社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－AVM－54 (5)
2006／9／28

視点位置情報の利用による内容画像検索の効率化

田中 完爾† 平山 満† 近藤 英二†

情報端末ロボットが，通常の家庭用パソコンと最も大きく異なる点の一つは，自ら実世界を動き回りながら様々な視覚情報（実世界情報）を獲得できることにある．本研究では，情報端末ロボットの利便性・操作性を向上させることを目的として，ロボットの視覚画像群に対する画像検索技術を開発してきた．本論文では，視覚画像に固有の視点位置情報を利用することで，ユーザの関心に適合する画像（適合画像）および適合画像に写っている物体（適合物体）を効率よく検索する方法を提案する．また，そのために，ユーザとの間のインタラクション履歴に基づいて，適合物体位置の空間分布を推定する Query-based Occupancy Map (QOM) という新しい手法を提案する．

Improving Efficiency of CBIR by Using Viewpoint Information

KANJI TANAKA,† MITSURU HIRAYAMA† and EIJI KONDO†

Images play important role in robot user-interface systems. A set of images taken by an image sensor equipped on a mobile robot is helpful for potential users to recognize the conditions of the robot as well as its operating environments. We view such an image set as a large-scale image database, and propose an efficient image retrieval technique on the robot image database. Our technique utlizes viewpoint information of images as a cue for the retrieval, in order to search relavant objects in the robot's workspace as well as to search relavant images in the image feature space. For the purpose, we also propose a novel technique named Query-based Occupancy Map (QOM) that represents the posterior density of the relevant object locations given a history of interactions between the user and the system.

1. はじめに

情報端末ロボットが，通常の家庭用パソコンと最も大きく異なる点の一つは，自ら実世界を動き回りながら様々な視覚情報（実世界情報）を獲得できることにある．この実世界情報に対して検索・可視化などの問合せを行う，可視化インタフェースを実現することができれば，情報端末ロボットの利便性・操作性が大きく向上すると期待される．本論文では，ロボットの最も標準的な視覚である画像センサに焦点を当て，視覚画像群を対象とした画像検索問題について考える．

視覚画像は，視点位置情報付き画像の一種ととらえることができる．ロボティクス分野においては，2000年前後から，ランドマーク観測に基づく移動体測位技術（SLAM: Simultaneous Localization And Mapping）[1]が急速な進展を遂げており，視点位置を高速・高精度に推定することが可能になりつつある．その一方で，画像検索技術は，コンピュータビジョンの分野において盛んに研究がなされているが[2]，視点位置情報付き画像群を対象としたものは殆どなかった．しかし，今後，GPS 機能付きカメラの普及などにともない，視点位置情報付き画像群は，画像検索技術の重要なアプリケーションとなることが予想される[3]．

一般に，視点位置情報は，画像投影面を決定する重要なパラメータであり，画像特徴との間に相関があるだけでなく，実空間の物体位置情報をも含んでいる．本論文では，この視点位置情報を利用して，画像特徴空間および物体位置空間を探索することで，効率的に適合画像および適合物体を絞り込む方法を提案する．また，そのために，ユーザと画像検索システムの間のインタラクション履歴が与えられたときに，適合物体位置の空間分布（条件付確率密度分布）を推定するための方法として，Query-based Occupancy Map (QOM) という新しい物体位置推定方法を提案する．以上の提案手法を，代表的なサポートベクタマシン (Support Vector Machine: SVM) に基づく内容画像検索に適用し，試作システムによる実機実験およびシミュレーションによる性能検証を行った結果を示す．

2. 内容画像検索問題

先ず，本論文で扱う内容画像検索問題を定式化し，

† 九州大学
Kyushu University

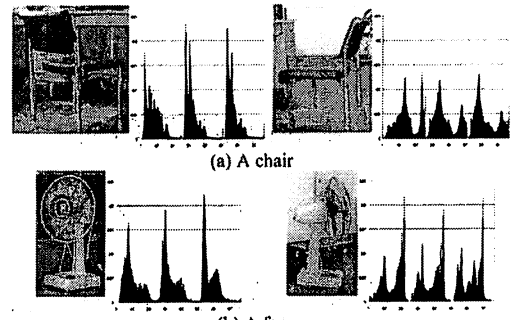
問題点を整理する．内容画像検索の目的は，システムが，ユーザとインタラクションを繰り返すことで，次第に，ユーザの関心に適合する画像（適合画像）の画像特徴を学習していくことにある．これは，画像特徴空間において，未知の適合画像と不適合画像を分離するような分類関数を学習していくことと等価である．ただし，一般的な仮定として，データベース内の全画像に対し，予め，色や模様などの特徴ベクトルが抽出されているものとする[4]．また，各データベース画像は，高々一個の物体を含むものとする．すなわち，予め，各センサ画像に対し，なんらかの物体領域分割を施し，複数のデータベース画像を生成しておく[5]．さらに，インタラクションにおいて，ユーザは嘘をつかない，あるいは，誤った判断を行わないものとする．

一回のインタラクションは，具体的に，以下のステップからなる．

(1) システムは，データベース内から一定数（N個）の例題画像を選出し，ユーザに提示する．
(2) ユーザは，各例題画像を，適合と不適合のいずれかに分類する．
(3) システムは，現時点までの例題画像と分類結果をもとに，最適な分類関数を学習する．

以上の手順からも分かるように，ステップ（1）の例題選出は，画像検索の成否を左右する，最も重要な計画問題である．すなわち，ステップ（1）において適切な例題を選出できてはじめて，ステップ（2）においてユーザから提供される情報量を最大化し，ステップ（3）において良質の学習結果を得ることが可能となる．この例題選出問題において，重要なポイントが2つある．一つは，画像特徴空間において適合画像と不適合画像のクラスタをできるだけ多く検出することである．通常，全データベース画像に占める適合画像の割合は非常に小さく，単一の大きな不適合画像クラスタと複数の孤立した比較的小さな適合画像クラスタが存在する．二つ目のポイントは，検出したクラスタの中から，分類関数を学習するのに最適な例題を選出することである．

この二番目のポイントに関しては，最適な例題選出手法が存在する．本論文では，その一手法としてサポートベクタマシン能動学習（SVM-AL: Support Vector Machine -Active Learning）[6]を利用する．この手法は，最適な2クラス分類器であるサポートベクタマシン（SVM）に基づいており，SVM本来の特性から，ステップ（3）において最適な分類関数を学習することができる．さらに，SVMの双対性を利用して，分類関数のパラメータが張るヴァージョン空間を最も大きく絞り込む例題選出手法が存在し[6]，ステップ（2）でユーザから提供される情報量を最大化できるという利点がある．


(a) A chair

(b) A fan

図1 適合画像と画像特徴の例．
Fig. 1 Example relevant images and features.

一方，一番目のポイント，すなわち，新たな適合画像クラスタを検出する問題では，アプリケーション特有の制約条件など，何らかの事前知識が必要になる．本問題の場合，3次元実環境における視点の違いにより，適合物体の多様な見え方，すなわち，複数の適合画像クラスタが存在する．図?? (a) および図?? (b) は，適合物体の2つの異なる視点からの見え方（適合画像）およびその画像特徴（色ヒストグラム）を示したものであり，視点の違いが画像特徴に大きく影響していることが分かる．異なる適合クラスタ間では画像特徴の相関が小さく，このため，新たな適合画像クラスタを検出するための手がかりとして，既に検出した適合画像との類似度を利用することは難しい．この問題に対し，本論文では，視点位置情報を事前知識として用いるアプローチを提案する．

3. 視点位置情報の利用

3.1 適合物体地図の作成

提案するアプローチでは，具体的には，視点位置情報を2つの手続きに利用する．

(1) ある時点までのインタラクション履歴が与えられたときに適合物体の空間分布（適合物体地図）を推定する手続き．
(2) 適合物体地図が与えられたときに適合画像である条件付確率が高い例題画像を選出する手続き．

この内，（2）の手続きについては，適合物体地図に記されている適合物体位置を視野範囲に含めるようなデータベース画像を探索すればよく，比較的簡単に実装できる．以下では，残る（1）の適合物体地図作成問題について考える．

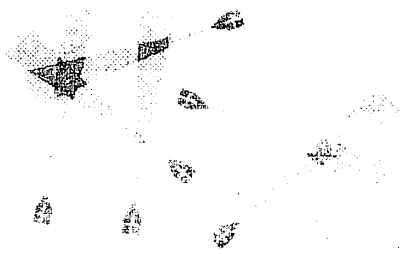


図 2　視点位置情報を用いた物体位置推定.
Fig. 2　Object localization using viewpoing information.

いま，t 回目のインタラクションにおける例題画像の画像特徴を x_t，対応する分類ラベルを y_t とし，t 回目のインタラクションを

$$H_t = (x_t, y_t) \tag{1}$$

と表記する．このとき，適合物体地図作成問題は，インタラクション履歴

$$H^t = (H_1, \cdots, H_t) \tag{2}$$

が与えられたときの適合物体位置 L の条件付確率密度 $P(L|H^t)$ を推定する問題と定式化できる．図 2 に，この確率密度分布を模式的に示す．この図では，6 つの異なる適合画像について，各視点における視野範囲を，三角形領域で示している．図中で，複数の三角形が重なり合う，色の濃い領域は，物体存在確率が高い領域とみなすことができる．以降では，この物体存在確率が高い領域を記した適合物体地図を作成する方法について検討する．

まず，簡単のために，適合物体位置 L が画像特徴 x_t に依存しないと仮定する．このとき，推定すべき条件付確率密度分布は，

$$P(L|H^t) \simeq P(L|y_1, \cdots, y_t) \tag{3}$$

となり，これは，観測結果 y_t が与えられたときの適合物体位置の条件付確率密度である．これは，ロボティクス分野において物体位置推定問題に用いられる Occupancy Map (OM)[7] により表現できる．OM は，2 次元のグリッド地図であり，このグリッドの各セルは，床面上の対応する位置に物体が存在する条件付確率を表す．OM を構築するには，物体を視野に含む画像が与えられる度に，その画像の視野範囲内にある全セルの値を一定値だけ増やすという処理を繰り返せばよい．その結果，画像群から物体位置を推定することができる．この方法を利用して，適合物体を視野に含む適合画像群から，OM を作成し，セル値が非零となる全てのセル位置を適合物体地図に記録する．以上の方法を，OM 手法と呼ぶことにする．

3.2　Query-based Occupancy Map (QOM)

3.1 節で提案した OM 手法は，既に検出した適合物体についてのみ，実空間位置を推定する．したがって，OM 手法では，同じ適合画像クラスタに属する新たな適合画像を検出することはできても，新たな適合物体を検出することは期待できない．そこで，本節では，既に検出された適合物体だけでなく，適合物体である確率の高い物体の空間分布を記録した，Query-based Occupancy Map (QOM) と呼ぶ新しい地図を提案する．QOM は，OM と同じフォーマットのグリッド地図であるが，その各セルは，インタラクション履歴 H^t が与えられたときの適合物体位置 L の条件付確率密度 $P(L|H^t)$ を表す．この確率密度 $P(L|H^t)$ は，適合画像の各仮説 I を用いて，

$$P(L|H^t) \propto \sum_I P(L|H^t, I) P(I|H^t) \tag{4}$$

のように表すことができる．式 (4) 右辺第二項の $P(I|H^t)$ は，H^t の下での適合画像 I の条件付確率であり，現時点の分類関数の出力 $M_t(I|H^t)$ で近似することができる．ただし，関数 M_t は，2 値関数であり，I を分類関数に入力した結果により，

$$M_t = \begin{cases} 1 & \text{(適合の場合)} \\ 0 & \text{(不適合の場合)} \end{cases} \tag{5}$$

のように異なる 2 種類の値をとるものとする．また，各インタラクション H_t が前回までのインタラクション履歴 H^{t-1} および適合画像 I にのみ依存することに着目すると，式 (4) 右辺第一項は，

$$P(L|H^t, I) = P(L|I) \tag{6}$$

となり，これは，適合画像 I の下での適合物体位置 L の条件付確率密度である．そこで，視野範囲 $R(I)$ 内で $P(L|I)$ が一様であると仮定し，

$$P(L|I) \propto \begin{cases} ||R(I)||^{-1} & (L \in R(I)) \\ 0 & (L \notin R(I)) \end{cases} \tag{7}$$

のように算出する．ただし，$||R||$ は，実空間における領域 R の面積を表す．以上のようにして生成される QOM の全セルをある閾値で二値化し，閾値を越えるセル群を適合物体地図に記録する．

以上の QOM による空間分布推定および例題選出は，以下の手順により，効率よく実行することができる．（1）QOM の各セルを 0 に初期化する．（2）画像データベース内の各画像 I について，$P(L|I)M_t(I|H^t)$ を算出し，視野範囲 $R(I)$ の全てのセル値に足し合わせる．（3）QOM を二値化し，適合物体地図を作成する．

4.　プロトタイプシステムの構築

提案システムのプロトタイプを構築し，移動ロボッ

(a) mobile robot (b) environment

図 3 移動ロボットと実験環境.
Fig. 3 Mobile robot and environment.

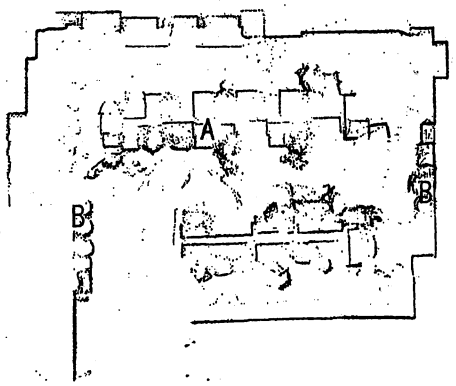
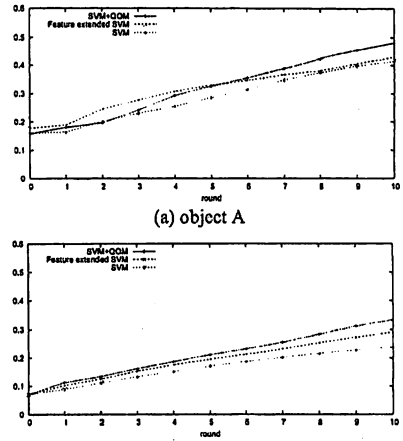


図 4 実験環境地図.
Fig. 4 Environment map.

(a) object A (PC) (b) object B (trash can)

図 5 適合画像.
Fig. 5 Relavant images.

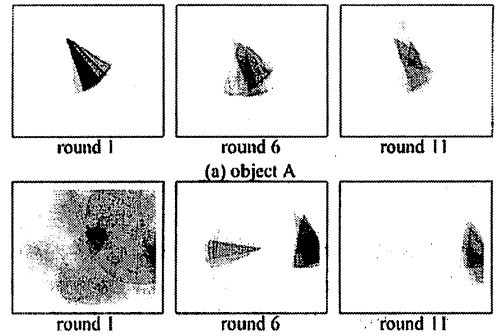
トに実装した．図 3(a) に示すように，ロボット（リバスト Pioneer3）の上に，全方位ビジョンセンサ（末陰産業 SOIOS 55-Cam）および全方位レーザ距離計（SICK LMS200）を搭載し，これらのセンサをオンボード PC と接続した．図 3(b) に示す一辺 10[m] の室内環境で，ロボットは，障害物を回避しながら自立走行し，計 298 枚の全方位画像を取得した．同時に，レーザ距離計および車輪エンコーダを利用して，各画像の視点位置を推定した．この視点位置推定の方法としては，スキャンマッチング法[1] を使用した．ただし，スキャンマッ



(a) object A

(b) object B

図 6 実験結果.
Fig. 6 Experimental results.



(a) object A

(b) object B

図 7 QOM が生成されている様子.
Fig. 7 Generating QOM.

チング法とは，最も基礎的な SLAM 技術であり，ロボット周囲のランドマークを検出・追跡することで，視点位置を推定する方法である．図 4 に，推定した視点位置列（'+' 印）およびランドマーク地図を示す．取得した全方位センサ画像に，物体領域分割を施し，計 1000 枚の水平方向視野角 40[deg] の透視投影画像を生成し，それらをデータベース画像とした．各データベース画像について，HV 色ヒストグラムを抽出し，特徴ベクトルとして用いた．

以上のようにして構築した画像データベースに対し，人間ユーザが，10 回のインタラクションからなる画像検索を実行する．ただし，本（図 5(a)）およびバケツ（図 5(b)）という 2 種類の適合物体について，画像検索をそれぞれ 100 回ずつ実行した．ただし，例題選出手法としては，2 で述べた視点位置情報を利用しな

い従来の SVM-AL（'SVM'）および 3.2 で述べた視点位置情報を利用する QOM 手法（'SVM+QOM'）をそれぞれテストした．各適合物体の正解位置を図 4 の A および B に示す．

検索性能を評価するために，標準的な尺度である，適合率（P: Precision），再現率（R: Recall）および F 値を用いる．適合率 P は，適合画像に分類された画像のうちで正しく分類されたものの割合であり，検索ノイズの少なさを表す．また，再現率 R は，データベース中の適合画像のうちで正しく分類されたものの割合であり，検索漏れの少なさを表す．F 値は，P と R の調和平均であり，総合評価に用いる．

図 6(a) および (b) に，各適合物体の画像検索の性能を，インタラクション毎に平均したものを示す．この結果から，実機実験において，視点位置情報を利用する提案手法（'SVM+QOM'）は，高い検索性能を示していることが分かる．

5. 評価実験

多様な実験環境を模擬するシミュレータを用いて性能検証を行った．このシミュレータは，計算機内に仮想的な実験環境を構築し，この仮想環境内で，ロボットは，移動観測を行い，各地点において画像特徴を取得する．ただし，各環境のサイズは，一辺 10[m] の正方形とする．また，物体の種類（物体クラス）は，50 種類とし，各クラスにつき，最低 1 最高 20 個のランダムな個数の物体を，環境内のランダムな位置に配置する．ロボットの観測地点数は 200，観測範囲は半径 3[m] の円形領域とし，画像特徴ベクトルの次元数は 100 とした．

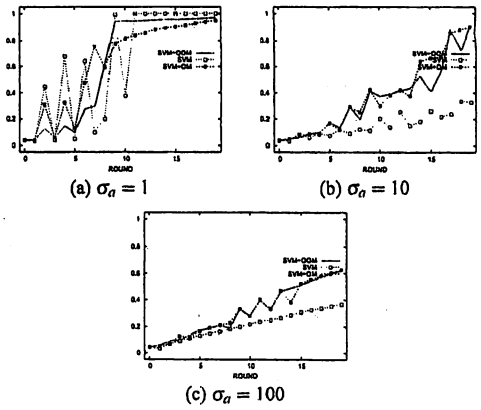
このシミュレータにおいて，物体の種類（物体クラス）および視線方向による，見え方のバラツキの大きさを表す，それぞれ σ_c および σ_a という 2 つのパラメータを導入した．この内，σ_c は，物体クラス毎の画像特徴のバラツキを表し，各クラス c の画像特徴（平均値）は，

$$f_c(c) = N(0, \sigma_c) \tag{8}$$

に従い，決定する．ただし，$N(0,\sigma)$ は，各成分の標準偏差が σ で平均 0 のガウス分布に従う 100 次元のベクトルとする．一方，σ_a は，視線方向が 1[deg] ずれる度に，どの程度見え方が変化するかを表すパラメータであり，各視線方向 a[deg] の画像特徴（平均値）は，

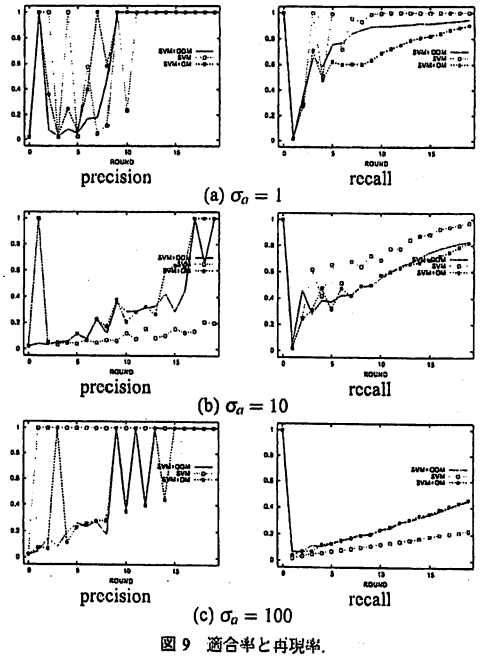
$$f_a(a) = \begin{cases} 0 & (a = 0) \\ N(0,\sigma_a) + f_a(a-1) & (a \in [1,359]) \end{cases} \tag{9}$$

に従い，決定する．さらに，センサノイズの標準偏差



(a) $\sigma_a = 1$　(b) $\sigma_a = 10$　(c) $\sigma_a = 100$

図 8　シミュレーション結果．
Fig. 8　Simulation results.



(a) $\sigma_a = 1$　(b) $\sigma_a = 10$　(c) $\sigma_a = 100$

図 9　適合率と再現率．
Fig. 9　Precision and recall.

を σ_d で表し，最終的に，画像特徴を，

$$f(c,a) = f_c(c) + f_a(a) + N(0,\sigma_d) \tag{10}$$

に従い，決定する．通常，$\sigma_c > \sigma_a, \sigma_d$ であることから，$(\sigma_c, \sigma_d) = (1000, 10)$ と定義し，σ_a の値は 1, 10, 100 の 3 通りを試した．

以上の方法で，100 通りのデータセット（環境，視点群，画像特徴群，適合画像）を構築し，各データセットについて，3 つの例題選出手法について，画像検索

を実行した．ただし，3つの例題選出手法とは，視点位置情報を利用しない従来のSVM手法（'SVM'），視点位置情報を利用するOM手法（'SVM+OM'）およびQOM手法（'SVM+QOM'）であるとした．図8に，インタラクション毎にF値の平均を求めた結果を示す．この結果から，$\sigma_a < \sigma_d$のケース（図8(a)）では，'SVM'と'SVM+QOM'の性能がほぼ同程度であり，'SVM+OM'よりも良い性能となっている．一方，$\sigma_a \geq \sigma_d$のケース（図8(b),(c)）では，'SVM+QOM'と'SVM+OM'の性能がほぼ同程度であり，'SVM'よりも良い性能となっている．

これらの結果をまとめると，'SVM+QOM'は，全てのケースについて，'SVM+OM'および'SVM'と比べて，同程度かより良い性能を示している．図9に示す適合率（precision）および再現率（recall）のグラフを用いて，その理由を考察する．

$\sigma_a < \sigma_d$のケース（図9(a)）では，適合率に関しては3つの手法がほぼ同程度の性能となっている．これに対し，再現率に関しては'SVM'と'SVM+QOM'が'SVM+OM'よりも優れており，この再現率の違いがF値の違いに表れていることが分かる．このケースは，画像特徴空間に単一の適合画像クラスタしか存在せず，SVM+ALにより，最適な例題選出が可能な状況にある．しかし，'SVM+OM'は，適合画像の空間分布に基づいて例題候補を著しく絞り込むため，最適な例題さえも例題候補から除外してしまう可能性が大きかったと考えられる．

$\sigma_a = \sigma_d$のケース（図9(b)）では，'SVM+QOM'および'SVM+OM'はほぼ同程度の適合率・再現率であるのに対し，'SVM'は適合率が著しく劣っている．このケースは，画像特徴空間において，複数の適合画像クラスタが存在するものの，適合画像クラスタ間の平均距離が比較的短いため，適合画像間の相関を利用して新たな適合画像クラスタを検出することは可能である．しかし，図9(b)の結果から，画像特徴だけでなく視点位置情報を利用した方が，より効率的に適合画像クラスタを検出可能であったと考えられる．

$\sigma_a > \sigma_d$のケース（図9(c)）では，'SVM+QOM'および'SVM+OM'はほぼ同程度の適合率・再現率であるのに対し，'SVM'は再現率が著しく劣っている．このケースは，適合画像クラスタ間の平均距離が長いため，適合画像間の相関を利用しても，新たな適合画像クラスタを検出困難な状況にある．すなわち，'SVM'は，新たな適合画像クラスタを検出することはなく，殆どの適合画像を検出できないため，再現率が著しく劣っていると考えられる．なお，このケースは，クラスタの平均サイズが小さく，'SVM'の適合率がほぼ1となっていることからも，個々の適合画像クラスタを分離する分類関数を学習することは比較的容易であったと考えられる．以上の結果から，視点画像検索問題に対する提案手法の有効性を確かめることができた．

6. 結　論

本論文では，移動ロボットによるセンサ画像群の内容画像検索問題を定式化し，センサ画像に特有の視点位置情報を利用して，画像検索を効率化する手法を提案した．特に，システムとユーザのインタラクションの履歴をもとに，ユーザの関心に適合する物体（適合物体）の実空間分布を推定することで，適合画像である条件付確率の高いデータベース画像を検出する手法を示した．本手法により，従来技術のように画像特徴空間において適合画像を探索するだけでなく，実空間においても適合物体を探索することで，より効率的に適合画像（適合物体）を絞り込むことが可能となった．

謝辞　本研究開発に対して，日本学術振興会科学研究費補助金若手研究(B)17700200，および，スズキ財団科学技術研究助成より一部助成を受けた．

参考文献

1) Thrun S. Robotic mapping: A survey. *CMU-CS-02-111*, 2002.
2) Sameer Antani, Rangachar Kasturi, and Ramesh Jain. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recognition*, 35:945–65, 2002.
3) 藤田 秀之，有川 正俊，岡村 耕二．高精度な空間情報付き写真の3次元実空間マッピング．信学論，J87-A(1):120–131, 2004.
4) Sanjoy Kumar Saha, Amit Kumar Das, and Bhabatosh Chanda. Cbir using perception based texture and colour measures. *17th International Conference on Pattern Recognition*, pages 985–988, 2004.
5) C.Dagli and T.S. Huang. A framework for grid-based image retrieval. *17th International Conference on Pattern Recognition*, pages 1021–1024, 2004.
6) Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2:45–66, 2001.
7) Tanaka K., Hirayama M., and Kondo E. Query-based occupancy map for svm-cbir on mobile robot image database. *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 2986–2992, 2005.