

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(10) 国際公開番号

WO 2011/021605 A1

(43) 国際公開日

2011年2月24日(24.02.2011)

PCT

- (51) 国際特許分類:
G06T 7/00 (2006.01) G06T 1/00 (2006.01)
- (21) 国際出願番号: PCT/JP2010/063822
- (22) 国際出願日: 2010年8月16日(16.08.2010)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2009-189367 2009年8月18日(18.08.2009) JP
- (71) 出願人 (米国を除く全ての指定国について): 公立大学法人大阪府立大学 (OSAKA PREFECTURE UNIVERSITY PUBLIC CORPORATION) [JP/JP]; 〒5998531 大阪府堺市中区学園町1-1 Osaka (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): クリンクィット マーティン (KLINKIGT, Martin) [DE/JP]; 〒5998531 大阪府堺市中区学園町1-1 公立大学法人大阪府立大学内 Osaka (JP). 黄瀬 浩一

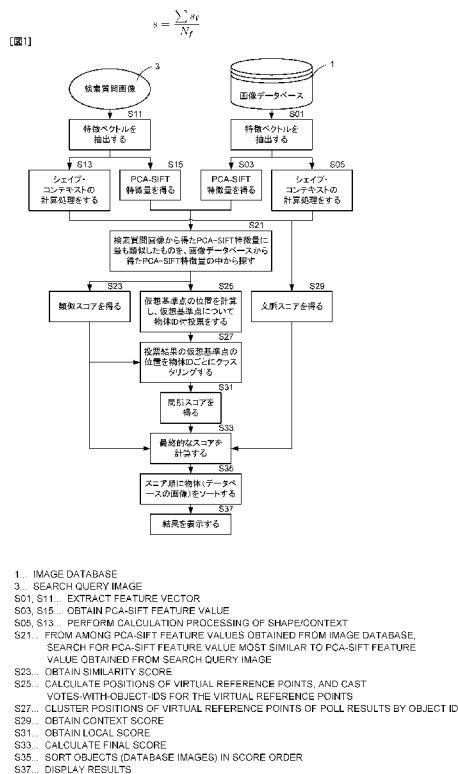
(KISE, Koichi) [JP/JP]; 〒5998531 大阪府堺市中区学園町1-1 公立大学法人大阪府立大学内 Osaka (JP). マウス ヘイコ (MAUS, Heiko) [DE/DE]; ディー-67663 カイザーラウテルン、トリップシュタッター シュトラーセ 122 ドイツ人工知能研究センター内 Kaiserslautern (DE). デンゲル アンドレアス (DENGEL, Andreas) [DE/DE]; ディー-67663 カイザーラウテルン、トリップシュタッター シュトラーセ 122 ドイツ人工知能研究センター内 Kaiserslautern (DE).

- (74) 代理人: 野河 信太郎 (NOGAWA, Shintaro); 〒5300047 大阪府大阪市北区西天満5丁目16-3 西天満ファイビル 野河特許事務所 Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR,

[続葉有]

(54) Title: OBJECT DETECTION METHOD

(54) 発明の名称: 物体検出方法



(57) Abstract: Disclosed is an object detection method that is provided with a step for extracting a plurality of reference feature vectors related to a local region from an image representing an object, and extracting a plurality of query feature vectors related to the local region from a search query image; a step for matching each query feature vector against each reference vector, and calculating a similarity score having a value that is higher the closer the distance between both vectors, the larger the local region for which the query feature vector has been extracted, and the larger the local region for which a matching reference vector has been extracted; a step for determining a reference vector for which a similarity score is highest as the similarity vector for each query feature vector; and a step for acquiring a final score by object associated with the similarity vectors, and setting the object returning the highest score as the detection result; and characterized in that the score is calculated by dividing the similarity score for each similarity vector by the number of feature vectors that have matched the object.

(57) 要約: 物体を表す画像からその局所的領域に係る複数の参照特徴ベクトルを抽出し、検索質問画像からその局所的領域に係る複数のクエリー特徴ベクトルを抽出する工程と、各クエリー特徴ベクトルと各参照ベクトルとの照合を行い、両者の距離が近いほど大きく、前記クエリー特徴ベクトルが抽出された局所領域が大きいほど大きな値を有し、かつ照合される参照ベクトルが抽出された局所領域が大きいほど大きな値を有する類似スコアを算出する工程と、前記類似スコアが最大となる参照ベクトルを各クエリー特徴ベクトルの類似ベクトルとしてそれぞれ決定する工程と、前記類似ベクトルに関連付けられた物体別の最終的なスコアを得、最高のスコアを与える物体を検出結果とする工程とを備え、前記スコアは、各類似ベクトルに係る類似スコアを、物体に照合された特徴ベクトルの数で割ることで算出されることを特徴とする物体検出方法。

WO 2011/021605 A1



LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称： 物体検出方法

技術分野

[0001] この発明は、画像中の物体検出方法に関し、特に複雑な背景の下で物体を精度よく検出する手法に関する。

背景技術

[0002] 物体検出とは、画像から物体を発見するという、コンピュータビジョンの分野でよく知られた研究課題である。最近の手法では、物体に関する情報は、局所特徴量の組で表される。局所特徴量とは、画像の部分領域を記述する特徴量である。画像から得られる局所特徴量の数と質は、処理に用いるアルゴリズムに大きく依存する。

[0003] 物体検出手法に共通する問題の一つに、背景から得られる局所特徴量の扱いがある。ここで、背景とは、物体以外の領域を表す。物体が異なっても同じ背景を持つことは十分考えられるため、背景から得られた局所特徴量を、前景から得られたものと区別せずに用いると、信頼性の高い結果が得られない。したがって、物体検出には、背景から得られたそれらの局所特徴量を用いないようにする仕組みが必要となる。

この問題を解決するための多くの手法は、教師あり学習を用いるものである。具体的には以下の3つのアプローチが提案されている。

[0004] 第一の方法は、背景を一切含まない画像を用いて物体の局所特徴量を学習するものである。背景を含まない物体だけの画像は、例えば、物体の領域を背景から切り出すことによって得られる。また、物体を撮影する際に単色の背景を用いておき、後でその色を用いないようにする方法（クロマキー）もある。いずれにせよ、これらの手法は手間がかかるため、検出対象となる物体の数が十分に少ないときにのみ用いることができる。

[0005] 第二の方法は、検出対象の物体を含まない画像を用いて、背景の特徴を学習するものである。これにより、背景からどのような局所特徴量が得られ、

物体からはどのような局所特徴量が得られるのかを知ることができる。ただしこの方法には次のような問題点がある。例えば、人物検出の場合、本棚は背景となる。しかし、本棚の検出にも興味がある場合には、本棚はもはや背景とはなり得なくなる。すなわち、背景の定義は検出対象に依存するため、容易ではないといえる。

- [0006] 第三の方法は、多様な背景を伴う検出対象の画像を多数用いるものである。この手法では、それらのすべての画像から共通して得られる局所特徴量を、検出対象物体から得られたものであると仮定する。この手法でも、必要な画像数は膨大となり、それらはユーザによって与えられなければならない点の問題となる。
- [0007] 複雑な背景が問題となる一つの理由は、画像から抽出された局所特徴量は、それが背景から得られたものか物体から得られたものかに関係なく、同様に扱う点にある。これを解決する一つの方法は、個々の局所特徴量を人手によって選別することであるが、局所特徴量の数などから現実的な方法とはいえない。もう一つの可能な方法は、背景のない画像を与えて、物体を容易に取り出す方法である。他にも単一の物体に対して様々な背景を持つ画像を与える方法もある。
- [0008] もう一つの問題は、物体認識や検出によく用いられる「bag-of-featuresモデル」という方法から生じるものである。このモデルでは、各局所特徴量をクラスタリングして代表ベクトル（visual wordと呼ばれる）を得、さらに得られた代表ベクトルから物体の特徴を表す一つのベクトル量（特徴量）を生成する。クラスタリングの過程で各局所特徴量は互いの関連性を捨象して、個々独立なものとして扱われる。「bag-of-featuresモデル」の方法により、識別に有効で安定した特徴量を得ることが可能となる。一方で、各局所特徴量の位置や大きさ、方向などの情報が失われるという問題も生じる。
- [0009] この問題に対処する方法として、これまでに星座モデル(constellation model)と暗黙的形状モデル(implicit shape mode)の2つが提案されている。即ち、星座モデルや暗黙的形状モデルは、前述のクラスタリング処理によっ

て失われてしまった、各局所特徴量の配置に関する情報を補償するためのものといえる。ただし、前記情報の扱い方はそれぞれのモデルで異なる。以下で各々について述べる。

《星座モデル》

[0010] 星座モデルでは、物体から得た局所特徴量は、2次元の確率空間上の位置を伴って保存される。結論からいえば、星座モデルは、検出対象の物体を示す検索質問画像とそれを用いて検出されるべき画像との間で主要なvisual wordの配置があまり変わらない場合に有効なモデルである。星座モデルでは、結果として得られるグラフ（局所特徴量をノード、局所特徴量間の位置関係をエッジで表すグラフ）を計算可能なものとするため、5個程度の局所特徴量（グラフのノード）しか用いることができない。

一つの局所特徴量を基準として正規化することにより、このモデルをスケールに不変とすることができる。

[0011] Fergusらが述べているように、このモデルの一つの弱点は特徴の検出器（detector）に大きく依存する点にある（例えば、非特許文献1参照）。もし、画像の比較的大きな領域に対して定義された局所特徴量（例えば、自転車の車輪など）を検出することができなければ、もはやこのモデルを用いることはできない。もう一つの弱点は、このモデルが回転不変ではなく、また、視点の移動を扱うことができない点である。

《暗黙的形状モデル》

[0012] Leibeらは、暗黙的形状モデルと呼ばれるモデルを提案した（例えば、非特許文献2参照）。このモデルでは、visual word間の相対的な配置によって形状が表されているわけではない。各visual wordは、予め定められた重心との相対的な位置によって表現される。物体検出時には、検索質問画像から取り出されたvisual wordは、画像データベース中のvisual wordと比較され、可能性のある重心の位置が提示される。これを可能性のある物体への投票と考え、可能性のある重心をクラスタリングして集めることにより、可能性のある物体がわかる。非特許文献3では、このモデルをスケールに不変とする手

法が提案されている。このモデルは物体クラス内の大きなバリエーションを表すことができるような、十分柔軟性のあるモデルとなっている。これは、異なる画像から得られた局所特徴量を用いることによって実現されている。

先行技術文献

非特許文献

[0013] 非特許文献1 : R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 2, pages 264–271, June 2003.

非特許文献2 : B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In ECCV workshop on statistical learning in computer vision, pages 17–32, 2004.

非特許文献3 : B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. Int. J. Comput. Vision, 77(1–3):259–289, 2008.

発明の概要

発明が解決しようとする課題

[0014] 前述のように、局所特徴量の互いの関連性についての情報を考慮することができれば識別に有効で安定した局所特徴量を得ることができるが、従来の星座モデルでは通常何千、何万という局所特徴量を扱う物体検出には不適であり、その他にも弱点がある。これに対し暗黙的形状モデルは、より柔軟性があるが、このモデルも視点の移動を扱えず、また、回転に対して不変ではないという弱点を持つ。背景の影響を排除するため、局所特徴量の配置情報を用いることのできる物体検出手法が望まれている。

[0015] この発明は、以上のような事情を考慮してなされたものであって、制限のない背景領域を伴う少数の画像（理想的な場合は1枚の画像）からでも物体

の特徴を学習しその物体を検出できる手法を提供するものである。

課題を解決するための手段

- [0016] 前述の課題を解決すべくなされたこの発明の特徴は、局所特徴量だけではなく、その配置も用いる点にある。単体の局所特徴量であれば、偶然に他の物体あるいは背景から共通のものが得られる可能性はあるが、多くの局所特徴量が配置を含めて一致するという状況は、偶然には起こりえない。この発明による手法はこのような情報を用いて物体検出を行う。
- [0017] この発明の特徴的な一側面は、物体を表す画像を、予め登録された多数の画像データから検出する方法であって、物体を示す画像としての検索質問画像からその局所的な特徴を表す複数のクーリエ特徴ベクトルを抽出し、予め登録された画像データは、その局所的な特徴を表す複数の参照特徴ベクトルと一つの基準点に対する各参照特徴ベクトルの距離と方向とが関連付けて格納され、各クーリエ特徴ベクトルにつき各参照ベクトルを比較して類似する参照ベクトルを類似ベクトルとして決定し、類似ベクトルに関連付けられた前記基準点からの距離と方向とを用いて検出すべき画像データを決定する点にある。
- [0018] また、他の特徴的な一側面は、各類似ベクトルにつき、それと関連付けられた距離及び方向を用いて仮想の基準点の位置をそれぞれ計算し、その位置につき物体ID付の投票処理を行い、投票結果としての仮想の基準点の位置が近いもの、換言すれば類似のものを集めてなる複数のクラスタを得、各クラスタの類似の度合いを表す局所スコアを所定の処理によって得、得られた局所スコアに係る評価に基づき所定数の物体IDを検出結果として採用する点にある。
- [0019] ここで、投票処理は、得られた証拠に基づいて選択肢の何れかにスコアを与え、すべての証拠を集計した結果最高あるいは上位のスコアを得た選択肢を選択する処理を指す。この技術分野では一般によく用いられている処理である。
- [0020] さらに、他の特徴的な一面は、各クエリー特徴ベクトルについて各参照特

徴ベクトルとの照合を行うために、両者の距離が近いほど大きく、前記クエリー特徴ベクトルが抽出された局所領域が大きいほど大きな値を有し、かつ照合される参照ベクトルが抽出された局所領域が大きいほど大きな値を有するように定められてなる類似スコアを算出し、前記類似スコアが最大となる参照ベクトルを各クエリー特徴ベクトルの類似ベクトルとしてそれぞれ決定し、各類似ベクトルにつき、各類似スコアに係る評価に基づき所定数の物体IDを検出結果として採用する点にある。

[0021] より詳細には、この発明は、物体を表す参照画像からその局所的領域に係る特徴を表す複数のベクトルを参照特徴ベクトルとして抽出し、それらの参照特徴ベクトルが対応する参照画像およびその参照画像が表す物体の物体IDと関連付けられて登録されてなる画像データベースの中から、検索質問画像に表された物体と同一の物体を表す画像を見出すべく、検索質問画像からその局所的領域に係る特徴を表す複数のベクトルをクエリー特徴ベクトルとして抽出する工程と、各クエリー特徴ベクトルについて各参照特徴ベクトルとの照合を行い、両者の距離が近いほど大きく、前記クエリー特徴ベクトルが抽出された局所領域が大きいほど大きく、かつ前記参照ベクトルが抽出された局所領域が大きいほど大きな値を有するように定められてなる類似スコアを算出する工程と、各クエリー特徴ベクトルについて前記類似スコアが最大となる参照ベクトルを類似ベクトルとしてそれぞれ決定する工程と、前記類似ベクトルに関連付けられた物体ID別の最終的なスコア s を所定の計算処理によって得、最高のスコアを与える物体IDに係る物体を少なくとも検出結果とする工程とを備え、前記スコア s は、各類似ベクトルに係る類似スコアを s_f 、前記クエリー特徴ベクトルのうち、画像データベースに登録された参照画像が表す物体に対応する特徴ベクトルとのマッチングが得られたものの数を N_f とするとき、

[0022] [数1]

$$s = \frac{\sum s_f}{N_f} \quad (1)$$

により算出されることを特徴とする物体検出方法を提供する。

発明の効果

- [0023] この発明の物体検出方法において、両者の距離が近いほど大きく、前記クエリー特徴ベクトルが抽出された局所領域が大きいほど大きく、かつ前記参照ベクトルが抽出された局所領域が大きいほど大きな値を有するように定められてなる類似スコアを算出する工程と、最高のスコアを与える物体IDに係る物体を少なくとも検出結果とする工程とを備えるので、特徴ベクトルが抽出された局所領域の大きさを考慮しない場合と比較した場合、複雑な背景から得られる特徴ベクトルの影響をより小さくする一方で物体から検出される特徴ベクトルを重視して物体の検出を行うことができる。これは、複雑な背景から得られた特徴ベクトルは、通常、その特徴ベクトルが抽出された局所領域の大きさが小さく、一方で、物体から得られる特徴ベクトルはその特徴ベクトルが抽出された局所領域の大きさが比較的大きいことが多いという性質を利用しているからである。
- [0024] 最終的なスコアをこのような単純な計算処理で得ても、有効な物体検出を行うことができる。即ち、局所領域の大きさを考慮しない従来の手法に比べ、複雑な背景から得られる特徴ベクトルの影響を小さくする一方で物体から検出される特徴ベクトルを重視して物体の検出を行うことができる。
- [0025] この発明において、画像からその局所的な特徴を表す特徴ベクトルを抽出する手法としては、公知のSIFT (Scale-Invariant Feature Transform) が適用可能である。SIFTの詳細は口述する。この実施形態では、SIFTの一種であるPCA-SIFTの手法により特徴ベクトルを抽出しているが、必ずしもこれに限定されるものではない。ただし、この発明においては、前述のように複雑な背景から得られた特徴ベクトルは、通常、小さい局所領域を持つという性質を利用しているので、この性質を満たす手法が極めて好ましい。SIFTやPCA-SIFTはその好適な一例である。

図面の簡単な説明

- [0026] [図1] この発明の物体検出方法の処理の流れを示すフローチャートである。

[図2] この発明に係る局所特徴量の基準点からの距離 t と角度 α の定義を示す説明図である。

[図3] この発明に係るシェイプ・コンテキストの一例を示す説明図である。

[図4] この発明に係る局所特徴量とシェイプ・コンテキストとの関係を示す説明図である。

[図5] この発明の有効性を示す実験に用いた画像の例を示す説明図である。

発明を実施するための形態

[0027] 以下、この発明の好ましい態様について説明する。

前記類似スコアは、特徴ベクトル f が抽出された局所領域の大きさを $w(f)$ とし、クエリー特徴ベクトル f_q とそれに対応する類似ベクトル f_d とのユークリッド距離を $d(f_q, f_d)$ とするとき、

[数2]

$$s_f = \frac{w(f_q) \cdot w(f_d)}{d(f_q, f_d)} \quad (2)$$

[0028] により算出されてもよい。このようにすれば、類似スコア s_f は、クエリー特徴ベクトルと参照特徴ベクトルの距離がより近ければより大きくなるだけでなく、クエリー特徴ベクトル、参照特徴ベクトルが抽出された局所領域の大きさがより大きければより大きくなる。よって、クエリー特徴ベクトル、参照特徴ベクトルの局所領域の大きさを考慮して類似ベクトルを決定することができ、また、検出結果としての物体を決定することが実現できる。

[0029] 前記画像データベースは、各参照画像に対して予め定められたそれぞれの基準点からの各参照ベクトルの距離および方向が、対応する参照特徴ベクトルに関連付けられてさらに登録されてなり、各類似ベクトルにつき、対応する参照ベクトルと関連付けられて格納された距離および方向を用いて前記類似ベクトルに対する仮想の基準点の位置を求める工程と、それぞれが互いに近接した仮想の基準点を有しかつ同じ物体IDに関連付けられた類似ベクトルから構成される複数のクラスタのうち各類似ベクトルが属すべきクラスタを

決定するクラスタリング工程と、前記クラスタリング工程により得られた各クラスタの類似の度合いを表す局所スコアを所定の処理によって得る工程とをさらに備え、前記最終的なスコアは、前記局所スコアを r_c 、クラスタの総数を N_c とすると、前式(1)に代えて、

[0030] [数3]

$$s = \frac{\sum s_f}{N_f} \sqrt{\frac{\sum r_c}{N_c}} \quad (3)$$

により算出されてもよい。

[0031] このようにすれば、類似スコアと局所スコアを用いて得ることによって、有効な物体検出を行うことができる。即ち、局所領域の大きさを考慮しない従来の手法に比べ、複雑な背景から得られる特徴ベクトルの影響をより小さくする一方で物体から検出される特徴ベクトルをより重視して物体の検出を行うことができる。

[0032] 即ち、この態様によれば、検索質問画像と画像データベース中の画像とをそれらの画像から抽出された局所特徴量を用いて照合する際に基準点とその基準点に対する相対位置の情報を用いる。これによって、単に対応する局所特徴の数が多いか少ないかだけではなく、局所特徴量の相対位置、即ち、局所特徴量の画像中での配置関係が類似しているか否かを考慮して照合を行うことができる。

[0033] ここで、画像データベースに登録される画像は、参照特徴ベクトルに関連付けられて登録されるだけでなく、各参照特徴ベクトルが基準点からの距離および方向に関連付けられて登録される。この基準点からの距離および方向の情報から、前記類似ベクトルに係る仮想の基準点を計算し、クラスタリングを行うのである。

[0034] 基準点は、各局所特徴量を表現するための基準である点において前述の非特許文献2、3における重心と同様の役割を担うが、物体の重心である必要はなく、また、物体の内部になければならないなどの制約条件はない。基準

点は、画像中のある決まった点であればよい。その具体的な態様は、例えば、画像の領域が矩形の場合、矩形領域 2 本の対角線の交点（矩形領域の中心）である。

[0035] 前記局所スコア r_c は、各クラスに含まれる仮定の基準点の総数として定義されてもよい。

[0036] ある特徴ベクトルを中心とする領域を複数の小領域であるビンに区分し、前記特徴ベクトルの周辺にあっていずれかのビンに属する他の特徴ベクトルの数をビンごとに集計してなるヒストグラムとして定義されるシェイプ・コンテキストの手順を適用して特徴ベクトルについてのシェイプ・コンテキストを得る工程と、前記シェイプ・コンテキストを用いた所定の計算処理を行って文脈スコアを算出する工程とをさらに備え、前記最終的なスコアは、前記文脈スコアを u_{sc} とするとき、前式（3）に代えて、

[0037] [数4]

$$s = \frac{\sum s_f}{N_f} \sqrt{\frac{\sum r_c}{N_c}} \sqrt{\sum u_{sc}} \quad (4)$$

により算出されてもよい。

[0038] このようにすれば、前記類似スコア、前記局所スコアおよび前記文脈スコアを用いて最終的なスコアを得、検出結果としての物体を決定できるので、複雑な背景から得られる特徴ベクトルの影響をより確実に小さくする一方で物体から検出される特徴ベクトルをより重視して物体の検出を行うことができる。即ち、局所特徴量の周囲にある局所特徴量の配置関係が類似しているか否か否かを局所スコアとは異なった文脈スコアの観点から考慮したうえで検索質問画像と対応する画像を見出すことができる。同じビンに入っているかどうかを見ることによって点の配置関係が同じか否かを評価することができるからである。

[0039] 前記シェイプ・コンテキストは、ある特徴ベクトルについて、その周辺にある特徴ベクトルとの相対的な配置の情報を与えるものである。対象とする

特徴ベクトルの周辺領域を複数の小領域であるビンに区分し、前記特徴ベクトルの周辺にあっていずれかのビンに属する特徴ベクトルの数をビンごとに集計してなるヒストグラムとして定義されるものである。

また、前記文脈スコアは、いずれかのビンに属する他の特徴ベクトルにつきビンごとの和を求めて各ビンの代表ベクトル

[0040] [数5]

$$v_i = (v_{ix}, v_{iy}) \quad (5)$$

とし m をシェイプ・コンテキストにおけるビンの数とすると、各代表ベクトルの要素を連結してなるベクトル、

[0041] [数6]

$$SC = \begin{pmatrix} v_{ix} \\ v_{iy} \\ \vdots \\ v_{mx} \\ v_{my} \end{pmatrix} \quad (6)$$

の長さとして定義されてもよい。

[0042] このようにすれば、シェイプ・コンテキストの中心の特徴ベクトルが物体の境界線付近に配置されている場合は、その周辺の特徴ベクトルが通常は一定方向を表わす傾向にあり、これに対して中心の特徴ベクトルが複雑な背景に配置されている場合はその周辺の特徴ベクトルの方向が一定でないという特性から、対象の特徴ベクトルが物体の境界線付近にある場合は複雑な背景にある場合に比べて大きな文脈スコアの値が得られる。従って、前記クラスタリング工程、あるいはさらに検出結果を得る工程において、この文脈スコアを用いることにより、複雑な背景から得られる特徴ベクトルの影響をより確実に小さくする一方で物体から検出される特徴ベクトルをより重視して物体の検出を行うことができる。

[0043] N_f 、 r_o および N_c について例を用いて説明する。クラスタリングを行った結果

、第1～第3の合計3つのクラスタが得られたとする。N_cはクラスタの総数であるからN_c = 3である。

[0044] r_cは、一つのクラスタについての仮想の基準点の数を表す。例えば、第1クラスタは仮想の基準点が3つであるとする。すると、第1クラスタについてはr_c = 3である。第2クラスタは仮想の基準点が2つであるとする。すると、第2クラスタについてはr_c = 2である。さらに、第3クラスタは仮想の基準点が1つであるとする。すると、第3クラスタについてはr_c = 1である。よって、それらのクラスタについての仮想の基準点の合計は、 $\sum r_c = 6$ である。クラスタリングが上手く行われない場合、少数の仮想の基準点を含む多数のクラスタが得られる。従って、 $\sum (r_c / N_c)$ の値は1に近くなる。一方、クラスタリングが上手く行われた場合、多数の仮想基準点を含む少数のクラスタが得られる。よって、 $\sum (r_c / N_c)$ の値は1よりも大きくなる。

[0045] N_fは、検索質問画像の特徴ベクトルのうち、画像データベースに登録された物体の特徴ベクトルとのマッチングが得られたものの数である。例えば、いずれの物体とも関連しない背景から抽出された特徴ベクトルは無視される。つまり、1000個の特徴ベクトルが検索質問画像から抽出され、それらのうち750個が物体Aとマッチし、600個が物体Bとマッチしたとする。この場合、物体AについてはN_f = 750、物体BについてはN_f = 600である。

[0046] ここで示した種々の好ましい態様は、択一的なものを除き、それら複数を組み合わせることもできる。

以下、図面を用いてこの発明をさらに詳述する。なお、以下の説明は、すべての点で例示であって、この発明を限定するものと解されるべきではない。

この発明による手法では、いくつかの異なるアイデアを組み合わせることによって、様々な問題に対処する。図1に処理の概要を示す。

[0047] この発明による手法では、画像データベース1に格納すべき画像および検索質問としての検索質問画像3の双方からPCA-SIFT特徴量とそのシェイプ・コンテキスト表現を抽出し（図1のステップS01、S03、S05、S11、S13、S15

参照)、画像を表現する。

[0048] 物体検出に係る第一のステップでは、検索質問画像3から抽出されたPCA-SIFT特徴量(前記S11、S13参照)と、画像データベース1に格納されている各画像に係るPCA-SIFT特徴量(前記S01、S03参照)とを比べ類似スコア(similarity score)を計算する処理(類似性処理、similarity processing、図1のステップS21、S23参照)を行う。なお、画像データベース1中の各画像のPCA-SIFT特徴量は、画像データベース1への登録時に各画像から抽出される(前記S01、S03参照)。この類似性処理の段階では、この発明の一つの特徴として提案する「基準点」に基づき仮想の基準点の位置をさらに計算する(図1のステップS25参照)。

[0049] 第二のステップとして前述のシェイプ・コンテキストから、文脈スコア(context score)を計算する(前記S05、S15およびステップS29参照)。第一のステップで得られた類似スコア、仮想の「基準点」(前記S25参照)、および第二のステップで得られた文脈スコアの値(前記S29参照)は、第三のステップであるクラスタリングステップ(図1のステップS27参照)で用いられ、局所スコア(locality score)が算出される(図1のステップS31参照)。第四のステップとして最終的に得られるスコアは、PCA-SIFTの類似スコア、文脈スコア、局所スコアから計算される(図1のステップS33参照)。

以下では、最初に局所特徴量のスコアを個別に紹介し、その後、仮想の基準点のクラスタリングについて述べ、最後にシェイプ・コンテキストについて述べる。

[0050] 1. 1 距離と個別のスコア付け(類似スコア)

我々は画像記述のための局所特徴量としてSIFTを選ぶ。SIFTは、Loweによって提案された局所特徴量の抽出手法であり(文献、D. G. Lowe. Object recognition from local scale-invariant features. In ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2, page 1150, Washington, DC, USA, 1999. IEEE Computer Society. 参照)、一つの局所特徴量は、128次元のベクトル(特徴ベクトル)で表現される。その手法

により得られる局所特徴量は識別性が高く、優れた性能を発揮することで知られている。計算速度を向上させるため、オリジナルの128次元のベクトルを、PCA (principal component analysis:主成分分析) によって次元削減した特徴ベクトルを用いるPCA-SIFTが知られており、この実施形態では36次元のPCA-SIFTを用いている。PCA-SIFTを用いると、精度も向上することが知られている(例えば、文献、Y. K. Rahul, Y. Ke, and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. pages 506-513, 2004. 参照)。SIFTおよびPCA-SIFTにより得られる局所特徴量のよい点の一つは、局所特徴量が前述したように特徴ベクトルとして表現され、局所領域の大きさと方向が特徴ベクトルによって与えられる点である。ここでいう局所特徴量の大きさと方向は、特徴ベクトルのベクトル量、即ち、ノルム(長さ)および方向と直接的に対応するものではない。つまり、特徴ベクトルのベクトル量とは別に、局所特徴量を抽出する元になった局所領域の大きさと向きを指している。

[0051] この局所領域の大きさをを用いることにより、複雑な背景の問題を解決する糸口をつかむことができる。複雑な背景から得られた局所特徴量は、通常、小さい局所領域を持つ。一方で、物体から得られる局所特徴量は比較的大きな局所領域を持つことが多い。ただし、これは単なる傾向であって、常に成り立つ訳ではない。例えば、物体の表面が細かい模様するとき、物体から得られる局所特徴量は小さい局所領域を持つ。

以下、この発明による手法を具体的に述べる。

[0052] いま、 $w(f)$ によりPCA-SIFT特徴量 f の大きさを表し、 $d(f_1, f_2)$ により、2つの局所特徴量 f_1, f_2 に対する特徴ベクトル間のユークリッド距離を表す。また、 f_q, f_d により、検索質問画像3から得た局所特徴量、画像から得た局所特徴量を表す。このとき、検索質問画像3から抽出された一つの局所特徴量と、画像データベース1中のある画像から抽出された一つの局所特徴量との類似スコア s_f は、以下で定義されるものである(前記S21、S23参照)。

[0053]

[数7]

$$s_f = \frac{w(\mathbf{f}_q) \cdot w(\mathbf{f}_d)}{d(\mathbf{f}_q, \mathbf{f}_d)} \quad (7)$$

[0054] 式(7)の定義によれば、類似スコア s_f は、検索質問画像3から得た特徴ベクトルと画像から得た特徴ベクトルの距離がより近ければより大きくなり、また、クエリー特徴ベクトル、参照特徴ベクトルの局所領域の大きさがより大きければより大きくなる。

この段階で、我々は各局所特徴量を独立に評価しており、最大スコアのものを照合結果として選択する。

[0055] 1. 2 局所的情報

この節では、PCA-SIFT特徴量をどのように用いて局所的情報を得るのかについて述べる。各局所特徴量は画像平面上で位置を持つ。もし、この情報だけを直接用いるのであれば、結果として得られるモデルは従来の星座モデルと似通ったものになってしまう。このような表現は柔軟性に欠けており、また、物体検出の際には、局所特徴量の組み合わせを全て試さなければならぬため、局所特徴量の数に対して指数関数的に計算量が増加してしまう。これに対して、この発明による局所的情報は、前述の暗黙的形状モデルと同程度の計算量に留めることができる。

[0056] この発明における局所的情報の取り扱いは以下のような着想に基づくものである。

まず、一つの基準点を画像中にとる。基準点の位置は基本的にどこであってもよいが、計算誤差の影響を小さくする観点から、物体の重心（あるいは画像の中心）に基準点を取ることが好ましい。

[0057] 局所特徴量を記録する際には、局所特徴量それ自体の大きさと方向だけではなく、基準点との相対位置、即ち、距離 t と角度 α をさらに記録しておく。基準点に対する相対位置の情報を記録しておくことで、検索質問画像と画像データベース中の画像とを照合する際に局所特徴量の相対位置を考慮するこ

とができる。

[0058] 例えば、検索質問画像の局所特徴量A1が画像データベースの局所特徴量B1と対応づけられたとする。B1の大きさと方向をA1の大きさと方向にあわせると、それに伴ってB1に対する基準点の位置も変換される。これを検索質問画像の各局所特徴量について実行すると、検索質問画像の各局所特徴量に対応付けられた画像データベース中の局所特徴量の数だけ、変換された基準点が得られる。もし、局所特徴量の対応づけが正しいものであって、同じ物体に対応付けられたものであれば、変換された基準点は同じ点を表すはずである。即ち、基準点は検索質問画像のどこか1点に集中するはずである。以上の考えに基づいて局所スコアを求めている。詳細を以下に説明する。

[0059] 1. 2. 1 基準点を用いた投票

学習時、即ち、画像データベース1へ各画像を予め登録する際には、登録する画像から抽出される各局所特徴量に対して、図2に示すように、基準点からの距離 t と角度 α を計算し画像と関連付けて登録する。

[0060] 図2の意味は以下の通りである。四角形で表された局所特徴量の領域（中心 (x, y) 、大きさ l （エル）、方向 θ ）は、基準点 (x_0, y_0) から距離 t 、角度 α の位置にある。逆にいえば、局所特徴量が与えられて、その局所特徴量の座標、大きさと方向が、それぞれ (x, y) 、 l 、 θ であるならば、それと t 、 α によって定まる位置に仮想の基準点が存在する。

[0061] 図1のステップS25の投票処理について、もう少し具体的に述べる。検索質問画像3から得た局所特徴量に対して、式(7)で定義されるスコアが最も大きい局所特徴量を画像データベース1から求める。この局所特徴量は、前述したように学習時に基準点からの距離 t と角度 α が計算されて画像データベース1に登録された画像から得られたものである。

ここで、その画像中の物体のIDを i とする。いま、検索質問画像3から得た局所特徴量の位置、大きさ、方向を、それぞれ、 (x', y') 、 l' 、 θ' とする。また、これに照合した画像データベース1中の局所特徴量の大きさ、方向を、 l 、 θ 、基準点の角度と距離を α 、 t とする。このとき、検索質問画像の座標

系において、仮想の基準点 (x_0', y_0') の存在する位置は、

[0062] [数8]

$$(x_0', y_0') = \left(x' - \frac{l'}{l} t \cdot \cos(\alpha - \theta + \theta'), y' - \frac{l'}{l} t \cdot \sin(\alpha - \theta + \theta') \right) \quad (8)$$

[0063] となる。投票は2次元空間中の仮想の基準点 (x_0', y_0') に対して行われる。具体的には、仮想の基準点の座標 (x_0', y_0') の箇所に前記物体IDに係る*i*というラベル付きの票を保存する。全ての局所特微量から投票が終了すると、後述のごとく2次元空間中でクラスタリング処理を適用する（図1のステップS27参照）。これは、実際には、すべての投票、即ち仮想の基準点がただ1点に集中することは稀なためである。

[0064] もし、検索質問画像3が画像データベース1中の同じ物体*i*を含むものであれば、仮想の基準点に対して物体*i*の票が集中するはずである。逆にいえば、物体*i*に対する仮想の基準点への投票が一定数を上回れば、それは、検索質問画像3中に、物体*i*が存在することを意味する。複雑背景の局所特微量の一部は、画像データベース1中の無関係な画像の局所特微量と誤って照合するかもしれない。しかし、このような偶然の照合結果から得られる票が仮想の基準点に集中するためには、偶然が何度も重なる必要があり、確率としては極めて低くなる。

[0065] 実際には、種々の誤差のため、票が仮想の基準点の場所に正確に集中することは希であり、仮想の基準点周辺に分散する。そこで、次に述べるクラスタリング処理（前記S27参照）によって、票をまとめてから比較する必要がある。

[0066] 1. 2. 2 票のクラスタリングと局所スコア

クラスタリングとは、与えられた空間において、点を集めるために適用される。このためには2つの関数が必要となる。一つは、類似度（あるいは距離）を計測するための関数、もう一つは2つのクラスタを併合する関数である。これらの関数が定義されると、クラスタリングアルゴリズムが適用可能となり、密な領域を発見できる。

- [0067] この発明では、Leibeらが前述の非特許文献3で用いた方法で、可能性のある票をクラスタリングすることにする。中心的なアイデアは、ランダムに選択した初期点群を用いて、相互最近傍対（reciprocal nearest neighbor (RN N) pair）を生成し、それを整列されたリストに格納し、ある定められた閾値を超えた類似度を持つペアについて、クラスタを形成するという方法にある。なお、クラスタリングの対象となる票は、物体IDのラベルを持つ点に留意すべきである。この発明では、同じラベルの票のみがクラスタリングの対象となる。いくら近い位置にあらうとも、異なるラベルの票が同じクラスタに属することはない。
- [0068] この発明における枠組みでは、クラスタリングの動作を以下のように変更して用いる。変更の目的は、同じラベルを持つ票について、より類似スコアの高いものを優先してまとめることである。このために、各クラスタは、クラスタリング値 ξ を持つこととする。このクラスタリング値は引力を表し、これが大きいとクラスタが形成されやすく、クラスタ同士の距離が互いに遠くなり、クラスタの数が少なくなる。逆に値が小さいと、クラスタが形成されにくくなる。
- [0069] クラスタリング処理の後には、各クラスタに対してクラスタ中の票の数を局所スコア r_c として算出する（図1のステップS31参照）。クラスタに含まれる票の数が多ければ多いほど、それに対応する物体の局所性が高いことになる。この段階で、一つの票しか含まないクラスタは消去される。
- [0070] 以上の処理は、前記非特許文献2, 3におけるLeibeらの暗黙的形状モデルの手法と類似している部分もある。しかし、本発明の手法が回転などを含む相似変換に対して不変な投票を行っているのに対して、Leibeらの手法にはそのような工夫が見られない点が最大の相違点である。また、本発明の手法が局所特徴量をそのまま扱うのに対して、Leibeらの手法は局所特徴量をそのまま扱うのではなく、visual wordと呼ばれる代表ベクトルに変換したのちに扱う点でも大きく異なる。
- [0071] 1. 2. 3 シェイプ・コンテキストと文脈スコア

仮定の基準点を用いた投票の考えは、これまでの手法と比べて十分な改善効果をもたらすものであるが、局所特微量間の相対的な配置の情報を捨象した処理となっている点に注意が必要である。即ち、局所スコアは、基準点と局所特微量の相対位置を規定したものであって、局所特微量間の相対位置を規定するものではない。そうすると、異なる物体が似た局所スコアを持つことがないとはいえない。

[0072] そのような懸念に対する解決手段として、シェイプ・コンテキスト(shape context)と呼ばれる方法を導入することが好ましい。シェイプ・コンテキストに基づいて、更なる評価指標である文脈スコアを導入することができる。文脈スコア u_{sc} は、ある局所特微量の周囲をみて、周囲にある局所特微量の配置が同じかどうかを評価する。つまり基準点と局所特微量の相対的な位置関係だけではなく、ある局所特微量の周囲にある局所特微量の配置関係が類似しているか否かを評価するのである。

[0073] シェイプ・コンテキストとは、位置に関する情報を保持するものであるが、星座モデルや暗黙的形狀モデルとは異なるものである。シェイプ・コンテキストは、点の具体的な局所性を保持するのではなく、空間をいくつかの局所的な小領域（ビン）に分割し、ビンに入る点の数を用いて局所性を表現する。ビン中の点の数は、点の局所的な配置を表すヒストグラムを作成するために用いられる。このような表現を用いると、点の絶対的な位置情報がビンへの帰属で近似される。

[0074] 同じビンに入っているかどうかを見ることによって点の配置関係が同じか否かを評価することができる。このような手法は、各点の相対位置を直接評価する手法に比べると各点の位置の細かい変動に対して頑健となり、より柔軟性の高い表現となる。Belongieらは、物体の輪郭点に対してこの手法を適用し、形状の検出を実現している（例えば、文献、S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell., 24(4):509-522, 2002. 参照）。

[0075] ただし、シェイプ・コンテキストは強い制約であるため、例えば、背景などから得られる対象物体と無関係な局所特徴量がシェイプ・コンテキストの領域に多数入ると照合がとれないといったように、いわゆる「外乱」に対して弱い面がある。従って、一つのビンの領域を大きくとり過ぎると好ましくない結果になる。つまり、大きな領域に対してシェイプ・コンテキストが「同じ」であることを要求することはできない。従って、限られた面積を一つのビンとしてシェイプ・コンテキストを適用する必要がある。

[0076] シェイプ・コンテキストの中心となる局所特徴量は、自身の大きさと方向を持っている。シェイプ・コンテキストの扇状の領域の大きさと方向を、前記局所特徴量の大きさと方向に整合させることで、前記局所特徴量の大きさと方向が安定している限り同じシェイプ・コンテキスト領域を得ることができる。このようにして、回転不変及びスケール不変なシェイプ・コンテキストの領域が得られる。そして、検索質問画像と画像データベース中の画像のシェイプ・コンテキストが照合可能になる。

[0077] シェイプ・コンテキストを回転不変にすることは扇形形状の採用により容易に実現できる。

一方で、スケール不変の実現手法については一考を要する。シェイプ・コンテキストの例を図3に示す。図3で区切られた空間がそれぞれビンを表す。図3はスケールの異なる2つのシェイプ・コンテキストを表している。左側のシェイプ・コンテキストは適切なスケールの場合であり、多くの特徴点が異なるビンに属している。一方、右側のシェイプ・コンテキストでは、殆どすべての点が内側の円内にあるため、点が所属するビンの数は少なくなる。したがって、このような場合、他の形状と区別することは困難となる。このようにシェイプ・コンテキストはスケールによって大きく変わるため、学習と検出の際に、同じスケールのものが照合されるように、検索質問画像と画像データベース中の各画像から適切なスケールのシェイプ・コンテキストを抽出するようしなければならない。

[0078] この発明による手法では、各PCA-SIFT特徴量について、そのシェイプ・コ

ンテキストを計算する（前記S05、S13参照）。図3の点に相当するものは、PCA-SIFT特徴を表す局所領域となる。シェイプ・コンテキストの計算にあたっては、PCA-SIFT特徴量の位置、局所領域の大きさ、方向の情報のみを用い、記述子の情報（36次元の特徴ベクトル）は用いない。ここでは、Moriらによって提案された拡張版のシェイプ・コンテキストを用いる（文献、Mori, Greg and Belongie, Serge and Malik, Jitendra. Efficient shape matching using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell., 27(11): 1832-1837, 2005. 参照）。このモデルでは、特徴点ではなく、正規化されたベクトルがビンに収められる。そして、ビンに収められたベクトルの和によってビンに対する代表ベクトルを求める。我々は、この正規化されたベクトルとして、局所特徴量の方向を表す単位ベクトルを用いる。

[0079] 図4(a)に、画像空間中に分布するPCA-SIFT特徴の局所領域とそのシェイプ・コンテキストを表す。いくつかの局所領域はシェイプ・コンテキストの外側にあるため、シェイプ・コンテキストには反映されない。局所領域の大きさの情報をここで用いる。ここでは、正規化したベクトルに局所領域の大きさの情報を用いて重みを付ける。いま、

[0080] [数9]

$$(n_{1x}, n_{1y})^T \dots (n_{ix}, n_{iy})^T \quad (9)$$

[0081] によって、局所特徴量 f_1, \dots, f_i に対する正規化されたベクトル（局所特徴量の方向を表すもの；ベクトルのノルムを1に正規化）を表し、 $w(f)$ によって局所特徴量 f の局所領域の大きさを表す。すると、ビンに対する代表ベクトル v は、

[0082] [数10]

$$v = \sum_{s=1}^i w(f_s) \cdot \begin{pmatrix} n_{sx} \\ n_{sy} \end{pmatrix} \quad (10)$$

[0083] の式で表されるものとなる。もし、ビンが物体の安定した境界線付近に配置

されれば、シェイプ・コンテキストは一定の方向を表すようになる。一方、複雑な背景から得られた局所特徴量については、方向が一定ではないため、和によって相殺され、ビンの代表ベクトルはゼロベクトルに近くなる。

[0084] 以上に述べたように、回転不変およびスケール不変なシェイプ・コンテキストを得る手法は、次のように実現される。まず回転不変については、局所特徴量の方向を用いてシェイプ・コンテキストの向きを定める。スケール不変については、局所領域の大きさを用いて定める。具体的には、シェイプ・コンテキストの外側の円の半径を、図2に示す局所特徴量のスケール l （エル）の100倍としている。

[0085] シェイプ・コンテキストの計算は、検索質問画像から得られた局所特徴量に対して、以下のように行われる。図4(b)にシェイプ・コンテキストの値として得られるヒストグラムの例を示す。このヒストグラムをベクトル SC とする。いま、 m をシェイプ・コンテキストにおけるビンの数とすると、各ビンにおいて、式(10)で定められるベクトル v を計算し、それらを連結することによって、 SC を

[0086] [数11]

$$SC = (v_{1_x}, v_{1_y}, \dots, v_{m_x}, v_{m_y})^T \quad (11)$$

と表現する。また、 u_{sc} によりベクトル SC の長さを表し、これをシェイプ・コンテキストの文脈スコアとする。

[0087] 以上のように、文脈スコア SC は、ある局所特徴量、即ち特徴ベクトルの始点を中心に配置された複数のビンをそれぞれ代表する各代表ベクトルを連結してなるベクトルの長さとして規定される。前述のごとく、物体の境界線付近に配置されたビンの代表ベクトルは長くなり、複雑な背景に配置されたビンの代表ベクトルはゼロベクトルに近くなるという一般的な特性を有することから、検索質問画像から得られた局所特徴量が複雑な背景から得られたものであるかどうかの指標を与える。

[0088] 1. 3 最終的なスコア

この発明で最も重要な点の一つは、前述のクラスタリング値 ζ をいかに定めるかである。以下の実施形態では、次のように定める。 s_f をPCA-SIFT特徴量の比較によって得られた類似スコア（1. 1節参照）とし、 $d(SC_1, SC_2)$ により、シェイプ・コンテキストから得られるベクトル（式（11））のユークリッド距離を表すものとする。 SC_q により、検索質問画像3から得られたシェイプ・コンテキスト、 SC_d により、画像データベース1中の画像から得られたシェイプ・コンテキストを表すものとする。このとき、クラスタリング値 ζ は、類似スコア s_f および文脈スコア u_{sc} を用いて、

[0089] [数12]

$$\zeta = s_f \frac{\sqrt{t} \cdot u_{sc}}{d(SC_q, SC_d)} \quad (12)$$

によって定義される。ここで、 \sqrt{t} を乗じている理由は、距離が遠くなるにつれてその特徴量が仮想の基準点を指し示す信頼性が低くなるためである。

[0090] この発明において、最終的なスコア s として、3つの態様が考えられる。第1の態様は、類似スコアだけを用いるものである。この場合、一つの物体に対して定められる最終的なスコア s は、

[数13]

$$s = \frac{\sum s_f}{N_f} \quad (13)$$

で定められる。

[0091] 最終的なスコア s の第2の態様は、類似スコアと局所スコアを共に用いるものである。この場合、スコア s は、

[数14]

$$s = \frac{\sum s_f}{N_f} \sqrt{\frac{\sum r_c}{N_c}} \quad (14)$$

で定められる。

[0092] 第3の態様では、前述の類似スコア s_f （式（7）参照）、局所スコア r_c および文脈スコア u_{sc} をすべて組み合わせて最終的なスコア s を得る（図1のステップS33参照）。 r_c を仮定の基準点のクラスタリングによって得られた局所スコアとすると、一つの物体に対して定められる最終的なスコア s は、

[0093] [数15]

$$s = \frac{\sum s_f}{N_f} \sqrt{\frac{\sum r_c}{N_c}} \sqrt{\sum u_{sc}} \quad (15)$$

で定められる。

[0094] このスコア s は、次のようなプロセスを経て適用される。

第1に、参照特徴ベクトルとクエリー特徴ベクトルを照合して、距離が最も近いものを求める。ただし、膨大な数の参照特徴ベクトルに対処すべく提案手法では近似最近傍探索を用いているため、距離が本当に「最も」近いかどうかを保証できない。

次に、その照合結果について、（15）式を用いてスコア s を決める。ここで、類似スコア s_f の算出に用いられる $w()$ は参照特徴ベクトルあるいはクエリー特徴ベクトルを抽出する元になった領域の大きさ（正方形の1辺の長さ）に比例する値である。

以上の処理によって、後述する実験例で比較対象とされた従来手法のWTAと比べて柔軟な処理が可能になっている。WTA法は最近傍がすべて同じ1票を持つが、この発明の方法は重みに応じた票を与えることができる。

[0095] ここで、式（13）、（14）、（15）中の類似スコア s_f は、検索質問画像3から得られた各PCA-SIFT特徴量について画像データベース1から得られた各PCA-SIFT特徴量との間で計算されるものである（1.1節参照）。 $\sum s_f$ は、同一物体IDに係る s_f の総和である。 N_f は、比較に用いられたPCA-SIFT特徴量のうち、いま考えている物体に照合された特徴ベクトルの数（現在、スコア s を求めようとしている物体の総得票数）である。

[0096] また、局所スコア r_c は、検索質問画像3から得られた各PCA-SIFT特徴量につ

いて最大の類似スコア s_f を与える画像データベース 1 に係るPCA-SIFT特徴量との対をそれぞれ求め、得られた各PCA-SIFT特徴量に係る仮想の基準点について物体IDのラベル付き投票を行い、同一ラベルの票についてクラスタリング処理を行うことによってクラスタを決定し、決定された各クラスタについての総得票数として規定されるものである（1. 2. 2 節参照）。

[0097] $\sum r_c$ は、同一物体IDに係る局所スコア r_c の総和である。 N_c はこの物体に対する仮想の基準点に係るクラスタの数である。文脈スコア u_{sc} は、局所特徴量が物体の境界線付近に配置されている場合は大きな値が得られ複雑な背景に配置されている場合は小さな値になる。 $\sum u_{sc}$ は、同一物体IDに係る u_{sc} の総和である。このように s_f に係る項を N_f で正規化し r_c に係る項を N_c で正規化することによって、検索質問画像 3 が小さい領域の局所特徴量を多数含む場合の悪影響を排除している。ルートをとる処理は値の影響を減らすためのものであり、本質的ではない。

最終的には、式（13）、（14）または（15）のいずれかの態様を最終スコアとし、その最終スコアの大きい順にデータベースの画像、即ち物体を並べ（図1のステップS35参照）、上位の物体を認識結果として出力する（図1のステップS37参照）。

[0098] 2. 実験結果

手法の評価においては、専門知識のない一般ユーザが本システムを利用するシナリオを考える。相手は一般ユーザであるため、物体がユーザによって切り出されている状況や、物体の大きさや位置が既知である状況、あるいは画質が十分高いという状況を仮定することはできない。すなわち、従来法で仮定されていた事項は成り立たないといえる。

[0099] 本実験では、提案システムとWinner-Take-All (WTA) アプローチの手法を比較する。WTAアプローチでは、検索質問画像 3 から得られた各局所特徴量は、その局所特徴量と最も類似している画像データベース 1 中の局所特徴量と照合され、その局所特徴量が属する物体に 1 票を与えるものである。そして、最も多くの票を集めた物体が勝者 (winner) となり、認識結果として出力さ

れる。ここでは、局所的な配置に関する情報は一切利用されていない点に注意する。なお、本実験では、検索質問画像3が、画像データベース1中に登録されている物体を含む場合のみを扱い、画像データベース1中の物体を全く含まない場合を対象としない。

[0100] 一般に、ユーザは結果を閲覧するため、正解が上位にランクされていればよいと考えられる。したがって、本実験では、第一位の結果だけが良ことを唯一の評価基準とせず、10位までに含まれているかどうかを評価基準とする。10位以降に正解が含まれている場合は、ユーザによって閲覧される可能性が低いと考えて失敗とみなす。

[0101] 実験に用いたデータは、Oxford building dataset（インターネット、〈URL : [http://www.robots.ox.ac.uk/~vgg/data/oxbuildings.](http://www.robots.ox.ac.uk/~vgg/data/oxbuildings)〉）である。画像の例を図5に示す。これらの画像はFlickrという画像共有サイトから、一般ユーザが撮影した建物の画像を集めたデータセットであるため、我々の評価の目的に合致する。このデータセットには、11個の物体（建物）に関する様々な画像が収められている。また、画像中に存在する対象物体の数は常に一つである。本実験では、これらの画像を検索質問画像3として用いた。

[0102] 一方、画像データベース1を構築する段階（学習段階）では、Oxford building datasetの画像ではなく、同じ建物についてWikipediaから得た画像、あるいはWikipediaに画像が存在しない場合にはGoogle画像検索で得た第一位の画像を用いた。画像の枚数としては、11個の建物のそれぞれに対して、1枚だけ用いた。したがって、画像データベース1のサイズは小さいものとなった。本実験では、これを第一の実験で用いた。第二の実験では、49個の新しい物体を画像データベース1に加えて、検出結果を比較した。

[0103] 実験結果を下の表1に示す。表1の第一列はOxford building datasetの11個の建物（物体）だけの画像データベース（Oxford DB）を用いたときの結果である。WTAとはこの発明に対する比較手法としての公知のWinner-Take-Allアプローチを表す。「類似スコア+局所スコア+文脈スコア」はこの発明に係る「最終的なスコア」として前述の第3の態様を適用したときの結果を表

す。第二列は、画像データベース 1 に 49 の無関係な物体のデータを加えた場合の結果を表す。第二列左側は先程と同じであり、左側はこの発明に係る「最終的なスコア」として前述の第 1 または第 2 の態様を適用した場合の結果である。具体的には、「類似スコア」は PCA-SIFT 単体での類似スコアだけを用いた第 1 の態様、「類似スコア+局所スコア」は、PCA-SIFT 単体での類似スコアに加えて、票のクラスタリングに係る局所スコアを加えた第 2 の態様の結果である。なお、「類似スコア+局所スコア+文脈スコア」は、類似スコア、局所スコアに加えて、文脈スコアを用いる第 3 の態様の結果である。

[0104] [表1]

	Oxford DB		Oxford DB + 49物体			
	WTA	本発明	WTA	本発明	類似スコア	類似スコア +局所スコア
第1位	47%	65%	4%	64%	45%	60%
第10位まで	99%	96%	77%	92%	79%	90%
失敗	1%	4%	23%	8%	21%	10%

[0105] 本実験で用いた検索質問画像 3 の数は 272 個であった。この検索質問画像 3 は、様々な角度、位置、大きさ、照明条件のもとに物体を撮影したものである。表 1 の 1 行目は、第一位の正解率を表す。2 行目は、上位 10 位までに正解が入っていた割合（10 位までの累積正解率）を表す。

[0106] 画像データベース 1 を拡張しない場合（表 1 の 1 列目）、この発明による手法（「類似スコア+局所スコア+文脈スコア」）、比較手法（WTA）ともによりよい結果を得た。上位 10 位の結果については、比較手法（WTA）の方がこの発明による手法（「類似スコア+局所スコア+文脈スコア」）より少し良かった。

[0107] しかしながら、Oxford DB を拡張し、それに 49 個の無関係な物体を追加した場合（Oxford DB + 49 物体）、表 1 の 2 列目左側に示すように、比較手法（WTA）では正解率が極端に下落した。一方、この発明による手法（「類似スコア+局所スコア+文脈スコア」）では、それほど低下することはなかった。

。比較手法（WTA）で結果が低下した理由は、画像データベース 1 に加えられた物体にある。新たに加えられた物体が他に比べて多数の局所特徴量を持っていたため、多くの票がその物体に対応付き、誤って出力された。これはシンプルなWTA法が、得票数の正規化を行っていないために生じる問題点であるといえる。

[0108] 次に、この発明による手法の最終的なスコア s の各態様の効果的な差異を見るため、態様ごとにどの程度の結果が得られるのかを計測した。結果を表 1 の 2 列目右側に示す。第 1 の態様を表す PCA-SIFT 特徴量の個別の類似スコア（類似スコアの列）は、比較手法（WTA）と比べて、主に第一位となる結果の精度を向上させていることがわかる。これに対して上位 10 位までの結果についてはあまり改善されていない。第 2 の態様を表す「類似スコア+局所スコア」の列に示されているように、仮想の基準点を用いて局所的な情報、即ち局所スコアを加味することにより、第一位の結果、上位 10 位の結果が共に改善された。第 3 の態様（表 1 の「Oxford DB+49 物体」の第二列目「類似スコア+局所スコア+文脈スコア」）は、これにシェイプ・コンテキストに係る文脈スコアを加えたものであり、さらに精度が改善されていることが分かる。ただし、シェイプ・コンテキストに係る文脈スコアを加えたことによって、それほど大きく物体のランクが変動することはなかった。

[0109] 前述した実施の形態の他にも、この発明について種々の変形例があり得る。それらの変形例は、この発明の範囲に属しないと解されるべきものではない。この発明には、請求の範囲と均等の意味および前記範囲内のすべての変形とが含まれるべきである。

[0110] この発明では、物体検出の精度を改善するため、複雑な背景の影響を緩和する手法を提案した。従来法の多くは、複雑背景の悪影響を緩和するため、人手による何らかの作業が必要なものであり、適用性に問題が残されていた。これに対して、この発明による手法はこの問題を自動的に緩和することに成功した。様々な能力を持つ異なる手法を組み合わせることによって、WTA法を明確に上回る成果を得ることができた。シェイプ・コンテキストの利用は

それほど改善効果を示さなかったが、基準点の利用は極めて効果的であった。また、処理の結果として得られる画像のランキングは、WTA法のそれと比べて意味のあるものであり、ユーザが画像の内容を把握する上でより良いヒントを与えるものである。

符号の説明

- [0111] 1 : 画像データベース
3 : 検索質問画像

請求の範囲

[請求項1] 物体を表す画像からその局所的領域に係る特徴を表す複数のベクトルを参照特徴ベクトルとして抽出し、各参照特徴ベクトルが前記画像およびその画像が表す物体の物体IDと関連付けられて登録されてなる画像データベースの中から、検索質問画像に表された物体と同一の物体を表す画像を見出すべく、

検索質問画像からその局所的領域に係る特徴を表す複数のベクトルをクエリー特徴ベクトルとして抽出する工程と、

各クエリー特徴ベクトルについて各参照特徴ベクトルとの照合を行い、両者の距離が近いほど大きく、前記クエリー特徴ベクトルが抽出された局所領域が大きいほど大きな値を有し、かつ照合される参照ベクトルが抽出された局所領域が大きいほど大きな値を有するように定められてなる類似スコアを算出する工程と、

前記類似スコアが最大となる参照ベクトルを各クエリー特徴ベクトルの類似ベクトルとしてそれぞれ決定する工程と、

前記類似ベクトルに関連付けられた物体ID別の最終的なスコアを所定の計算処理によって得、最高のスコアを与える物体IDに係る物体を少なくとも検出結果とする工程とを備え、

前記スコアは、各類似ベクトルに係る類似スコアを s_f 、前記検索質問画像の特徴ベクトルのうち、画像データベースに登録された物体を表す特徴ベクトルとのマッチングが得られたものの数を N_f とするとき、

[数1]

$$s = \frac{\sum s_f}{N_f} \quad (1)$$

により算出されることを特徴とする物体検出方法。

[請求項2] 前記類似スコアは、特徴ベクトル f が抽出された局所領域の大きさを $w(f)$ とし、クエリー特徴ベクトル f_q とそれに対応する類似ベクトル

f_d とのユークリッド距離を $d(f_q, f_d)$ とするとき、

[数2]

$$s_f = \frac{w(f_q) \cdot w(f_d)}{d(f_q, f_d)} \quad (2)$$

により算出される請求項 1 に記載の方法。

[請求項3]

前記画像データベースは、登録された各画像に対して予め定められたそれぞれの基準点からの距離および方向が、前記画像から抽出された各参照特徴ベクトルに関連付けられて登録されてなり、各類似ベクトルにつき、対応する参照ベクトルと関連付けられて格納された距離および方向を用いて前記類似ベクトルに対する仮想の基準点の位置を求める工程と、
 求めた仮想の基準点の位置が近くかつ同じ物体IDが関連付けられた類似ベクトルどうしを集めて複数のクラスタを得るクラスタリング工程と、
 前記クラスタリング工程により得られた各クラスタの類似の度合いを表す局所スコアを所定の処理によって得る工程とをさらに備え、
 前記最終的なスコアは、前記局所スコアを r_c 、各クラスタに属する類似ベクトルの総数を N_c とするとき、前式（1）に代えて、

[数3]

$$s = \frac{\sum s_f}{N_f} \sqrt{\frac{\sum r_c}{N_c}} \quad (3)$$

により算出される請求項 1 または 2 に記載の方法。

[請求項4]

前記局所スコア r_c は、各クラスタに含まれる仮想の基準点の総数として定義される請求項 3 に記載の方法。

[請求項5]

ある特徴ベクトルを中心とする領域を複数の小領域であるビンに区分し、前記特徴ベクトルの周辺にあっていずれかのビンに属する特徴ベクトルの数をビンごとに集計してなるヒストグラムとして定義され

るシェイプ・コンテキストを各参照特徴ベクトルに適用して参照特徴ベクトルのシェイプ・コンテキストを得、前記シェイプ・コンテキストに所定の計算処理を行って文脈スコアを算出し、前記シェイプ・コンテキストを各クエリー特徴ベクトルに適用してクエリー特徴ベクトルに係る文脈スコアを算出する工程をさらに備え、

前記最終的なスコアは、前記文脈スコアを u_{sc} とするとき、前式（3）に代えて、

[数4]

$$s = \frac{\sum s_f}{N_f} \sqrt{\frac{\sum r_c}{N_c}} \sqrt{\sum u_{sc}} \quad (4)$$

により算出される請求項3または4に記載の方法。

[請求項6]

前記文脈スコアは、いずれかのビンに属する周辺の特徴ベクトルにつきビンごとの和を求めて各ビンの代表ベクトル

[数5]

$$v_i = (v_{ix}, v_{iy}) \quad (5)$$

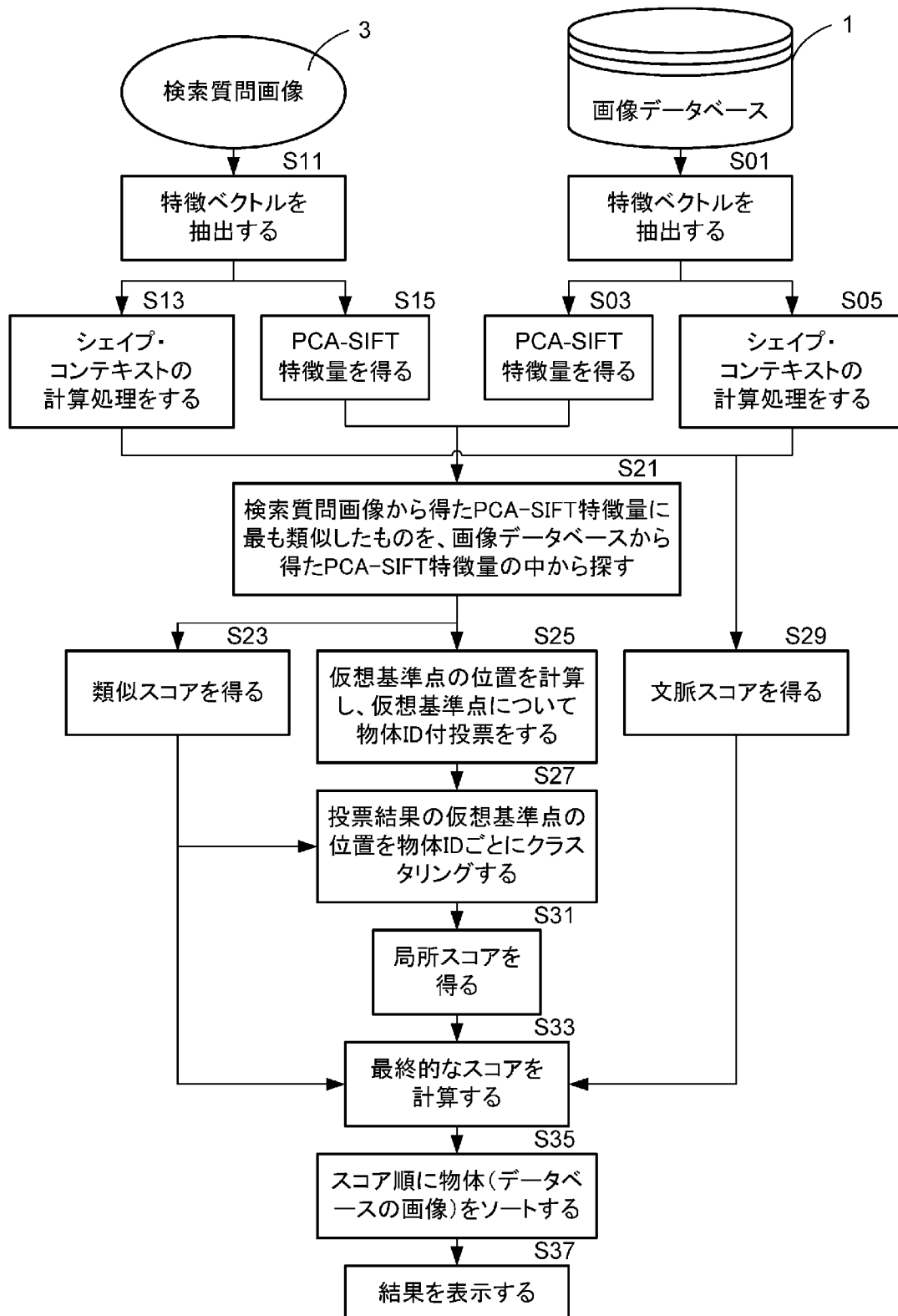
とし、 m をシェイプ・コンテキストにおけるビンの数とするとき、各代表ベクトルの要素を連結してなるベクトル、

[数6]

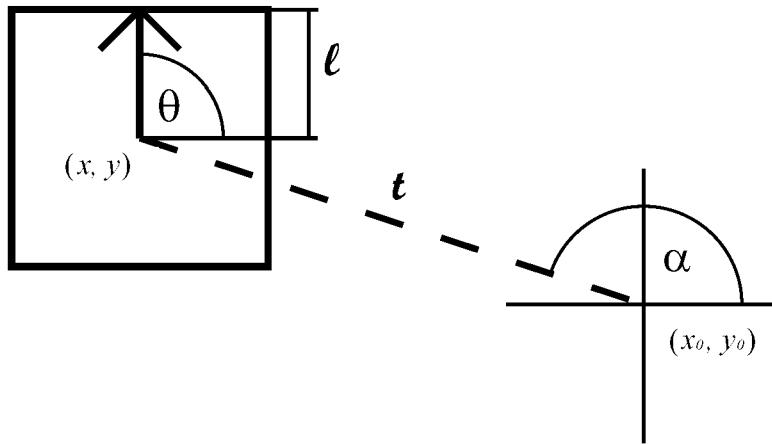
$$SC = \begin{pmatrix} v_{ix} \\ v_{iy} \\ \vdots \\ v_{mx} \\ v_{my} \end{pmatrix} \quad (6)$$

の長さとして定義される請求項5に記載の方法。

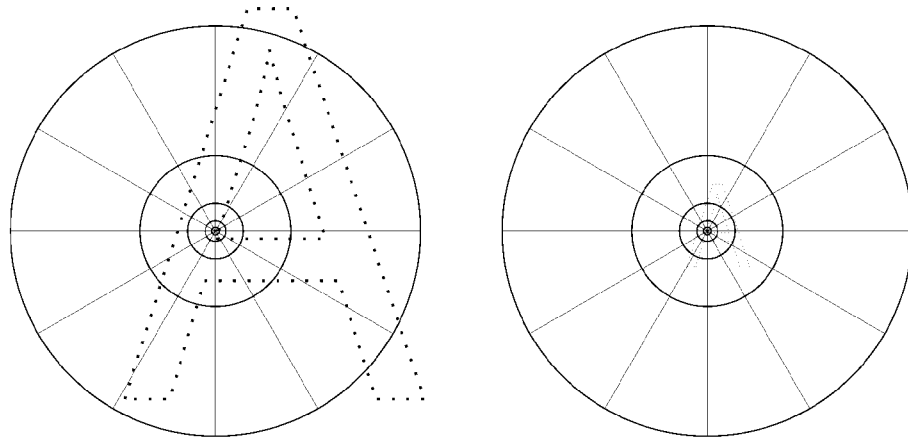
[図1]



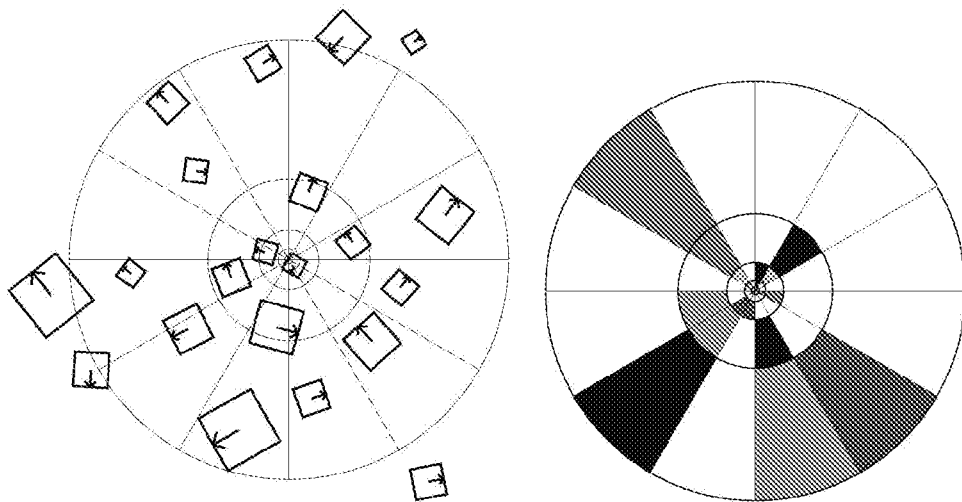
[図2]



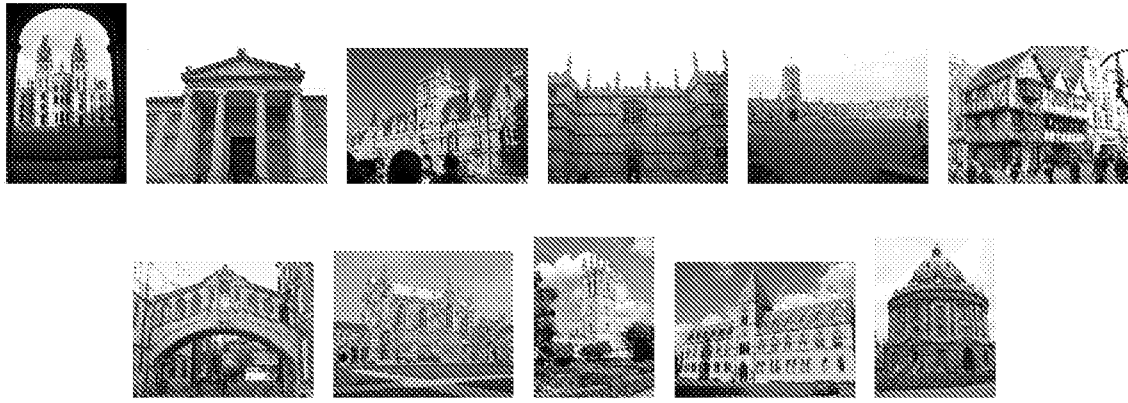
[図3]



[図4]



[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/063822

A. CLASSIFICATION OF SUBJECT MATTER

G06T7/00(2006.01) i, G06T1/00(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T7/00, G06T1/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Takayuki HONDO, Koichi KISE, "Daikibo Gazo Ninshiki no Tameno Kyokusho Tokuchoryo no Seino Hikaku", Meeting on Image Recognition and Understanding (MIRU2008), The Institute of Electronics, Information and Communication Engineers PRMU Kenkyukai, 2008.07	1-6
A	Bastian Leibe, Ales Leonardis, Bernt Schiele, Robust Object Detection with Interleaved Categorization and Segmentation, International Journal of Computer Vision, Vol.77, No.1, 2008.05.01, p.259-289	3-6
A	Greg Mori, Serge Belongie, Jitendra Malik, Efficient Shape Matching Using Shape Contexts, Pattern Analysis and Machine intelligence, Vol.27, 2005, p.1832-1837	5-6

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
03 September, 2010 (03.09.10)

Date of mailing of the international search report
14 September, 2010 (14.09.10)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

特許協力条約

PCT

国際調査報告

(法8条、法施行規則第40、41条)

[PCT18条、PCT規則43、44]

出願人又は代理人 の書類記号 0F5422PC	今後の手続きについては、様式PCT/ISA/220 及び下記5を参照すること。	
国際出願番号 PCT/J P 2010/063822	国際出願日 (日.月.年) 16.08.2010	優先日 (日.月.年) 18.08.2009
出願人(氏名又は名称) 公立大学法人大阪府立大学		

国際調査機関が作成したこの国際調査報告を法施行規則第41条(PCT18条)の規定に従い出願人に送付する。
この写しは国際事務局にも送付される。

この国際調査報告は、全部で 4 ページである。

この調査報告に引用された先行技術文献の写しも添付されている。

1. 国際調査報告の基礎

a. 言語に関し、この国際調査は以下のものに基づき行った。

出願時の言語による国際出願

出願時の言語から国際調査のための言語である _____ 語に翻訳された、
この国際出願の翻訳文(PCT規則12.3(a)及び23.1(b))

b. この国際調査報告は、PCT規則91の規定により国際調査機関が認めた又は国際調査機関に通知された明らかな誤りの訂正を考慮して作成した(PCT規則43.6の2(a))。

c. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでいる(第I欄参照)。

2. 請求の範囲の一部の調査ができない(第II欄参照)。

3. 発明の単一性が欠如している(第III欄参照)。

4. 発明の名称は 出願人が提出したものを承認する。

次に示すように国際調査機関が作成した。

5. 要約は 出願人が提出したものを承認する。

第IV欄に示されているように、法施行規則第47条第1項(PCT規則38.2)の規定により国際調査機関が作成した。出願人は、この国際調査報告の発送の日から1月以内にこの国際調査機関に意見を提出することができる。

6. 図面に関して

a. 要約書とともに公表される図は、

第 1 図とする。 出願人が示したとおりである。

出願人は図を示さなかったため、国際調査機関が選択した。

本図は発明の特徴を一層よく表しているため、国際調査機関が選択した。

b. 要約とともに公表される図はない。

第IV欄 要約 (第1ページの5の続き)

物体を表す画像からその局所的領域に係る複数の参照特徴ベクトルを抽出し、検索質問画像からその局所的領域に係る複数のクエリー特徴ベクトルを抽出する工程と、各クエリー特徴ベクトルと各参照ベクトルとの照合を行い、両者の距離が近いほど大きく、前記クエリー特徴ベクトルが抽出された局所領域が大きいほど大きな値を有し、かつ照合される参照ベクトルが抽出された局所領域が大きいほど大きな値を有する類似スコアを算出する工程と、前記類似スコアが最大となる参照ベクトルを各クエリー特徴ベクトルの類似ベクトルとしてそれぞれ決定する工程と、前記類似ベクトルに関連付けられた物体別の最終的なスコアを得、最高のスコアを与える物体を検出結果とする工程とを備え、前記スコアは、各類似ベクトルに係る類似スコアを、物体に照合された特徴ベクトルの数で割ることで算出されることを特徴とする物体検出方法。

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06T7/00(2006.01)i, G06T1/00(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G06T7/00, G06T1/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2010年
日本国実用新案登録公報	1996-2010年
日本国登録実用新案公報	1994-2010年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	本道貴行、黄瀬浩一、大規模画像認識のための局所特徴量の性能比較、画像の認識・理解シンポジウム (MIRU2008), 電子情報通信学会 PRMU 研究会, 2008.07	1-6
A	Bastian Leibe, Ales Leonardis, Bernt Schiele, Robust Object Detection with Interleaved Categorization and Segmentation, International Journal of Computer Vision, Vol. 77, No. 1, 2008.05.01, p. 259-289	3-6

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」口頭による開示、使用、展示等に言及する文献
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」同一パテントファミリー文献

国際調査を完了した日

03.09.2010

国際調査報告の発送日

14.09.2010

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

鹿野 博嗣

電話番号 03-3581-1101 内線 3531

5H

4063

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	Greg Mori, Serge Belongie, Jitendra Malik, Efficient Shape Matching Using Shape Contexts, Pattern Analysis and Machine intelligence, Vol.27, 2005, p.1832-1837	5-6