

局所記述子に基づく 3 次元物体認識のためのメモリ削減

局所記述子の取捨選択によるアプローチ

井上 勝文[†] 三宅 弘志[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科 〒 599-8531 大阪府堺市中区学園町 1-1

E-mail: †{inoue,miyake}@m.cs.osakafu-u.ac.jp, †kise@cs.osakafu-u.ac.jp

あらまし 画像中に存在する 3 次元物体を認識する手法として、局所記述子を用いるものがある。この手法は、大量の画像から抽出される局所記述子を用いてモデルを構築し、このモデルを用いて画像中の物体を認識する。この手法には、モデル構築に用いる局所記述子の数が莫大になり、これらを記憶しておくメモリ容量が膨大になるという問題点がある。本稿では、できるだけ少ない局所記述子で、物体を高精度で認識できるように、局所記述子を取捨選択し、メモリ容量を削減する手法を提案する。提案手法では、モデル構築用の画像を用いて物体認識シミュレーションを行い、認識率を落とす可能性の低い局所記述子をモデルに登録することで、高精度の物体認識を実現する低容量モデルを作成する。11 物体を用いた実験の結果、約 100 分の 1 に局所記述子の数を削減したモデルで、98%以上の認識率を得ることができた。また COIL-100(Columbia Object Image Library-100)を用いた実験の結果、約 6 分の 1 に局所記述子の数を削減したモデルで、96%以上の認識率を得ることができた。

キーワード 3 次元物体認識, 局所記述子, メモリ容量削減, COIL

A Memory Reduction Method for 3D Object Recognition Based on Local Descriptors

An Approach by Selecting Local Descriptors

Katsufumi INOUE[†], Hiroshi MIYAKE[†], and Koichi KISE[†]

[†] Graduate School of Engineering, Osaka Prefecture University

1-1 Gakuencho, Naka, sakai, Osaka, 599-8531 Japan

E-mail: †{inoue,miyake}@m.cs.osakafu-u.ac.jp, †kise@cs.osakafu-u.ac.jp

Abstract 3D object recognition methods based on local descriptors have been investigated. These methods employ local descriptors extracted from a lot of images to construct object models and recognize 3D objects with these models. An important problem of these methods is extensive memory requirement due to a large number of local descriptors that constitute the models. In this report we propose a simple method to reduce the amount of memory by selecting local descriptors stored in the models. The proposed method is characterized by the mechanism of selection which employs an estimate of the "contribution" of each local descriptor to object recognition. From experimental results for 11 objects, we achieved the recognition rate over 98% with about 1% of local descriptors extracted from all images for models. Experimental results with COIL-100(Columbia Object Image Library-100), show that 1/6 of the total local descriptors allowed us the recognition rate over 96%.

Key words 3D object recognition, local descriptor, reduction of memory size, COIL

1. はじめに

近年、デジタルカメラの普及・高性能化に伴い、これらの機器が新たな情報デバイスとして注目を集めてきている。また、

ハードディスクの大容量化により、個人が大量の画像データを所持できるようになってきている。これに伴い、大量のデジタル画像や動画を扱う研究が盛んに行われている。その一分野として画像中に存在する 3 次元物体を認識する研究がある。

画像から3次元物体を認識する手法は、大きく物体のクラスを認識するもの、インスタンスを認識するものに分類できる。前者は、椅子、自動車などの物体のクラスを結果として返すのに対して、後者は例えば自動車の特定のモデルなど、インスタンスを識別するものである。本稿では、後者のインスタンスの識別に焦点をあてて議論を進める。中でも特に、SIFT(Scale-Invariant Feature Transform) [1]などの局所記述子を用いる3次元物体認識に着目する。従来法の中には、様々な角度から撮影した物体の画像をもとに、局所記述子の照合を通して物体の3次元表面モデルを構築し、認識に用いるもの[2],[3]のほか、3次元モデルを用いず、画像から抽出した局所記述子をモデルとして未知の画像と照合するもの[4],[5]などがある。本研究は後者のアプローチに関する。

このようなアプローチの手法で、最も単純なものは、多様な条件で撮影された物体の画像から多数の局所記述子を抽出し、それらをすべてモデルに登録しておく手法である。この単純なアプローチの利点は、高精度な認識が容易に実現できることであるが、得られる局所記述子の数が膨大となるため、局所記述子のマッチング処理に莫大な時間がかかるという問題点と、認識する際に必要なメモリ容量が莫大になり大規模な物体認識を行うことが困難という問題点がある。

前者の問題では、局所記述子の最近傍探索の効率向上が必須となる。そこでこの問題を解決する一手法として、局所記述子の近似最近傍探索を用いる手法がある。この手法を物体認識に導入することにより、高速かつ高精度に物体を認識できることが野口ら[6]によって報告されている。一方、後者の問題では、認識に必要なメモリ容量の内、モデル容量(モデルに必要なメモリ容量)が支配的であるため、モデル容量を削減することが主な問題となる。この問題の解決策として、モデルに登録する局所記述子を取捨選択し、モデル容量を削減することが考えられる。このとき、物体の認識率を保ちつつモデルに登録する局所記述子の数を削減する必要がある。そこで本稿では、認識率を低下させる可能性の低い局所記述子を取捨選択し、高精度の物体認識を実現する低容量モデルを作成する手法を提案する。

本手法の特徴は、大量の画像を用いた物体認識シミュレーションを通して、各局所記述子が物体認識に貢献する度合い(プラス関与性)、ならびに障害となる度合い(マイナス関与性)を求め、それに基づいて局所記述子を取捨選択する点にある。提案手法を用いて11物体のモデルを構築し、認識実験を行った結果、約100分の1に削減したモデル容量で、98.7%の認識率を得ることができた。また、COIL-100(Columbia Object Image Library-100)の画像データを用い、100物体のモデルを構築し、認識実験を行った結果、約6分の1に削減したモデル容量で、96.5%の認識率を得ることができた。

2. 関連研究

局所記述子を用いた3次元物体認識法のうち、物体の3次元モデルの構築を行わない手法は、物体を撮影した画像があれば、局所記述子を抽出することによって簡単にモデルを構築できるという利点がある。このようなシンプルな手法で3次元物体認

識の精度を得るためには、様々な条件で撮影した多数の画像をモデル構築に用いる必要がある。一般に1枚の画像から数十から数千の局所記述子が抽出されるため、1つの物体のモデル化には極めて多数の局所記述子が関与することになり、その扱いが中心的な課題となる。

従来法の多くは、局所記述子をベクトル量子化して、visual wordと呼ばれる代表ベクトルに置き換える方法を用いている。未知の画像を認識する場合も、それから得た局所記述子をvisual wordに置換し、照合する。認識対象にもよるが、物体インスタンスの識別の場合は、特にvisual wordの数が増加すればそれだけ認識率も向上することが知られている。例えば、Nistèらは、1600万個のvisual wordを用いた例について報告している[4]。大量のvisual wordを用いる場合、局所記述子とvisual wordの照合に必要な計算時間が無視できず、木構造など様々なデータ構造を用いて高速化する必要がある[4],[5]。

このような大量のvisual wordを用いる手法のうち、最も極端なものは、ベクトル量子化を行わず、局所記述子の「事例」をすべて用いる手法である。このアプローチでは、高い認識率が期待できるものの、モデルの記録に必要なメモリ容量は膨大になるという問題点が生じる。

さて、認識手法の内、最も単純なものは、上記の事例となる多数の局所記述子に物体のラベルを付与しておき、質問画像から得た局所記述子との照合で、物体のラベルに投票するものである。照合は通常、最近傍探索によって行われる。このようなプロセスでは、質問画像から得た局所記述子に正しいラベルが割り当てられればよいから、すべての局所記述子を記録しておく必要はない。

すべての局所記述子を記録した場合と全く(あるいはほぼ)同じ効果を保証しつつ、不要なものを削除する方法としては、condensingと呼ばれるものが提案されている。例えば和田らは高次元空間においても効率的に適用可能な手法[7]を提案している。

さらに、多数決で物体が認識されることを考えると、すべての投票が正しくなくてもよいことがわかる。2次元平面物体の場合ではあるが、97%以上の認識率を得るために必要な照合の精度は高々15%程度であることが知られている[6]。この点に着目すると、condensingで得られるものより、さらに少数の局所記述子で物体モデルを構築可能となることが考えられる。本稿で提案する手法はこのような観点に基づくものである。

3. 提案手法

局所記述子をすべてモデルに登録すれば、3次元物体を高精度で認識できるが、局所記述子の数が膨大なため、モデルに必要なメモリ容量が莫大になるという問題点がある。本稿では、この問題を解決するために、高精度の物体認識を実現する低容量モデルを作成する手法を提案する。

3.1 考え方

3次元物体を高精度で認識できることは、物体がどのような視点から撮影された画像を質問画像としても、この画像中に存在する物体を認識することができることである。そして、前

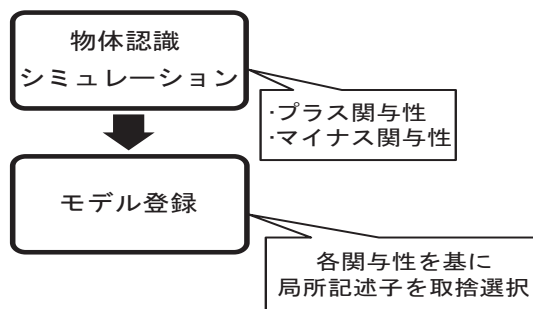


図1 処理の流れ

節で述べた投票処理に基づく認識手法に着目すると、高精度な3次元物体認識とは、どのような画像を質問画像としても、正解物体が他の物体よりも1票でも多く得票できることであると言い換えることができる。そこで本提案手法では、投票処理に基づく認識手法を用い、高精度な物体認識を実現しつつ、メモリ容量を削減する手法を提案する。

提案手法の処理の流れを図1に示す。投票処理を用いた認識手法では、モデルに登録されている各局所記述子に物体のラベルを付与しておき、質問画像から得られる局所記述子と照合し、対応付いた局所記述子に付与されているラベルを持つ物体に投票する。このため局所記述子の中には、モデルに登録しておくことで、質問画像から得られる局所記述子に正しい投票を行わずことのできるものや、誤投票を行わせるものが存在する。本研究では、物体認識に対して、質問画像の局所記述子に正しい投票を行わせるように関与することを「プラス関与性」、誤投票を行わせるように関与することを「マイナス関与性」と呼ぶ。プラス関与性を持つ局所記述子を必要最低限登録し、マイナス関与性を持つ局所記述子を可能な限りモデルに登録しなければ、高精度な認識を維持しつつ、メモリ容量を削減することができると思われる。そこで提案手法では、モデル構築用画像を用い物体認識シミュレーションを行い、どのような局所記述子が物体認識に対してプラス関与、またはマイナス関与するかを調べ、有効性を吟味するために必要な情報を得る。

局所記述子の中には、物体認識シミュレーション時に複数の画像の局所記述子に対してプラス関与するものがある。このような局所記述子をモデルに登録しておく方が、同じ枚数の画像の認識に必要な局所記述子の数が少なく済むと考えられる。そこで本研究では、このような局所記述子は、高精度な物体認識を実現する低容量モデルを作成するという目的を達成するうえで、有効な局所記述子であると考えられる。このとき、このような局所記述子の中には、一部の画像に対してマイナス関与するものがある。マイナス関与性を持つ局所記述子とは、物体認識シミュレーションにおいて誤対応を起こしたものである。このため、このようなマイナス関与性を持つ局所記述子をモデルに登録すると、未知の画像を認識する際に、誤投票を引き起こす可能性が高いと考えられる。そこで本研究では、マイナス関与性の低い局所記述子ほど、誤った投票を起こす可能性が低く、有効な局所記述子であると考えられる。

以上の考えから、提案手法は物体認識シミュレーションの結

果を基に、局所記述子の関与性とその性質を調べ、マイナス関与性を持つ局所記述子をできるだけ削除し、目的達成のために有効なプラス関与性を持つ局所記述子の組み合わせを探るものである。しかしこの組み合わせを探る問題は、大規模な組み合わせ最適化問題であるため、最適解を求めることが困難である。また、物体をどのような角度から撮影する場合にも、安定して認識できるようにするためには、あらゆる視点から得られる局所記述子を満遍なくモデルに登録する必要があり、さらに組み合わせ最適解を求めることは困難となる。そこで提案手法では、目的達成のために有効な局所記述子を Greedy にモデルに登録する手法を提案する。

提案手法では、認識対象物体を回転させて撮影した画像をフレーム画像と呼び、モデル構築用画像に用いる。そして、局所記述子の抽出には、PCA-SIFT [8] を用いる。この手法ではまず、SIFT [1] を用いて特徴点を抽出し、特徴点の近傍領域より、輝度勾配情報を基に128次元の特徴ベクトルを求める。次に、特徴ベクトルを主成分分析し、36次元の部分空間へ射影する。本研究では、この特徴点と特徴ベクトルを合わせて局所記述子と呼ぶ。モデル構築用画像から得られる局所記述子には、どの物体の画像から抽出されたかを示す物体IDと、その物体のどのフレーム画像から抽出されたかを示す画像IDと、その画像のどの局所記述子であるかを示す記述子IDを付与しておく。以下に、モデルが作成されるまでに行われる各処理について具体的に説明する。

3.2 物体認識シミュレーション

提案手法では全フレーム画像を用いて物体認識シミュレーションを行い、各局所記述子がどのフレーム画像の局所記述子と対応付いたかを求め、局所記述子が物体認識に対してプラス関与、またはマイナス関与するかを調べる。本節では、この物体認識シミュレーションについて説明する。

本研究では、物体認識シミュレーションに野口らの手法 [6] を用いる。この手法ではまず、局所記述子の近似最近傍探索を行うために、モデルを構築し、ハッシュ表にデータを登録する。次に、質問画像から局所記述子を求め、この局所記述子をクエリとし、近似最近傍探索でハッシュ表より対応する局所記述子を求める。このとき、最近傍探索に用いる距離計算には、ユークリッド距離を用いる。そして、対応付いた局所記述子が抽出された物体に投票を行う。これを、質問画像から得られる全ての局所記述子に対して行い、最も得票数の多い物体を認識結果とする。物体認識シミュレーションでは、ある1枚のフレーム画像から得られる局所記述子をクエリとし、残りのフレーム画像から得られる局所記述子を全てモデルに登録して認識を行う。このとき、クエリの最近傍となる局所記述子がクエリと同じ物体を撮影したフレーム画像から得られるものである場合、この局所記述子は、クエリの局所記述子にプラス関与する局所記述子であると言える。また、クエリの最近傍となる局所記述子がクエリと異なる物体を撮影したフレーム画像から得られるものである場合、この局所記述子は、クエリの局所記述子にマイナス関与する局所記述子であると言える。そこで物体認識シミュレーションでは、各局所記述子にどの画像の局所記述子とプラ

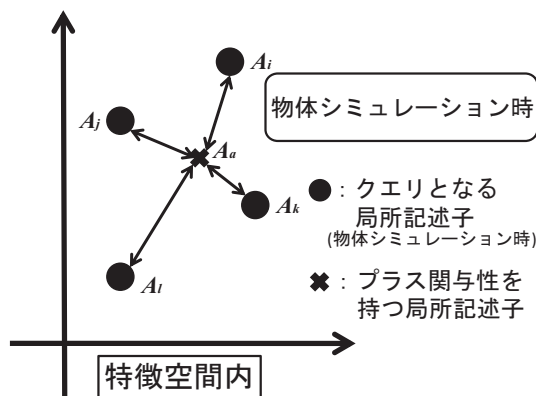


図 2 複数の局所記述子とプラス関与する局所記述子

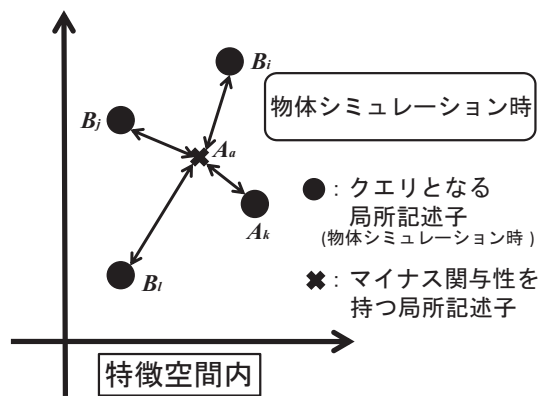


図 3 一部の局所記述子とプラス関与する局所記述子

ス、またはマイナス関与したかを記憶させ、有効性を吟味するために必要な情報を得る。以上の処理を全フレーム画像に対して行い、全局所記述子の関与性を調べる。

3.3 物体認識に対する局所記述子の関与性とその性質

3.3.1 プラス関与性とその性質

プラス関与性を持つ局所記述子の性質について説明する。物体認識シミュレーションの結果より、プラス関与性を持つ局所記述子の中には、図 2 に示すように、特徴空間内で複数の局所記述子の最近傍となるものが存在する。ここで、図 2 の A_a 等の添え字は、大文字がその局所記述子の物体 ID を、小文字がその局所記述子の画像 ID を示している。このように局所記述子の中でも、より多くの画像の認識にプラス関与できるものをモデルに登録する方が、同じ枚数の画像の認識に必要な局所記述子の数が少なくて済むと考えられる。つまり、 n 枚の画像を認識するために、1 枚にしかプラス関与できない局所記述子を n 個モデルに登録するよりも、 n 枚の画像にプラス関与する局所記述子を 1 個登録する方が、登録する局所記述子の数が少なくて済む。また、このような局所記述子を近傍の局所記述子の代表記述子としてモデルに登録すれば、メモリの容量を削減することができると考えられる。

以上の考えから本研究では、物体認識シミュレーションにおいて、より多くの画像の認識に対してプラス関与できると判断している局所記述子ほど、目的達成のために有効な局所記述子とする。そこで提案手法では、プラス関与性を持つ局所記述子の有効性を判断する基準として、何枚の画像に対してプラス関与できるかを判断基準とし、この枚数を評価値 E_1 とする。つまり、 E_1 が大きい局所記述子ほど、物体認識シミュレーションにおいて多くの画像に対してプラス関与性を持つことから、低容量モデルを作成するために有効な局所記述子と言える。但し本研究では、プラス関与性を持つ局所記述子が抽出された画像も、この局所記述子がプラス関与すると考え、評価値 E_1 が示す画像の枚数に、この画像の分も加えておく。

3.3.2 マイナス関与性とその性質

マイナス関与性を持つ局所記述子の性質について説明する。マイナス関与する局所記述子の近傍には、他の物体から得られる局所記述子が存在することを意味する。このため、このような局所記述子をモデルに登録すると、質問画像の認識の際に、

誤投票を行う可能性が高くなると考えられる。このような考えから本研究では、マイナス関与する局所記述子は、目的達成のためにあまり有効でないと考える。

しかし、図 3 に示すように、マイナス関与性を持つ局所記述子の中にも、ある画像に対しては、プラス関与性を示すものも存在する。このため、いくつかの局所記述子に対してマイナス関与性を示す局所記述子の全てが目的達成のために有効でないとはいえない。例えば、ある局所記述子が n 枚の画像に対してプラス関与し、1 枚の画像に対してマイナス関与するものとする。このとき、 n 枚の画像を認識するために必要な局所記述子の数は、この局所記述子とマイナス関与している画像に対してプラス関与している局所記述子 2 個をモデルに登録する方が、1 枚しかプラス関与しない局所記述子を n 個モデルに登録するよりも、少なくて済む。

以上の考えから提案手法では、マイナス関与性を持つ局所記述子の中で有効な局所記述子を探すために、何枚の画像に対してマイナス関与するかを判断基準として、局所記述子の有効性を求める。この枚数を評価値 E_2 とする。つまり、評価値 E_2 が小さい局所記述子ほど、局所記述子の近傍に他の物体から得られる局所記述子が存在しないことを意味する。このような局所記述子は、誤認識を起こす可能性が低いと考えられるため、目的達成に有効な局所記述子と言える。

3.4 モデルへの登録

高精度の物体認識を実現する低容量モデルを作成するためには、マイナス関与性を持つ局所記述子をできるだけ削除し、目的達成のために有効なプラス関与性を持つ局所記述子の組み合わせを探す必要がある。また、本研究で用いる物体認識システムでは、クエリ画像から得られる各局所記述子とモデルに登録されている局所記述子との対応を調べ、対応付けられた局所記述子の物体に投票を行い、最も得票数の多い物体を認識結果とする。つまり、他の物体より 1 票でも多く得票した物体が認識結果となる。このため、できるだけ少ない局所記述子で、正解物体の得票数が最も多くなるような局所記述子の組み合わせを探す必要がある。さらに、物体をどのような角度から撮影した場合にも、安定して認識するために、あらゆる視点から得られる局所記述子を満遍なくモデルに登録する必要がある。このため、最適な局所記述子の組み合わせを求めることは困難である。

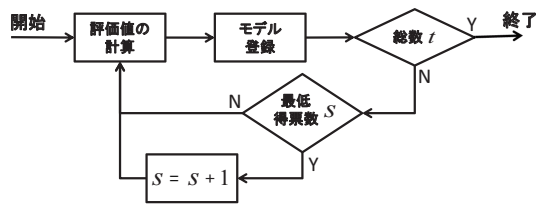


図 4 モデル登録処理の流れ

そこで本節では、物体認識シミュレーションの結果から得られる局所記述子の関与性と有効性を基に、Greedy な方法で目的を達成する局所記述子の組み合わせを求める手法について詳しく説明する。

3.4.1 処理の流れと前提条件

モデル登録処理の流れを図 4 に示す。まず、最低得票数 s を 1 に初期化する。そして、物体認識シミュレーションの結果を基に、最低得票数を満たすように、モデルに登録するための登録評価値を計算する。最低得票数 s については、後で詳しく述べる。次に、登録評価値の最も大きな局所記述子をモデルに登録する。提案手法では、モデルに登録する局所記述子の総数に閾値 t を設け、全ての物体が最低得票数を満たしたとき、モデルに登録されている局所記述子の数が t 個未満であれば、最低得票数 s を 1 ずつ増やす。そして、再び全ての物体が最低得票数を満たすように、局所記述子をモデルに登録していく。以上の処理を局所記述子の総数が t 個になるまで繰り返す。ただし、提案手法は、次に示す前提条件が成り立つと仮定して、モデルを作成する。物体認識シミュレーションで得られた E_1, E_2 は、クエリ画像以外のフレーム画像から得られる局所記述子を全てモデルに登録した場合に得られる値である。このため、一部の認識に有効な局所記述子のみをモデルに登録すると、物体認識シミュレーションで正しく対応付いていた局所記述子が、誤って対応付くようになり、 E_1, E_2 の値が変化する可能性がある。提案手法では、前提条件として、このような評価値変動の影響は小さいものと仮定し、 E_1, E_2 の値は変化しないものとしてモデル構築の評価基準として用いる。

3.4.2 登録するための評価基準 1

本節では、物体認識シミュレーションの結果から得られる局所記述子の関与性と有効性を基に、モデルに登録するための具体的な登録評価基準について述べる。

まず、局所記述子の関与性に基づいて、局所記述子をモデルに登録するための登録評価基準について述べる。物体認識シミュレーションにより、プラス関与性を持つ局所記述子の有効性を示す評価値 E_1 と、マイナス関与性を示す局所記述子の有効性を示す評価値 E_2 が求まっている。このとき、 E_1 の値が大きい局所記述子ほど、物体認識シミュレーションにおいて多くの画像の認識に対してプラス関与性を持ち、目的を達成するために有効な局所記述子である。また、 E_2 の値が小さい局所記述子ほど、この局所記述子の近傍に他の物体から得られる局所記述子が存在せず、誤認識を引き起こしにくいことから、目的を達成するために有効な局所記述子である。つまり、 $E_1 - E_2$

の値が大きい局所記述子ほど、有効な局所記述子であると言える。

次に、 $E_1 - E_2$ の値を考慮し、あらゆる視点から得られる局所記述子を満遍なくモデルに登録するための登録評価基準について述べる。どのような画像が質問画像となっても、正しく認識を行えるようにするためには、少なくともフレーム画像を全て正しく認識できる必要があると考えられる。各フレーム画像に対してプラス関与する局所記述子は、物体認識シミュレーションにおいて求まっている。この情報を基に、提案手法ではフレーム画像を全て正しく認識できるようにモデルを構築する。そこで、ある局所記述子をモデルに登録することで、どのフレーム画像を認識することができるようになったかがわかるように、物体ごとに画像表 $X = (X_1, \dots, X_N)$ を設定する。ここで N は、ある物体のモデルを構築する際に用いた画像の枚数で、画像表 X は、 ID_j の画像をクエリとしたときに、現在登録されているモデルで何票他の物体より多く票を獲得できるかを示すものである。このとき、ある局所記述子をモデルに登録するごとに、物体認識シミュレーションにおいて、この局所記述子がプラス関与するフレーム画像に対して 1 票ずつ投票を行う。これは、この局所記述子をモデルに登録することで、この局所記述子がプラス関与するフレーム画像が 1 票正解票を獲得することができたことを意味する。また、登録した局所記述子がマイナス関与性も持つものである場合、物体認識シミュレーション時に誤って投票を行ってきた物体の画像に対する画像表から 1 票減票する。これは、ある画像が 1 票減票されたならば、その画像に関して正しく認識するためには、2 票以上正しく認識できる票を得ることができるよう局所記述子をモデルに登録する必要があることを意味する。このため、あらゆる角度から撮影された場合でも正しく認識するためには、画像表に票が入っていないフレーム画像や、画像表の表がマイナスのフレーム画像を優先的に認識する必要がある。

そこで提案手法では、画像表の最低得票数 s を $s = 1$ と設定し、全ての画像表の得票数が s 以上になるように、登録評価基準 E を以下のように設定する。

$$E = E_1 - E_2 - E_3 \quad (1)$$

ここで、 E_3 は、ある局所記述子がプラス関与する画像の内、現時点で画像表の値が s 以上の画像の枚数を表す。提案手法では、この登録評価基準 E を現時点でモデルに登録されていない全局所記述子に対して求め、最も値の大きい局所記述子から順に Greedy にモデルに登録する。

具体例を図 5 に示す。物体 A から抽出された局所記述子 f が、物体認識シミュレーション時に物体 A の画像 $ID1, 2, 4, 5$ に対してプラス関与し、物体 B の画像 $ID1, 3$ の画像に対してマイナス関与する局所記述子であるとする。この f をモデルに登録したとすると、図 5 の (1) に示すように、物体 A の画像表 X^A に対して投票を行い、物体 B の画像表 X^B に対して減票を行う。次に、物体 A から抽出された局所記述子 g が、物体認識シミュレーション時に物体 A の画像 $ID1, 2, 3$ に対してプラス関与し、物体 B の画像 $ID1$ の画像に対してマイナス関与する局所記述子であるとする。ここで、 g がプラス関与でき

f		g	
物体 ID	画像 ID	物体 ID	画像 ID
A	1	A	1
A	2	A	2
A	4	A	3
A	5	B	1
B	1		
B	3		

正しく認識できる画像 ID

誤認識を起こす画像 ID

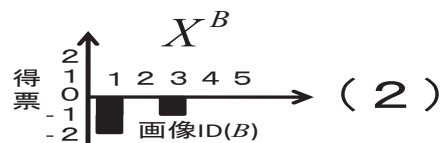
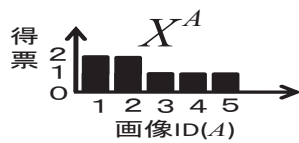
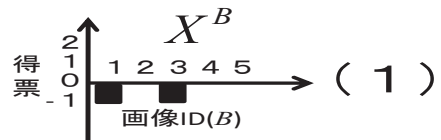
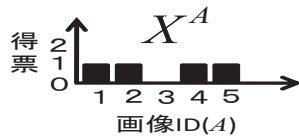


図 5 Greedy な特徴点の選択手法

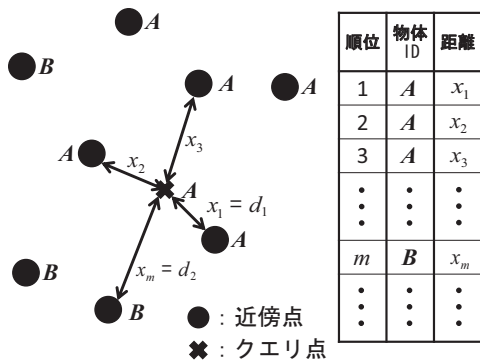


図 6 近傍点との距離計算

る画像 ID の内、現時点で作成されているモデルでプラス関与できない画像 ID は、物体 A の画像 ID3 のみである。よって、 g の新たな評価値 E_3 は、 $E_3 = 1$ となる。ここで、仮に g を次にモデルに登録したとすると画像表は図 5 の (2) のようになる。

3.4.3 登録するための評価基準 2

提案手法では、 E の値を評価基準として局所記述子をモデルに登録する際に、 E の値が等しい局所記述子が存在する場合がある。本節では、このような場合に局所記述子をモデルに登録する順番を決める評価基準について説明する。

提案手法では、物体認識シミュレーション時に、近似最近傍探索において求まるクエリの局所記述子の全近傍点と、クエリの局所記述子との距離を計算して保存しておく。この距離計算により、クエリとなる局所記述子に対して、特徴空間内でクエリと同じ物体 ID を持つ局所記述子との距離が近ければ、この局所記述子は正しく投票を行う可能性が高くなると考えられる。また、特徴空間内でクエリと異なる物体 ID を持つ局所記述子との距離が遠ければ、誤投票を行う可能性が低くなると考えられる。そこで提案手法では、この距離情報を用い、局所記述子をモデルに登録する順番を決める。以下に具体的に説明する。

提案手法では、図 6 に示すように、物体シミュレーションにおいて、クエリとなる局所記述子の全近傍点とクエリとなる局所記述子との距離を計算する。このとき、クエリとなる局所記述子の物体 ID と、この局所記述子の最近傍となる局所記述子の物体 ID が等しいとき、この局所記述子間の距離を評価値 d_1 とする。図 6 では、 x_1 を d_1 とする。この d_1 の値が小さいも

のほど、 E が等しい局所記述子の中でも、よりクエリとなる局所記述子の近傍に同じ物体 ID を持つ局所記述子が存在することを意味する。このため、他の物体を映した未知の画像から得られる局所記述子が、見つかった最近傍となる局所記述子よりも、クエリの最近傍となる可能性が低いと考えられる。また、クエリとなる局所記述子と異なる物体 ID を持つ局所記述子の中で、最も距離の近い局所記述子との距離を評価値 d_2 とする。図 6 では、 x_m を d_2 とする。この d_2 の値が大きいものほど、クエリとなる局所記述子の近傍に他の物体 ID を持つ局所記述子が存在しないことを意味する。このため、同じ物体を映した未知の画像から得られる局所記述子が、他の物体 ID を持つ局所記述子よりも、クエリの最近傍となる可能性が高いと考えられる。そこで提案手法では、この d_1 と d_2 を用い、

$$D = \frac{d_2}{d_1} \quad (2)$$

を、局所記述子の登録評価基準として設定し、 E の値が等しい時、評価値 D の値が大きい順にモデルに登録する。ただし、 D の値を求める際に以下の 3 つのことを考慮する必要がある。

第一に、ハッシュを用いて近似的にクエリとなる局所記述子の近傍点を求めるため、近傍に他の物体 ID を持つ局所記述子が存在しない場合がある。このような局所記述子は、モデルに登録しても誤認識を起こす可能性が低いと考えられる。そこでこのような場合、 d_2 の値を ∞ とする。第二に、クエリとなる局所記述子と、最近傍となる局所記述子の物体 ID が異なる場合がある。このような局所記述子は、モデルに登録すると誤認識を起こす可能性が高いと考えられる。そこでこのような場合、 $D = 0$ とする。第三に、クエリとなる局所記述子の近傍となる局所記述子が全く求まらない場合がある。このような局所記述子は、大量にあるフレーム画像のどの画像の認識にも関与していないことを意味する。このため、このような局所記述子は、認識率に影響を及ぼす可能性が低いと考えられる。そこでこのような場合、 $D = 0$ とする。

3.4.4 モデルへの登録手法

提案手法では、前述のモデル登録基準 E と D の大きいものから順にモデルに登録する。ここで、画像表の値が全て s 以上になったとき、モデルに登録されている局所記述子の数が t 個未満であれば、より安定して物体を認識できるように、画像表の最低得票数 s の数を 1 ずつ増やし、全ての画像表の値が s 以

物体名	人形1	人形2	人形3	人形4	人形5	人形6	デジタルカメラ	ホットキス	マウス	コップ	腕時計
物体											
画像枚数[枚]	1166	1254	1191	1174	1157	1151	1189	1222	1184	1210	1188
平均抽出数[個]	242	205	144	141	332	261	145	168	64	206	224

図 7 フレーム画像枚数及び、1 フレーム画像当たりの局所記述子の平均抽出数



図 8 COIL-100 のフレーム画像の例

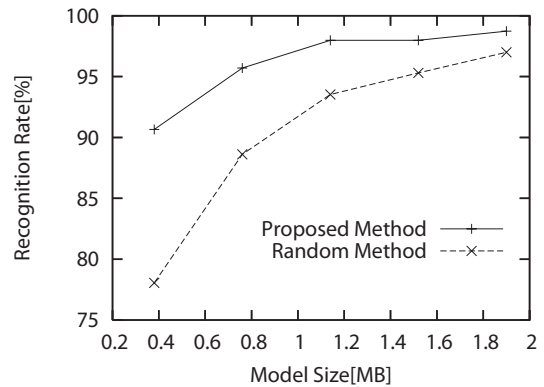


図 9 実験結果 (11 物体)

上になるように、 E 、 D の値の大きい局所記述子から順に登録する。この処理をモデルへの局所記述子の登録数が t 個に達するまで繰り返す。

4. 実験

本実験は、図 7 に示す 11 物体と、COIL-100 を用いて行った。まず、11 物体に関して行った実験で用いたデータセットについて説明する。本実験では、モデル構築に用いた画像として、3次元物体を一回転させ、ハイビジョンビデオカメラ^(注1)で撮影した動画から得られる全フレーム画像を用いた。全フレーム画像は 13086 枚であり、得られた局所記述子数は約 250 万個、全モデル容量は 191.3MB である。無削減状態のモデルを用いた場合の物体の認識率は 100% である。図 7 に用いた物体、各物体のモデル作成に用いたフレーム画像の枚数と、1 枚のフレーム画像より抽出される局所記述子の平均抽出数を示す。クエリ画像としては、モデル構築用の画像を撮影した時と照明条件を変え、3次元物体を 10° ずつ回転させて撮影したものを、1 物体あたり 36 枚用意した。

次に、COIL-100 に関して行った実験について説明する。COIL^(注2)とは、100 物体を 5° ごとに回転させて撮影した画像データベースである。画像の解像度は 128×128 である。本実験では、 5° ごとに撮影されている 72 枚の画像の内、角度が



図 10 11 物体で認識に失敗した質問画像

$0^\circ, 10^\circ, 20^\circ, \dots, 350^\circ$ の画像一物体あたり 36 枚をモデル構築用の画像に用いた。クエリ画像としては、角度が $5^\circ, 15^\circ, 25^\circ, \dots, 355^\circ$ の画像一物体あたり 36 枚を用いた。モデル構築用の全画像から得られた局所記述子数は約 18 万個で、そのときのモデルの容量は 13.6MB である。無削減状態のモデルを用いた場合の物体の認識率は 98.86% である。フレーム画像 1 枚当たりから抽出される局所記述子数の平均は 100 物体で約 50 個で、最も多い物体で平均 174 個、最も少ない物体で平均約 2 個であった。図 8 に用いたフレーム画像の例を示す。図 8 の上段のフレーム画像の物体は、局所記述子数の平均個数の最も多かった 3 物体を左から順に示している。また、図 8 の下段のフレーム画像の物体は、局所記述子数の平均個数の最も少なかった 3 物体を左から順に示している。

両実験において 3 次元物体認識手法には、野口らの手法を用いた。提案手法で行う物体認識シミュレーションと本実験では、用いたパラメータは同じである。

(注1): SONY HDR-HC1, 解像度:740×480

(注2): <http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php>

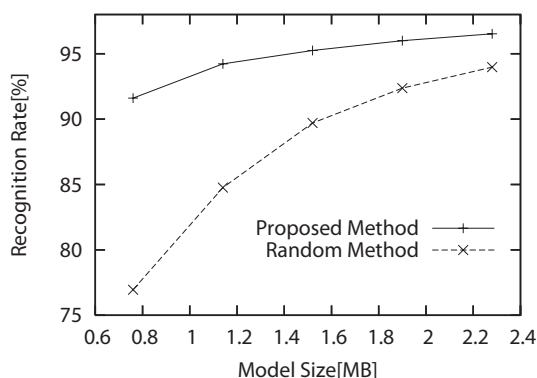


図 11 実験結果 (COIL-100)

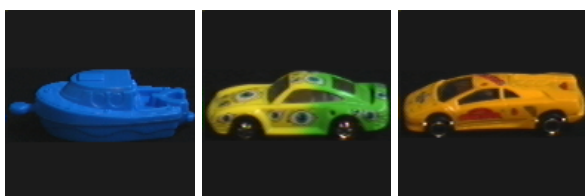


図 12 COIL-100 で認識率 80%未満の物体

本実験では、提案手法を用いて作成したモデルと、全局所記述子からランダムに取捨選択したモデルを用い、認識率を比較した。ランダム的手法では、両実験において、登録総数 t ごとにモデルを 10 種類作成した。

11 物体を用いて行った実験の結果を図 9 に示す。ただし、ランダム手法は 10 モデルの平均認識率を示している。モデル容量は、局所記述子そのものを記憶するために必要なメモリ容量と、その局所記述子の物体 ID を記憶するために必要なメモリ容量の合計である。実験結果より、提案手法はランダムに選択したモデルより、容量が少なくなればなるほど有効であることがわかった。ここで、認識率が 98.73% と最も高かった $t = 25000$ のモデルの結果について注目する。このモデルでは、「人形 2」と「腕時計」でそれぞれ 1 枚誤認識を起こし、「人形 6」で三枚誤認識を起こした。図 10 に認識に失敗した質問画像を示す。誤認識を起こした画像は、正解物体よりも他の物体への投票が多い、あるいは全く得票できないといった理由で誤認識していた。正解物体への票より、他の物体の票の方が多くなるという問題は、局所記述子の取捨選択により、登録された各局所記述子の最も距離の短い局所記述子が、無削除状態のときから変化するために起こると考えられる。

次に、COIL-100 を用いて行った実験結果を図 11 に示す。ただし、ランダム手法は 10 モデルの平均認識率を示している。モデル容量は、局所記述子そのものを記憶するために必要なメモリ容量と、その局所記述子の物体 ID を記憶するために必要なメモリ容量の合計である。ここで、認識率が 96.52% と最も高かった $t = 30000$ のモデルの結果に注目する。このモデルを用いた場合認識率が 90% 未満となった物体が 16 物体あった。この内、認識率が 80% 未満の物体を図 12 に示す。これら 3 物体の内、「obj15 (図 12 右図)」が最も認識率が低かった。「obj15」で誤認識を起こしていた質問画像は、他の物体の方が得票数が

多かった質問画像が 7 枚、全く票を得ることができなかった質問画像が 1 枚存在した。また 16 物体の内、7 物体が車関係の物体であった。この 7 物体の質問画像で、誤認識を起こした画像の認識結果を調べたところ、誤認識を起こした質問画像 40 枚中、他の車と誤認識を起こした質問画像は 1 枚しかなかった。認識率の低かったこれらの物体で、誤認識を起こした原因の多くは、質問画像から得られる局所記述子の数が少なく、全く得票することができなかったことが挙げられる。このため、得られる局所記述子の数が少ない物体に対して、多くの局所記述子をモデルに登録し、得票しやすくなるような評価基準を設定する必要があると考えられる。

5. ま と め

本稿では、局所記述子を用いた 3 次元物体認識において、できるだけ少ない局所記述子で物体を高精度で認識できるように、局所記述子を取捨選択し、モデルに登録する局所記述子の総数を減らすことで、メモリ容量を削減する手法を提案した。11 物体を対象とした実験の結果、容量を約 100 分の 1 に削減したモデルで 98.73% の認識率を得た。また、COIL-100 を対象とした実験の結果容量を約 6 分の 1 に削減したモデルで 96.52% の認識率を得た。

今後の課題として、認識に用いる物体数を増やす、モデル構築に用いる画像の枚数を減らすといったことが挙げられる。また、得られる局所記述子が少ない物体に対して、多くの局所記述子をモデルに登録し、得票しやすくなるような評価基準を設定するといったことが挙げられる。

6. 謝 辞

本研究の一部は、科学研究費補助金 (基盤研究 (B) 19300062) の補助による。

文 献

- [1] D.Lowe: "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91–110 (2004).
- [2] F.Rothganger, S.Lazebnik, C.Schmid and J.Ponce: "3D Object Modeling and Recognition from Photographs and Image Sequences", Ponce et al., Eds., Toward Category-Level Object Recognition, LNCS4170, Springer, pp. 105–126 (2006).
- [3] D. Lowe: "Local Feature View Clustering for 3D Object Recognition", Proc. CVPR2001, Springer, pp. 682–688 (2001).
- [4] D. Nistér and H. Stewénus: "Scalable Recognition with a Vocabulary Tree", Proc. CVPR2006, pp. 775–781 (2006).
- [5] Š. Obdržálek and J. Matas: "Sub-linear Indexing for Large Scale Object Recognition", British Machine Vision Conference (BMVC), pp. 1–10 (2005).
- [6] 野口和人, 黄瀬浩一, 岩村雅一: "近似最近傍探索の多段階化による物体の高速認識", 画像の認識・理解シンポジウム (MIRU2007) 論文集, OS-B2-02, pp. 111–118 (2007).
- [7] 加藤丈和, 和田俊和: "近接性グラフに基づく効率的 condensing のアルゴリズムと評価", 信学技報 PRMU, Vol. 103, No. 96, pp. 19–24 (2003).
- [8] Y.Ke and R.Sukthankar: "PCA-SIFT: A more distinctive representation for local image descriptors", Proc. of CVPR2004, Vol. 2, pp. 506–513 (2004).