物体の認識率と局所記述子の照合精度の関係

2 項分布に基づく投票モデルと実験的検討

野口　和人†　　黄瀬　浩一†　　岩村　雅一†

† 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: †noguchi@m.cs.osakafu-u.ac.jp, ††{kise,masa}@cs.osakafu-u.ac.jp

あらまし　局所記述子を用いた物体認識の最も単純な手法は，物体モデルに保存された局所記述子と未知の画像から抽出した局所記述子を最近傍探索によって照合し，投票するものである．一般に，局所記述子の数は膨大であるため，このアプローチの鍵は，いかに照合の計算コストを削減するかにある．近似最近傍探索の技術を用いると，照合に必要なコストが劇的に削減されるが，どの程度の照合精度があれば，高精度な物体認識が可能なのかという疑問点が残る．本稿では，この点について次の 2 つの成果を示す．まず，局所記述子の照合精度はそれほど高くなくてもよいことを実験的に示す．例えば，10 万物体に対して，97%の認識率は 14%の照合精度で得られる．もう一つは，単純な 2 項分布を用いた，投票による認識のモデルを提案し，実験結果と良く合うことを示す．

キーワード　物体認識，局所記述子，近似最近傍探索，投票，２項分布

Relation Between the Object Recognition Rate and the Matching Accuracy of Local Descriptors

A Binomial Voting Model and Its Experimental Evaluations

Kazuto NOGUCHI†, Koichi KISE†, and Masakazu IWAMURA†

† Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Naka, Sakai, Osaka 599-8531, Japan
E-mail: †noguchi@m.cs.osakafu-u.ac.jp, ††{kise,masa}@cs.osakafu-u.ac.jp

Abstract　A simple method of object recognition with local descriptors is voting by matching local descriptors with nearest neighbor (NN) search. Because the number of local descriptors is so large, a key technology for this approach is to reduce the computational cost of matching. It is known that the cost of NN search can be drastically cut by approximation, though it poses another question of how accurate the matching of local descriptors should be for accurate object recognition. The major contribution of this report is twofold. First, we experimentally show that it is not necessary for matching to be accurate. For example, an object recognition rate of 97 % with 100,000 object models is achieved by the matching accuracy of only 14%. Secondly, we propose a model of recognition by voting based on a simple binomial distribution, which agrees with experimental results.

Key words　Object recognition, Local descriptors, Aproximate nearest neighbor, Voting, Binomial distribution

1. ま え が き

SIFT(Scale-Invariant Feature Tarnsform) [1] に代表される局所記述子を用いた物体認識が注目を集めている [2]. このような認識手法のうち，最も単純なものは，物体モデルと未知画像の局所記述子を照合し，物体に投票するものである．認識結果は，最大得票数を得た物体となる．局所記述子を用いた物体認識には，隠れや照明変動に対するロバスト性という利点がある反面，局所記述子の数が多い（画像あたり数百から数千）ため，照合にかかる計算量が膨大であるという問題点もある．

局所記述子を照合する一つの手法は，最近傍探索によるものである [3]. 未知画像の局所記述子は，物体モデルの局所記述子

のうち，最も距離の短いものに対応つけられる．これにより，局所記述子を単位として，物体への票が得られる．局所記述子の数が膨大であることを考えると，このようなアプローチの鍵は，いかに最近傍探索を高速化するかにある．

物体認識に限らず，最近傍探索は様々な場面で用いられる基本的な処理であるため，これまでにも多数の効率化手法が提案されてきた．その一つは，最近傍探索に近似を導入するものであり，例えば，ANN (approximate nearest neighbor) [4] や LSH (locality sensitive hashing) [5] が知られている．近似最近傍探索では，最近傍を見つける確率が下がることと引き替えに，劇的な高速化を実現することができる [6]. したがって，ここでの問題は，物体認識で一定の認識率を達成するために，どの程度の照合精度が必要なのかという点となる．

本稿で述べるポイントの一つは，「照合はそれほど頑張らなくても，高精度な物体認識は達成可能である」という点にある．例えば，写真，ポスター，本の表紙などといった平面物体の認識の場合，10 万物体のモデルを用いた場合の認識率 97%が，局所記述子の照合精度 14%で実現できる．これは主に「投票」という処理の効果によるものである．ある程度多数の物体モデルを用いる場合，近似最近傍探索によって得られる票は，正解の物体に集中しやすく，誤った物体には集中しない．本稿のもう一つのポイントは，上記のような「投票による物体認識」のモデルを提案することである．提案モデルは投票を 2 項分布でモデル化するものである．ただし，2 項分布だけでは単純すぎて実際のデータとは合致しないため，本モデルでは，局所記述子数の分布，局所記述子の照合精度の分布の 2 つを用いて 2 項分布モデルを改訂する．これにより，実験結果とよく合致するモデルが得られる．

本稿の構成は以下の通りである．まず，**2.** において局所記述子を用いた物体認識ならびに近似の導入について簡単にまとめる．次に，**3.** において，2 項分布に基づく投票モデルを提案する．**4.** では，10 万個の物体モデルと 2000 の未知画像を用いた実験により，提案した物体認識の投票モデルが，実際の実験データと良く合うことを示す．

2. 局所記述子を用いた物体認識

2.1 局所記述子

局所記述子は，撮影角度や照明変動に対してあまり影響を受けない形で画像の局所領域を記述するものである．記述の対象領域が局所に限られるため，隠れに対しても自然に頑健な認識が実現できる．

これまでに多数の記述子が提案され，優劣が検証されてきた [7]. その中でも，SIFT は性能がよく，広く使われている記述子である．本稿では，高速化のためには低次元化も重要であるとの観点から，SIFT で得られる局所領域の特徴に対して主成分分析を施した PCA-SIFT [8] を用いる．

2.2 最近傍探索と投票による認識

局所記述子を用いて物体認識を実現する手法は幾つかのアプローチに分類できる．

最も一般的なアプローチは，局所記述子の配置を捨象した

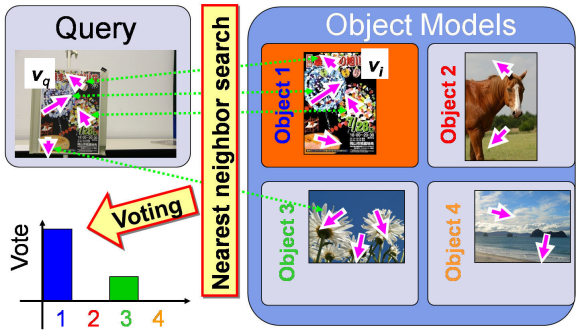


図 1　投票による物体認識.

“bag of words” あるいは “bag of features” と呼ばれるものである．このアプローチでは，まず，サンプル画像から得た大量の局所記述子をベクトル量子化し，“visual word” と呼ばれる符号語を定める [9]. そして，visual word を用いて画像を索引付けする．具体的には，出現の有無やその頻度を特徴として捉える．このようなアプローチは，情報検索で用いられているものと基本的に同じであり，そのため，一旦，索引が得られると，非常に効率的な照合処理が実現可能である．画像を対象とした手法では，さらに木構造を用いて照合などを高速化した手法 [10] や，visual word の定め方に関する様々な検討が行われている [11].

このようなアプローチでは，物体認識の精度や速度は，visual word の識別性や安定性（再現性ともいわれる）に大きく影響を受ける．もし，visual word が粗い量子化によって定められていると，識別性が十分ではなくなるため，物体認識の精度が低下する．識別性は量子化を細かくすれば改善するが，今度は安定性に問題が生じる．すなわち，元の局所記述子が十分似ていたとしても，同じ visual word が割り当てられるとは限らず，結果として，照合できないことになる．

以上の問題は，visual word を用いない手法によって解決できる．visual word を用いない手法とは，(1) visual word ではなく元の局所記述子そのものを物体モデルに用い，(2) 局所記述子の類似度あるいは距離を計測して，照合する，という手法である．ここで (2) のプロセスは，最近傍探索と呼ばれるものである．

最近傍探索に基づく最も単純な認識法は，投票によるものである．図 1 に概要を示す．以下では，認識対象となる未知の画像を検索質問，検索質問から得られる局所記述子の特徴ベクトルを検索質問ベクトルと呼ぶ．また，物体モデルは，元となる画像から抽出された局所記述子の集合として記録されている．認識の過程では，個々の検索質問ベクトルを物体モデル中のすべての局所記述子と照合し，その中で最近傍となる物体に投票する．図 1 の例では，3 つの検索質問ベクトルが物体 1 に対応し，一つが物体 3 に対応している．認識結果は，得票数が最大の物体となる．このとき，最低得票数の閾値 s(s 以上の得票の場合に認識結果とする) を設けることにより，得票が不十分な場合にリジェクトすることが可能となる．なお，$s = 1$ はリジェクトなしの認識を表す．

上記のような投票によるアプローチでは，正しい物体の得票数が他に比べて 1 票でも多ければ，認識結果は正しくなる．言

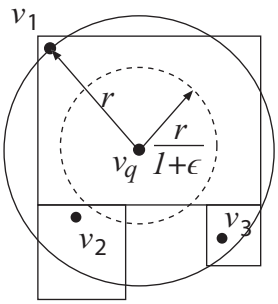


図 2　ANN における近似.

い換えれば，すべての検索質問ベクトルが正しく照合されなくてもよいとも考えられる．

2.3　近似による高速化

このような投票によるアプローチは有望と思えるが，良いことばかりでもない．一般に，一枚の画像から得られる特徴ベクトルは多数となるため，計算量の観点から，単純な全数照合を行うことはできない．従って，このアプローチの鍵は，いかに最近傍探索を効率化するかにある．

幸い，最近傍探索の高速化については，これまでに様々な手法が提案されている．現在利用可能な手法は大まかに，近似を用いないものと用いるものの 2 つに分類できる．

近似を用いない手法は，最近傍を発見するということを保証しつつ，探索を効率化するものである．一方，近似を用いる手法は，そのような保証を行わないものである．手法によっては確率的な保証を行うものがある．また，大半の手法では，近似の程度を制御するためのパラメータが与えられている．

近似を用いない手法に比べて，近似を用いるものは，より大幅な効率化が可能である．代表的な手法には，ANN (approximate nearest neighbor) [4] や LSH (locality sensitive hashing) [5] がある．本研究では，ANN を用いて処理を行う．

ANN は k-d 木と呼ばれるデータ構造に基づく近似最近傍探索法である．葉以外のノードは，特徴空間を 2 つに分割するテストを表すものであり，葉ノードは一つの特徴ベクトルを納めた特徴空間上の領域（セルと呼ばれる）を表す．ANN は，このような木構造を探索することにより，検索質問ベクトルとの距離計算の対象となる特徴ベクトルを収集する．ANN では，このベクトルの数を制御することによって，効率化を図る．

図 2 に ANN による最近傍探索の近似方法を示す．まず，ANN は木を根から葉に向かってたどることにより，検索質問ベクトル $\boldsymbol{v}_q$ が含まれるセルを発見する．セルには物体モデルの特徴ベクトル $\boldsymbol{v}_1$ が一つ記録されている．それと検索質問ベクトルとの距離を r とするとき，最近傍は必ず半径 r の超球の中に存在する．したがって，ANN では，その超球と領域が交差するセルをすべて求め，そのセルに存在する特徴ベクトルを距離計算の対象とする．

近似は，超球の半径 r をパラメータ ϵ を用いて $\frac{r}{1+\epsilon}$ に縮小することによって達成される．図 2 の場合では, 近似によって v_3 との距離計算を省略できる．この場合は依然として真の最近傍 v_2 を発見することができている．一般には，ϵ の値が大きくなればそれだけ，真の最近傍を発見することは難しくなる．

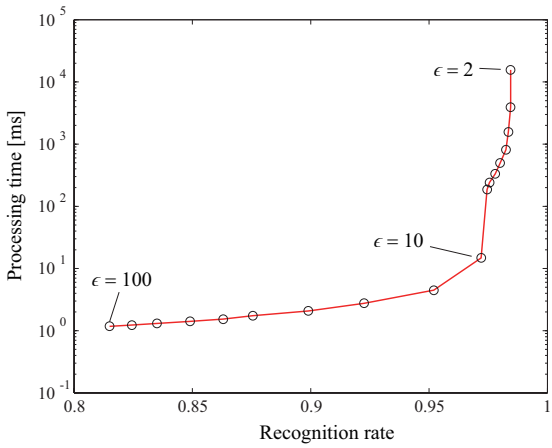


図 3　認識における近似の効果.

次に近似による効率化について見てみよう．図 3 は，物体認識の実験結果として得られた，認識率と処理時間 (未知の画像 1 枚を認識するのに要した時間; 特徴抽出の時間は含まれない) を示すものである．認識タスクはポスターなどの平面物体の認識，物体モデル数は 10 万，検索質問数は 2,000，ϵ の値は，2,3,...,10,20,30,...,100 の 18 通りを試している．図に示すように，ϵ が 2〜10 程度までは，認識率をあまり低下させることなく，処理時間が劇的に短縮されている．例えば，$\epsilon = 10$ とは，最初に定めた半径 r の 9%(=1/11) の領域しか探索しないので，かなり大幅な近似であると言える．しかしながら，得られる認識率は 97.2%であり，$\epsilon = 2$ の場合と比べて 1.3%しか失っていない．一方，処理時間については，約 1/1000（$\epsilon = 10$ で 14.7 ミリ秒，$\epsilon = 2$ で 15 秒）となっており，大幅な高速化が達成されている．

この結果は，高精度な物体認識を実現する上で，局所記述子の照合精度はそれほど必要ないことを表している．むしろ不正確さとの引き替えに，多くの距離計算を回避でき，処理時間を十分稼ぐことが可能であると言える．

3.　2 項分布を用いた投票モデルの構築

前節の議論により，高い認識率を保ちつつ物体認識を高速化するための鍵となる技術は，近似であることが分かった．それでは，どの程度の近似を導入することが可能であろうか．可能な近似の程度は，物体モデルの数と無関係であろうか．あるいは検索質問との関係はどうであろうか．

このような疑問に答える一つの方法は,「投票による認識」のモデル（投票モデル）を構築し，物体の認識率と局所記述子の照合精度との関係を支配する要因を見定めることである．

3.1　単純な投票モデル

まず，単純な投票モデルから始めよう．いま，N を物体モデルの数，m を検索質問に含まれる局所記述子の数，r_i を物体 i が得た得票数とする．ここで，$\sum_i r_i = m$ である. 認識プロセスのモデル化のため，物体 i が一票得票する可能性（局所記述子の照合精度）は，確率 p_i によって表せると仮定する. すると，物体 1,...,N が $r_1, ..., r_i, ..., r_N$ の票を得る確率は，多項分布に

よって次のように表せる．

$$\frac{m!}{\prod_{i=1}^{N} r_i!} \prod_{i=1}^{N} p_i^{r_i}. \tag{1}$$

正しい認識結果が得られるのは，正解物体 c の得票数が最大であること，すなわち，

$$c = i^* = \arg\max_i r_i, \tag{2}$$

の場合である．この事象が生じる確率 $P_M(i^* = c)$ は，

$$P_M(i^* = c) = \sum_{r_c} \sum_{r_i : r_i < r_c} \frac{m!}{r_c! \prod_{i \neq c} r_i!} p_c^{r_c} \prod_{i \neq c} p_i^{r_i}. \tag{3}$$

のように表すことができる．しかし残念ながら，m と N が大きいときに P_M を求めるのは容易ではなく，この式はモデル化にあまり役に立たない．

そこで本稿では，上記の式を 2 項分布により単純化するという方針を採る．物体モデルの数 N が十分大きい場合，誤った物体に投票される確率 $p_i (i \neq c)$ は，正しい物体に投票される確率 p_c と比べて極めて小さいと言える．したがって，特定の誤った物体が多数の票を得ることは，ほぼ生じ得ないと仮定する．この仮定に基づけば，正しい物体が得た票数を見るだけで，認識結果が正解であるかどうかを判定することができる．換言すれば，正しい物体が s 票以上の票を得ると，正しく認識できると仮定する．

いま，r を正しい物体の得票数，p を正しい物体が 1 票得票する確率とする．上記の仮定によると，物体を正しく認識する確率 P_B は，物体が s 票以上を得る事象の和事象として，2 項分布に基づいて次のようにモデル化できる．

$$P_B(i^* = c; p, m, s) = \sum_{r=s}^{m} \binom{m}{r} p^r (1-p)^{m-r} \tag{4}$$

この 2 項分布の投票モデルを評価するため，式 (4) により得られる値と，ANN を用いて実際に測定した値を比較した．物体モデルの数は 10 万，検索質問の数は 2 千，検索質問の局所記述子数 m の値には平均 588 を用い，最低得票数 s は 3 とした．ANN を用いた実測値を得る際には，ϵ の値を 2〜1,000 に変化させ，局所記述子の照合精度 p として様々な値を発生させた．p の値は検索質問によって異なるため，比較にはその平均 p_{ave} を用いた．

図 4 に示すように，上記の単純な 2 項分布による投票モデルは，実際の認識結果とかけ離れた結果となる．実際の認識は投票モデルの予測と比べて，かなり難しいタスクであることが分かる．ただし，依然として，近似最近傍探索による照合精度 p が低くても，高い物体認識率を得ることが可能であると分かる．例えば，認識率 97.2%を得るために必要な照合精度は，13.8%($\epsilon = 10$) であった．

3.2 検索質問や物体モデルに依存する因子の考慮

それでは，上記の単純な 2 項分布の投票モデルが実測値と合わない理由としては，どのようなものが考えられるであろうか．上記のモデルは，以下の 2 つの暗黙の仮定に基づいている．

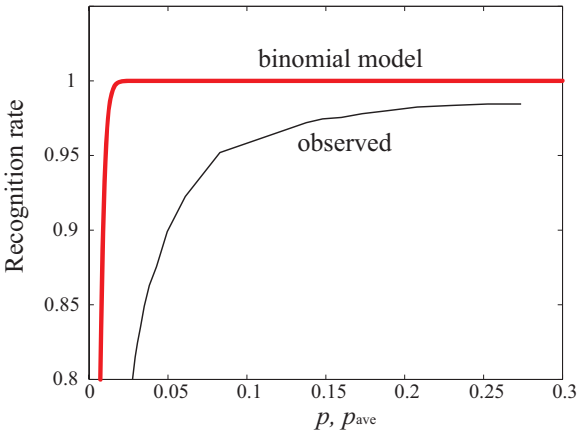


図 4　2 項分布による投票モデルと ANN による実測値の比較．

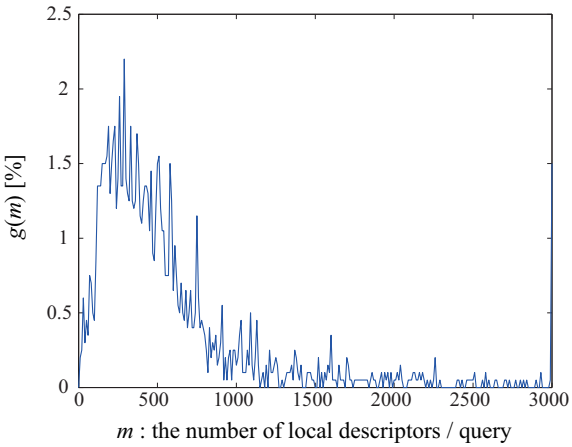


図 5　検索質問あたりの局所記述子数の分布．

仮定 1　検索質問はすべて m 個の局所記述子で表現される．
仮定 2　p の値は検索質問に依存せず，ANN における近似の程度 ϵ だけに影響を受ける．また，平均値 p_{ave} により十分代表できる．
しかしながら，これらの仮定は適切ではないことが分かった．

仮定 1 から順に考察する．図 5 は，実験に用いている 2 千の検索質問から得た局所記述子数の分布である．横軸が検索質問あたりの局所記述子数，縦軸 $g(m)$ は m 個の局所記述子を持つ検索質問の割合であり，$\sum_m g(m) = 1$ である．千以上の局所記述子を持つ検索質問がある一方で，平均値 (588) よりも少ない局所記述子しか持たない検索質問も数多く存在する．局所記述子の数が少なければ少ないほど，認識システムが十分な数の得票を得ることは難しくなる．例えば，$p = 0.14$ の場合，最低得票数 $= 3$ を実現するには，少なくとも 22 個の局所記述子が必要である．図 5 の分布では，0.5%程度の検索質問がこの数を下回っている．

次に仮定 2 について見てみよう．照合精度 p の値は，近似の程度 ϵ に依存するだけではなく，物体モデルの数 N や検索質問にも大きく依存する．図 6 に，$N = 100,000$ の場合の，p の値の分布を示す．横軸は p の値，縦軸 $h(p; \epsilon, N)$ は，照合精度が p となる検索質問の割合を示す．

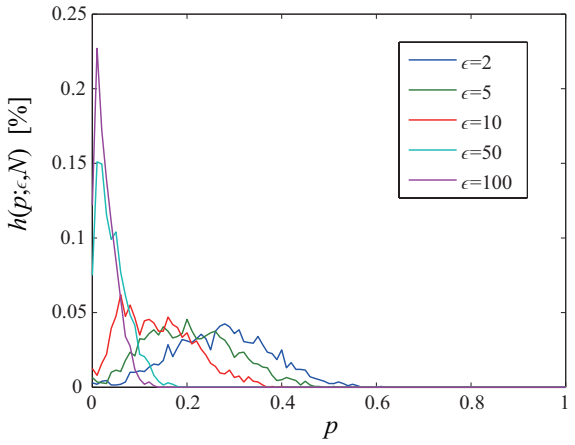


図 6　照合精度 p の分布.

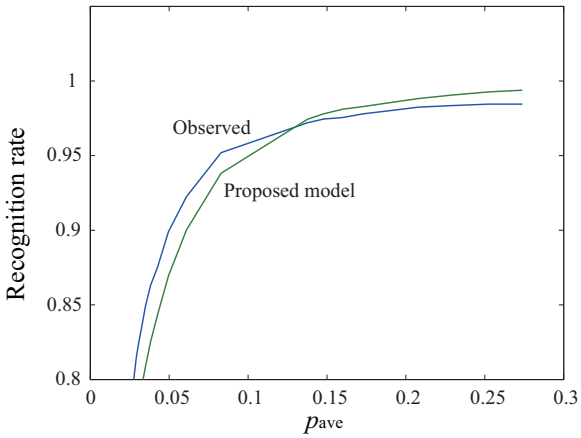


図 7　改訂された投票モデルと ANN による実測値の比較.

ϵ が小さいとき，p の分布は広く，小さい p の値を持つ検索質問は殆ど存在しない．一方，ϵ の値が大きいときには，分布は $p = 0$ に近づいていく．実際の例を見てみよう．100 個の局所記述子を持つ検索質問の場合，最低得票数 $s = 3$ を満たすためには，$p \geq 0.03$ でなければならない．$\epsilon = 2$ の場合の分布では，99%の検索質問がこの条件を満たす．ところが $\epsilon = 100$ の場合には，42%の検索質問しか条件を満たさない．残りの 58%は $p < 0.03$ であり，これが認識率を低下させる原因となっている．

そこで本研究では，先に示した 2 項分布の投票モデルを上記の 2 つの因子，すなわち，検索質問あたりの局所記述子数，ならびに，照合精度 p の分布を考慮して改訂する．その結果，式 (4) は次のような形になる：

$$P(i^* = c; \epsilon, N, s) = \sum_{m=0}^{\infty} g(m) \int h(p; \epsilon, N)\ P_B(i^* = c; m, p, s) dp. \quad (5)$$

ここで，$P_B(i^* = c; m, p, s)$ は元の 2 項分布モデルである．式 (5) は，P_B の期待値を，単純に，$h(\cdot)$ ならびにすべての可能な m を用いて得るものである．

図 7 に，上記のモデルで得られた曲線と，図 4 に掲載した実測値を示す．横軸は ϵ によって決定される照合精度の平均 p_{ave} であり，式 (5) の場合は次のように定義される．

$$p_{\text{ave}} = \int p h(p; \epsilon, N) dp. \quad (6)$$

この図に示すように，2 つの曲線はほぼ一致している．

4.　実験による検証

4.1　デ　ー　タ

提案した投票モデルを評価するため，実際のデータを用いた大規模な実験を行った．物体モデルには，写真，ポスター，本の表紙などの平面物体 10 万枚を用いた．使用した画像は，Google 画像検索，Flickr，および PCA-SIFT のサイト (注1) からダウンロードした．例を順に図 8 に示す．

(注1)：http://www.cs.cmu.edu/~yke/pcasift/

(a) Google

(b) Flickr

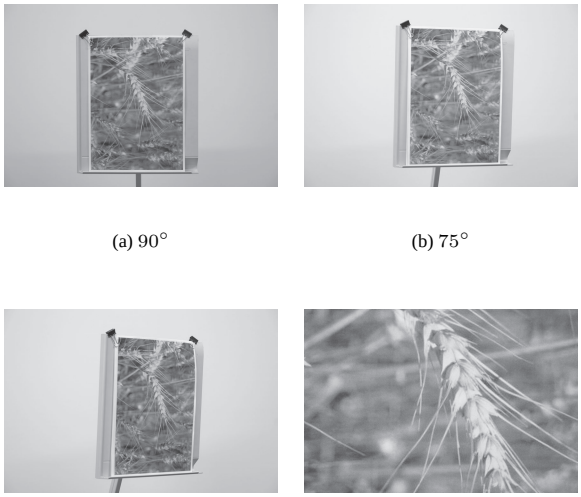
(c) PCA-SIFT

図 8　物体モデルのための画像の例.

物体モデルの作成には，長辺が 640 画素以下になるように縮小した画像を用いた．抽出された局所記述子の数は平均で約 2 千個程度であった．従って，物体モデルに用いられた局所記述子の総数は 2 億となる．

検索質問として用いた画像は次のように撮影した．まず物体モデル用の画像から 500 サンプルを取り出し，それを A4 の用紙に印刷した．そしてデジタルカメラで様々な角度から図 9 のように撮影することにより検索質問とした．4 通りの撮影を行ったため，検索質問は合計 2000 個得られた．検索質問の画像サイズを 512 × 341 画素に縮小したのち，PCA-SIFT を適用して局所記述子を抽出した．平均の局所記述子数は 588 であった．

式 (5) で定義された確率の値を求めるためには，$g(m)$ と $h(p; \epsilon, N)$ の分布が必要となる．本実験では，上記の検索質問 2000 個を用いて計測された値 (図 5 と図 6 に示したもの) を用いた．



(a) 90°　(b) 75°

(c) 60°　(d) Part

図 9　検索質問画像の例．(a)〜(c) は物体全体を撮影したものであり，(d) は 1/4 程度の領域を撮影したものである．

表 1　実験に用いたパラメータの値．

ϵ	2,3,...,10,20,...,100,1000
N	500,1000,5000,10000,50000,100000
s	1,2,...,5

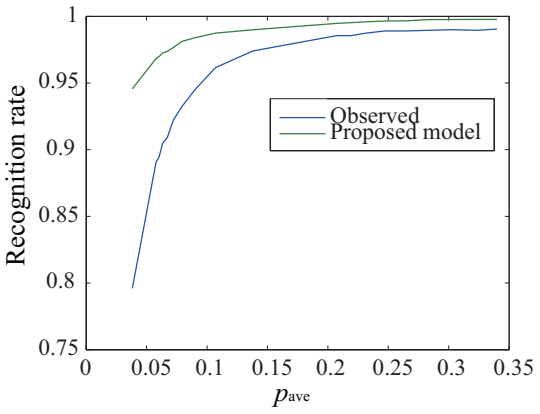
4.2　結果と考察

提案した投票モデル（式 (5)）は 3 つのパラメータ ϵ, N, s を持つ．実験の目的は，これらのパラメータを様々に変化させたとき，どの程度，投票モデルと実測値が合うかを確かめることである．本実験では，パラメータの値として表 1 に示す種類のものを用い，すべての組み合わせを試した．また，実測値は ANN を用いて物体認識を実際に行うことによって計測した．

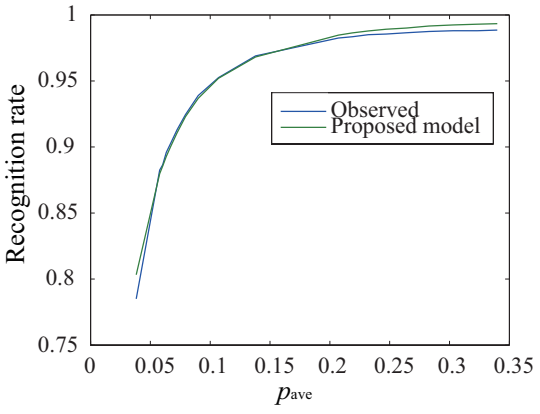
結果は次の通りである．認識実験によって得られる実測値と，投票モデルの値のずれが最も大きかったのは，物体モデルの数が少なく (500)，最低得票数の閾値が小さく ($s = 1$)，かつ ϵ の値が大きいときであった．このずれは，物体モデル数と閾値 s の増加，ならびに ϵ の減少 (即ち照合精度 p_{ave} の増加) に伴って小さくなった．以下では，s, N, ϵ の効果について順に述べる．

4.2.1　最低得票数 s

物体モデル数 500, $s = 1$ の場合の結果を図 10(a) に示す．2 つの曲線の大きなずれは，主に投票モデルには組み込まれていない「多数決の効果」によるものである．提案した投票モデルで $s = 1$ のとき，ある物体が 1 票以上を得ると，その物体が正しく認識されると仮定している．ところが言うまでもなく，この場合，仮定は間違っており，他にも 1 票以上獲得した物体は多数存在するはずである．一方，実際の認識では，最低得票数を満たす物体のうち，得票数が最大のものが結果となる．したがって，このような場合にも投票モデルを適合させるためには，多数決の仕組みをモデルに導入する必要があると言える．図 10(b) は同じ 500 物体モデルを用いているが $s = 5$ の場合である．s が大きくなるとずれは解消されることが分かる．



(a) $s = 1$



(b) $s = 5$

図 10　投票モデルと実測値の比較．物体モデル数 500 の場合．

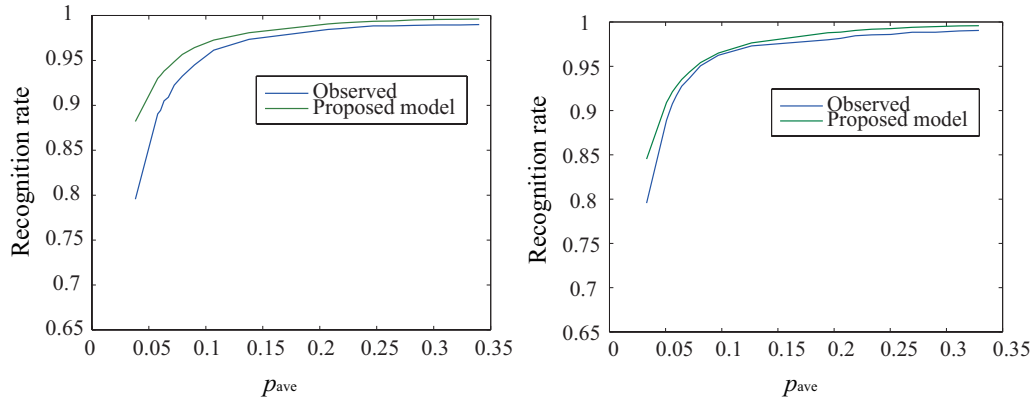
4.2.2　物体モデル数 N

図 11 に，$s = 3$ に固定し，物体モデル数を変化させた結果を示す．図 11(a) に示すように，物体モデル数が 500 の場合は，$s = 3$ の場合でも依然としてずれが大きかった．しかしながら，このずれは物体モデル数が 1,000 以上になると減少した．このことから，提案した投票モデルは，N が一定以上のときに適用可能であることが分かる．この図に示した $s = 3$ の場合であれば，物体モデル数は 1,000 で十分である．一方，5,000 物体モデルの場合には，$s = 2$ でもよく適合した．

図 12 に，別の視点からの比較を示す．図 12 は，ANN の ϵ を 10 に固定し，各々の最低得票数 s について，物体モデル数 N と認識率の関係を表したものである．最低得票数 s の値が小さいときには，すべての N について，認識率に 1%程度の差が見られるが，s の値が大きくなるにつれてその差は減少し，$s = 5$ では投票モデルと実測値の差が 0.1%〜0.2% にまで減少している．

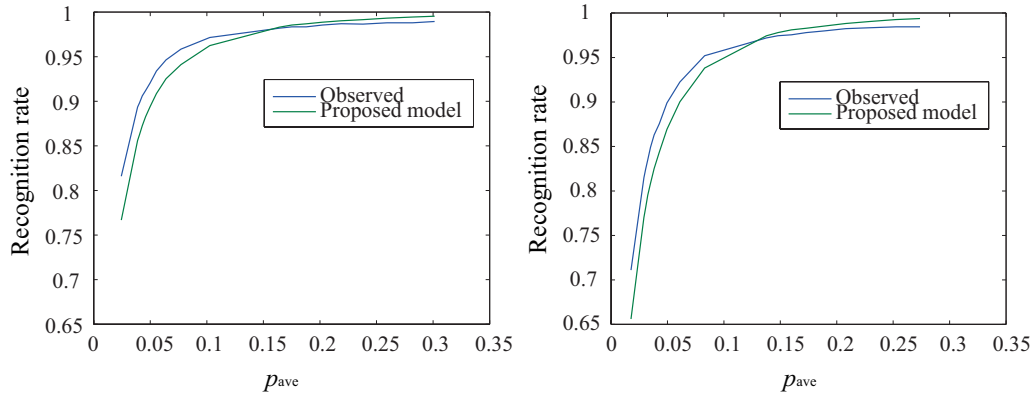
4.2.3　近似の程度 ϵ

全実験を通して，ϵ の値が大きくなるに伴って，投票モデルと実測値の値が離れていった．ϵ の値が大きくなると，p_{ave} の値が小さくなり，それゆえ，正解の票数が減少する．モデルとの



(a) 500 object models

(b) 1,000 object models



(c) 10,000 object models

(d) 100,000 object models

図 11　投票モデルと実測値の比較．$s = 3$ として物体モデル数を変化させた場合．

ずれは，票数が少ないことによる誤差に加え，次の理由があったと考えている．提案した投票モデルの式 (5) では，$h(p; \epsilon, N)$ と局所記述子の数 $g(m)$ には相関がないことを仮定している．しかし厳密にはこれは正しくない．これにより，実測値よりもよい値が投票モデルにより予測されたものと考えられる．

4.2.4　$g(m)$ と $h(p; \epsilon, N)$ の設定

最後に，投票モデルで重要な役割を果たす $g(m)$ と $h(p; \epsilon, N)$ について述べておく．提案した投票モデルは，これら 2 つの分布を導入することによって，2 項分布モデルを実測値に近づけたものといえる．この投票モデルは，例えば認識システムの設計に用いることが考えられる．大規模で網羅的な物体認識実験を行わずに，希望する認識率を実現するために必要な近似のパラメータ ϵ を予測することが期待できる．ただし，このためには，2 つの分布 $g(m)$ と $h(p; \epsilon, N)$ を得ておく必要がある．これらの値は検索質問や物体モデルに依存する点に注意が必要である．

$g(m)$ は検索質問に依存する．ただし，認識対象の画像の傾向が大幅こ異ならず，また撮影条件も類似である場合には，それほど変動する分布ではないと考えられる．一方，$h(p; \epsilon, N)$ については，近似のパラメータ ϵ だけではなく物体モデルの規模 N にも依存するため，認識システムに応じて定める必要がある．現在は，実際の検索質問を用いて認識実験を行い，分布を計測しているが，この分布に対しても何らかの妥当なモデルを定め，認識実験を行わずに分布を得る必要がある．

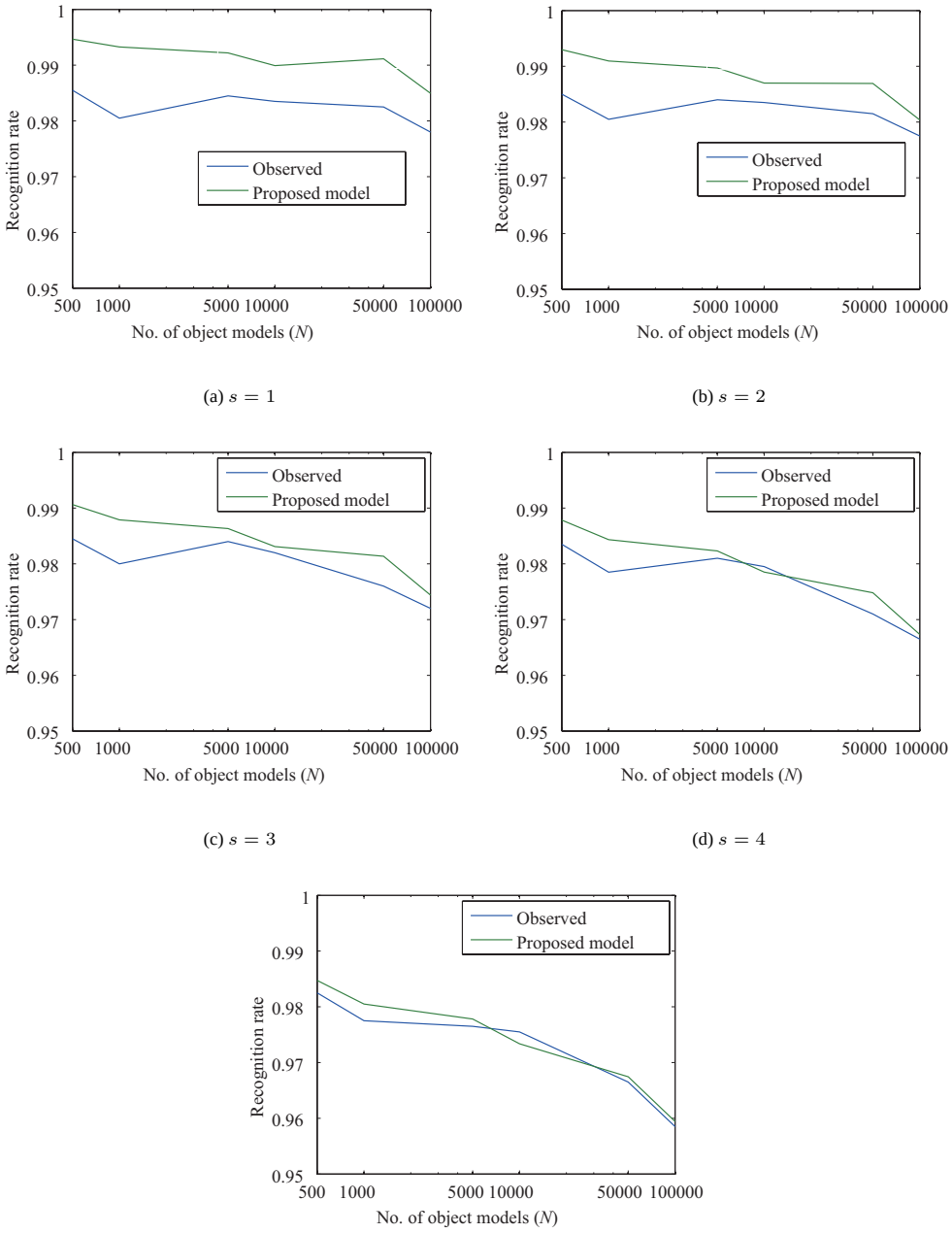
5. むすび

本稿では，局所記述子を用いた物体認識に対して次の 2 つの点を述べた．第一は，高い認識率を達成するために必要な局所記述子の照合精度は，それほど高くないということである．投票処理により票は正解の物体に集まりやすく，誤った票は多数の不正解の物体に分散されるため，照合精度が低くてもあまり悪影響を及ぼさない．第二は，投票に基づく物体認識が，2 項分布によってモデル化できるという点である．実測値に合うモデルは，検索質問における局所記述子数の分布，ならびに，局所記述子の照合精度の分布を加味することによって得られた．

10 万の物体モデル，2 千の検索質問を用いた実験により，物体モデル数と最低得票数が一定以上であれば，提案した投票モデルが ANN を用いた実測値とほぼ合うことが分かった．

今後の課題には，より精度の高い投票モデルのための改良に加えて，投票モデルを用いた物体認識の改善が挙げられる．

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．



(a) $s = 1$ (b) $s = 2$

(c) $s = 3$ (d) $s = 4$

(e) $s = 5$

図 12　投票モデルと実測値の比較．$\epsilon = 10$ として最低得票数 s を変化させた場合．

文　　献

[1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol.60, no.2, pp.91–110, 2004.

[2] J. Ponce, M. Hebert, C. Schmid and A. Zisserman Eds., Toward Category-Level Object Recognition, Springer, 2006.

[3] G. Shakhnarovich, T. Darrell and P. Indyk Eds., Nearest-neighbor methods in learning and vision, The MIT Press, 2005.

[4] S. Arya, D. M. Mount, R. Silverman and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching,” Journal of the ACM, vol.45, no.6, pp.891–923, 1998.

[5] A. Andoni, M. Datar, N. Immorlica, P. Indyk and V. Mirrokni, Locality-sensitive hashing using stable distributions, Nearest-Neighbor Methods in Learning and Vision (Eds. by G. Shakhnarovich, T. Darrel and P. Indyk), The MIT Press, pp.61–72, 2005.

[6] K. Kise, K. Noguchi and M. Iwamura, Simple representation and approximate search of feature vectors for large-scale object recognition, Proc. BMVC2007, pp.182–191, 2007.

[7] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” IEEE PAMI, vol.27, no.10, pp.1615–1630, 2005.

[8] Y. Ke, R. Sukthankar and L. Huston, Efficient near-duplicate detection and sub-image retrieval, MM2004, pp.869–876, 2004.

[9] J. Sivic and A. Zisserman, Video google: A text retrieval approach to object matching in videos, Proc. ICCV2003, Vol. 2, pp.1470–1477, 2003.

[10] D. Nistér and H. Stewénius, Scalable recognition with a vocabulary tree, Proc. CVPR2006, pp.775–781, 2006.

[11] E. Nowak, F. Jurie and B. Triggs, Sampling strategies for bag-of-features image classification, Proc. ECCV2006, Vol. 4, pp.490–503, 2006.