と、前記画像変換部により設定された各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算する認識計算部と、前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度の合計値と、前記重複する複数の候補領域の面積に応じた係数とをパラメータとし、前記面積が大きいほど値が大きくなる評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成するマップ生成部と、前記マップ生成部により作成されたマップに基づいて対象の位置を特定する位置算出部と、を有する、対象認識装置である。

【0014】

(4) 本発明の一態様は、上述の対象認識装置であつて、前記画像変換部は、前記認識対象画像の特徴量を抽出し、当該抽出した特徴量に基づいて、前記認識対象画像内に含まれる特徴の大きさに応じた候補領域を設定し、前記認識対象画像に設定された複数の候補領域のうち、所定の領域面積より小さい領域面積の候補領域を削除する。

【0015】

10

20

30

40

50

(5) 本発明の一態様は、上述の対象認識装置であって、学習画像に基づいて対象を学習する学習処理部を有し、前記学習処理部は、外部から供給された学習画像の全体を用いて学習結果を生成し、前記認識計算部は、前記画像変換部により設定された候補領域の画像と、前記学習結果とを用いて対象候補の確信度を計算する。

【0016】

(6) 本発明の一態様は、認識対象画像に任意面積の候補領域を複数設定するステップと、前記各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算するステップと、前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度を合計した評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成するステップと、前記マップに基づいて対象の位置を特定するステップと、を有する、対象認識方法である。

10

【0017】

(7) 本発明の一態様は、コンピュータに、認識対象画像に任意面積の候補領域を複数設定させ、前記各候補領域に含まれる画像情報に対して認識処理を行い、前記各候補領域に対象候補ごとの確信度を計算させ、前記認識対象画像の各位置において、当該位置で重複する複数の候補領域について、前記認識対象画像に含まれる対象候補を表す各クラスIDごとに、前記確信度を合計した評価値を算出し、前記各クラスIDごとに、前記認識対象画像の各位置における前記評価値の分布をマップとして生成させ、前記マップに基づいて対象の位置を特定させる、ためのプログラムである。

20

【発明の効果】

【0018】

本発明の一態様によれば、より高い認識精度で画像内の主な対象および当該対象の位置を判定することができる。

【図面の簡単な説明】

【0019】

【図1】本発明を適用した実施形態の対象認識装置1の機能的な構成を示すブロック図である。

【図2】本発明を適用した実施形態の対象認識装置1の一構成例を示すブロック図である。

30

【図3】タワー状建造物を近距離且つ下方から撮像した学習画像P10を示す図である。

【図4】建造物を遠距離且つ上方から撮像した学習画像P11を示す図である。

【図5】建造物を夜間において撮像した学習画像P12を示す図である。

【図6】本発明の実施形態として示す対象認識装置1における認識処理の手順を示すフローチャートである。

【図7】本発明を適用した対象認識装置1に入力される認識対象画像に設定された候補領域と、候補領域ごとの確信度の一例を示す図である。

【図8】本発明を適用した対象認識装置1における確信度および面積と候補領域との関係を示す図である。

【図9】本発明を適用した対象認識装置1において、マップ生成部16により作成されたマップを例示した図である。

40

【図10】比較例としての認識対象画像P1に設定された候補領域の一例を示す図である。

【図11】比較例としての認識対象画像P1に設定された他の候補領域の一例を示す図である。

【発明を実施するための形態】

【0020】

以下、本発明を適用した対象認識装置、対象認識方法、およびプログラムを、図面を参照して説明する。

まず、本発明を適用した実施形態の対象認識装置1の機能的な説明を行う。

50

図1は、本発明を適用した実施形態の対象認識装置1の機能的な構成を示すブロック図である。対象認識装置1は、認識対象画像に含まれる対象のうち主たる対象を認識し、認識した対象の属性(名称、種類)および位置を出力するものである。

対象認識装置1は、画像変換部12と、認識計算部14と、マップ生成部16と、位置算出部18とを有する。

【0021】

画像変換部12は、外部から認識対象画像が供給される。画像変換部12は、任意の手法により認識対象画像に複数の候補領域を設定し、複数の候補領域はおのおの重なっていてもよい。画像変換部12は、認識対象画像の特徴に基づいて候補領域を設定するものとして説明する。しかし、候補領域の設定手法は、これに限らず、いくつかの大きさの枠をずらしていくなどの手法を採用してもよい。

10

【0022】

認識計算部14は、画像変換部12により設定された各候補領域に含まれる画像情報に対して認識処理を行う。認識計算部14は、学習結果データベース140に記憶された学習結果を参照して、各候補領域に含まれる画像情報が対象候補に類似するほど確信度を高くする。これにより、認識計算部14は、各候補領域に対象候補ごとの確信度を計算する。対象候補は、認識対象画像に含まれる物体、風景等の画像を見る者が認識可能なものであり、認識対象画像に含まれる可能性がある候補となる対象である。また、確信度とは、対象候補の確からしさであり、任意の画像とラベルとともに学習された画像の主体の類似度が、他のクラス(対象候補)の類似度より相対的にどの程度高いかを示す量である。

20

認識計算部14は、候補領域に含まれる画像について複数の対象候補が計算された場合には、各対象候補について確信度を得る。これにより、認識計算部14は、各候補領域について、一又は複数の対象候補と、当該各対象候補の確信度を得る。

【0023】

マップ生成部16は、クラスIDおよび確信度に基づいて、認識対象画像の位置ごとに評価値を算出する。マップ生成部16は、例えば認識対象画像のピクセルごと、あるいは、いくつかのピクセル集合をひとつの位置として見立てた代表位置ごとに評価値を算出する。マップ生成部16は、確信度に対して所定の統計処理を実施し評価値を算出する。この評価値は、認識対象画像に含まれる主な対象がクラスIDに対応する対象であるという評価が高いほど、高い値となる。

30

統計処理は、例えば、認識対象画像の位置ごとに、クラスIDごとに確信度を合計して評価値を算出する。この統計処理によれば、同じクラスIDが算出された候補領域が重複している位置ほど、高い評価値を算出する。

また、統計処理は、候補領域の面積に応じた係数と確信度とを用い、候補領域の面積が大きいほど評価値を高くしてもよい。

さらに統計処理は、クラスIDおよび確信度を合計すると共に候補領域の面積に応じた係数を用いて評価値を算出してもよい。この変換処理によれば、同じクラスIDが算出された候補領域が重複しており、且つ候補領域の面積が大きいほど、高い評価値を算出する。

なお、統計処理は、認識対象画像に含まれる主な対象がクラスIDに対応する対象であるという評価が高くなる評価値が算出できればよく、その他の処理を行ってもよい。

40

これにより、マップ生成部16は、認識対象画像の位置ごとに、クラスIDごとの評価値を含むマップを作成することができる。つまり、クラスIDに対応した対象が存在する可能性が高いほど高い評価値が分布されたマップを生成する。

【0024】

位置算出部18は、マップ生成部16により作成されたマップを参照して、認識対象画像に主に含まれる対象(クラスID)の位置を計算する。位置算出部18は、認識対象画像のうち、クラスIDの評価値が高い領域に、クラスIDに対応した対象が存在すると判定する。

位置算出部18は、例えば、評価値が所定の閾値を超えた認識対象画像の領域を抽出し

50

、抽出された領域にクラスIDに対応した対象が存在すると判定してもよい。また、位置算出部18は、マップに対してエッジ抽出処理を実施し、評価値が段階的に高くなったエッジ線に囲まれる領域を抽出し、抽出された領域にクラスIDに対応した対象が存在すると判定してもよい。

これにより、位置算出部18は、クラスIDと、当該クラスIDに対応した対象の位置情報とを得ることができる。この位置情報は、対象の輪郭位置であってもよく、対象の中心位置であってもよい。

【0025】

以下、図1に示した対象認識装置1の具体的な構成および動作について説明する。

図2は、本発明を適用した実施形態の対象認識装置1の一構成例を示すブロック図である。対象認識装置1は、例えば携帯型通信端末に実装される。対象認識装置1には、画像処理装置200および画像表示部300と接続される。

【0026】

画像処理装置200は、画像生成部202とAR(Augmented Reality)処理部204とを有する。

画像生成部202は、対象認識装置1に供給される認識対象画像を生成する。画像生成部202は、例えばカメラ装置、外部から画像データを受信する通信部、または画像データを記憶する記憶部により構成される。画像処理装置200は、認識対象画像に含まれる対象を認識させる場合に、画像生成部202により生成された認識対象画像を対象認識装置1に供給する。

【0027】

AR処理部204は、画像生成部202により生成された認識対象画像に対して画像処理を行うことにより、AR処理を行う。このAR処理とは、認識対象画像により表される環境に情報を付加、削除、強調または減衰させる処理を含む。AR処理には、例えば、カメラ装置により撮像された認識対象画像に、当該認識対象画像に含まれる対象の説明画像を付加する処理が含まれる。AR処理部204は、認識対象画像に含まれる対象を対象認識装置1に識別させる場合に、画像生成部202から対象認識装置1に認識対象画像を出力させる。

本実施形態において、対象認識装置1は、認識対象画像に含まれるランドマーク画像を認識し、ランドマーク画像についての説明を認識対象画像に付加するAR処理を行うもの

【0028】

なお、本実施形態において、対象認識装置1により対象が認識された結果(ランドマーク)を、AR処理に用いることについて説明するが、対象認識装置1により対象が認識された結果は、他の用途に用いてもよいことは勿論である。

例えば、画像処理装置200は、SNS(social networking service)アプリケーションから出力された要求に応じて、SNSアプリケーションにより表示させる画像に含まれる対象を対象認識装置1に識別させてもよい。SNSアプリケーションは、対象認識装置1により対象が認識された結果(クラスIDおよび位置)に基づき、認識対象画像に予め設定した対象が含まれる場合に、当該認識対象画像から対象を削除または減衰する、他の情報を付加するといった画像処理を実施する。

【0029】

対象認識装置1は、画像変換部12と、認識計算部14と、マップ生成部16と、位置算出部18と、インターフェース部20と、学習画像蓄積部22と、学習処理部24とを有する。

なお、対象認識装置1は、ソフトウェアを実行する制御装置および記憶部を有するコンピュータである。対象認識装置1は、CPU(Central Processing Unit)などのプロセッサが記憶部(不図示)に記憶されているプログラムを実行することによって機能するソフトウェア機能部である。また、対象認識装置1は、LSI(Large Scale Integration)やASIC(Application Specific Integrated Circuit)等のハードウェア機

10

20

30

40

50

能部であってよい。

また、対象認識装置 1 における記憶部は、例えば、HDD (Hard Disc Drive)、フラッシュメモリ、EEPROM (Electrically Erasable Programmable Read Only Memory)、ROM (Read Only Memory)、またはRAM (Random Access Memory) 等により実現される。また、記憶部には、ファームウェアやアプリケーションプログラム等の各種プログラム、各種機能部による処理結果、インターフェース部 20 が外部から取得した情報などが記憶される。

【0030】

インターフェース部 20 は、携帯型通信端末に実装された画像処理装置 200 との間で情報を授受すると共に、対象認識装置 1 における各部との間で情報を授受する。

10

インターフェース部 20 は、画像処理装置 200 から認識対象画像が供給されたことに応じて、当該認識対象画像を処理対象として画像変換部 12 に出力する。インターフェース部 20 は、認識対象画像に対する対象の処理結果としてクラスIDおよび位置が位置算出部 18 から供給され、当該クラスIDおよび位置を画像処理装置 200 に出力する。このクラスIDおよび位置は、認識対象画像に含まれるランドマーク画像の分類を表す情報である。クラスIDとは、認識対象画像に含まれる主な対象を特定するための情報であり、例えば、建造物名、市街地名、風景名等が挙げられる。

【0031】

学習画像蓄積部 22 は、クラスIDが付加された学習画像を蓄積する機能を有する。学習画像蓄積部 22 は、画像変換部 220 と画像データデータベース 222 とを有する。

20

学習画像蓄積部 22 は、学習素材としての画像を記憶するサーバ装置等（不図示）から学習画像をダウンロードする。学習画像蓄積部 22 は、例えば、対象認識装置 1 において実行されるAR用アプリケーションの命令に従って学習画像をダウンロードする。

この学習画像には、ラベルが付加される。このラベルには、学習画像の属性情報が書き込まれる。属性情報としては、学習画像に含まれるランドマークの分類が含まれる。属性情報は、クラスIDとして付加される。

本実施形態において、学習画像蓄積部 22 は、ランドマークを含む画像素材を学習画像としてダウンロードする。この学習画像に含まれるランドマーク画像としては、人が見てランドマークであることが認識可能な画像であることが望ましい。また、学習画像におけるランドマーク画像は、画像面積に対して任意の面積であってもよく、画像内位置も任意の位置であってもよい。

30

【0032】

画像変換部 220 には、外部から学習処理を行うために用いられる学習画像が供給される。画像変換部 220 は、後段の学習処理の処理方式に合わせて学習画像に前処理を行う。画像変換部 220 は、例えば学習画像の面積変更処理やアスペクト比の調整等を行い、学習処理が実施可能な形式に変換する。画像変換部 220 は、学習画像の全体を用いて学習処理部 24 により学習させることが望ましいが、学習画像を分割または領域切り出して学習を学習処理部 24 により行わせてもよい。

画像データデータベース 222 は、外部から供給された変換前の学習画像、および画像変換部 220 により変換された学習画像を蓄積する。画像データデータベース 222 は、画像変換部 220 により変換された学習画像が学習処理部 24 において学習処理が実行可能なバッチファイルとして記憶される。また、画像データデータベース 222 は、ラベルにより分類されたランドマークごとに変換前の学習画像および変換後の学習画像を蓄積する。

40

【0033】

学習処理部 24 は、画像データデータベース 222 に記憶されたラベル（クラスID）付きの画像に基づいて学習を行う。学習処理部 24 により実行される学習処理は、画像データをベクトル等の数値データに変換して行う。この学習処理により得られた学習結果は、認識計算部 14 における学習結果データベース 140 に記憶される。

学習処理部 24 により得られた学習結果は、認識計算部 14 により参照される。

50

【0034】

画像変換部12、認識計算部14、マップ生成部16および位置算出部18は、図1を参照して説明したように動作する。これにより、対象認識装置1は、認識対象画像が供給されたことに応じて、クラスIDおよび位置をインターフェース部20に出力する。

【0035】

なお、学習処理部24による学習手法は、畳み込みニューラルネットワークを作成してもよく、その他の機械学習手法(SVM等)でもよく、学習結果データベース140には、学習手法に応じた学習結果が記憶される。認識計算部14は、学習結果データベース140に記憶された学習結果に基づいて認識対象画像に含まれる主な対象に対応したクラスIDおよび確信度を得ることとなる。

10

【0036】

クラスIDおよび確信度の算出に畳み込みニューラルネットワークを用いる場合、畳み込みニューラルネットワークは、認識対象画像を任意の個数のクラスに分類するよう学習される。学習処理部24は、クラスID:1の学習画像が供給されたことに對し、クラスID:1の学習結果が得られるように畳み込み層におけるフィルタ係数およびバイアス値が学習される。具体的には、クラスを「海岸」、「秋葉原」、「渋谷」、「東京タワー」、「都庁」に分類して学習した畳み込みニューラルネットワークは、「海岸」、「秋葉原」、「渋谷」、「東京タワー」、「都庁」を主な対象として含む認識対象画像が供給されたことに對じて、「海岸」、「秋葉原」、「渋谷」、「東京タワー」、「都庁」に分類することができる。

20

また、この学習処理において、畳み込みニューラルネットワークには、学習画像として、例えば図3~図5に示すようなものが挙げられる。図3に示した学習画像P10はタワー状建造物(ランドマーク)を近距離且つ下方から撮像した画像であり、図4に示した学習画像P11はタワー状建造物を遠距離且つ上方から撮像した画像であり、図5に示した学習画像P12はタワー状建造物を夜間において撮像した画像である。

この学習処理において、対象認識装置1は、特定の物体を抽出して候補領域を切り出すことなく、学習画像の全体について学習処理を実施する。

【0037】

次に、上述したように構成された対象認識装置1における対象の認識処理について説明する。図6は、本発明の実施形態として示す対象認識装置1における認識処理の手順を示すフローチャートである。

30

【0038】

まず、画像変換部12は、認識対象画像が入力されるまで待機し(ステップS200)、認識対象画像が入力されたことに對じて(ステップS200:YES)、認識対象画像に任意面積の領域を複数設定する(ステップS202)。このとき、画像変換部12は、認識対象画像の特徴量を抽出し、当該抽出した特徴量に基づいて、認識対象画像内に含まれる特徴の大きさに對じた任意面積の領域(候補領域)を複数設定する。画像変換部12は、候補領域毎の画像を認識計算部14に供給する。

【0039】

このとき、画像変換部12は、認識対象画像に設定された複数の候補領域のうち、所定の領域面積より小さい領域面積の候補領域を削除してもよい。この所定の領域面積は、認識対象画像のうち主な対象が含まれる割合が設定されることが望ましい。例えば、認識対象画像のうち画像を見る者が注目するような対象の大きさであればよい。この所定の領域面積は、予め設定されていてもよく、認識対象画像について設定された複数の候補領域の面積に基づいて設定してもよい。

40

【0040】

図7は、本発明を適用した対象認識装置1に入力される認識対象画像に設定された候補領域と、候補領域ごとの確信度の一例を示す図である。

図7に示すように、タワー状建造物が略中央位置に大きな画像面積で含まれる候補領域の画像として供給された場合、画像変換部12は、タワー状建造物Aを含む複数の候補領

50

域と、タワー状建造物以外の建造物Bに設定された複数の候補領域と、建造物Cに設定された複数の候補領域とが設定される。

なお、画像変換部12は、図7に示すように、認識対象画像に含まれる物体について漏れが少なく複数の候補領域を設定できる手法であれば、候補領域を設定する既存の手法のどのような手法であればよい。

【0041】

認識計算部14は、画像変換部12から領域候補を受け取り、学習結果データを用いて、画像変換部12から渡された領域画像に対して、クラスIDおよび確信度を算出する(ステップS204)。認識計算部14は、画像変換部12により設定された全ての候補領域について確信度の計算が完了するまで、次の候補領域の選択(ステップS208)と、

10

クラスIDおよび確信度の計算(ステップS204)とを繰り返して行う。

認識計算部14は、全ての候補領域について確信度の計算が完了したことに応じて(ステップS206: YES)、マップ生成部16によって認識対象画像の位置毎に評価値を算出して、評価値のマップを生成する(ステップS210)。そして、位置算出部18は、マップ生成部16により生成されたマップに基づいて、認識対象画像に含まれる主な対象を表すクラスIDおよび位置をインターフェース部20に出力する(ステップS212)。

【0042】

図7に示す例において、複数の候補領域ごとにクラスIDごとの確信度が算出された様子を示す。図8は、本発明を適用した対象認識装置1におけるマップ生成部が評価値を算出する際に用いる候補領域ごとの確信度および面積との関係を例示した図である。図8において、面積は、百万ピクセルを超えた候補領域をカウントしている。タワー状建造物Aについては、図8(a)に示すように、確信度が高く且つ面積が大きい候補領域が多い。建造物Bについては、図8(b)に示すように、確信度が高いが、面積が小さい候補領域が多い。建造物Cについては、図8(c)に示すように、確信度が小さく、面積が小さい候補領域が多い。

20

図8に示すような特性を利用し、対象認識装置1は、クラスIDごとに認識対象画像の位置の確信度を統計処理し、統計処理した結果得た評価値が高い領域を主な対象が存在する領域であるとしてもよい。

【0043】

図9は、本発明を適用した対象認識装置1において、認識対象画像Pを認識計算部14により認識し、マップ生成部16により作成されたマップが表す位置と評価値の関係を説明するために可視化した例を示す図であるこのマップは、クラスIDごとのTa(タワー型建造物)、Tb(山)、Tc、Te、Tfについて作成される。この図9が示す例では、クラスIDごとのマップは、統計処理の結果が高くなった評価値の位置ほど、高い輝度となるよう作成したものである。なお、マップは、評価値を演算する統計処理の相違により(a)~(d)のように異なるが、図9に示した(a)~(d)のマップのうちいずれを用いて対象の位置を算出してもよい。

30

位置算出部18は、それぞれのクラスIDのマップに対してエッジ抽出処理を実施し、評価値が段階的に高くなったエッジ線に囲まれる領域を抽出し、抽出された領域にクラスIDに対応した対象が存在すると判定してもよい。図9によれば、クラスIDがTa(タワー型建造物)のマップに対してエッジ抽出処理を実施することにより、認識対象画像の略中央位置に存在するTa(タワー型建造物)を抽出することができる。

40

これにより、対象認識装置1は、タワー型建造物を表すクラスID(Ta)と抽出したタワー型建造物の位置とを含む処理結果を、インターフェース部20から出力させることができる。

【0044】

なお、図8に示した説明では、マップ生成部16により評価値のマップをクラスIDごとに作成したが、認識対象画像に含まれる可能性が低いクラスIDについては評価値のマップを作成しなくてもよい。確信度を統計処理して演算された評価値が低い場合には評価

50

値のマップから主な対象が抽出されない可能性が高く、マップ生成部16は、このような評価値が演算されたクラスIDについてはマップを作成しないと判定する。

マップ生成部16は、マップの作成前に、マップの作成処理よりも少ない計算量で演算が可能な評価指標を演算する。マップ生成部16は、例えば、候補領域の面積および確信度をパラメータとした評価関数の合計値をクラスIDごとに求め、クラスID間で合計値を比較して、評価値のマップを作成するクラスIDと評価値のマップを作成しないクラスIDとを決定する。これにより、対象認識装置1は、主な対象の認識精度の低下を抑制しつつ、マップの作成についての処理量を抑制できる。

【0045】

以上説明したように、本発明を適用した対象認識装置1によれば、認識対象画像内に任意面積の領域を複数設定し、各領域に含まれる対象候補の確信度を演算し、対象候補の確信度と、対象候補の位置とに基づいて対象候補ごとのマップを生成し、マップから対象の位置を特定するので、単に確信度が高い領域に含まれる対象が認識対象画像に含まれる主な対象であるという誤認識を抑制でき、より高い認識精度で認識対象画像内の主な対象を判定することができる。

10

【0046】

例えば、多数の建造物が含まれる市街地の風景や、多数の建造物が含まれているが単一のランドマークが含まれる風景が認識対象画像として入力されても、認識対象画像内で明確に含まれている単一の対象が認識対象画像における主な対象であるという分類結果を出力することを抑制できる。より具体的には、本発明を適用した対象認識装置1によれば、象徴的な町並みの市街地の名称を主な対象として認識でき、さらに、広い面積を占めるランドマークを主な対象として認識できる。

20

【0047】

図10は、比較例としての認識対象画像P1に設定された候補領域の一例を示す。図11は、比較例としての認識対象画像P1に設定された他の候補領域の一例を示す。

比較例として、図10に示した候補領域A1、B1、C1の画像を畳み込みニューラルネットワークに入力して得た確信度のみに基づいて、畳み込みニューラルネットワークから出力されたランドマークを分類結果として得ると、認識対象画像P1について設定した候補領域A1、B1、C1のうち、候補領域B1の確信度が最も高くなる。しかし、認識対象画像P1に含まれる主な対象は建造物T1であり、候補領域A1の画像に含まれる。

30

【0048】

同様に、比較例として、図11に示した候補領域A2、B2、C2の画像を畳み込みニューラルネットワークに入力して得た確信度のみに基づいて、畳み込みニューラルネットワークから出力されたランドマークを分類結果として得ると、認識対象画像P2について設定した候補領域A2、B2、C2のうち、候補領域B2の確信度が最も高くなる。しかし、認識対象画像P2に含まれる主な対象は建造物T2であり、候補領域A2の画像に含まれる。

【0049】

このように、比較例のように、認識対象画像に複数の対象が含まれている場合、確信度のみに基づいて認識対象画像に含まれる主な対象を認識すると誤認識が発生する場合がある。

40

これに対し、本発明を適用した対象認識装置1によれば、候補領域ごとにクラスIDおよび確信度を演算し、クラスIDごとに確信度を統計処理するので、例えば図10または図11に示したように、候補領域のうち面積の広い候補領域に含まれる対象が主な対象として認識できる確率を高めることができる。

【0050】

さらに、本発明を適用した対象認識装置1によれば、学習処理部24により、外部から供給された学習画像の全体を用いて学習するので、認識対象画像の全体に含まれる対象を認識するように学習することができ、より高い認識精度で認識対象画像内の主な対象を判定することができる。

50

【 0 0 5 1 】

すなわち、本発明を適用した対象認識装置 1 によれば学習処理において正解となる対象を切り出す必要がないので、学習画像のうち正解となる対象が学習画像のどの位置に含まれていようとも、学習画像の全体について学習処理を行うことができ、認識処理において画像全体において主な対象がどのような形態に含まれていても、高い認識精度で認識対象画像内の主な対象を判定することができる。

【 0 0 5 2 】

具体的には、既存の技術のように特定の物体を認識できるように教師データを作成してにおいて特徴をマッチングすると、風景（例えば市街地、大通りなどのような、物体の領域が区別されにくい画像や、近景から遠景に亘って見え方が大きく異なる大型の建造物には

10

【 0 0 5 3 】

さらに本発明を適用した対象認識装置 1 によれば、学習処理のために正解となる対象を含む画像を切り出す必要がないので、学習処理の前処理の処理量を抑制でき、例えば、インターネット等において利用されている一般的な画像を学習画像として利用でき、学習処理の効率を高めることができる。また、本発明を適用した対象認識装置 1 によれば、1つの学習画像について1つのラベルとなるように学習画像を編集する必要がないため、学習処理の前処理のために専用の教師画像を用意する必要が無く、処理量をさらに抑制できる

20

【 0 0 5 4 】

さらに、本発明を適用した対象認識装置 1 によれば、画像変換部 1 2 により設定された複数の領域のうち、重複した複数の領域についての複数の確信度を統計処理するので、高い確信度であって重複した領域が多いほど、認識対象画像に含まれる主な対象であるという誤認識を抑制でき、より高い認識精度で認識対象画像内の主な対象を判定することができる。

【 0 0 5 5 】

さらに、本発明を適用した対象認識装置 1 によれば、画像変換部 1 2 により設定された複数の領域のうち、重複した複数の領域についての複数の確信度に加えて候補領域の面積に基づいて統計処理するので、高い確信度であって面積が広い候補領域ほど、認識対象画像に含まれる主な対象であるという誤認識を抑制でき、より高い認識精度で認識対象画像内の主な対象を判定することができる。

30

【 0 0 5 6 】

さらに、本発明を適用した対象認識装置 1 によれば、認識対象画像に設定された複数の領域のうち、所定の領域面積より小さい領域面積の領域を削除するので、各領域の領域面積に基づいて認識対象画像に含まれる対象を特定することを利用して、処理負荷を少なくしても、高い認識精度で認識対象画像内の主な対象を判定することができる。

【 0 0 5 7 】

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれると同様に、特許請求の範囲に記載された発明とその均等の範囲に含まれるものである。

40

【 0 0 5 8 】

また、上述した対象認識装置 1 の機能を実現するためのコンピュータプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行するようにしてもよい。なお、ここでいう「コンピュータシステム」とは、OS や周辺機器等のハードウェアを含むものであってもよい。

50

また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、フラッシュメモリ等の書き込み可能な不揮発性メモリ、DVD (Digital Versatile Disk) 等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。

【0059】

さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（例えばDRAM (Dynamic Random Access Memory)）のように、一定時間プログラムを保持しているものも含むものとする。

また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。

また、上記プログラムは、前述した機能の一部を実現するためのものであっても良い。

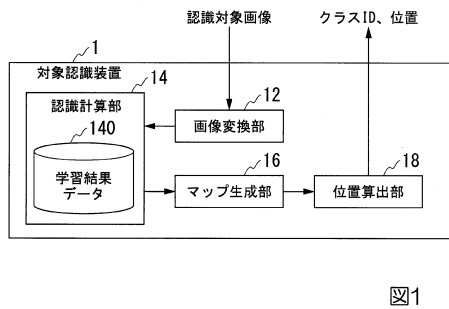
さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

【符号の説明】

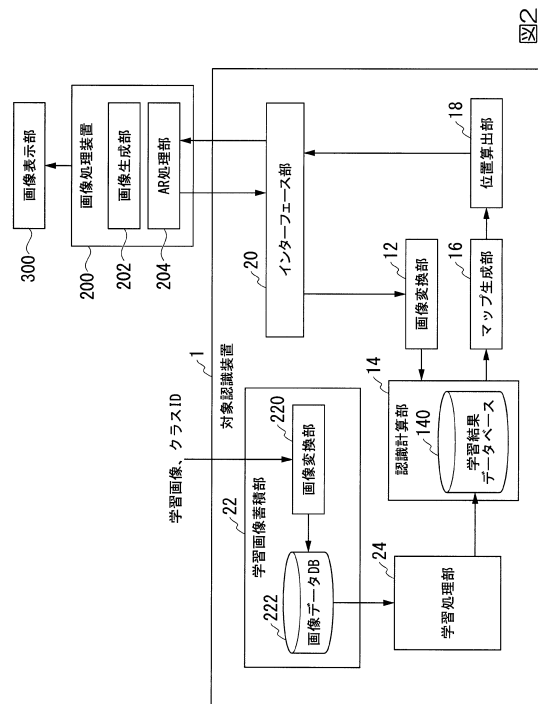
【0060】

1...対象認識装置、12...画像変換部、14...認識計算部、16...マップ生成部、18...位置算出部、20...インターフェース部、22...学習画像蓄積部、200...画像処理装置、202...画像生成部、204...AR処理部、202...画像表示部、220...学習画像蓄積部、222...画像データベース、24...学習処理部

【図1】



【図2】



10

20

図2

図1

【図3】

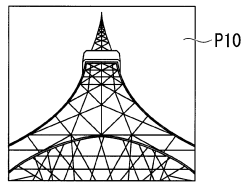


図3

【図4】

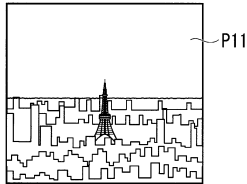


図4

【図5】

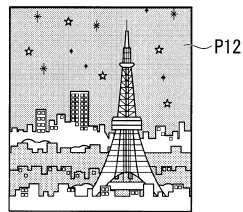


図5

【図6】

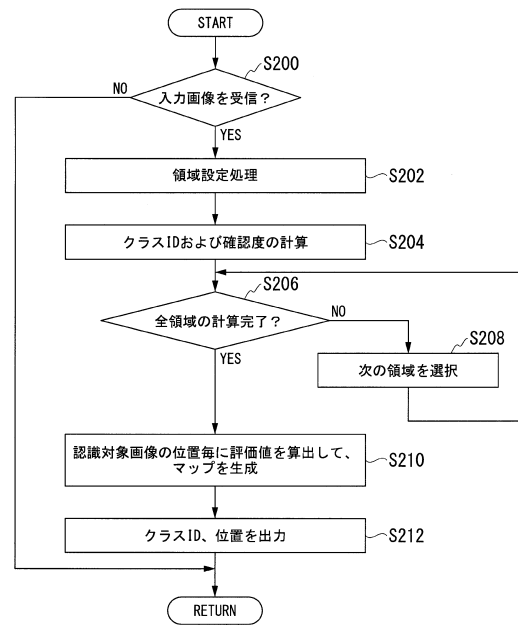
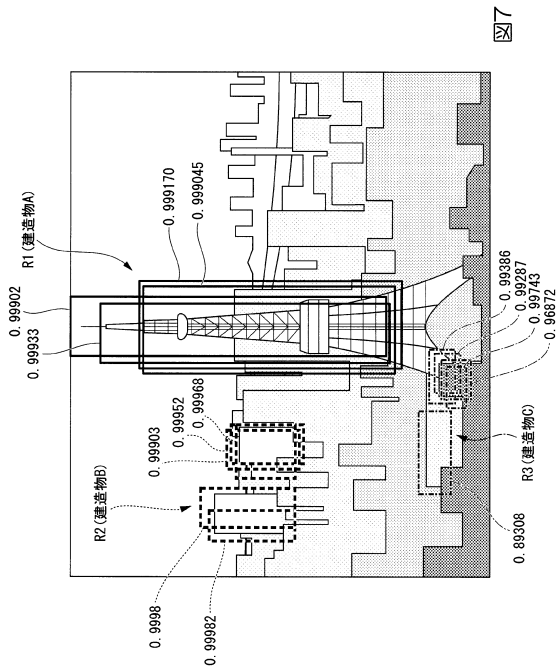


図6

【図7】



【図8】

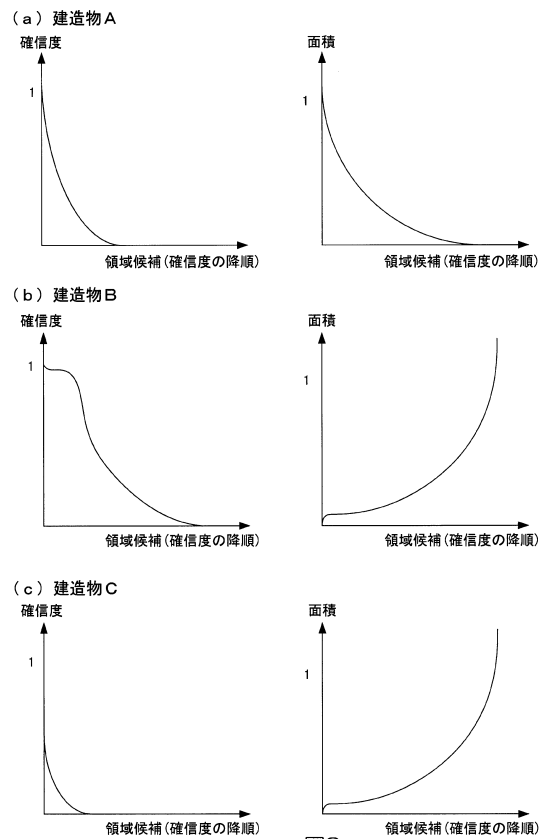
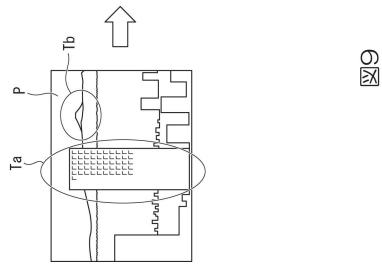
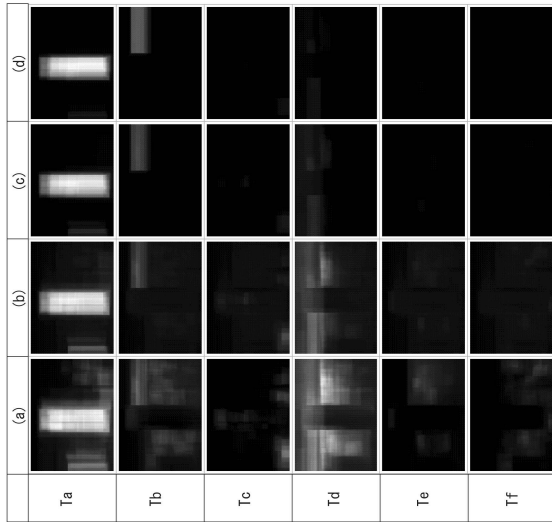
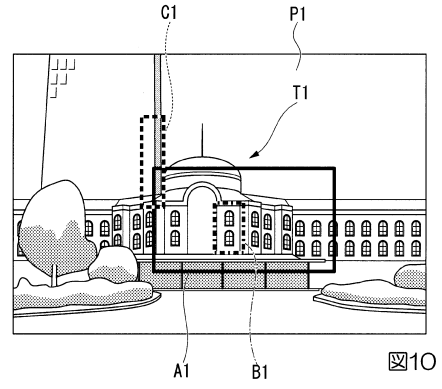


図8

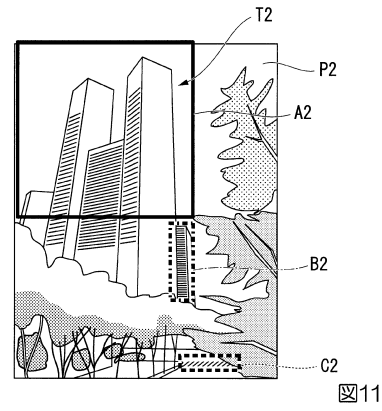
【 9 】



【 10 】



【 11 】



フロントページの続き

- (72)発明者 兒玉 松男
東京都港区港南一丁目9番1号 エヌ・ティ・ティ・コムウェア株式会社内
- (72)発明者 木全 英明
東京都港区港南一丁目9番1号 エヌ・ティ・ティ・コムウェア株式会社内

審査官 松浦 功

- (56)参考文献 特開2014-191400(JP,A)
特開2013-114596(JP,A)
特開2013-182478(JP,A)
国際公開第2012/124000(WO,A1)
特開2010-160640(JP,A)
特開2014-186524(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06T 7/00 - 7/90
G06T 1/00