

画像認識のための生成型学習

村 瀬 洋†

画像認識は、入力パターンの多様な変動や変形を吸収し、モデルあるいは学習パターンと照合することにより実現できる。認識の前処理として用いられる入力パターンの正規化は、変動や変形を吸収するための1つの手法である。一方、変動や変形に対処するための別の手法として、学習パターンを変動や変形により様々に生成し、これと入力パターンと照合する手法も考えられる。この学習パターンを生成する処理は、認識の事前処理である学習段階に行うことが可能であり、生成型の学習といってもよい。またこの手法は、学習パターンを多数生成することが可能になるので、収集できる学習パターン数が少ない場合に有効である。本論文では、著者自身がこれまで行ってきた画像認識の研究例を用いて、生成型学習の観点で体系的に整理してみた。

Generative Learning for Image Recognition

HIROSHI MURASE†

This paper describes a generative learning method for image recognition. Generally, an input pattern is normalized in the preprocessing stage of a recognition system to cope with the deformation of the pattern. In contrast, our generative learning method synthesizes the learning patterns by image generation techniques such as adding noise, interpolation, generating function. This method can improve the recognition accuracy even if we have a small number of learning samples. This paper introduces several examples that authors applied in their recognition systems so far.

1. はじめに

画像認識は、画像中の入力パターンの多様な変動や変形を吸収し、標準パターン（以下、学習パターンと呼ぶ）と照合することにより実現できる。変動や変形を吸収するための代表的な手法には、入力パターンに対する前処理がある。たとえば、入力パターンの位置や大きさの正規化、形状の平滑化、画素値の正規化などがあげられる。これらの前処理により、入力パターンの各種変動を正規化した後に学習パターンとの照合を行うことにより、変動や変形に対してロバストな認識が実現できる。

入力パターンの変動や変形に対処するもう1つの手法として、学習段階で可能な限り様々に変動や変形したパターンを収集し、その収集したすべてを学習パターンとして登録し、それらと入力パターンとを照合することにより認識する手法が考えられる。またこれらの多様な学習パターンを用いて精度の高い識別器を設計することも可能である。しかし、観測されるす

べての可能性を含むように、変動や変形したパターンを収集することは一般に困難である。それに対して、収集した少数のパターンを基に様々に変動や変形させた学習パターンを生成した後に、入力パターンと照合する手法が考えられる。この生成処理は認識の事前処理である学習段階に行うことも可能であり、生成型学習といってもよい。本論文では、著者自身がこれまで行ってきた画像認識の実例を用いて、生成型学習の観点で整理してみた。同様な考え方は、パターン認識の分野で古くから多数の研究者が、暗黙にあるいは明示的に利用している手法でもある。さらには、そもそも学習とは少数の学習パターンから汎化能力の高い識別器を構成する処理であるという議論もある。しかし、これまで生成型学習という観点から体系的に整理はされていない。

生成型学習では、いかに少数の学習パターン（以下、これを基準学習パターンと呼ぶことにする）から入力パターンに含まれるような変形や変動した学習パターンを生成するかが重要な鍵となる。これまで著者自身が利用した学習パターン生成方法を、以下のように整理してみた。

① 基準学習パターンにランダムな微小変動を加え

† 名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

ることにより学習パターンを生成する（例：変動吸収特性核 [1981]）。

- ② 基準学習パターンにフィルタをかけることにより学習パターンを作成する。このフィルタを生成関数と呼び、生成関数自体は基準学習パターンの一部から推定する（例：低解像度文字認識 [2004]）。
- ③ 複数の基準学習パターンを特徴空間で補間や外挿することにより学習パターンを生成する（例：パラメトリック固有空間法 [1993]）。
- ④ 複数の基準学習パターンを画像上で形状の補間や外挿をすることにより学習パターンを生成する（例：パラメトリック固有空間法の改良 [1996]）。
- ⑤ 入力時の各種情報から動的に学習パターンを生成する（例：画数・筆順自由手書き線図形認識 [1982]）。

本論文では、これらの学習パターンの生成法を具体的な認識手法を用いて説明する。

2. 生成型学習の特性

上述したように、入力パターンの変動や変形を吸収するためには、入力パターンを前処理で正規化する正規化法と、モデルや少数の基準学習パターンから変形や変動を生成する生成型学習法が考えられる。ここでは、各手法の長所、短所を考察してみる。

正規化法は、大きさ、位置、形状変形などの情報をボトムアップに規格化する手法である。この手法は、入力図形をいったん正規化してしまえば、その後に単純に学習パターンと照合すればよい。ただし、正規化のうち、大きさや位置などの正規化は基準が明確なために比較的簡単に実現できるが、形状などの正規化は、形状を抽出する処理が必要なうえに正規化するための基準が曖昧なため、簡単には実現できない。また、変動や変形が大きい場合には、簡単な正規化では変動や変形が吸収しきれない。変動を正規化するためにはその変動の逆関数を適用することになるが、一般的にその逆関数を求めることは困難である。

生成型学習法では、生成するための指針さえ与えれば、とにかく多様な変動が容易に生成できる。また変動の発生する機構がモデル化できれば、生成関数を推定でき、逆関数を求める必要はない。さらに、仮にランダムに生成していても入力パターンに近いものが偶然に生成される可能性もある。しかし、様々な画像が生成されるために、生成したパターンをどのように学習に利用するか工夫が必要である。またあまり大き



図1 ボトムアップな正規化は困難な例
Fig.1 An example that is difficult to preprocess.

な変動幅で生成してしまうと、他クラスとのパターンの重なりが大きくなり誤認識の原因にもなる。

我々の思考の過程においては、一般に正規化より生成のほうが考えやすい。その例として、図1に示す画像を用いて説明する。この画像はトップダウンに処理すると犬が知覚できるという図であるが、この入力画像から犬がいるという知識なしに犬の画像を抽出し、正規化することは簡単ではない。しかし、犬がここにいると仮定すると、この絵と類似した絵を生成できるかもしれないという気になる。しかし、実際には完全に同一の絵が偶然に生成される確率はほとんどないことも事実である。つまり、まったく手がつかない正規化法よりも、確率的にはほとんど同じものは生成されないものの、何らかの処理ができる方が考えやすい場合も多い。

3. 生成型学習の具体例

本章では、前章で示した生成型学習法を、具体的な研究例で説明する。取り上げた研究例には古いものもあるが、考え方として分かりやすい例を取り上げた。もちろん、これ以外にも類似した手法は種々あると思われる。

3.1 微小な変動を学習パターンに加える手法

まず、少数の基準学習パターンに適当な微小変動を加える手法が考えられる。その例には「変動吸収特性核」と名付けられた手法¹⁾がある。これは部分空間法²⁾において、学習パターンに微小変動を加えることで、認識精度を向上させる手法である（図2参照）。

手書き文字認識の1つの課題は、多様な手書き変動や変形をいかに吸収するかである。もちろん、これを解決するために文字認識で一番重要なことは、手書き変動に安定な特徴を抽出することであり、実際に、これまで多数の識別に有効な特徴が提案されてきた。しかし、大きな変動がある場合や、十分な学習パターンが得られない場合には十分な精度が得られないこともある。しかしここで変動パターンをうまく生成すること

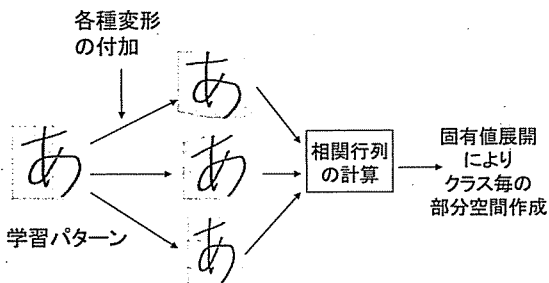


図2 部分空間法における変形の付加

Fig.2 Adding deformation to the learning patterns in the subspace method.

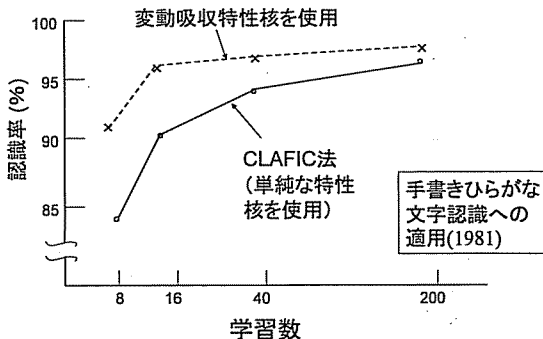


図4 変動吸収特性核の実験結果

Fig.4 Experimental results when using a deformation matrix.

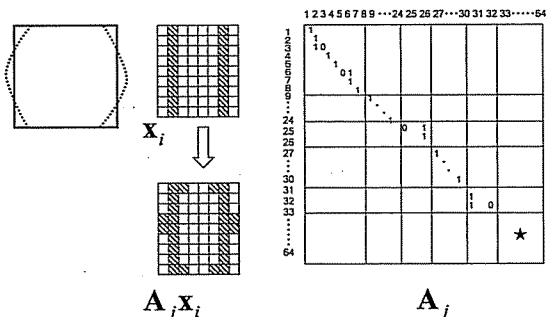


図3 変形行列 A_i の例

Fig.3 Deformation matrix.

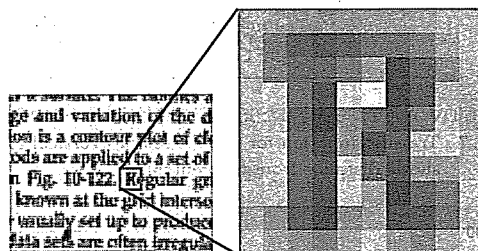


図5 低解像度の文字

Fig.5 A low resolution character.

ができれば、認識率の向上が望める。ここでは、部分空間法において学習パターンに変動を加える手法である変動吸収特性核について紹介する。まず文字画像を変形させる変形行列 A_i を定義し、これを多数用意する。

部分空間法では、最初に基準学習パターンから自己相関行列(特性核) R を作成する。その際に、変形行列 A_i を特性核 R に対して掛けることにより新しい特性核 $\tilde{R} = \sum_i A_i R A_i^T$ を作成する。変形行列 A_i をうまくデザインすれば変動を吸収する部分空間を作成することができる。この特性核のことを変動吸収特性核と呼んだ。変形行列の例を図3に示す。

手書き平仮名文字認識に適用したところ、図4に示すように認識率は大幅に向上している。この実験では21種類の変形行列を用意した。特に学習パターンが少ないときに有効であることが分かった。古い実験であり、現在の文字認識の水準からすると認識率の絶対値は高いとはいえないが、変動吸収特性核の効果は確認できる。

3.2 生成関数を用いた生成

この例は、少数の基準学習パターンからボケなどのフィルタ(これを生成関数と呼ぶことにする)を用いて学習パターンを生成する手法である。著者らは、低解像度の文字を認識するための生成型学習法²⁾を提案

した。認識の枠組みは同じく部分空間法を用いた。

近年、カメラ付き携帯やデジカメの普及は著しい。これらで撮影した画像から文字を認識するニーズは高い。カメラの解像度は年々向上しているが、どれだけカメラの解像度が向上しても、一度にたくさんの文字情報を入力したいという要求はあり、その段階で低解像度の画像を認識したいという問題は発生する。低解像度文字の例を図5に示す。

低解像度の文字を認識するためには、たとえば超解像のように複数の低解像度の画像を集めることにより解像度の高い画像を生成してから認識する手法が考えられる。もう1つの手法が生成型学習のアプローチである。もしある文字が低解像度で観測されるとすれば、それはどのような画像になるであろうかを考え、そのような画像を生成して認識する手法²⁾である。後者のアプローチについて紹介する。

低解像度の画像が得られる過程(図6参照)を考えると、光学的なボケ、撮像系でのボケや画素に対応するセンサの配置にともなうボケなど様々な要因が考えられる。ここでは、第1近似として、これらを一括して生成関数という名前のフィルタで表現する手法を試した。生成関数は、点広がり関数(PSF)と画素のサンプリングの配置をモデル化したものである。

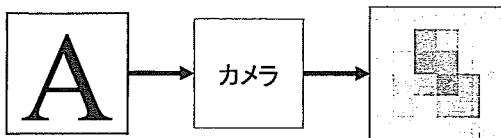


図 6 画像の生成モデル
Fig. 6 A camera model.

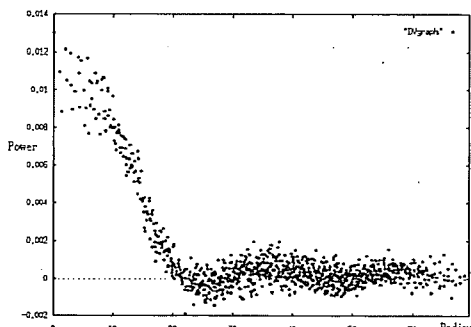


図 7 DV カメラから推定した伝達関数 (横軸は中心からの距離)
Fig. 7 Transfer function for a DV camera (Vertical axis shows distance from the center).

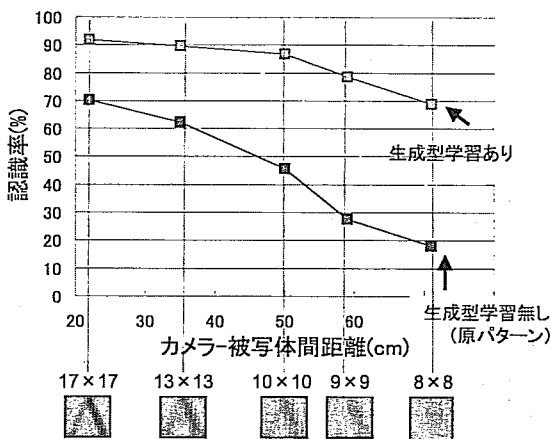


図 8 生成型学習の効果
Fig. 8 An experimental results when using generative learning.

まず、事前処理として高解像度の画像とそれをカメラで撮影した際に発生する劣化した画像を用いて、カメラに対応した点広がり関数を推定する (図 7 参照)。さらに解像度に合わせたサンプリング位置を決めこれにより劣化関数を構成する。次に、高解像度の基準学習パターンから生成関数を用いて多数の学習パターンを生成した。多数の学習パターンとの照合には部分空間法を用いた。その理由は、生成した多数のパターンは相互に高い相関を持つこと、相関を持つ多数のパターンとの照合を効率的に適用できるためである。その結果を図 8 に示す。

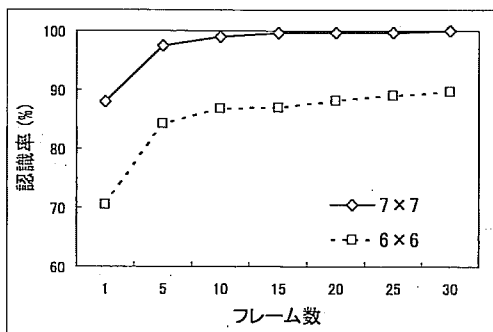


図 9 入力フレーム数に対する認識率
Fig. 9 Recognition results vs. a number of frames.

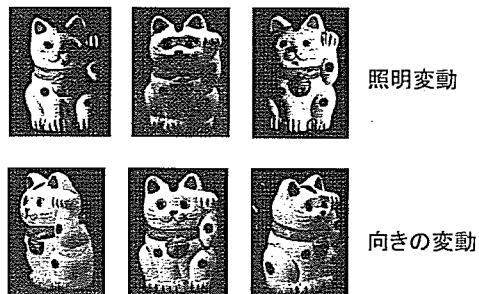


図 10 照明や向きの変動による見かけの違い
Fig. 10 Appearances when changing illumination and object pose.

生成型学習のポイントは、いかに学習パターンを生成するかと、生成したパターンをいかに効率的に照合するかによる。ちなみに、ここでは複数の入力フレームを利用して高い認識精度を実現している (図 9 参照)。本手法では部分空間法で得られた類似度を複数の入力フレームに対して平均化することにより新しい類似度を計算する手法を提案したが、相互部分空間法³⁾なども利用することが可能である。

また、生成パターンの作成には、たとえば Baird が提案しているコピーなどともなう文字の品質劣化モデル⁴⁾などを併用することによりさらに応用範囲が広がる。

3.3 特徴空間上での補間による生成

この例は、複数の基準学習パターンを補間することにより新しい学習パターンを生成する手法である。これは 1992 年に提案のパラメトリック固有空間法^{5),6)}の中に利用されている。パラメトリック固有空間法では、物体の向き、照明の方向を補間することにより、物体の向きや照明の変動 (図 10) にロバストな物体認識を実現している。

パラメトリック固有空間法は 2 次元画像の照合により 3 次元物体を認識する手法である。3 次元物体は見

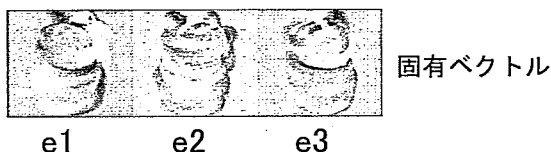


図 11 固有ベクトルの例 (図 10 の画像)
Fig. 11 Eigenvectors for the object shown in Fig. 10.

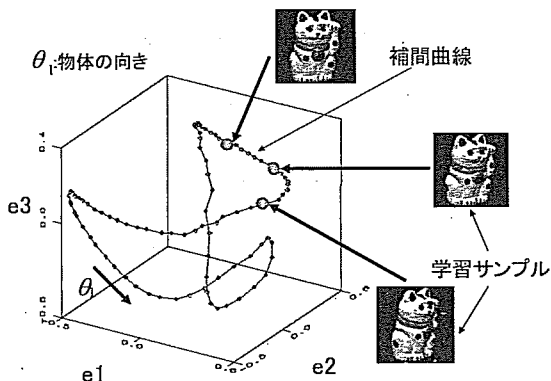


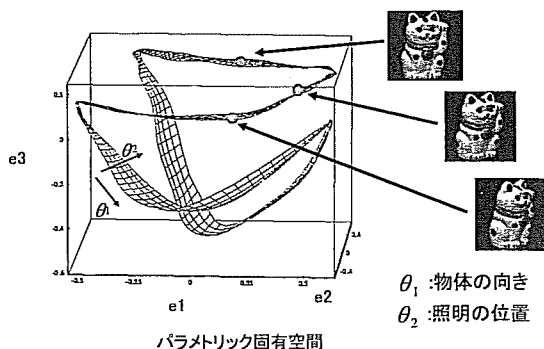
図 12 学習パターンの特徴空間上での補間
Fig. 12 Interpolation among learning patterns in the feature space.

る方向や照明により見かけが大きく変化する。そこで様々に変化する見かけ画像をパラメトリック固有空間表現で表し、認識に用いる。パラメトリック固有空間法では、学習段階では、最初に、認識すべき物体を様々な方向、様々な照明条件⁷⁾で体系的に撮影する。それを学習パターンとして固有空間を用いて画像圧縮を行う (図 11)。

次に各学習パターンを固有空間に投影し、その特徴点を記憶する。学習パターンのそれぞれの見かけ画像が大きく異なる場合には、特徴点も固有空間上でまばらに分布する点となる。一般的に物体の見かけ画像の連続的な変化は、固有空間上の点の連続的な軌跡になる (図 12)。つまり特徴点をスプライン関数などで補間することで、中間的な画像の認識を実現することができる。また照明条件についても同様に固有空間上で補間を行い、中間的な照明条件 (図 13) を表現することができる。

認識段階では、未知の画像をこの固有空間に投影し、その特徴点があらかじめ作成した軌跡の上に存在すれば、その物体とし、そうでなければその物体ではないとする。軌跡は物体の向きや照明条件の情報を持っているので、その情報を抽出することにより、物体の向きの検出も行うことができる。

パラメトリック固有空間法では、学習パターンを、固有空間 (特徴空間) の中で補間していることになる。



パラメトリック固有空間
図 13 向きと照明の両パラメータの補間
Fig. 13 Interpolation for both illumination and object-pose parameters.

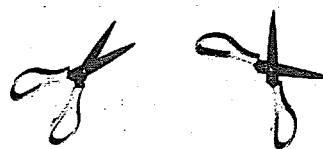


図 14 観測された 2 枚の画像
Fig. 14 Two observed images.

つまり、特徴空間上で学習パターンを補間した特徴点と照合を行っていることになる。

3.4 画像上での形状の補間およびアフィン変換

前節で述べたような特徴空間上での補間ではなく、直接、画像空間で少数の基準学習パターンを補間する手法が考えられる。その例として改良型のパラメトリック固有空間法⁸⁾がある。前節のパラメトリック固有空間法では、2つの学習パターン間の見かけ画像の生成を、特徴空間 (固有空間) 上での補間により実現していた。しかし、もし学習パターンの数が非常に少なく、特徴空間が画像を直接、主成分分析したものであった場合には、必ずしも補間が正しく機能するとは限らない。その理由は、画像の主成分分析は、基本的にはもとの画像の線形結合により表現する特徴空間であるためである。その空間の中でどのように補間を実現しても、もともと存在しない中間的な画像は生成されない。

もし特徴空間上ではなく、画像空間上で2枚の学習パターン (図 14 参照) を補間した画像が生成 (図 15 参照) できれば、より補間の精度が向上することが期待できる。画像を補間する技術としてはモーフィングが知られる。ここでは、図 14 に示す画像を用いて、特徴空間上で補間をする場合と、画像空間上で補間する場合の比較を行う。

ここで利用した画像補間の手法 (モーフィング) を簡単に紹介する。まず、2枚の基準学習パターンの間の

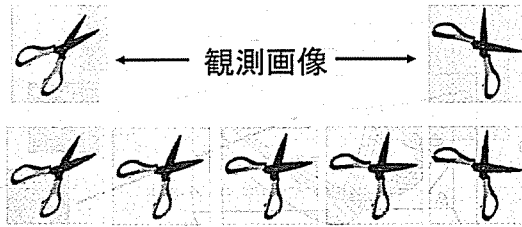


図 15 画像の補間

Fig. 15 Interpolation between two images.

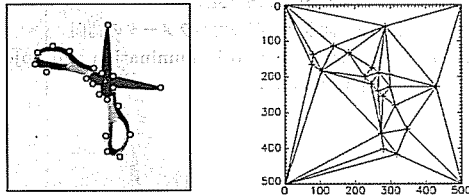


図 16 補間に用いた制御点と三角分割

Fig. 16 Control points for interpolation, and triangulation.

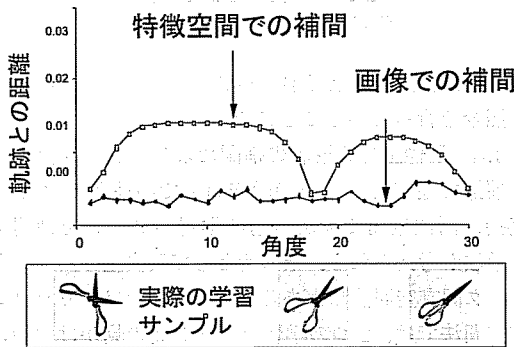


図 17 特徴空間での補間と画像での補間の比較

Fig. 17 Comparison between interpolation in feature space and interpolation in image space.

対応点を人手で与える。この対応点を用いて三角パッチを生成する（ドローネー網）。各三角パッチごとに学習パターン間の中間画像を生成する（図 16）。図 17 に特徴空間で補間した場合と、画像空間で補間した場合の、固有空間上の軌跡と入力パターンとの距離を示す。特徴空間上での補間は基準学習パターンから形状が離れるに従い距離が大きくなること分かる。

実験では、はさみやペンチなどの 4 種類の物体を用いて認識率を評価した。学習では、画像の補間に加えて画像の回転のアフィン変換も同時に行い学習パターンを生成した（図 18）。図 19 に 3 個の基準学習パターンを用いた場合の補間の効果を示す。3 個の基準学習パターンを特徴空間上で補間すると 92.5% 程度の認識率であったのが、画像上で補間することによりほとんど認識できるようになることが分かる。一方で、

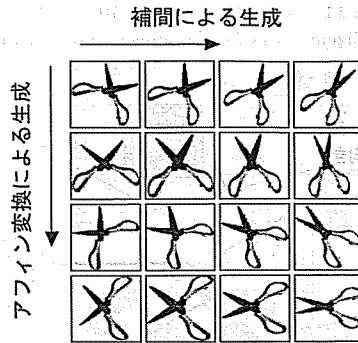
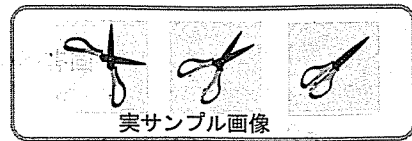


図 18 画像空間での補間とアフィン変換で学習パターンを生成

Fig. 18 Generating learning patterns using interpolation and affine transformation.

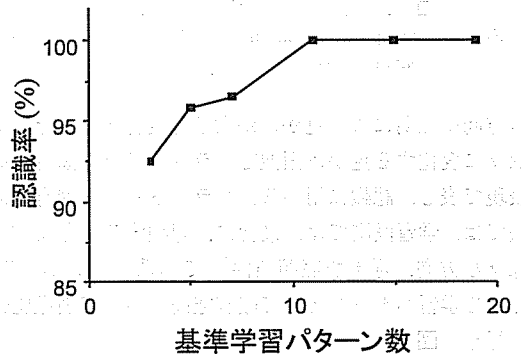


図 19 認識実験結果（画像上で補間する効果）

Fig. 19 Recognition rates vs. a number of the interpolated patterns.

このような認識対象の場合、12 個以上の学習パターンが用意できれば、あるいはこれらが画像空間で補間生成できれば、あとは特徴空間上での補間で十分であることも分かる。

関連研究として、顔画像認識で、モーフィングで生成した様々な顔を用いて認識する手法^{9),10)} などもある

3.5 学習パターンの動的な生成

上述の生成型学習と少し異なるが、認識時に動的に学習パターンを生成する手法がある。ここでの処理は厳密には学習とは呼べないが、前節との整合性のためにあえて学習パターンと呼ぶ。以下の例は、画数筆順に依存しないオンライン手書き線図形認識¹¹⁾ である。

手書き線図形認識は、優れたマンマシンインタフェースを実現するための技術である。一般にオンライン型の認識ではペンのストローク（軌跡）を照合に使用する

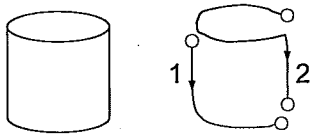


図 20 シンボル例とその手書きストローク例

Fig. 20 An example of hand-writing stroke sequence.

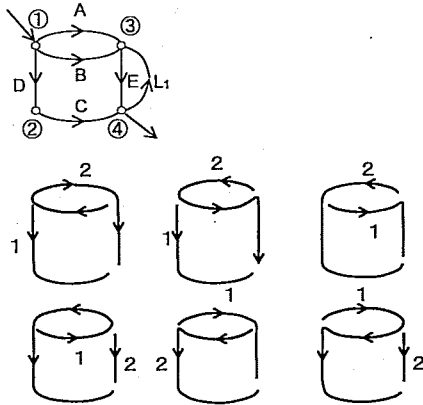


図 21 動的に学習パターンを生成した例

Fig. 21 Generating stroke sequences.

るために、学習パターンとしてストローク系列を用意する必要がある。漢字などの手書き文字の場合には、標準的な画数と筆順がある。また、義務教育でこれらを学ぶ。そのため、正しい画数や筆順が正確に守られていないにしてもそのバリエーションはあまり大きくなく、各文字に対してある程度のバリエーションを事前に用意しておけば実用上はほとんどの画数や筆順に対応できる。

しかし、線図形には定められた画数や筆順がない。たとえば、図 20 に示す磁気ディスクをとっても、10名の人の書き方を確認したところ、20種類以上の書き方のバリエーションが存在した。このように書き方のルールのない図形の場合、ストロークのバリエーションすべてを学習パターンとして登録しておくことはあまり現実的ではない。そこで考え出された手法が動的なストローク生成である。

入力時に入力ストローク列から抽出した簡単な情報を基に、その情報から得られた制約を持つストローク列を自動的に生成する手法である。たとえば、入力図形として図 20 (右) のように書かれたとする。このストロークの開始点と終了点のみを用いて、このように書けるストローク列の生成を行う。これは N 筆書きの問題(一筆書き問題の拡張問題)となり、グラフ探索で実現できる。この例の場合、図 21 に示す 6 種類の図形が生成された。次の段階ではこの 6 種類の生成

された標準ストローク列と、入力ストローク列を照合すれば認識できる。これは、入力の情報により生成パターンを制約し、可能性のある少数の学習パターンのみを生成する。つまり生成空間を小さくすることで、処理量を効率化した例である。

4. む す び

本論文では、画像認識における生成型学習についての様々な方法を、著者自身の研究を例に紹介した。画像認識は、入力パターンの多様な変動や変形を吸収し、モデルあるいは学習パターンと照合することにより実現できる。変動や変形を吸収するための代表的な手法としては、画像認識における前処理があるが、もう 1 つの手法として、基準学習パターンを様々に変動や変形させた後に、入力パターンと照合する手法である。この処理は認識の学習段階に行うことが可能であり、特に収集できる基準学習パターン数が少ない場合に有効である。

生成型学習では、いかにモデルから、入力パターンに含まれるような変動や変形した学習パターンを生成するかが重要な鍵となる。生成の例として、(1) 基準学習パターンに微小変動を付加することにより生成、(2) 生成フィルタなどの生成関数により生成、(3) 基準学習パターンの特徴空間でのデータの補間で生成、(4) 2 次元パターンの補間やアフィン変換などにより生成、(5) 入力時の各種情報から動的に生成する手法などが考えられる。本論文では、これらを具体的に紹介した。

同様な考え方は、古くから多くの人によりパターン認識の分野で、暗黙にあるいは明示的に利用されている。今後も同様な研究がますます発展していくことを期待している。

参 考 文 献

- 1) 村瀬, 木村, 吉村, 三宅: パターン整合法における特性核の改良とその手書き平仮名文字認識への応用, 電子通信学会論文誌, Vol. J64-D, No.3, pp.276-283 (1981).
- 2) 石田, 柳詰, 目加田, 井手, 村瀬: 部分空間法による低解像度文字認識のための生成型学習法, 信学技法, PRMU2004-7, pp.37-42 (2004).
- 3) Yamaguchi, O., Fukui, K. and Maeda, K.: Face recognition using temporal image sequence, *ICAFGR*, Nara, Japan, pp.318-323 (1998).
- 4) Baird, H.: Document Image Defect Models and Their Uses, *ICDAR*, pp.20-22 (1993).
- 5) 村瀬, Nayar: 2 次元照合による 3 次元物体認

- 識—パラメトリック固有空間法, 信学論, Vol.J77-D-II, No.11, pp.2179–2187 (1994).
- 6) Murase, H. and Nayar, S.K.: Visual Learning and Recognition of 3-D Objects from Appearance, *International Journal of Computer Vision*, Vol.14, pp.5–24 (1995).
- 7) Murase, H. and Nayar, S.K.: Illumination Planning for object recognition using parametric eigenspace, *IEEE Trans. PAMI*, Vol.16, No.12, pp.1219–1227 (1994).
- 8) Murase, H. and Nayar, S.K.: Learning by a generation approach to appearance-based object recognition, *IAPR ICPR96*, pp.24–29 (1996).
- 9) Vetter, T. and Poggio, T.: Linear Object Classes and Image Synthesis from a Single Example Image, *IEEE Trans. PAMI*, Vol.19, No.7, pp.733–742 (1997).
- 10) Okada, K., Akamatsu, S. and Malsburg, C.: Analysis and synthesis of pose variations of human faces by a linear PCMAP model and its application for pose-invariant face recognition system, *ICAFGR*, pp.142–149 (2000).
- 11) 村瀬, 若原, 梅田: 候補ラティス法による手書きフローチャートのオンライン認識, 信学論 (D), Vol.J66-D, No.6, pp.675–682 (1983).
- 12) 石井, 上田, 前田, 村瀬: わかりやすいパターン認識, オーム社 (1998).
(平成 17 年 1 月 10 日受付)
(平成 17 年 7 月 18 日採録)

(担当編集委員 佐藤 洋一)



村瀬 洋 (正会員)

昭和 30 年生。昭和 55 年名古屋大学大学院工学研究科電気電子工学専攻修士課程修了。同年日本電信電話公社 (現在の NTT) に入社。平成 4 年から 5 年にかけて米国コロンビア大学客員研究員。平成 15 年より名古屋大学大学院情報科学研究科教授。文字認識, 画像認識, マルチメディア認識の研究に従事。工学博士。平成 6 年 IEEE-CVPR 最優秀論文賞, 平成 7 年山下記念研究賞, 平成 13 年高柳記念奨励賞, 平成 14 年電子情報通信学会業績賞, 平成 15 年文部科学大臣賞, 平成 16 年 IEEE Trans. MM 論文賞ほか受賞。電子情報通信学会, IEEE-CS 各会員。