参照特徴ベクトルの増大による特定物体認識の高速化と高精度化

黄瀬　浩一†　　野口　和人†　　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: †{kise,masa}@cs.osakafu-u.ac.jp, ††noguchi@m.cs.osakafu-u.ac.jp

あらまし　局所特徴量の照合による特定物体認識を考える．一般に，生成型学習を用いて，索引付けのための局所特徴量（参照特徴ベクトル）の数を増やせば，それだけ認識率が向上する．本論文では，認識率の向上だけではなく，処理時間も短縮可能であることを示す．参照特徴ベクトルの数が増えると処理時間が短縮される，という逆説的な効果は，認識器の多段階化による認識処理の早期終了によって得られる．これは，生成型学習によって照合に必要な探索空間を制限できるという効果による．1 万画像を用いた認識実験の結果，6.6 倍の参照特徴ベクトルを用いることで，処理時間を 2/3, 認識率を 12.2%改善できること，ならびに 26 億個の参照特徴ベクトルを用いて索引付けされた 100 万物体を，59ms/query, 90%の認識率で認識可能であることを示す．

キーワード　特定物体認識，局所特徴量，生成型学習，多段階化，低品質画像

Improving Efficiency and Robustness of Specific Object Recognition by Increasing Reference Feature Vectors

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

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

Abstract　This paper concerns specific object recognition based on local features. In general, increasing the number of local features for indexing (reference feature vectors) by generative learning enables us to improve the recognition rate. In this paper, we show that generative learning is also effective for shorten the processing time. This paradoxical effect (i.e., shorter time with more reference feature vectors) is achieved by cascading recognizers. Generated local features allow us to terminate the recognition process earlier. In other words, generative learning is effective to reduce the search space for finding nearest neighbors. From the experimental results using 10,000 reference images, 6.6 times reference feature vectors enabled us both to reduce the processing time to 2/3 from the original, and to improve the recognition rate by 12.2%. Another experiment with 1 million reference images indexed by 2.6 billion reference feature vectors yielded the recognition rate of 90% in 59ms/query.

Key words　Specific object recognition, Local feature, Generative learning, Cascade, Low-quality image

1. はじめに

学習に用いるデータが増えれば，それだけタスク実行の精度が向上するとともに必要な時間も短くなる．人間の学習では半ば当然であるこの性質は，計算機の処理では一般には成り立たない．通常は，学習データの増加により，処理精度が向上しても処理時間が短くなることはない．本論文では，最近傍探索に基づく物体認識において，学習データを増やすことにより精度向上のみならず処理時間を短縮するという効果を持つ手法を提案する．

本論文で対象とする物体認識は，特定物体認識と呼ばれる範疇に属する．これは，画像中に存在する物体と同じものがどれであるのかを認識するというインスタンスレベルの認識である．このような画像認識のタスクは，例えば，カメラ付き携帯電話で撮影した画像をもとに関連サイトに誘導するなどの，様々なサービスを起動するための仕組みとなり得る．従来サービスとの関連でいえば，画像認識技術を用いたバーコードの代替といえる．

このようなサービスを真に実用的なものとするには，少なくとも，(1) 大規模，(2) 省メモリ，(3) 高効率，(4) 頑健，という 4 つの条件を満たす必要がある．認識対象となる物体数の大規模化は，様々な物体をサービスに繋

げるために必要である．物体数が増えれば必要なメモリ量も増加するため，メモリ量を削減する技術も求められる．処理時間増加を抑えるには高効率な処理も必要である．頑健な認識は，特にカメラ付き携帯電話で得られる低品質画像の認識に必要となる．具体的には，照明条件の変化，オクルージョン，低解像度，ボケやブレなどに対処しなければならない．特に，携帯デバイスでの撮影に対しては，ボケとブレへの対処が必須となる．

これまで，多くの研究者が特定物体認識の解決策を求めて研究を行ってきた．大きなブレークスルーは，局所特徴量に基づく手法，すなわち Schmid ら [1] ならびに Lowe [2] による先駆的な研究によりもたらされた．その後，これらを基礎とし，上記の条件をより良く満たすために，様々な拡張が試みられている [3]〜[5]．しかしながら，これらの努力にもかかわらず，条件を十分満足する手法は得られていない．

本論文では，同様に局所特徴量の照合による認識という方針に基づいて，上記の条件をより良く満たす手法を提案する．局所特徴量としては PCA-SIFT，照合方法としては近似最近傍探索，認識方法としては照合結果に基づく投票というシンプルな手法を用いる．本研究の最も重要なポイントは，画像の索引付けに用いる局所特徴量（参照特徴ベクトル）を増大させることにより，認識をより頑健かつ効率的にする点にある．局所特徴量の増大には生成型学習を用いる．すなわち，画像が被るボケやブレをシミュレートして数多くの局所特徴量を抽出する．生成型学習を用いて参照特徴ベクトルを増加させると，照合に時間がかかったりメモリ量を圧迫したりすることが問題となる．これらの問題には，認識の多段階化による高速化，ならびにスカラー量子化によって対処する．特に多段階化を生成型学習と組み合わせて用いることにより，認識率を改善するだけではなく，処理効率も向上する点は常識に反するものであり，全く新しい結果といえる．また，本論文では，100 万物体のデータベースを用いた認識実験を通して，提案手法の有効性を検討する．

2. 関 連 研 究

本論文では，局所特徴量に基づく物体認識に焦点を当てる．物体認識のタスクは，大きく，特定物体認識と一般物体認識に分類できる．特定物体認識がインスタンスレベルの認識であるのに対して，一般物体認識はクラスレベルの認識である．ここでは，一般物体認識のアプローチと対比しつつ，物体の表現，頑健性，効率ならびにメモリ量の観点から特定物体認識の手法を概観する．

物体の表現とは，局所特徴量を用いて物体を記述する方法であり，visual word(VW) [3] を用いる手法と局所特徴量をそのまま用いる手法がある．VW を用いる手法は一般物体認識において支配的であるが，特定物体認識に適用すると多数の VW を用いる必要があるという問題が生じる．Nistér らは 1 つの VW が高々2,3 個の局所特徴量しか表さないという状況で高い認識率が得られると述べている [4]．また，VW に基づく物体の表現としては，情報検索で提案されているベクトル空間モデルがよく用いられる．ベクトルの各次元は個々の VW に対応し，その出現頻度を元に物体を表すベクトルが定められる．ただし，このベクトルは物体全体を表す大域的なものであるため，オクルージョンの問題を扱うことができない．一方，局所特徴量をそのまま用いる手法は，物体から抽出した局所特徴量の集合として物体を表現するものであり，特定物体認識の初期（例えば [2]）から利用されている．VW 以上に処理時間とメモリ量の問題は生じるものの，表現が局所的であるため，オクルージョンの問題を解決することができる．すなわちオクルージョンによって一部分を失っても，他の部分から得た局所特徴量により依然として認識が可能となる．

次に頑健性について述べる．VW によるベクトル表現を用いた一般物体認識では，頑健性を得るために SVM などの高い能力を持つ識別器や EMD(Earth Mover's Distance) などの柔軟な距離尺度が導入される．しかし，これらの手法を特定物体認識に導入することは難しい．これは，ベクトル表現が高次元になったり，多数の特徴ベクトルを扱う必要があるからである．特定物体認識では，生成型学習 [6] により，様々な変動を加えた多数の画像から局所特徴量を抽出して認識に用いることにより，頑健性を得る手法が提案されている．例えば，Random Fern と呼ばれる手法では，認識対象物体は少数ではあるが，幾何学的変動に対する有効性が実証されている [5]．

効率が重要な問題になるのは，特定物体認識で局所特徴量をそのまま用いる場合である．このとき，物体の認識は，局所特徴量が最も良く照合する物体を見つける処理となる．照合の処理としては，例えば最近傍探索を考えることができる．局所特徴量の数は，画像あたり数十から数千と膨大となるため，照合の効率化は必須課題である．Lowe は Best-Bin-First アルゴリズムと呼ばれる近似最近傍探索法を用いて，効率化を実現している [2]．Ke らは Locality Sensitive Hashing (LSH) [7] という近似最近傍探索を用いた手法を提案している [8]．

メモリ量は特定物体認識におけるもう一つの大きな問題である．対策の鍵は，高次元実数値ベクトルとして表される局所特徴量をいかに圧縮するかにある．PCA-SIFT [8] は，SIFT の次元を圧縮できるため，この目的に適している．もう一つの方法は量子化である．一般物体認識では，VW を得るためにベクトル量子化が用いられる．しかし前述の通り，これは特定物体認識には有効でない．一方，スカラー量子化は特定物体認識に有効な手法である．例えば，2bit/dim. まで圧縮しても，ほとんど認識率に悪影響を及ぼさないことが知られている [9]．

上記の一般物体認識と特定物体認識のアプローチの差は，射撃のアナロジーによって上手く説明できる．一般物体認識の手法は，注意深く設計された弾丸（ベクトル

図 1　カメラ付き携帯電話で撮影された検索質問画像の例.

表現）と精密な銃（識別器，距離尺度）を用いて，一発の射撃の精度を高めるものである．他の様々な対象の認識も，この方策に基づくものが多い．一方，特定物体認識の手法は，弾（未知の画像から得た局所特徴量）と的（参照する画像から得た局所特徴量）の数を増やすことにより，個々の弾の命中精度は低くても，全体として当たる確率を高めるものであり，比較的新しい方策である．この方策では，頑健性，効率，メモリ量をいかに高い次元でバランスするかが重要な課題となる．本論文では，従来研究で成果を挙げている，生成型学習，近似最近傍探索，スカラー量子化を導入し，低品質画像の省メモリ，高速，高精度な認識法を提案する．

3.　生成型学習による低品質画像の認識

3.1　タ　ス　ク

本論文で扱う認識のタスクは，平面特定物体の認識である．認識対象となる画像の例を図 1 に示す．これらの画像は，写真の小さいプリントをカメラ付き携帯電話で撮影して得られたものである．以後，認識対象画像を検索質問画像と呼ぶ．この図にも示されているように，カメラ付き携帯電話で得た画像には，認識を困難にする様々な問題が含まれている．具体的には，低解像度 (QVGA)，不均一な照明，射影歪み，ボケ，ブレなどである．加えて，いくつかの写真は全体が画像に含まれておらず，一方で，他の物体（他の写真や背景）を画像の一部に含む場合もある．

本論文の認識タスクは，このような低品質画像を効率的かつ高精度に認識することである．また，タスクの性質上，検索質問画像と照合して識別される画像（データベース中の画像）は，十分多い（例えば 100 万画像）ものとする．ボケやブレが画像に加えられる場合，多くの画像は似てくるため，データベースのサイズ（参照する画像数）が大きいと認識はより困難となる．

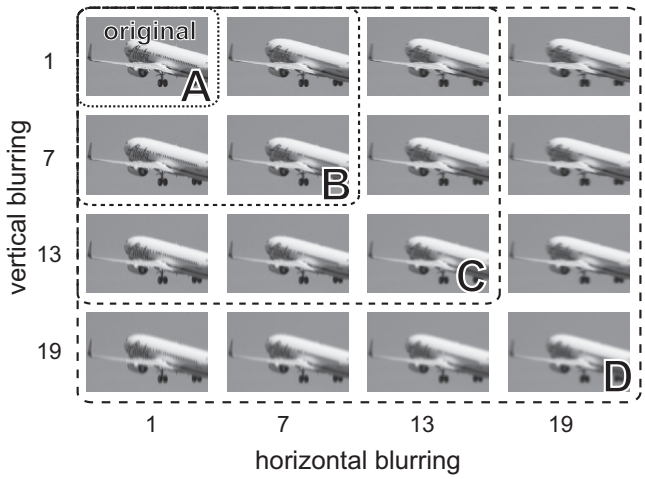
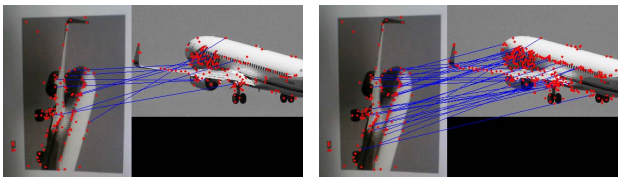
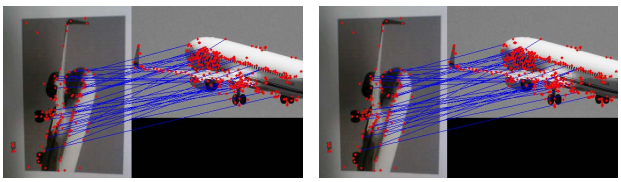


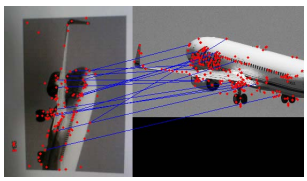
図 2　生成された画像.



(a) 学習セット A (9/199)　　(b) 学習セット B (21/575)



(c) 学習セット C (36/839)　　(d) 学習セット D (36/1059)



(e) 学習セット D_{diag} (16/342)

図 3　生成型学習の効果.

3.2　生成型学習

上記のタスクを実行するため，本手法では生成型学習の考え方を導入する．具体的には，生成型学習によって，ボケやブレを受けた画像を生成し，データベースに加えることによって，認識精度を高めるという方策である．図 2 に画像生成の方法を示す．原画像 A を，ガウス関数の標準偏差を変化させてたたみ込むことによって，様々にぼかす．図に示された，7, 13 などの水平方向，垂直方向のボケを表すパラメータ w は，ガウス関数の標準偏差

と $\sigma = 0.3(w/2-1)+0.8$ という関係にある．水平，垂直方向に独立にパラメータを変化させることにより，ボケだけではなくブレもシミュレートできる．

生成型学習の効果を図 3 に示す．(a)–(e) の小さい図において，左側の画像が検索質問画像，右側の画像がデータベースの画像を表す．また，(a)–(d) はそれぞれ図 2 の A–D に対応している．例えば (c) の場合，図 2 の C に対応する生成された画像 (9 枚) を用いて局所特徴量を取り出し，データベースに収める．(e) は，図 2 の対角部分のみを含む学習セットを用いた場合である．

左右の画像に示された点は，局所特徴量が得られた特徴点を表している．左の検索質問の画像からは 134 個の特徴点が得られている．また，両者の間を結ぶ線は，局所特徴量の間の照合結果を表している．後に述べる照合方法によって，対応する局所特徴量が得られた場合に照合されたと判定している．(a)–(e) のキャプションに示された (9/199) などの数字は，/の左が照合された局所特徴量の数，右がデータベース中の画像から得られた局所特徴量の数である．この結果から分かるように，生成型学習はデータベース中の画像から得る局所特徴量の数を増加させるものであり，それによって，照合される局所特徴量の数も増加している．画像の認識の基本は局所特徴量の照合であるため，生成型学習によって，認識精度の向上が期待できる．一方で，より多くの局所特徴量を用いるため，メモリ量と処理時間の双方に悪影響が及ぶ懸念もある．

4. スカラー量子化と多段階化による解決

4.1 問題解決のための戦略

生成型学習を用いるためには，メモリ量と処理時間の問題を解決しなければならない．データベースが大規模な場合，これらの問題は手法の適用性を左右する極めて重要なものである．その際に，生成型学習で得られた認識率を犠牲にするようなことがあってはならない．すなわち，認識率，メモリ量，処理時間について，望まれるバランスを実現可能な手法が望まれる．

認識率については，先に述べたように生成型学習によって向上させることを考える．メモリ量の問題については，スカラー量子化 [9] を導入する．通常，局所特徴量は実数を要素とする特徴ベクトルとして表現される．スカラー量子化では，これを個々の次元について限られたビット数（実数より少ないビット数）で表現する．これによりメモリ量は削減されるが，一方で特徴ベクトルの識別能力を損なう恐れもある．

処理時間の問題については，近似最近傍探索を導入する．一般に，10〜15 次元を超える次元を持つ特徴ベクトルに対して最近傍探索を行う場合，ほぼ線形探索と同程度の時間がかかることが知られている．これに対して近似を導入すると処理時間を短縮可能である．短縮の効果は，近似の程度によって異なる．より大幅な近似を行えば，それだけ高速な処理が実現できる．ただし，その代償として，正しい最近傍が得られる可能性も下がる．

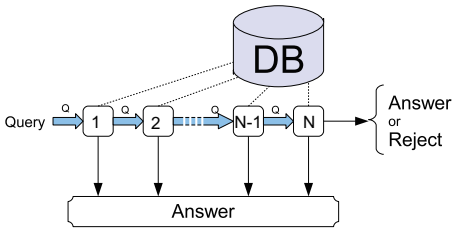


図 4　多段階に接続された認識器.

ここでの問題は近似の程度をどのように決定するかにある．大幅な近似を用いても正しく認識可能な画像がある反面，正しい認識結果を得るためにはあまり近似を行えない画像もある．本論文では，前者を認識容易な画像，後者を認識困難な画像と呼ぶことにする．このような状況で処理時間を短縮するためには，画像に応じて近似の程度を適応的に変化させる仕組みが必要となる．このような条件を満たす手法に多段階化がある [10]．そこで，本手法でも多段階化を導入し，問題解決を図る．

4.2 構　　成

多段階化に基づく認識手法の構成を図 4 に示す．図中の 1,...,N の番号が付けられた四角は認識器を表す．各認識器は，検索質問画像から得た全ての局所特徴量を用いてデータベース (DB) から対応する画像を探し出す処理を行う．具体的には，以下の通りである．まず，検索質問画像から得た各局所特徴量について，近似最近傍探索を用いてデータベースから最近傍を求める．そして，最近傍となった参照特徴ベクトルが属する画像に投票する．すなわち，認識器とは，全ての局所特徴料を用いて投票処理による認識を実行するモジュールである．ただし，各認識器は同じ能力のものではなく，番号が若いほど大幅な近似を行う．近似の程度が大幅であればそれだけ低精度ではあるが高速な処理が可能なため，前段では迅速に，後段では慎重に処理を行うという構成になっている．

処理の流れは以下の通りである．まず最も大幅な近似を伴う認識器で検索質問画像を認識する．この段階で十分な証拠が得られれば，認識処理を停止し，結果を出力する．一方，十分な証拠が得られなければ次の段に進み，近似の程度を少し弱めてより慎重に認識する．最終段まで処理が進んでも十分な証拠が得られていないと判断された場合（図 4 の N 段目右側の矢印) には，設定に応じて，判定不能としてリジェクトする，あるいは最大得票数のものを強制的に認識結果とする，のいずれかを行う．本論文では，常に強制的に認識を行うものとする．

このような多段階の認識器は，Viola らによる多段階処理 [11] とは次のように思想が異なるものである．Viola らの手法は，リジェクトを高速に行うことにより，全体の処理を高速化するものである．これは，認識処理の対

象の大部分がリジェクトされるもの（非顔領域）であり，その多くは簡単に判定できることによる．一方，本論文の手法は，リジェクトではなく認識のための多段階である．即ち，簡単に判定できる認識対象については早期に認識処理を打ち切るという方策によって，全体の処理を高速化するものである．

さて，処理を後段に進めるか否かの判定は，認識率や処理時間を左右する重要なポイントとなる．ただし，この判定に長い時間を要すると本末転倒となるため，簡便な方法が望ましい．そこで次のような手法を用いる．第 s 段における，得票数第 1 位と第 2 位の画像の得票数を，それぞれ $v_1(s)$, $v_2(s)$ とする．これらが以下の条件を満たすとき，第 s 段で認識処理を終了し，結果を出力する．

$$v_1(s) > t, \tag{1}$$

$$rv_1(s) > v_2, \tag{2}$$

ここで，t と r はパラメータである．この式は，第 1 位の得票数が十分大きく，かつ第 2 位との差が十分あれば，十分な証拠が得られていると判定するものである．

4.3 処理時間短縮の仕組み

検索質問画像から得た i 番目の特徴ベクトルを $\boldsymbol{q}_i$ とする．また，データベースに収められた参照特徴ベクトル全体の集合を P とする．s 段目の認識器では，$\boldsymbol{q}_i$ の各々に対して次の 2 処理からなる近似最近傍探索を行う．

（1） $\boldsymbol{q}_i$ に対して近似最近傍探索の候補となる参照特徴ベクトルの集合 $P_i^{(s)} \subset P$ を求める．ここで，$|P_i^{(s)}| \ll |P|$ である．

（2） $\boldsymbol{p} \in P_i^{(s)}$ に対して距離計算を行い，その中での最近傍（以後暫定最近傍と呼ぶ）を結果 $\hat{\boldsymbol{p}}^{(s)}$ とする．

$$\hat{\boldsymbol{p}}^{(s)} = \arg \min_{\boldsymbol{p} \in P_i^{(s)}} \|\boldsymbol{p} - \boldsymbol{q}_i\| \tag{3}$$

なお，本論文では距離計算にユークリッド距離を用いる．

最も単純な最近傍探索が P の要素全てを対象として距離計算を行うのに対して，近似最近傍探索ではそれより大幅に少ない $P_i^{(s)}$ の要素に限定することによって処理時間を短縮する．ただし，その代償として，$P_i^{(s)}$ に真の最近傍が含まれない場合が生じ得る．

さて，多段階化によって処理時間がさらに短縮される仕組みは，次のように説明できる [10]．図 4 中の四角で表された認識器の性能は，その中で行われる近似最近傍探索によって決まる．先に述べたように，N 個の認識器には，近似の程度が異なる N 個の近似最近傍探索器が備わっている．多段階の段 s が後段になればそれだけ，近似の程度が低くなる（近似をしない最近傍探索に近づく）．近似の程度の差は，距離計算の対象となる参照特徴ベクトルの数で表すことができる．すなわち，

$$\forall i \ \forall s \ |P_i^{(s-1)}| \leqq |P_i^{(s)}| \tag{4}$$

が成り立つ．

いま認識器で実行される近似最近傍探索に次の性質が成り立つとしよう．

［Definition 1］（単調性） 近似最近傍探索が次の性質を満たすとき，単調性があるという．

$$\forall i \ \forall s \ P_i^{(s)} \supseteqq P_i^{(s-1)}. \tag{5}$$

［Definition 2］（差分検索性） 近似最近傍探索が差集合

$$P_i^{(s)} - P_i^{(s-1)}. \tag{6}$$

を効率的に求められるとき，差分検索性があるという．

近似最近傍探索が単調性を満たす場合，s 段目における距離計算の対象は $P_i^{(s)}$ ではなく，$P_i^{(s)} - P_i^{(s-1)}$ としてもよい．これにより，s 段目の認識器を単独で適用するときの距離計算の対象と，1 段目から s 段目までの多段階を適用する場合の対象が次のように等しくなる．

$$P_i^{(s)} = \bigcup_{k=1}^{s} \left(P_i^{(k)} - P_i^{(k-1)} \right) \tag{7}$$

ここで，$P_i^{(0)} = \phi$ である．さらに，差分検索性を満たす場合，差集合 $P_i^{(s)} - P_i^{(s-1)}$ の計算時間を無視できるため，効率が悪化することもない．

各段での処理の受け渡しは次のように行われる．$\boldsymbol{q}_i$ に対して $(s-1)$ 段までの処理で得られた暫定最近傍を $\hat{\boldsymbol{p}}_i^{(s-1)}$ とする．s 段では，$X = P_i^{(s)} - P_i^{(s-1)}$ の要素に対して $\boldsymbol{q}_i$ との距離計算を行い，その中の最近傍

$$\boldsymbol{p}_i^* = \arg \min_{\boldsymbol{p} \in X} \|\boldsymbol{p} - \boldsymbol{q}_i\| \tag{8}$$

を得る．そして，s 段の暫定最近傍を，

$$\hat{\boldsymbol{p}}_i^{(s)} = \begin{cases} \boldsymbol{p}_i^* & \text{if } \|\boldsymbol{p}_i^* - \boldsymbol{q}_i\| \leqq \|\hat{\boldsymbol{p}}_i^{(s-1)} - \boldsymbol{q}_i\| \\ \hat{\boldsymbol{p}}_i^{(s-1)} & \text{otherwise,} \end{cases}, \tag{9}$$

とする．

以上から，早期に認識処理を打ち切ることが可能な認識容易な画像については大幅に効率が向上するとともに，最終段まで判定がもつれる認識困難な画像についても，最終段を単独で適用した場合と同じ効率を得ることができる．

4.4 登　　録

以下では，より具体的な処理について述べる．提案手法では，局所特徴量として PCA-SIFT を用いる．PCA-SIFT で得られる特徴ベクトル $\boldsymbol{p} = (p_1, ..., p_n)$ は，実数値を要素とする 36 次元ベクトル $(n = 36)$ である．ただし要素の値の範囲は概ね 16bit で収まるため，ここでは元のベクトルを 16bit で表現することとする．また，スカラー量子化を行う場合には，p_j を 2bit で表現する．p_j は平均値 θ_j がほぼ 0, 平均値に対してほぼ対称の分布を

持つため，量子化された値 0,1 は負，2,3 は正に対応する．

提案手法では，近似最近傍探索の手法として，ハッシュに基づくもの [10] を用いる．参照特徴ベクトル $\boldsymbol{p}$ をハッシュ表に格納するために，まず d 次元ビットベクトル $\boldsymbol{u} = (u_1, ..., u_d)$ に変換する．ここで，

$$u_j = \begin{cases} 1 & \text{if } p_j - \theta_j \geqq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{10}$$

であり，$d < n$ である．そして，次のハッシュ関数

$$H_{\text{index}} = \left(\sum_{j=1}^{d} u_j 2^{(j-1)} \right) \mod H_{\text{size}} \tag{11}$$

により参照特徴ベクトル $\boldsymbol{p}$ のハッシュ値を計算し，ハッシュ表に格納する．ここで，H_{size} はハッシュ表のサイズである．ハッシュ表への格納に際しては，参照特徴ベクトルと画像 ID を組にして記録する．また，同じハッシュ値を持つ参照特徴ベクトルが格納されている場合（衝突が生じた場合）には，チェイン法によって追加記録する．

多数の衝突が起きる場合には，距離計算の回数が増えることが考えられる．また，多数の衝突が生じるような参照特徴ベクトルは，互いに類似している可能性が高く，その場合には認識に必要な識別能力が乏しい．そこで，チェイン法によって記録される要素数が閾値 c を上回るとき，そのハッシュ値のエントリをすべて削除するとともに，以後の追加記録を禁止する．

以上の処理を全画像の全参照特徴ベクトルに適用することによって，データベースへの登録が終了する．

4.5 認　　識

認識の際には，検索質問画像から同様に PCA-SIFT の特徴ベクトル（検索質問ベクトル）$\boldsymbol{q}_i = (q_1, ..., q_n)$ を取り出し，ハッシュ表から距離計算の対象となる参照特徴ベクトルの集合を得る．

基本的には，登録と同じ処理によって得たビットベクトルを用いてハッシュ表にアクセスし，距離計算の対象となる参照特徴ベクトルの集合 P_i を得る．ところがこのような方法では，検索質問ベクトルの値が変動して異なるハッシュ値が得られる場合に対処できない．そこで，

$$|q_j - \theta_j| \leqq e, \tag{12}$$

を満たす場合には，q_j の値が閾値を超えて異なるビットに変換される可能性も考える．ここで，e は閾値である．この場合，u_j だけではなくビットを $u'_j = 1 - u_j$ のように反転してハッシュ値を計算し，それも用いてハッシュ表にアクセスする．これは，複数のハッシュ関数を用いることにより，変動の影響を受けやすい次元 j を無視して参照特徴ベクトルを検索することを意味する．距離計算の対象は得られた参照特徴ベクトルの和集合とする．

多段階との対応は以下の通りである．まず 1 段目では，上記のビット反転を適用しない．2 段目ではビット反転を 1 つの次元に対して適用する．複数の次元が式 (12) の条件を満たす場合には，次元 j のより大きいもの (より小さい固有値に対するもの) を優先する．s 段目では $(s-1)$ 個の次元についてビット反転を適用する．これにより，$2^{(s-1)}$ 通りのハッシュ値を計算し，ハッシュ表にアクセスする．

このようなプロセスを全ての次元に適用すると，データベース中の全参照特徴ベクトルを対象に距離計算を行うことになり，効率が大幅に悪化する．そこで，上限 b を設けて，それを超える場合には適用しないことにする．なお，多段階の段数 N との関係は，$N = b + 1$ である．

以上のビット反転の処理は，単調性，差分検索性を満たすため，先に述べた効率的な処理が可能である．

表 1　学習セット

Training set	# of images (ratio to A)	Ave. # of feature vectors (per image) [ratio to A]
A	1	5.0×10^2 [1.0]
B	4	1.6×10^3 [3.2]
C	9	2.6×10^3 [5.2]
D	16	3.3×10^3 [6.6]
D_{diag}	4	9.8×10^2 [2.0]

5. 実　　験

5.1 実 験 条 件

まず実験条件について述べる．

データベースの画像としては，画像共有サイト Flickr から得た 100 万画像を用いた．例を図 5 に示す．画像取得に用いたキーワードは，“birthday”, “food” などの語のほか，“2007.01.01” などの日付である．これらの画像は，データベースに格納する前に長辺が 320 画素以下になるように縮小した．

局所特徴量の抽出に際しては，生成型学習を適用し，図 2 に示すようなボケやブレを伴う画像を生成した．表 1 に，用いた画像数と抽出された参照特徴ベクトルの数を示す．ボケやブレの程度を大きくすると，それだけ抽出される特徴ベクトルの数は少なくなるため，画像数が増加するほどは，特徴ベクトルの数は増えていない．

検索質問画像については，データベースに収めた画像から無作為に取り出した 1,000 枚を印刷し，撮影することにより用意した．印刷に際しては，A4 の用紙に 16 枚の画像を印刷する場合 (以後，1/16 と記す)，4 枚の画像を印刷する場合 (1/4) の双方を試した．印刷領域が小さければそれだけ，精細な画像を得ることは困難になる．印刷された画像は，8 名の学生によって，それぞれ異なるカメラ付き携帯電話を用いて撮影された．使用した携帯電話の中には，マクロモードのあるものとないものがある．マクロモードがないと，1/16 の印刷に対して鮮明な画像を得ることは極めて困難となる．画像のサイズは

図 5　データベース中の画像の例.

QVGA, 画像数の合計は 8,000 枚である.

処理に用いたハッシュ表のサイズは, $H_{\mathrm{size}} = 2^d$, すなわち剰余演算を行わず, ビットベクトル全体を収めるサイズとした. 以下の実験では, 特に断りのない限り, 以下のパラメータ値を用いた. $b = 10$, $c = 100$, $d = 28$, $e = 400$, $t = 4$, $r = 0.4$. また, 以下で示す処理時間は画像 1 枚をデータベースと照合するのにかかった時間を意味し, 特徴抽出に要する時間は含まれていない. 処理に用いたコンピュータは, CPU が AMD Opteron 2.8GHz, メモリが 64GB のものである.

5. 2　生成型学習の効果

最初に, 生成型学習の効果について述べる. 認識には, 多段階化, スカラー量子化は用いず, ビット反転の閾値を b とした認識器 を 1 つだけ用いる. 特徴ベクトルは, 次元あたり 16bit で表現されている. データベースに含まれるオリジナル画像数 (図 2 の A に相当する画像の数) は 1 万である.

表 2 の 1 行目に結果を示す. 学習に用いる画像数が増えるにつれて, 認識率が向上していることがわかる. 例えば, オリジナルのみ (A) を用いて得られる認識率 81%が, 学習セット D を用いると 93.3%(+12.3%) まで改善される. 特に, マクロモードを用いずに撮影された画像に対して, 生成型学習は効果的であった. ある撮影者のデータに対しては, A に対する認識率 57.0%が D で 88.7%(+31.7%) まで改善した. ブレを伴わない学習セット $\mathrm{D}_{\mathrm{diag}}$ を用いた場合でも, 改善効果は得られるものの, ブレを含む学習セット D に比べて限定的である.

処理時間は, 生成型学習を用いた方が増加した. 例えば, 学習セット C を用いた場合, A の 2 倍程度の時間を要している. メモリについても同様である. 学習セット A で 2.5GB であったものが, 3.5GB (B), 4.3GB (C), 4.5GB (D) と増加した. 従って, 認識率の向上は, 処理時間とメモリ量の犠牲の上に成り立っているといえる.

図 1 に示した検索質問画像は, 学習セット A で認識に失敗したものの C を用いて成功した画像の一例である. この例からも分かるように, 生成型学習はこのように極めて低品質な画像を認識する上で効果的な手法である.

5. 3　多段階化の効果

次に, 多段階化の効果について述べる. 用いたパラメータとデータベースは上と同じである. スカラー量子化は用いていない. 結果を表 2 の 2 行目に示す. 多段階化なしの場合と比べて処理時間が大幅に短縮されている. 最も重要な点は, 学習セットが A から D になるにつれ, 参照特徴ベクトルの数が 6.6 倍に増加しているにもかかわらず, 処理時間が減少していることである.

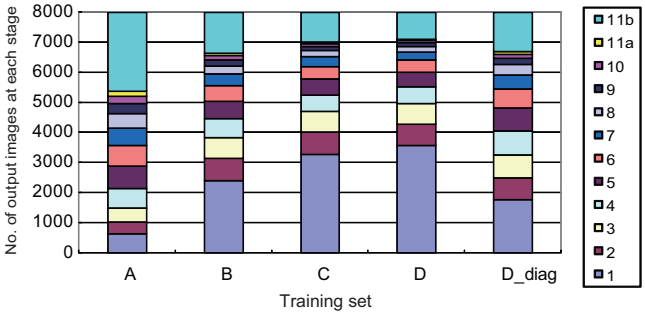


図 6　多段階の各段で出力された画像数.

このような現象が生じる原因を探るため, 検索質問画像が多段階のどの段階で認識されたのかを調べた. 図 6 に結果を示す. 個々の棒グラフは, 下から順に 1 段目, 2 段目,..., 最終段で出力された画像数を表す. ここで 11a は, 最終段 (11 段) で条件を満たして認識結果を得た場合, 11b は, 条件を満たさず最大得票数のものを強制的に認識結果とした場合を示す. グラフから, 生成型学習を用いると認識される段階が早まっていることがわかる. 認識容易な画像は多段階の早い段階で出力されることを考えると, 生成型学習は, 認識困難な画像を認識容易な画像に変換する効果を持つことがわかる. これにより, 処理時間を短縮することが可能となっている.

最近傍探索という観点からは, 次のように説明できる. 認識困難な画像の場合, 多段階の後段の方で, 探索範囲を広げて注意深く最近傍を探すことにより, 終了条件を満たす得票数の差を得ることができる. 生成型学習を用いると, 狭い探索範囲の中に, 生成型学習によって生成された, 最近傍となる参照特徴ベクトルが存在するため, 広範囲な探索を行う必要がない. これによって, 早い段階で十分な得票数差が得られ, 終了条件を満たすことになる. このように生成型学習と多段階を組み合わせることによって, 認識の高精度化だけではなく高速化という望ましい性質を得ることが可能となる.

5. 4　スケーラビリティ

最後にスケーラビリティについて述べる. データベースのサイズとしては, 100 万画像を最大として様々なものを試した. 生成型学習に用いた学習セットは C であり, スカラー量子化 (2bit/dim.) と多段階については, 適用する場合としない場合の双方を試した. また, $c = 250$ として, より認識率を重視した.

表 2　認識率 [%] と処理時間 [ms](両者とも平均)．データベースのサイズは 1 万画像.

DB	A(original)		B		C		D		D_{diag}	
cascade	recog. rate	time	recog. rate	time	recog. rate	time	recog. rate	time	recog. rate	time
なし	81.0	7.7	89.9	12.0	92.6	14.8	93.3	16.4	91.0	9.5
あり	81.0	2.3	89.9	1.7	92.5	1.5	93.2	1.5	90.9	1.6

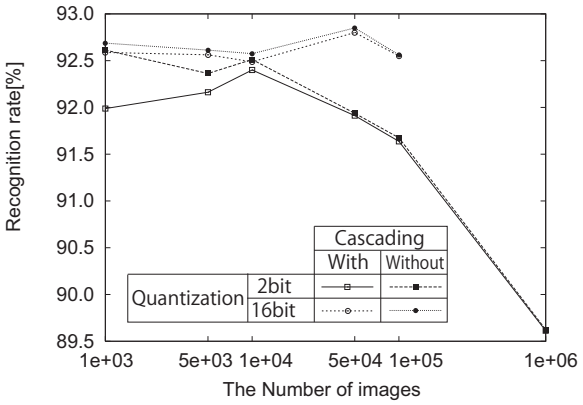


図 7　データベースのサイズと認識率の関係.

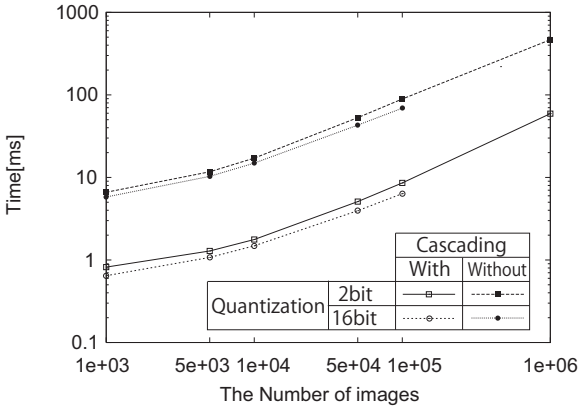


図 8　データベースのサイズと処理時間の関係.

まず，メモリ量について述べる．10 万画像のデータベースを用いる場合，スカラー量子化ありの手法で必要なメモリ量は 6.7GB, スカラー量子化なしの手法では 22.6GB である．100 万画像のデータベースの場合，スカラー量子化なしの手法は，計算機のメモリ (64GB) を超えるため適用できなかったが，スカラー量子化ありの手法は適用可能であった．実際に要したメモリ量は 31.6GB であった.

次に認識率と処理時間を見てみる．図 7 に，データベースのサイズと認識率の変化を示す．スカラー量子化によって 1%程度，認識率が低下していること，多段階は悪影響を及ぼしていないことが分かる．図 8 に処理時間との関係を示す．この図から，多段階により，それを用いない場合に比べて，10 倍程度の高速化を達成していることが分かる．スカラー量子化を導入すると量子化のための処理時間が少しかかるが，あまり大きな差にはなっていない.

最終的には，スカラー量子化と多段階化を用いると，100 万物体に対して，認識率 89.6%，処理時間 59.3ms/query を得ることができた．以上から，スケーラビリティを得る上で，多段階とスカラー量子化は有効であることがわかった.

6. おわりに

本論文では，低品質画像にも有効な特定物体認識手法を提案した．提案手法の特徴は，生成型学習により認識率を高めること，スカラー量子化によりメモリ量を圧縮すること，ならびに多段階化により処理時間を短縮することの 3 点にある．特に生成型学習と多段階化を組み合わせることによって，認識率の向上だけではなく高速化も達成可能となる点は重要である.

今後の課題には，より大量の画像を用いた実験に加えて，3 次元物体への拡張がある.

文　献

[1] C. Schmid and R. Mohr, “Local grayvalue invariants for image retrieval,” IEEE PAMI, vol.19, no.5, pp.530–535, 1997.

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

[3] J. Sivic and A. Zisserman, Video Google: a text retrieval approach to object matching in videos, Proc. of ICCV2003, pp.1470–1477, 2003.

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

[5] M. Özuysal, M. Calonder, V. Lepetit and P. Fua, “Fast keypoint recognition using random ferns,” IEEE PAMI. Accepted for publication.

[6] H. Ishida, T. Takahashi, I. Ide, Y. Mekada and H. Murase, Identification of degraded traffic sign symbols by a generative learning method, Proc. ICPR2006, pp.531–534, 2006.

[7] A. Andoni and P. Indyk, “Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions,” Comm. of the ACM, vol.51, no.1, pp.117–122, 2008.

[8] Y. Ke and R. Sukthankar, Pca-sift: A more distinctive representation for local image descriptors, CVPR2004, Vol. 2, pp.506–513, 2004.

[9] K. Kise, K. Noguchi and M. Iwamura, Memory efficient recognition of specific objects with local features, Proc. of the 19th International Conference of Pattern Recognition (ICPR2008), 2008.

[10] 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，pp.111–118，July，2007.

[11] P. Viola and M. Jones, Robust real-time object detection, Second Int’l Workshop on Statistical and Computational Theroies of Vision – Modelling, Learning, Computing, and Sampling, 2001.