

参照特徴ベクトルの増加による低品質画像の高速・高精度認識

黄瀬 浩一^{†a)} 野口 和人[†] 岩村 雅一^{†b)}Robust and Efficient Recognition of Low Quality Images
by Increasing Reference Feature VectorsKoichi KISE^{†a)}, Kazuto NOGUCHI[†], and Masakazu IWAMURA^{†b)}

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。1万平面物体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、並びに26億個の参照特徴ベクトルを用いて索引付けされた100万平面物体を、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

1. ま え が き

学習が進めばそれだけタスク実行の精度が向上するとともにタスク実行に必要な時間も短くなる。高速化、高精度化の双方が同時に実現されるというこのような理想的な学習は、残念ながら一般には成り立たない。通常は、学習を進めるためにデータを増加させると、精度は向上するものの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もつ手法を提案する。

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、という四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つな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、ぼけやぶれ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、ぼけとぶれ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。大きなブレイクスルーは、局所特徴量に基づく手法、すなわち Schmid ら [1] 並びに 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

[†] 大阪府立大学大学院工学研究科，堺市

Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuencho, Naka-ku, Sakai-shi, 599-8531 Japan

a) E-mail: kise@cs.osakafu-u.ac.jp

b) E-mail: masa@cs.osakafu-u.ac.jp

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．また，大量のデータが得やすく，カタログ認識などのサービスともつながるという理由から，認識対象としては平面物体（様々な写真；以後単に物体と呼ぶ）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）を増加させることにより，認識をより頑健かつ効率的にする点にある．局所特徴量の増加には生成型学習を用いる．すなわち，画像が被るばけやぶれ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，並びにスカラ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ることで，認識率を改善するだけでなく，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，効率並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 らは一つの VW がたかだか 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，その出現頻度をもと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 などの高い能力をもつ識別器や 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，局所特徴量が最もよく照合する物体を見つけたる処理となる．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万と膨大となるため，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．対策のかぎは，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．しかし前述のとおり，これは特定物体認識には有効でない．一方，スカラ量子化は特定物体認識に有効な手法である．例えば，2 bit/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



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。ぶれやぼけ、射影ひずみなどのために、認識が困難となっている。

Fig. 1 Examples of images in a database. They are hard to be correctly recognized due to motion-blur, defocus, perspective distortion, etc.

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、射影ひずみ、ぼけ、ぶれなどである。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。ぼけやぶれが画像に加えられる場合、多くの画像は似てくるため、データベースが大きいと認識はより困難となる。

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、ぼけやぶれ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畳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、垂直方向のぼけを表すパラメータ w は、ガウス関数の標準偏差と $\sigma = 0.3(w/2 - 1) + 0.8$ と

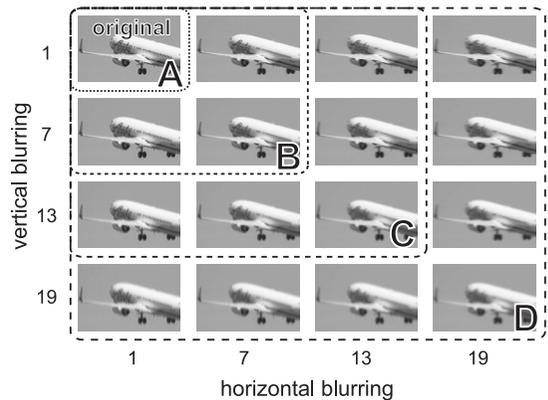


図2 生成された画像。左上がオリジナルの画像。右下にいくほど、ぶれやぼけが大きくなる。縦横の数字はぼけの程度を表すパラメータ。生成型学習では、オリジナルAに加えて、水平垂直のぼけを施して得られるB~Dの画像群を用いる。

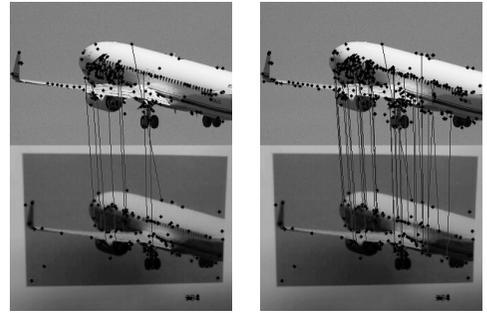
Fig. 2 Generated blurred images. The original image is shown at the upper left corner. The level of blurring increases for images from left to right, top to bottom. In generative learning, image sets B-D obtained by horizontal and vertical blurring are employed in addition to the original A.

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，ぼけだけではなくぶれもシミュレートできる．以下では，原画像のみを含む画像集合を学習セット A, それに 1 段階ぼかした画像を含めた画像集合 (B) を学習セット B と呼ぶ．学習セット C や D も同様に定義される．

適切な標準偏差の値は，検索質問画像が受けるぼけやぶれの程度に依存する．本研究では，図 1 のような画像に対処するため，予備実験により 1 から 19 という値を得た．一般に生成に用いる標準偏差の値を大きくすると，生成された画像から得られる局所特徴量の数が減少するとともに，相互に類似するようになる．その結果，誤認識が生じると予想される．したがって，生成ではこのようなぼかしすぎに注意しなければならない．なお，図 2 に示す程度のぼけやぶれでは全く問題にならない．

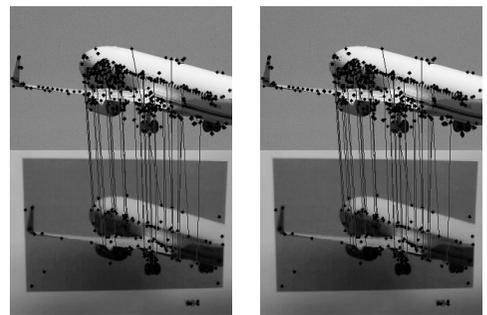
生成型学習の効果を図 3 に示す．(a)～(e) の小さい図において，下側の画像が検索質問画像，上側の画像がデータベースの画像を表す．また，(a)～(d) はそれぞれ図 2 の学習セット A～D に対応している．例えば (c) の場合，図 2 の C に対応する生成された画像 (9 枚) を用いて局所特徴量を取り出し，データベースに収める．(e) は，図 2 の対角部分のみを含む学習セットを用いた場合である．この学習セットにより，ぼけのみの生成型学習の効果が分かる．

上下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．下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，局所特徴量の間で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．ぶれを含まない (e) の場合であっても，生成型学習を用いない場合と比べてより多くの局所特徴量が照合できている．画像の認識の基本は局所特徴量の照合であり，検索質問画像の局所特徴量が正解の画像に数多く対応づけばそれだけ他の画像に誤認識される可能性が低くなる．したがって，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



(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)

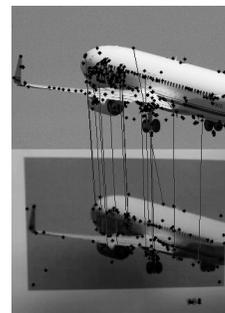
(e) 学習セット D_{diag} (16/342)

図 3 生成型学習の効果．(a)～(e) の画像組において，上側はデータベースの画像，下側はクエリ画像，線は特徴点間の照合結果を表す．

Fig. 3 Effect of generative learning. Upper and lower images in the pairs (a)–(e) represent database and query images, respectively, and the lines indicate matching of features.

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必要である。

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おそれもある。

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も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。

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すべ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、以下のとおりである。まず、検索質問画像から得た各局所特徴量につい

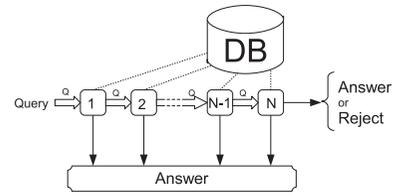


図 4 多段階に接続された認識器
Fig. 4 Cascaded recognizers.

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、すべての局所特徴量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のため、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とおりである。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。

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。すなわち、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 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、 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) 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) $p \in P_i^{(s)}$ に対して距離計算を行い、その中で最近傍(以後暫定最近傍と呼ぶ)を結果 $\hat{p}^{(s)}$ とする。

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

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 の要素すべ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更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、

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

が成り立つ。

今、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。

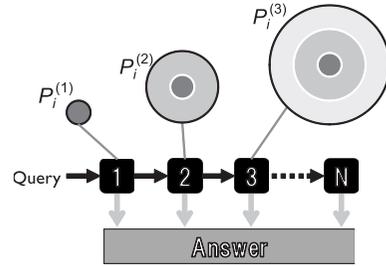


図 5 近似最近傍探索の単調性。円領域は各段で探索する空間の広さを表す。

Fig. 5 Monotonicity of approximate nearest neighbor search. Circles represent the search space explored at the corresponding steps.

[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。

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

[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。

これらの性質が処理時間短縮に果たす役割を以下に説明する。まず、単調性の概念を図 5 に示す。ここで、各段の上に示す円は、各段の認識器が探索する空間の広さを模式的に表している。単調性が成り立つということは、後段の認識器の探索空間が、前段のそれを含んでいることを意味する。

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。具体的には、図 5 に示す層状の探索空間において、新たに加わった最も外側のドーナツ状の部分だけを探索すればよい。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 段目までの多段階化を適用する場合の対象が次のように等しくなる。

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

ここで、 $P_i^{(0)} = \phi$ である。更に、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。

各段での処理の受渡しは次のように行われる。 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 $\hat{p}_i^{(s-1)}$ とする。 s 段では、 $X = P_i^{(s)} - P_i^{(s-1)}$ の要素に対して q_i との距離計算を行い、その中の最近傍

$$p_i^* = \arg \min_{p \in X} \|p - q_i\| \quad (8)$$

を得る。そして、 s 段の暫定最近傍を、

$$\hat{p}_i^{(s)} = \begin{cases} p_i^* & \text{if } \|p_i^* - q_i\| \leq \|\hat{p}_i^{(s-1)} - q_i\|, \\ \hat{p}_i^{(s-1)} & \text{otherwise,} \end{cases} \quad (9)$$

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 $p = (p_1, \dots, p_n)$ 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おおむね 16 bit で収まるため、ここではもとのベクトルを 16 bit で表現することとする。また、スカラ量子化を行う場合には、 p_j を 2 bit 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もつため、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 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 $u = (u_1, \dots, u_d)$ に変換する。ここで、

$$u_j = \begin{cases} 1 & \text{if } p_j - \theta_j \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

であり、 $d < n$ である。そして、次のハッシュ関数

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

により参照特徴ベクトル 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} はハッシュ表の大きさである。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、同じハッシュ値をも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、チェー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、チェーン法によって記録される要素数がしきい値 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。認識時に、削除されたハッシュ表のエントリに対応する特徴ベクトルが検索質問画像から得られても、ハッシュ表には記録が存在しないため、照合や投票の対象とはならない。これにより、識別能力が乏しい局所特徴量による誤投票を抑制することができる。

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） $q_i = (q_1, \dots, q_n)$ 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、

$$|q_j - \theta_j| \leq e, \quad (12)$$

を満たす場合には、 q_j の値がしきい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 はしきい値である。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とおりである。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一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すべ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．

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 6 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 に示すようなぼけやぶれ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．ぼけやぶれ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．なお，表 1 の D_{diag} は，図 2 の対角部分（ぼけ）のみを含む学習セットである．

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．画像の大きさは 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大きさは， $H_{size} = 2^d$ ，すなわち剰余演算を行わず，ビットベクトル全体を収め



図 6 データベース中の画像の例

Fig. 6 Examples of images in the database.

表 1 学習セット
Table 1 Training sets.

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]

る大きさとした．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 2.8 GHz，メモリが 64 GByte 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，スカラ量子化は用いず，ビット反転のしきい値を b とした認識器を一つだけ用いる．特徴ベクトルは，次元当り 16 bit 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分かる．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

表 2 認識率 [%] と処理時間 [ms] (両者とも平均). データベースの大きさは 1 万画像
 Table 2 Recognition rate[%] and processing time [ms] for each training set for the DB of 10,000 images and query set.

DB	A(original)		B		C		D		D _{diag}	
	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

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. ぶれを伴わない学習セット D_{diag} を用いた場合でも, 改善効果は得られるものの, ぶれを含む学習セット D に比べて限定的である.

処理時間は, 生成型学習を用いることにより増加した. 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 2.5 GByte であったものが, 3.5 GByte (B), 4.3 GByte (C), 4.5 GByte (D) と増加した. したがって, 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.

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 7 に結果を示す. 個々の棒グラフは, 下から順に 1 段階目, 2 段階目, ..., 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分かる. 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

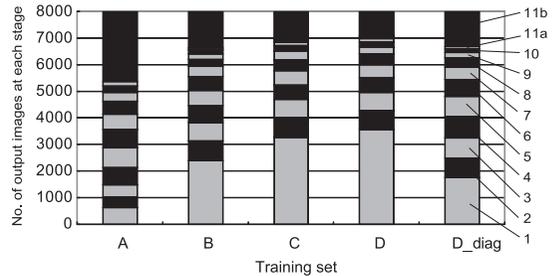


図 7 認識器の各段で出力された画像数. 棒グラフ左端の数字は, 認識器の段を表す.
 Fig. 7 The number of images outputted at each step. The numbers at the left side represents steps of cascaded recognizers.

識困難な画像を認識容易な画像に変換する効果をもつことが分かる. 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化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

次にスケーラビリティについて述べる. データベースの大きさとしては, 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, スカラ量子化 (2 bit/dim.) と多段階化については, 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.

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, スカラ量子化ありの手法で必

表 3 リジェクトをする場合の結果
Table 3 Results with rejection.

パラメータの決定基準 (学習セットに対して)		認識用検索質問				リジェクト用検索質問		
		C_1 [%]	E_1 [%]	R_1 [%]	処理時間 [ms]	E_2 [%]	R_2 [%]	処理時間 [ms]
基準 A	$\min(E_1 + E_2 + R_1)$, 処理時間 1 ms 以下	84.6	1.0	14.4	0.5	3.9	96.1	0.8
基準 B	$\min(E_1 + E_2 + R_1)$, 処理時間 10 ms 以下	88.1	0.4	11.5	1.3	0.9	99.1	5.5
基準 C	$\min(E_1 + E_2 + R_1)$, 処理時間 100 ms 以下	90.1	0.3	9.6	13.8	1.7	98.3	68.8

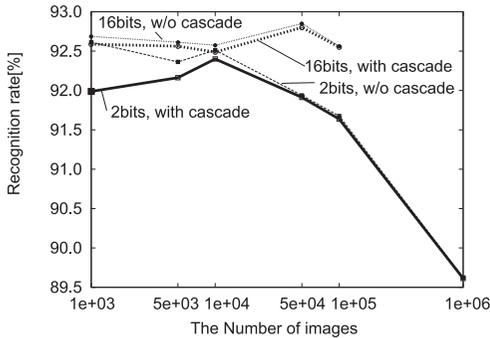


図 8 データベースの大きさと認識率の関係
Fig. 8 Recognition rate and the size of DB.

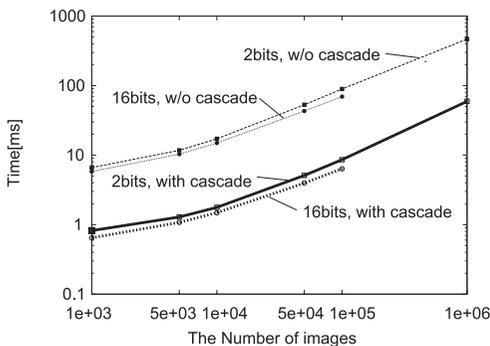


図 9 データベースの大きさと処理時間の関係
Fig. 9 Processing time and the size of DB.

要なメモリ量は 6.7 GByte, スカラ量子化なしの手法では 22.6 GByte である。100 万画像のデータベースの場合, スカラ量子化なしの手法は, 計算機のメモリ (64 GByte) を超えるため適用できなかったが, スカラ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 31.6 GByte であった。

次に認識率と処理時間を見つめる。図 8 に, データベースの大きさと認識率の変化を示す。スカラ量子化によって 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 9 に処理時間との関係を示す。この図から, 多段階化により, 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 59.3 ms/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, 多段階化とスカラ量子化は有効であることが分かった。

5.5 リジェクト

物体認識のタスクによっては, 正しく認識できることに加え, 誤った認識をしないことが重要な場合がある。そこで, 図 4 の最終 N 段目まで処理を行っても式 (1) と式 (2) の条件を満たさない場合に認識不能としてリジェクトする実験を行った。図 7 でいえば, 11b のクエリがリジェクトされることになる。

条件は以下のとおりである。データベースに含まれるオリジナルの画像数を 1 万枚, 学習セットを C とし, 多段階化とスカラ量子化の双方を用いた。また, リジェクト用の検索質問画像としては, データベースに含まれない画像 1,000 枚を同様に印刷し, 2 名の撮影者が異なる携帯電話 (マクロモードなし) で撮影した合計 2,000 枚を用いた。したがって, この実験で用いた検索質問画像数は合計 1 万枚である。

結果の評価尺度は次のように定める。まず, 対応する画像のある検索質問に対しては, 認識率 C_1 , 誤認識率 E_1 , リジェクト率 R_1 ($C_1 + E_1 + R_1 = 1$) とする。対応する画像のない検索質問に対しては, 誤認識率 E_2 , リジェクト率 R_2 ($E_2 + R_2 = 1$) とする。実験には, 10-fold cross validation を用いた。学習サンプルに対して, 表 3 の 3 通りの基準を用いてパラメータを求め, テストサンプルに適用した。パラメータについては, $b = 5, 10, 15$, $c = 2, 5, 10, 100$, $d = 20, 24, 28$, $e = 200, 400, 600$, $r = 0.2, 0.4, 0.6$, $t = 4, 8, 12$ のすべての組合せについて試した。結果を表 3 に示す。処理時間については, リジェクト用検索質問の方が長くなっている。これは, リジェクトとして回答するためには最終段まで到達する必要があるためである。ただ

し、表 2 に示した多段階化なしの処理時間と比べて大幅に増加することはなかった。処理時間を 10 ms、リジェクト率 R_1 を 10% 程度許容すると、誤認識率が 1% 以下という誤りにくい認識を実現できた。

6. む す び

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、並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を加えて、3 次元物体認識への適用がある。

謝辞 本研究の一部は、日本学術振興会科学研究費補助金基盤 (B) (19300062) の補助による。

文 献

- [1] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.5, pp.530–535, 1997.
- [2] D. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., 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. 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 Trans. Pattern Anal. Mach. Intell., vol.32, no.3, pp.448–461, 2010.
- [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," Commun. 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. 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 Theories of Vision – Modelling, Learning, Computing, and Sampling, 2001.

(平成 21 年 9 月 24 日受付, 22 年 2 月 27 日再受付)



黄瀬 浩一 (正員)

昭 61 阪大・工・通信卒。昭 63 同大学院博士前期課程了。同年同大学院博士後期課程入学。平 2 阪府大・工・電気助手。現在、同大学院工学研究科教授。博士 (工学)。平 12~13 ドイツ人工知能研究センター客員教授。文書画像解析, 情報検索, 画像認識などの研究に従事。平 18 本学会論文賞, 平 19 IAPR/ICDAR Best Paper Award, 平 22 IAPR Nakano Award (DAS 2010 Best Paper Award) 各受賞。現在, IAPR TC11 (Reading Systems) Vice Chair, IAPR ICPR Standing Committee, Conferences & Meetings Committee 委員。情報処理学会, 人工知能学会, 電気学会, IEEE, ACM 等各会員。



野口 和人

平 19 阪府大・工・情報卒。平 21 同大学院博士前期課程了。現在, ソニー (株) 勤務。在学中, 画像認識に関する研究に従事。平 21 本会 PRMU 研究奨励賞受賞。



岩村 雅一 (正員)

平 10 東北大・工・通信卒。平 15 同大学院博士課程了。同年同大学院工学研究科助手。平 16 阪府大大学院工学研究科助手。現在助教。博士 (工学)。パターン認識, 画像認識, 情報検索などの研究に従事。平 18 本学会論文賞, 平 19 IAPR/ICDAR The Best Paper Award, 平 22 IAPR Nakano Award (DAS 2010 Best Paper Award) 各受賞。IEEE, 情報処理学会, Pattern Recognition Society 各会員。