

風景画の理解に関するJAVA言語による RECOGNITRONの計算機シミュレーション

鈴木 昇一 川俣 博司 大槻 善樹

A Computer Simulation Carried Out by Program Language JAVA about a System RECOGNITRON Which can Understand a Scene Image

Shoichi Suzuki Hiroshi Kawamata Zenju Otuki

あらまし

本論文では、風景画像に関する、文献 [6] の理解理論をほぼ忠実に、JAVA言語でシミュレーションした結果が報告される。画像内容を理解する従来の方法と決定的に相違する点は、入力画像の意味はこの画像から想起される内容であるとしていることである。

風景画像を理解する本研究で構築された本システムは、パターンモデル $T\phi$ を見るとあたかも原パターン ϕ であるかのように錯覚する(パターンモデル $T\phi$ と原パターン ϕ と間の同一知覚原理)。

処理の対象とするパターン ϕ のパターンモデル $T\phi$ を次々と多段階にわたって連想的に変換しながら最終的には或るカテゴリの不動点パターンモデルを求め認識する認識法(不動点連想形多段階認識法)がS.Suzukiによって既に提案されている。この多段階認識法を画像内の各画素ごとにその各画素の帰属するカテゴリ(類概念)を決定すること(認識)に適用し、風景画像を理解するシステムRECOGNITRONが既に構築されているが[6]、本論文では、このRECOGNITRONをJAVA言語で表現し、1枚の風景画像の内容を理解する機能に関する計算機シミュレーション結果が検討される。

多段階認識法によって決定された各画素の帰属するカテゴリをすべての画素について集めると、本画像理解システムRECOGNITRONが一枚の画像をどのように理解しているかが判明することになる。RECOGNITRONが今注目している画素の近傍にある画像成分の情報を用いて、この画素の帰属するカテゴリを決定し、得られた帰属するカテゴリのラベルをその画素に貼りつけるとしよう。その結果、画像内の全ての画素について、その帰属するカテゴリラベルが付与される。このようにして、風景画像内に空、木、車、家、道路、電柱(6個のカテゴリ)が存在するかどうかをRECOGNITRONは理解する。

各画素にカテゴリラベルを付与する形式のため、領域分割(セグメンテーション)を必要としない認識処理を実現することができ、画素単位に一枚の画像についてその知識を詳細にかつ連想的に抽出できることが特色である。本研究において構築されたシステムは、その万能性を備えていることが証明されている"S.Suzukiのパターン情報処理に関するSS数理"を素直に画像理解に適用して得られたものである。

その性能を確かめた本画像理解法の特徴として、次の①、②、③が挙げられる：

- ①多段階にわたって、SS理論でのモデル構成作用素T、類似度関数SM、大分類関数BSCで構成される構造受精変換を用い、入力パターンのモデルを或るカテゴリの不動点パターンモデルへ連想形変換する。
- ②画素単位の認識なので、1枚の画像内容を画像の理解以前にセグメンテーションをする必要がない。
- ③画像内の個々の事物についてRECOGNITRONの持っている知識と学習能力とに応じ、認識可能(該当するカテゴリが唯1個あること)・認識不定(該当するカテゴリが複数個あること)・認識不能(該当するカテゴリが1個もないこと)という3結果が得られ、理解できない場合異常な結果を出力することがある。 □

本計算機シミュレーションでは、複数のモデル構成作用素T、複数の類似度関数SMと、ファジィ推論を行うニューラルネットとしての大分類関数BSCが構築されており、これらの組み合わせによって、画像理解システムとしての多種多様のRECOGNITRONの集合体得られる多重構成方式が採用されている。マウスをクリックすることによって、1つの構成方式を選定できるようになっている。計算機シミュレーション結果においては、ほぼ良好な理解性能が確保されている。

キーワード

パターン認識の数学的理論 (SS理論) モデル構成作用素 類似度関数 大分類関数
カテゴリ選択関数 構造受精変換 カテゴリ帰属知識 不動点連想形認識
画素単位の認識 風景画像 知識の抽出 セグメンテーション
ファジィ推論ニューラルネット 最急降下法

Abstract

A result of simulation carried out by using a JAVA language is presented here according to an image-understanding theory [6] proposed by S.Suzuki. A point of difference between S.Suzuki theory and traditional theories is that a meaning of an input image is represented by a model of a recalled image in S.Suzuki theory.

A system presented here which is intended so as to be able to understand a scene-image deals with a model $T\varphi$ corresponding to an input pattern φ as if $T\varphi$ were φ . We call this assumption an identical perceptual principle between $T\varphi$ and φ .

S.Suzuki proposed already a multi-stage transformational recognition method (MSTRM) of searching and recalling a fixed-point. MSTRM transforms a input pattern φ to be processed in question into $T\varphi$ in the first place, and then transforms $T\varphi$ into a sequence of pattern-models for the purpose of searching and recalling a fixed-point model of a prototypical pattern associated with a category. We can try to apply MSTRM to determining each category to which each pixel belongs. As the conclusion of such a trial we have already obtained RECOGNITRON which can understand a substance of an scene image in question [6]. In this paper we construct this RECOGNITRON by use of program language JAVA, and examine a result obtained about a faculty of understanding a given image.

If each category determined by MSTRM is called together concerning each pixel in the image, how much RECOGNITRON understood a substance of the image will become clear. Let RECOGNITRON stick on each pixel a label of a belonging category determined by using the neighbour of the pixel being watched to show what the image is. As this result all pixels in the image have belonging category-labels. By this means RECOGNITRON can understand whether sky, trees, cars, houses, roads and electric-light poles (six categories) exists in the image.

RECOGNITRON does not necessitate to segment the image because RECOGNITRON adopts a method of attaching each category-label to each pixel, and can fully extract knowledges from the image associatively. This fact is characteristic of RECOGNITRON. The system constructed here has been obtained by obediently applying a mathematical theory (SS theory) of recognizing patterns proposed by S.Suzuki to a field of image-understanding. SS theory has already been proven to be universal in recognizing patterns.

Three distinctive features of the method of image-understanding whose performance are made sure are explained as follows :

- ① RECOGNITRON transforms associatively the model of the input pattern to a fixed-point pattern-model of a prototypical pattern of a category through multi-stages by repeatedly using a structural fertilization transformation whose components are a model-construction operator T, a similarity-measure function SM and a rough classifier BSC.
- ② A region segmentation is not needed before understanding an image because of pixel-wise recognition.
- ③ According to how much RECOGNITRON has knowledges and learnability, RECOGNITRON sometimes has two ill feelings against a thing in the input image. RECOGNITRON may output 3 results of recognition for things in the input image : (I) definite recognizability (in the event of obtained categories being only one)
(II) indefinite recognizability (in the event of obtained categories being more than two)
(III) impossible recognizability (in the event of obtained categories being zero) □

RECOGNITRON has the plural model-construction operators Ts, the plural similarity-measure functions SMs, and the single rough classifier BSC which can perform a fuzzy inference as a neural network. We can select only one RECOGNITRON among these combinations by clicking a mouse. The obtained computer simulation came to conclusion that RECOGNITRON serves the purpose of an image-understanding.

Key words : a mathematical theory of recognizing patterns (SS theory)
model-construction operator similarity-measure function rough classifier category-selection
function structure-fertilization transformation categorical-membership knowledge
associative recognition of searching a fixed-point pixelwise recognition scenery image
method of steepest descent

1. はじめに

本論文では、文献 [6] の画像理解理論をほぼ忠実に、JAVA言語でシミュレーションした結果

が報告される。従来の画像を理解する方法と決定的に相違する点は、入力画像内の事物の意味はこの事物から想起されるパターンモデルであるとしていること(内容理解; content-based understanding)である。

一枚の画像内にいかなる事物が存在するかをコンピュータに理解させるのは、難しい。一枚の画像内容を理解するシステムを画像理解システム(Image-Understanding System)というが、画像理解システムの研究は現在にいたっても進展していない。

世界に存在する事物、事象、関係などが“類似性”によって分類されるにつれて、カテゴリ(類概念)が形成される。それが何という物であるかを見分けること、つまり事物の認識に最も基礎となる資料の一つは形状である。脳内で計算された結果に基づいて、知識によって記憶内に形成されたパターンモデルと、外界パターンモデルとのマッチングによって外界を認識すると想定してみよう。このときの認識の働きはカテゴリに関する帰納推論(一を聞いて十を知る働き)である。事物の存在する領域を抽出し(領域抽出; セグメンテーション)、抽出された領域内から事物の形状情報を抽出する。それが何という物であるかを見分けること(事物の認識)に最も役立つ知覚属性の一つは形状(shape characteristics)である。本研究では、灰色形状(gray scale shape)のみによって認識するシステムが構築される。

画像理解に関する他の研究に目を向けよう。

対象世界に関し完全な知識を予め備えることが困難な場合が画像理解では多くて、文献 [5] ではこういった場合に役立つ各種推論が解説されている。

文献 [8] では、Hermite functionsを用いた直交展開係数を特徴抽出量として用い、scene analysisに有用なtwo models of nearest-neighbour classifiersを提案している。

2文献 [9], [10] では、Gauss形状を前件部メンバシップ関数として使ったfuzzy 推論を行うneural networkを、画素単位での、風景画像からの知識抽出に用いている。この2文献の前件部メンバシップ関数からヒントを得たfuzzy推論neural networkは、本研究での大分類関数BSCの構成に使われている。

Scene analysisには、image segmentationと採用する分類器との両検討が必要であるが、文献 [11] では、各種のimage segmentation techniquesが解説されている。

文献 [12] では、gray-level, color, and textured imagesに適用できるgraph 構造を使ったsegmentationのalgorithmをhierarchical approachの立場から研究されている。

文献 [13], [14] では、対象のsegmentationが行われていない動画画像から複数の動作を同時に識別できるselective attentionの、系列の解析を行う非決定性有限オートマトン機構が提案されている。

文献 [15] では、交通監視動画画像から背景、自動車、自動車の影を、輝度とSobel値を観測シンボルとするHidden Markov Modelを用いて高精度に分類できる手法が提案されている。

本研究では、平面上の直交座標系 $\langle x_1, x_2 \rangle$ が付与された2次元画像面を見て、さまざまな外界の事物の知覚的認識を行う人間の視覚認識の機能をJAVA言語により作成したプログラムによって実現している。つまり、経験に基づいて最もありそうなことを想起し、最終的に想起されたパターンモデルが外界の事物のモデルであるとする「想起型認識(associative recognition)」の働き(外界の事物の帰属するカテゴリと、その外界の事物から想起されるパターンモデルとの双方を出力する働き)を基盤として、風景画像を解釈・理解するシステムRECOGNITRONが構築され、その計算機シミュレーションが実施されている。

画像中の任意の画素に視点を合わせ、視野の範囲に存在する形状情報を手がかりに認識を行い、

その結果得られた帰属するカテゴリラベルを各画素に貼り付けていき、1枚の画像を構成する全ての画素について適切なカテゴリラベルを付与することができれば、1枚の画像から、画素単位に画像から知識を抽出することができ、同時に画像を意味のある有限個の領域に分割する領域分割(region segmentation)の機能を実現することができる。

パターン情報処理に関する数理(SS理論) [1]~[4] を適用し、この領域分割の機能を備え、風景画像から事物の抽出を行う「画像理解システム」を構築したが [6], その際必要となる3基本要素(モデル構成作用素T, 類似度関数SM, 大分類関数BSC)の構成を説明し、得られた計算機シミュレーションの結果を本論文では検討する。Tには外界のモデルを形成する働きがあり、SMにはモデル同士の似ている程度を測る働きがあり、最後に、BSCには外界の事物が帰属するカテゴリの候補を複数個抽出する働きがある。

物体の存在する領域を大きさ $(2p+1)(2q+1)$ (画素の個数)の画像成分と仮定し、領域抽出をあらゆるさまに行わない認識方式を計算機シミュレーションで実現した本研究で構築された画像理解システムは、2次元平面に投影された物体の濃度値から得られる灰色形状情報のみを手がかりに認識・解釈を行うが、各カテゴリの代表パターンからの変形がある程度少ない画像であればかなり高い認識率を得られることが今回のシミュレーションにおいて確認されている。

本研究で提案されている処理方式の特徴のひとつに、処理画像が例えば、風景画像に限定されない利点があり、学習に用いる訓練パターン(training pattern)と、あらかじめ用意するカテゴリラベルを変更するだけで、顔画像の理解や、指紋解析、文字認識などに適用することも可能であるという一般性を挙げることができる。

なお、以下で引用している4つのaxiom 1~4については、2文献 [3], [4] で説明されている。

2. 入力画像 φ の2次元配列への変換, 右手系座標系への変換式, 画像間の内積・ノルムの定義

本章では、濃度値の内部表現を確保するため、まず、デジタルカメラ画像から入力画像へ変換する方法を説明する。その次に、入力画像データの1次元配列から2次元配列へ変換する方法が説明され、平面上の右手系座標系への変換式が確立される。最後に、画像間に内積を定義し、それによって画像のノルムを定義する。

2.1 デジタルカメラ画像から入力画像への変換

デジタルカメラで撮影した画像を、処理の対象となる入力画像 φ へと変換するために、以下の変換処理を施した。

2.1.1 変換の手順

デジタルカメラで撮影した画像を入力画像へ変換する手順は、次の通りである：

- ① デジタルカメラで撮影した画像を計算機に読み込む。
- ② 画像をデジタルカメラ固有の保存形式から無圧縮24ビットカラー形式へ変換する。
- ③ 画像サイズの調整(例：101×81)。
- ④ 単色モード(256階調モノクロ画像)への変換。 □

2.1.2 濃度値の単色モードへの変換

カラー画像を256階調単色画像へと変換するため、輝度値

$$Y = 0.299R + 0.587G + 0.114B$$

(Rは赤, Gは緑, Bは青成分)

(2.1)

を計算する。

ただし, 本研究で作成した画像理解システムでは, 単色画像であっても24ビットカラー情報として入力する必要があるため, 求めた輝度値を

$$\text{GrayImage} = 0\text{xffrrggbb} \quad (\text{rr} = \text{gg} = \text{bb} = Y)$$

(2.2)

の32ビット16進数の形に置き換え保存する。

2.2 濃度値の内部表現

2.1.2の処理をほどこし処理対象として入力された画像について, 32ビットの論理積

$$\text{LowEightBit} = 0\text{x000000ff} \ \& \ \text{GrayImage}$$

(2.3)

を計算して下位8ビットのみを有効にすることによって, 0(黒)~255(白)の256段階灰色値を抽出し, 次に, 0(白)~255(黒)へと変換するため, 計算

$$\text{Inverse} = 255 - \text{LowEightBit}$$

(2.4)

を行い白黒の度合いを反転し, 32ビット16進数の形に置き換え保存する。最後に

$$\text{phi} = \text{Inverse} - 128$$

(2.5)

の計算を行い, 処理対象画像は最終的に-128(白)~+127(黒)の値をとる濃度値の集合として表現される。

2.3 採用する右手系座標系

採用した座標系は図2.1, 図2.2での右手系整数座標系 $\langle x_1, x_2 \rangle$ である。

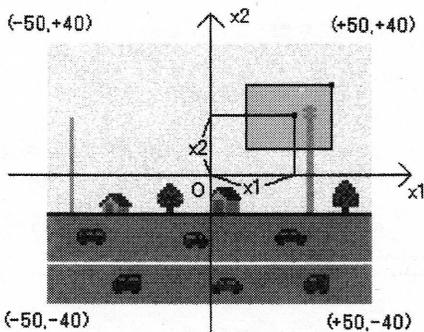


図2.1 採用した右手系座標系 $\langle x_1, x_2 \rangle$

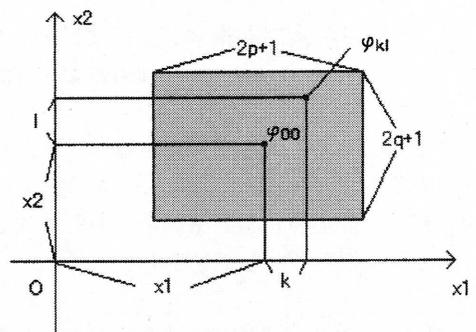


図2.2 $\varphi(x_1, x_2)$ の第 (k, l) 成分 φ_{kl}

面積が 100×80 であるような矩形領域を考え, x_1 軸, x_2 軸の値が

$$x_1 = 0, \pm 1, \dots, \pm 50, x_2 = 0, \pm 1, \dots, \pm 40$$

であるような整数の値をとる右手系座標系 $x = \langle x_1, x_2 \rangle$ を設ける. 処理対象とする一枚の画像 φ は2の8乗(=256)階調の灰色画像であり,

$$\varphi(x) = \varphi(x_1, x_2) \in \{0, \pm 1, \pm 2, \dots, \pm 127, -128\}$$

(2.6)

と表わされる. 以後 $x = \langle x_1, x_2 \rangle$ を固定し,

$$\varphi_{kl} \equiv \varphi(x_1 + k, x_2 + l)$$

(2.7)

において k, l を動かして得られるパターン

$$\varphi = \{\varphi_{kl} | k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (2.8)$$

を認識処理するシステム RECOGNITRON $\langle x_1, x_2 \rangle$ を構成する。このような認識システムの集合体 $\{ \text{RECOGNITRON} \langle x_1, x_2 \rangle | x_1=0, \pm 1, \dots, \pm 50, x_2=0, \pm 1, \dots, \pm 40 \}$ が画像理解システムである。

2.4 入力画像データの1次元配列から2次元配列への変換と、右手系座標系への変換式

画像理解システムは x_1, x_2 などをキーとし、2座標変換式(2.11), (2.12)を使って対応する2次元配列上の位置を計算することで、右手系座標系を仮想的に作り出している。

画像理解システムは、入力された画像 `msBitmapData` をまず1次元配列として読み込む。そして次の計算を行い2次元配列 `inputImageMap` に変換する：

$$\begin{aligned} \text{inputImageMap} [a_1][a_2] &= \text{msBitmapData} [\text{width} \times a_1 + a_2] \\ a_1 &= 0, 1, \dots, \text{height} - 1, a_2 = 0, 1, \dots, \text{width} - 1 \\ \text{width} &: \text{画像の横幅} \quad \text{height} : \text{画像の縦幅} \\ \text{ここに, width} &= 101, \text{height} = 81 \end{aligned} \quad (2.9)$$

□

変換された2次元画像データ `inputImageMap` を右手系座標系として扱う次の諸式を使って $\langle x_1, x_2 \rangle$ に対応する配列座標 $[a_1][a_2]$ を計算し、処理対象となる2次元配列 phi (2次元画像関数) を作成する(図2.3を参照)：

$$\text{phi} [b_1(l)][b_2(k)] = \text{inputImageMap} [a_1(x_2+1)][a_2(x_1+k)] \quad (2.10)$$

$$k = 0, \pm 1, \pm 2, \dots, \pm p, l = 0, \pm 1, \pm 2, \dots, \pm q \quad (2.10)$$

$$a_1(x_2) = x_2_{\text{max}} - x_2, \text{ここに, } x_2 > x_2_{\text{max}} \text{ のとき } x_2 = x_2_{\text{max}} \text{ として計算} \quad (2.11)$$

$$a_2(x_1) = x_1_{\text{max}} + x_1 \text{ここに, } x_1 > x_1_{\text{max}} \text{ のとき } x_1 = x_1_{\text{max}} \text{ として計算} \quad (2.12)$$

$$x_1_{\text{max}} = (\text{width} - 1) / 2 \quad (2.13)$$

$$x_2_{\text{max}} = (\text{height} - 1) / 2 \quad (2.14)$$

$$b_1(l) = q - l \quad (2.15)$$

$$b_2(k) = p + k \quad (2.16)$$

□

本シミュレーションでは式(2.13)の x_1_{max} , 式(2.14)の x_2_{max} は各々, $x_1_{\text{max}} = 50, x_2_{\text{max}} = 40$ となっている。

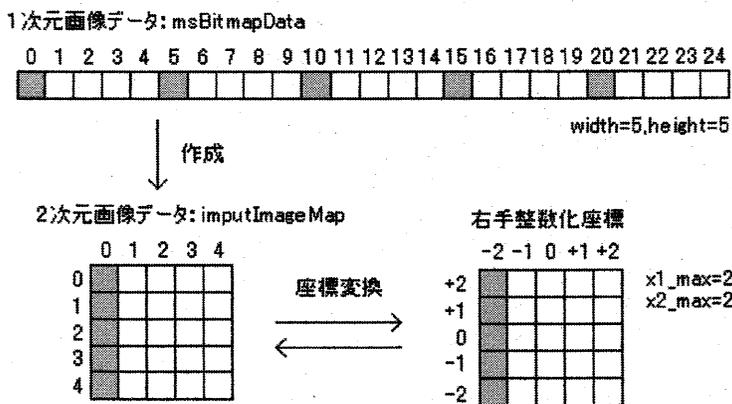


図2.3 データ形式変換の概念図

本シミュレーションで採用した画像の大きさは101×81であるが、本画像理解システムは基本的には、横幅×縦幅が1000000画素以下の画像であれば処理可能であるように、JAVA言語でプログラム化されている。しかし、原点を含み原点から上下左右対称の座標系を形成することを前提に設計されているため、画像の横幅、縦幅の大きさが共に奇数であることが必要である。

2.5 内積とノルムの定義

パターンの性質やパターン同士の関係を数値的に扱うためには、内積やノルムを導入しなければならない。

本認識システムRECOGNITRONが注目している直交2次元座標値 x_1, x_2 は固定されている状況下で、パターン φ と η の内積 (φ, η) は、

$$(\varphi, \eta) = \sum_{k=-p}^{+p} \sum_{l=-q}^{+q} \varphi_{kl} \cdot \eta_{kl} \quad (2.17)$$

と定義され、同様に、ノルム $\|\varphi\|$ は、

$$\begin{aligned} \|\varphi\| &= \sqrt{(\varphi, \varphi)} \\ &= \sqrt{\sum_{k=-p}^{+p} \sum_{l=-q}^{+q} (\varphi_{kl})^2} \end{aligned} \quad (2.18)$$

と定義される。この内積 (φ, η) 、ノルム $\|\varphi\|$ は、直交2次元座標 $\langle x_1, x_2 \rangle$ を持つ画素がいかなるカテゴリを表すパターンの成分であるかをRECOGNITRONが決定するとき用いられるものである。

3. 採用した3種類のモデル構成作用素T

本研究で構築した画像理解システムでは、2文献 [3], [4] でのaxiom 1を満たす3種類のモデル構成作用素Tを内蔵している。本章で説明されるパターン φ のモデル $T\varphi$ は $(2p+1) \cdot (2q+1)$ 個の振幅の集まりで、 x_1, x_2 を固定した条件の下で、

$$T\varphi \equiv \{(T\varphi)_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\}$$

と表示される。

パターン φ の意味は φ から多段階にわたり想起される或るカテゴリ \mathcal{C}_j の代表パターン $\omega[j]$ のモデル $T\omega[j]$ であるとしている。それ故、いかなるモデル構成作用素Tを採用するかは、認識システムRECOGNITRONの性能に直接、影響する。

3.1 振幅が3値のモデル構成作用素T

各 $(T\varphi)_{kl}$ については

$$(T\varphi)_{kl} \equiv \begin{cases} 1 \cdots h^+_{kl} < (S\varphi)_{kl} \leq 1 & \text{のとき} \\ 0 \cdots h^-_{kl} \leq (S\varphi)_{kl} \leq h^+_{kl} & \text{のとき} \\ -1 \cdots -1 \leq (S\varphi)_{kl} \leq h^-_{kl} & \text{のとき} \end{cases}$$

$$k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q \quad (3.1)$$

と定義され、閾値 h^+, h^- は、

$$h^+ \equiv \{h^+_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (3.2)$$

$$h^- \equiv \{h^-_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (3.3)$$

であり、 h^-_{kl}, h^+_{kl} は不等式

$$-1 < h_{kl}^- \leq 0 \leq h_{kl}^+ < +1 \quad (3.4)$$

を満たすものである。

本研究では、

$$h_{kl}^+ = 0.5 \quad h_{kl}^- = -0.5$$

と与えている。(Sφ)_{kl}については3.3の式(3.6)で説明されている。

3.2 振幅が5値のモデル構成作用素 T

各(Tφ)_{kl}については

$$(T\varphi)_{kl} \equiv$$

$$\begin{cases} 1 & \cdots & 2/3 < (S\varphi)_{kl} \leq 1 \\ 0.5 & \cdots & 1/3 < (S\varphi)_{kl} \leq 2/3 \\ 0 & \cdots & -1/3 \leq (S\varphi)_{kl} \leq 1/3 \\ -0.5 & \cdots & -2/3 \leq (S\varphi)_{kl} < -1/3 \\ -1 & \cdots & -1 \leq (S\varphi)_{kl} < -2/3 \end{cases}$$

$$k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q \quad (3.5)$$

と求める。(Sφ)_{kl}については3.3の式(3.6)で定義されている。

3.3 振幅の最大値によるパターン規格化作用素 S を使ったモデル構成作用素 T

各(Sφ)_{kl}については

$$(S\varphi)_{kl} \equiv$$

$$\begin{cases} 0 \cdots \max_{r,s} |\varphi_{rs}| = 0 \text{ の場合} \\ |\varphi_{kl}| / \max_{r,s} |\varphi_{rs}| \cdots \max_{r,s} |\varphi_{rs}| > 0 \text{ の場合} \end{cases}$$

$$k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q$$

$$r=0, \pm 1, \pm 2, \dots, \pm p, s=0, \pm 1, \pm 2, \dots, \pm q \quad (3.6)$$

と選び、Tφ_{kl}を

$$T\varphi_{kl} = S\varphi_{kl}$$

と定義する。写像Sは振幅の最大値max_{r,s} | φ_{rs} | によってパターンφを規格化するだけの機能を持つパターン変換作用素である。

3.4 パターンモデル Tφ を画像として表示する処理

本研究で構築された画像理解システムRECOGNITRONは式(2.8)の原パターンφに対応し、そのパターンモデルTφをシステム内部の表現として確保する機能を備えている。本システムは、パターンモデルTφを見るとあたかも原パターンφであるかのように錯覚する(パターンモデルTφと原パターンφ間の同一知覚原理)画像として表示する際の濃度値は、Tφの値にそれぞれ127を掛け整数化したものであり、-127~+127の濃度値を持つ画像として出力される。その1例が2図3.1, 3.2に示されている。

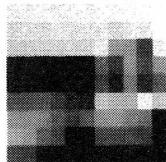


図3.1 パターンφ

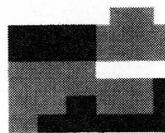


図3.2 φのパターンモデルTφ (3値)

4. 採用した5種類の類似度関数 SM

本章では、画素 $\langle x_1, x_2 \rangle$ を認識する RECOGNITRON $\langle x_1, x_2 \rangle$ が内蔵している、積集合 $\Phi \times \Omega$ から単位区間集合 $\{s \mid 0 \leq s \leq 1\}$ への、2文献 [3], [4] での axiom 2 を満たすの5種類の写像 SM (similarity-measure function)

$$SM: \Phi \times \Omega \rightarrow \{s \mid 0 \leq s \leq 1\} \quad (4.1)$$

が説明される。ここに、

$$\Omega = \{\omega [j] \mid j \in J\} \quad (4.2)$$

は、第 $j \in J$ 番目のカテゴリの代表パターン $\omega [j]$ の集まりであり、また、カテゴリの総数を m として、

$$J = \{1, 2, \dots, m\} \quad (4.3)$$

はカテゴリ番号の全集合である。

4.1 相互情報量 MI を使った SM

全カテゴリ集合内の第 $j \in J$ 番目のカテゴリ \mathcal{C}_j の持つ諸性質を典型的に代表する代表パターンを $\omega [j] = \{\omega [j]_{kl}\}$ とし、2つのパターンモデル $T\varphi, T\omega [j]$ について、 $T\varphi$ 内に $T\omega [j]$ が含まれる程度を情報量として計量する相互情報量 (mutual information) を

$$MI(T\varphi, T\omega [j]) \equiv - (1/2) \cdot \log_e [1 - |nip(T\varphi, T\omega [j])|^2] \quad (4.4)$$

と定義する。式(4.4)に登場している $nip(T\varphi, T\omega [j])$ は $T\varphi$ と $T\omega [j]$ との間の規格化内積 (normalized inner product) といわれるものであり、

$$nip(T\varphi, T\omega [j]) \equiv \begin{cases} 0 & \dots \quad \|T\varphi\| \cdot \|T\omega [j]\| = 0 \text{ の場合} \\ (T\varphi, T\omega [j]) / \|T\varphi\| \cdot \|T\omega [j]\| & \dots \quad \|T\varphi\| \cdot \|T\omega [j]\| > 0 \text{ の場合} \end{cases} \quad (4.5)$$

と定義される。そして、この相互情報量を用いて構成する SM が

$$SM(T\varphi, T\omega [j]) \equiv \begin{cases} \textcircled{1} \sum_{j \in J} MI(T\varphi, T\omega [j]) = 0 \text{ ならば} \\ \quad P(\mathcal{C}_j) \\ \textcircled{2} \sum_{j \in J} MI(T\varphi, T\omega [j]) > 0 \text{ ならば} \\ \quad MI(T\varphi, T\omega [j]) / \sum_{i \in J} MI(T\varphi, T\omega [i]) \end{cases} \quad (4.6)$$

と定義される。なお、 $P(\mathcal{C}_j)$ は \mathcal{C}_j の生起確率であり、等確率に

$$P(\mathcal{C}_j) = 1/m \quad (4.7)$$

と与えている。

4.2 各カテゴリに帰属する複数個のパターンを使った SM ①

ここでは、第 $j \in J$ 番目のカテゴリ \mathcal{C}_j に帰属するパターン $\psi [j; q]$ の集合

$$\Psi_j = \{\psi [j; 1], \psi [j; 2], \dots, \psi [j; r]\} \quad (4.8)$$

を導入し、

$$\psi [j; q] = \{\psi [j; q]_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (4.9)$$

ここに

$$\psi [j; q]_{kl} \equiv \psi [j; q] (x_1+k, x_2+l) \quad (4.10)$$

(x_1, x_2 は固定されていることに注意)

とする。次に、 $i \neq j$ であれば、

$$\|T\omega[i] - T\omega[j]\| > 0 \quad (4.11)$$

と選ばれており、 $T\psi[j]$ と $T\psi[i]$ とが共通の要素を持たないように各 $\psi[j; q]$ を選定する。

カテゴリ番号 $j \in J$ を助変数とする関数 $g_i(\varphi)$ を

$$g_i(\varphi) = \min \{ \|T\varphi - T\psi\| \mid \psi \in \psi[j; q], q=1 \sim r \} \quad (4.12)$$

と定義する。ここで、

$$\text{ndis}(T\varphi, T\psi) = \|T\varphi - T\psi\| \quad (4.13)$$

と定義される $\text{ndis}(T\varphi, T\psi)$ はノルム距離と呼ばれ、パターン $T\varphi$ と $T\psi$ 間の距離(相違の程度)を測定する1つの方法である。

次に、関数 $f_j(\varphi)$ を

$$f_j(\varphi) = \min_{i \in J - \{j\}} g_i(\varphi) \quad (4.14)$$

と定義する。その後、 $SM(\varphi, \omega[j])$ を

$$SM(\varphi, \omega[j]) \equiv \begin{cases} f_j(\varphi) / \sum_{i \in J} f_i(\varphi) & \cdots \quad \sum_{i \in J} f_i(\varphi) > 0 \text{ のとき} \\ P(\mathcal{Q}_j) & \cdots \quad \sum_{i \in J} f_i(\varphi) > 0 \text{ のとき} \end{cases} \quad (4.15)$$

と定義する。

このとき、すべての $j \in J$ について

$$SM(\psi[j; q], \omega[j]) = 1 \quad (4.16)$$

がすべての $q (= 1, 2, \dots, r)$ につき成り立つ。また、任意に固定した $j \in J$ につき

$$SM(\psi[i; q], \omega[j]) = 0 \quad (i \neq j) \quad (4.17)$$

がすべての q , すべての $i (\neq j)$ につき成り立っている。

4.3 量子化SM

本節では、類似度 $SM(\varphi, \omega[j])$ の値を2つの閾値 $s_0(j)$, $s_1(j)$ を使い、 $s_0(j)$ 以下は0に、また、 $s_1(j)$ 以上は1に量子化する方法を述べる。Axiom 2を満たす類似度関数 SM から得られた式(4.23)の量子化類似度関数 SM^* は axiom 2を満たす。この量子化により、パターンの変形を吸収し、多段階認識過程の収束を速めることができる。

4.3.1 SMを量子化する方法

本項では、閾値 $s_0(j)$, $s_1(j)$ を導入し、 $SM(\varphi, \omega[j])$ の値を量子化する方法(量子化方法)を述べる。第 $j \in J$ 番目のカテゴリに帰属しているパターン

$$\varphi[j; t] = \{\varphi[j; t]_k \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (4.18)$$

の集合

$$\Phi_j = \{\varphi[j; t] \mid t=1 \sim N_j\} \quad (4.19)$$

を導入し、閾値 $s_0(j) (< s_1(j))$, $s_1(j)$ を不等式

$$s_1(j) \leq \min_{\psi \in \Psi_k} SM(\psi, \omega[j]) \quad (4.20)$$

$$\max_{k \in J - \{j\}} \max_{\psi \in \Psi_k} SM(\psi, \omega[j]) \leq s_0(j) \quad (4.21)$$

を満たすように決める。その後、 $SM(\varphi, \omega[j])$ を

$$s(\varphi, j) \equiv \begin{cases} 1 & \cdots \quad s_1(j) \leq SM(\varphi, \omega[j]) \leq 1 \text{ のとき} \\ [SM(\varphi, \omega[j]) - s_0(j)] / [s_1(j) - s_0(j)] & \cdots \quad s_0(j) < SM(\varphi, \omega[j]) < s_1(j) \text{ のとき} \\ 0 & \cdots \quad 0 \leq SM(\varphi, \omega[j]) \leq s_0(j) \text{ のとき} \end{cases} \quad (4.22)$$

へと変換し,

$$SM'(\varphi, \omega[j]) \equiv \begin{cases} s(\varphi, j) / \sum_{k \in J} s(\varphi, k) \cdots \sum_{k \in J} s(\varphi, k) > 0 \text{ のとき} \\ P(\mathcal{C}_j) \cdots \sum_{k \in J} s(\varphi, k) = 0 \text{ のとき} \end{cases} \quad (4.23)$$

を求める. この $SM'(\varphi, \omega[j])$ の値が $SM(\varphi, \omega[j])$ を量子化した値である.

なお, $s_0(j) < s_1(j)$ が成り立つように選定されていなければならないことに注意する.

4.3.2 閾値 $s_0(j)$, $s_1(j)$ の選び方

前節4.2の第 $j \in J$ 番目のカテゴリに帰属するパターンの, 式(4.8)の系列 Ψ_j を4.3.1の Φ_j として採用する場合等を考慮し, 閾値 $s_0(j)$, $s_1(j)$ の与え方を以下のように決める.

① $a(j) < b(j)$

$$a(j) = \sum_{k \in J - |j|} \max_{\varphi \in \Phi_k} SM(\varphi, \omega[j]) \quad (4.24)$$

$$b(j) = \min_{\varphi \in \Phi_j} SM(\varphi, \omega[j]) \quad (4.25)$$

が成り立っていることを確認する.

② 上述の①の不等式がすべての $j \in J$ について成り立っていないならば, 量子化しないSMで行う.

③ 上述の①の不等式がすべての $j \in J$ について成り立っていて, しかも

$$b(j) < 1 \text{ かつ } 0 < a(j)$$

であれば,

$$s_0(j) = a(j) \cdot x_j \quad (4.26)$$

$$s_1(j) = b(j) \cdot y_j \quad (4.27)$$

とする. ここに

$$x_j = |J| / (|J| - 1) \quad (4.28)$$

$$y_j = \min \{ (|J| - 1) / |J|, 1/b(j) \} \quad (4.29)$$

$$|J| = m \text{ (カテゴリ総数)} \quad (4.30)$$

④ $b(j) = 1$ の場合

$$s_0(j) = a(j) \cdot x_j \quad (4.31)$$

$$s_1(j) = (|J| - 1) / |J| \quad (4.32)$$

⑤ $(1/|J|) \cdot (1/y_j) < b(j)$ かつ $a(j) = 0$ の場合

$$s_0(j) = 1 / |J| \quad (4.33)$$

$$s_1(j) = b(j) \cdot y_j \quad (4.34)$$

⑥ その他のとき

$$s_0(j) = a(j) \quad (4.35)$$

$$s_1(j) = b(j) \quad (4.36)$$

と選ぶ.

4.4 各カテゴリに帰属する複数個のパターンを使ったSM②

本節4.4から4.6は, 4.2と同じく各カテゴリに帰属する複数のパターンを使ってSMを求める式(4.15)による方法である. 従って, 4.2の, 式(4.12)の関数 $g_j(\varphi)$ の与え方を変えるだけで, 式(4.8)の Ψ_j や式(4.14)の関数 $f_j(\varphi)$ などはまったく同じであるため, $g_j(\varphi)$ の選び方のみ示す.

関数 $g_j(\varphi)$ は、

$$g_j(\varphi) = \begin{cases} \min_{\phi \in \Psi_j} -2^{-1} \cdot \log_e | \text{nip}(T\varphi, T\phi) |^2 \\ \quad \dots \text{すべての } \phi \in \Psi_j \text{ について} \\ \quad \quad \quad \| T\varphi \| \cdot \| T\phi \| > 0 \text{ かつ } | \text{nip}(T\varphi, T\phi) |^2 > \epsilon > 0 \text{ のとき} \\ -2^{-1} \cdot \log_e \epsilon \\ \quad \dots \text{ある1つの } \phi \in \Psi_j \text{ について} \\ \quad \quad \quad \| T\varphi \| \cdot \| T\phi \| = 0 \text{ かつ } | \text{nip}(T\varphi, T\phi) |^2 \leq \epsilon \text{ のとき} \end{cases} \quad (4.37)$$

と、定義される。定義式(4.37)に登場している閾値 ϵ は

$$0 < \epsilon < \min_{i \in J} \min_{j \in J} \min_{\phi \in \Psi[i], \psi \in \Psi[j], i \neq j} | \text{nip}(T\phi, T\psi) |^2 \quad (4.38)$$

$$\therefore \epsilon = (1/|J|) \min_{i \in J} \min_{j \in J} \min_{\phi \in \Psi[j], \psi \in \Psi[j], i \neq j} | \text{nip}(T\phi, T\psi) |^2 \quad (4.39)$$

と与える。

4.5 各カテゴリに帰属する複数個のパターンを使ったSM③

前節4.4と同じく、本節では、 $g_j(\varphi)$ を

$$g_j(\varphi) = \min_{\phi \in \Psi[j]} \{ 1 - | \text{nip}(T\varphi, T\phi) |^2 \} \quad (4.40)$$

と与える。ここで、

$$\text{idis}(T\varphi, T\phi) = 1 - | \text{nip}(T\varphi, T\phi) |^2 \quad (4.41)$$

と定義される $\text{idis}(T\varphi, T\phi)$ は内積相違度と呼ばれ、パターン $T\varphi$ と $T\phi$ 間の違いを計測する1つの方法である。

4.6 各カテゴリに帰属する複数個のパターンを使ったSM④

本節では、各カテゴリについて $a_j(j \in J)$ を、

$$a_j = 3^{-1} \cdot \min_{\phi \in \Psi[i], \psi \in \Psi[j], \Psi[i] \neq \Psi[j]} \{ \| T\psi - T\phi \|^2 \} \quad (4.42)$$

と求めた後、前節4.5と同じく $g_j(\varphi)$ を

$$g_j(\varphi) = \min_{\phi \in \Psi[j]} \{ 1 - \exp[-a_j^{-1} \cdot \| T\varphi - T\phi \|^2] \} \quad (4.43)$$

と選ぶ。

5. 各カテゴリの代表パターンの学習による決定法

本章では、第 $j \in J$ 番目の代表パターン $\omega[j]$ を式(5.7)の訓練パターンの集合に適応させる形式で決定する手法(学習による決定法；文献 [3] での付録Iでの学習アルゴリズム)が説明される。

5.1 各代表パターン $\omega[j]$ の学習アルゴリズム

本研究で構築される認識システムRECOGNITRONでは、入力されるパターンの意味はこのパターンから想起されるパターンである。そして、正しく認識される入力パターンについては、この入力パターンが帰属するカテゴリの代表パターンのモデルである。

m をカテゴリ総数とし、カテゴリ番号の集合 J を $J = \{1, 2, \dots, m\}$ とする。代表パターン ω

[j] は、RECOGNITRONが第j番目のカテゴリに帰属するパターンを正しく認識する場合に想起されるモデル $T\omega[j]$ に対応するパターンである。ここでは、第j番目のカテゴリの代表パターン $\omega[j] = \{\omega[j]_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\}$ (5.1) を学習によって求める方法を述べる。

なお、任意の訓練パターンを各 $\omega[j]$ に視察で選ぶことも可能である。

[代表パターン $\omega[j]$ の学習アルゴリズム]

① 初期化(initialization)

第t学習段階において求められた ω を $\omega[j; t]$ と表し、

$$\omega[j; t]_{kl} \mid t=0 = \omega'[j]_{kl} \quad (5.2)$$

とする。 $\omega'[j]$ は第j番目のカテゴリに帰属する訓練パターン(training pattern)のうち、最もそのカテゴリの特徴をよくあらわしているパターンである。

② 帰納段階(recursion)

$$\omega[j; t+1]_{kl} = \omega[j; t]_{kl} + \Delta\omega[j; t]_{kl} \quad (5.3)$$

($k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q$)

とする。 $\Delta\omega[j; t]_{kl}$ については5.2で述べる。

③ 終了(termination)

$$\|\omega[j; t+1] - \omega[j; t]\|^2 < \delta_1 \quad (5.4)$$

$$\delta_1 = (2p+1)(2q+1) + 1 \quad (5.5)$$

が満たされたとき、 $\omega[j]$ として、 $\omega[j; t]$ を採用する。

5.2 $\Delta\omega[j; t]_{kl}$ の求め方

式(5.3)の更新分 $\Delta\omega[j; t]_{kl}$ を

$$\Delta\omega[j; t]_{kl} = \alpha[j; t]_{kl} \cdot [\varphi[j; t]_{kl} - \omega[j; t]_{kl}] \quad (5.6)$$

とおく。

ここで、 $\varphi[j; t]$ はあらかじめ用意される訓練パターンの系列(sequence)

$$\varphi[j; 0], \varphi[j; 1], \dots, \varphi[j; t] \quad (5.7)$$

であり、

$$\varphi[j; t] = \{\varphi[j; t]_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (5.8)$$

と定義されるもので、 $\alpha[j; t]$ は

$$\alpha[j; t] = \{\alpha[j; t]_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q\} \quad (5.9)$$

は不等式

$$0 \leq \alpha[j; t] < 1 \quad (5.10)$$

を満たすように適切に与えられるものであり、

$$\alpha[j; t]_{kl} = C[j; t]_{kl} \cdot \exp[-\{(k^* - k)^2 + (l^* - l)^2\} / (V[j; t]_{kl})^2] \quad (5.11)$$

(便宜的に $C[j; t]_{kl} = 0.9$ とおく)

であり、 k^* , l^* は、

$$\max_{kl} |\varphi[j; t]_{kl} - \omega[j; t]_{kl}|^2 = |\varphi[j; t]_{k^*l^*} - \omega[j; t]_{k^*l^*}|^2 \quad (5.12)$$

を満たすものである。 $V[j; t]$ は、

$$V[j; t] = \frac{D_{kl}}{\sqrt{(2p)^2 + (2q)^2 \cdot \log_e[2+t]}} \quad (5.13)$$

と選ぶ。 D_{kl} はあらかじめ適切に与えられるものであり、本シミュレーションでは

$$D_{kl} = 1$$

とおいている。

5.3 終了条件の補正

各代表パターン ω の学習の終了条件は式(5.4)であるが、不等式(5.4)を満たす t が見つからず、循環・振動することもしばしば起こる。そこで、訓練時刻 t が100を超えたら以下の条件を終了条件に加える。

$$d = \|\omega[j; t+1] - \omega[j; t]\|^2$$

とおき、カテゴリ $j \in J$ に帰属する訓練パターンの総数を r として、 t が1つ進む度に、これまでの d の中での最小値 $mini_d$ との比較を行い、

$$d < mini_d$$

が成り立つとき、 $mini_d = d$ とし、 $mini_d$ と同じ値の d を $r+1$ 回繰り返したら終了させる。

もし、 $mini_d$ を最後に更新したときの t から訓練パターンを2順しても最小値を繰り返さないときは、訓練パターンを2順した時点での d を $mini_d$ と選び、最小値を再び求め直す。

6. 採用したファジィ推論ニュートラネットとしての大分類関数BSC

本研究で構成され、文献 [3], [4] での axiom 3 を満たす大分類関数BSC (binary state classifier) は、画像理解システム RECOGNITRON が持っている知識が処理の対象としている画像内に存在するかどうかを判定するための知識抽出ネットワーク (knowledge-extraction network) として働き、ファジィ推論規則を表現するニューラルネットとして、以下のごとく構成される。

6.1 BSCの構造形式

2値関数BSCは $\Phi \times J$ から集合 $\{0, 1\}$ への写像

$$BSC : \Phi \times J \rightarrow \{0, 1\} \quad (6.1)$$

であり、次のように定義される：

$$BSC(\varphi, j) \equiv psn(y(\varphi, j)). \quad (6.2)$$

ここに、1実数値 u の2値関数 psn (positive sign function) は

$$psn(u) = \begin{cases} 1 & \dots u \geq 0 \text{ の場合} \\ 0 & \dots u < 0 \text{ の場合} \end{cases} \quad (6.3)$$

と定義されており、現実出力 $y(\varphi, j)$ は

$$y(\varphi, j) \equiv Wmean(\varphi, j) - b[j] \quad (6.4)$$

と定義され、

$$Wmean(\varphi, j) \equiv \sum_{k=-p}^{+p} \sum_{l=-q}^{+q} W[j]_{kl} \cdot p_{kl}(\varphi, j) \quad (6.5)$$

$$p_{kl}(\varphi, j) \equiv q_{kl}(\varphi, j) / \sum_{i \in J} q_{kl}(\varphi, i) \quad (6.6)$$

$$q_{kl}(\varphi, j) \equiv \exp[-\{(\mathcal{T}\varphi)_{kl} - (\mathcal{T}\varphi[j]_{kl})\}^2 / (\sigma[j]_{kl})^2] \quad (6.7)$$

とする。

$y^{\wedge}(j)$ をカテゴリに関する理想出力 (desired output) とすると, 理想的な fuzzy 推論規則

if $\{T\varphi \text{ is } T\omega [j]\}$ then $\{y^{\wedge}(j)_{kl} \text{ is } W[j]_{kl} - b[j]\}$

を表すように, 式(6.2)の大分類関数 $BSC(\varphi, j)$ の構造が設定されている.

上述の neuro-fuzzy system としての構造形式を備えている BSC 内の助変数 $W[j]_{kl}$, $b[j]$, $\sigma[j]_{kl}$ を次節 6.2 の最急降下学習によって求める.

なお, 本 RECOGNITRON では注目しているどの画素 $\langle x_1, x_2 \rangle$ についても共通に学習しており, どの画素についても共通の学習結果を使って認識を行っている.

6.2 BSC 内の $W[j]_{kl}$, $b[j]$, $\sigma[j]_{kl}$ の最急降下法による学習

本節では, 6式(6.2)~(6.7)で定義される BSC 内の $W[j]_{kl}$, $b[j]$, $\sigma[j]_{kl}$ を最急降下法 (method of steepest descent) によって学習する方法を述べるが, そのため, あらかじめ訓練パターンの系列

$$\varphi[j_i, t] = \{\varphi[j_i, t]_{kl} \mid k=0, \pm 1, \pm 2, \dots, p, l=0, \pm 1, \pm 2, \dots, q\} \quad (6.8)$$

$$t=0, 1, 2, \dots$$

を用意する. j_i は時刻 t に入力される訓練パターンが帰属するカテゴリ番号である.

6.2.1 重み $W[j]_{kl}$ の最急降下法による学習

あらかじめ

$$\textcircled{1} \quad \delta_2 > 0 \quad (6.9)$$

$$\textcircled{2} \quad \varepsilon_2[j_i, t]_{kl} > 0 \quad \text{かつ} \quad \varepsilon_2[j_i, t]_{kl} \geq \varepsilon_2[j_{t+1}, t+1]_{kl} \quad (6.10)$$

$$\textcircled{3} \quad W[j_i, t]_{kl} \mid t=0 \equiv 0.1 \quad (6.11)$$

を用意する. そして $W[j_i, t]_{kl}$ を

$$W[j_{t+1}, t+1]_{kl} = W[j_i, t]_{kl} + \Delta W[j_i, t]_{kl} \quad (6.12)$$

と, 時刻 $t+1$ の $W[j_{t+1}, t+1]_{kl}$ へと修正していき, 不等式

$$\sum_{k=-p}^{+p} \sum_{l=-q}^{+q} |W[j_{t+1}, t+1]_{kl} - W[j_i, t]_{kl}|^2 < \delta_2 \quad (6.13)$$

が成立した時点 t で終了する.

このとき, $W[j]_{kl}$ として,

$$W[j]_{kl} = W[j_i, t]_{kl} \quad (6.14)$$

と, 終了した時刻 t における $W[j_i, t]_{kl}$ を採用する.

$\Delta W[j_i, t]_{kl}$ の求め方は, 最急降下法によれば, 次のとおりである.

$$\Delta W[j_i, t]_{kl} = -\varepsilon_2[j_i, t]_{kl} \cdot [y(\varphi[j_i, t], j_i) - y^{\wedge}(j_i)] \cdot p_{kl}(\varphi[j_i, t], j_i) \quad (6.15)$$

ここに, $y^{\wedge}(j_i)$ はカテゴリ t に関する理想出力である.

現実出力 $y(\varphi[j_i, t], j_i)$ の, 理想出力 $y^{\wedge}(j_i)$ からの違い $[y(\varphi[j_i, t], j_i) - y^{\wedge}(j_i)]$ の自乗の 1/2 を最小とするように, 更新式(6.12)を設けていることに, 注意しておく.

6.2.2 閾値 $b[j]$ の最急降下法による学習

あらかじめ

$$\textcircled{1} \quad \delta_3 > 0 \quad (6.16)$$

$$\textcircled{2} \quad \varepsilon_3[j_i, t] > 0 \quad \text{かつ} \quad \varepsilon_3[j_i, t] \geq \varepsilon_3[j_{t+1}, t+1] \quad (6.17)$$

$$\textcircled{3} \quad b[j_i, t] \mid t=0 \equiv 0 \quad (6.18)$$

を用意する. そして $b[j_i, t]$ を

$$b[j_{t+1}, t+1] = b[j_i, t] + \Delta b[j_i, t] \quad (6.19)$$

と, 時刻 $t+1$ の $b[j_{t+1}, t+1]$ へと修正していき, 不等式

$$|b[j_{t+1}, t+1] - b[j_t, t]| < \delta_3 \quad (6.20)$$

が成立した時点 t で終了する。

このとき、 $b[j]$ として、

$$b[j] = b[j_t, t] \quad (6.21)$$

と、終了した時刻 t における $b[j_t, t]$ を採用する。

ここで求められた $b[j]$ は仮の値であり、最終的に6.3の方法で選びなおす。

$\Delta b[j_t, t]$ の求め方は次のとおりである。

$$\Delta b[j_t, t] = \epsilon_3[j_t, t] \cdot [y(\varphi[j_t, t], j_t) - y^{\wedge}(j_t)] \quad (6.22)$$

現実出力 $y(\varphi[j_t, t], j_t)$ の、理想出力 $y^{\wedge}(j_t)$ からの違い $[y(\varphi[j_t, t], j_t) - y^{\wedge}(j_t)]$ の自乗の1/2を最小とするように、更新式(6.19)を設けていることに、注意しておく。

6.2.3 標準偏差 $\sigma[j]_{kl}$ の最急降下法による学習

あらかじめ

$$\textcircled{1} \delta_4 > 0 \quad (6.23)$$

$$\textcircled{2} \epsilon_4[j_t, t]_{kl} > 0 \quad \text{かつ} \quad \epsilon_4[j_t, t]_{kl} \geq \epsilon_4[j_{t+1}, t+1]_{kl} \quad (6.24)$$

$$\textcircled{3} \sigma[j_t, t] |_{t=0} = 1$$

を用意する。そして $\sigma[j_t, t]_{kl}$ を

$$\sigma[j_{t+1}, t+1]_{kl} = \sigma[j_t, t]_{kl} + \Delta \sigma[j_t, t]_{kl} \quad (6.25)$$

と、時刻 $t+1$ の $\sigma[j_{t+1}, t+1]_{kl}$ へと修正していき、不等式

$$\sum_{k=-p}^{+p} \sum_{l=-q}^{+q} |\sigma[j_{t+1}, t+1]_{kl} - \sigma[j_t, t]_{kl}|^2 < \delta_4 \quad (6.26)$$

が成立した時点 t で終了する。

このとき、 $\sigma[j]_{kl}$ として、

$$\sigma[j]_{kl} = \sigma[j_t, t]_{kl} \quad (6.27)$$

と、終了した時刻 t における $\sigma[j_t, t]_{kl}$ を採用する。

ここで、 $\Delta \sigma[j_t, t]_{kl}$ の求め方は次のとおりである：

$$\begin{aligned} \Delta \sigma[j_t, t]_{kl} = & \epsilon_4[j_t, t]_{kl} \cdot (-1) \cdot [y(\varphi[j_t, t], j_t) - y^{\wedge}(j_t)] \\ & \cdot W[j_t]_{kl} \cdot q_{kl}(\varphi[j_t, t], j_t) \\ & \cdot 2 \cdot \{(T\varphi[j_t, t]^{kl}) - (T\varphi[j_t]_{kl})\}^2 \cdot (\sigma[j_t, t]_{kl})^{-3} \\ & \cdot [1 - p_{kl}(\varphi[j_t, t], j_t)] \cdot \left[\sum_{i \in I} q_{kl}(\varphi[j_t, t], i) \right]^{-1} \end{aligned} \quad (6.28)$$

現実出力 $y(\varphi[j_t, t], j_t)$ の、理想出力 $y^{\wedge}(j_t)$ からの違い $[y(\varphi[j_t, t], j_t) - y^{\wedge}(j_t)]$ の自乗の1/2を最小とするように、更新式(6.25)を設けていることに、注意しておく。

6.3 式(6.4)内の閾値 $b[j]$ の最適化

式(6.2)の大分類関数BSCが2文献[3], [4]のaxiom 3を満たし、2文献[3], [4]でいうカテゴリ間の相互排除性BSC($\omega[i], j) = 0$ (ij)が達成されるように機能させるためには、学習によって最終的に求められた $W[j]_{kl}$, $\sigma[j]_{kl}$ に従って適切な値を選びなおす必要があり、その選び方によってBSCの様相も違ってくる。ここでは採用した式(6.4)の閾値 $b[j]$ の4種類の選び方を示す。

6.3.1 代表パターンを考慮した $b[j]$ の決定法

式(6.4)の閾値 $b[j]$ の値を、第 $j \in J$ 番目のカテゴリに帰属する代表パターン $\omega[j]$ を考慮して決定する方法について述べる。

$$b_1[j] = W_{\text{mean}}(\omega[j], j) \quad (6.29)$$

$$b_0[j] = \max_{i \in J - \{j\}} W_{\text{mean}}(\omega[i], j) \quad (6.30)$$

として、 $b[j]$ を

$$b[j] = \begin{cases} (b_1[j] + b_0[j]) / 2 & \cdots b_1[j] \geq b_0[j] \text{ の場合} \\ b_1[j] & \cdots b_1[j] < b_0[j] \text{ の場合} \end{cases} \quad (6.31)$$

と決める。このとき、カテゴリ間の相互排除性BSC($\omega[i], j$)= 0 ($i \neq j$)が満たされている。

6.3.2 訓練パターンを考慮した $b[j]$ の決定法

$b[j]$ の値を、第 $j \in J$ 番目のカテゴリに帰属する代表パターン $\omega[j]$ を含むパターンの系列 $\phi[j; q]$ ($q=1, 2, \dots, r$)を考慮して決定する方法について述べる。

$$c_1[j] = \min_{q=1, 2, \dots, r} W_{\text{mean}}(\phi[j; q], j) \quad (6.32)$$

$$c_0[j] = \max_{i \in J - \{j\}} \max_{q=1, 2, \dots, r} W_{\text{mean}}(\phi[i; q], j) \quad (6.33)$$

として、 $b[j]$ を

$$b[j] = \begin{cases} (c_1[j] + c_0[j]) / 2 & \cdots c_1[j] > c_0[j] \text{ の場合} \\ c_1[j] & \cdots c_1[j] \leq c_0[j] \text{ の場合} \end{cases} \quad (6.34)$$

と決める。このとき、少なくとも、カテゴリ間の相互排除性BSC($\omega[i], j$)= 0 ($i \neq j$)が満たされている。

6.3.3 全ての訓練パターンについてBSCを1にする $b[j]$ の決定法

$b[j]$ の値を、第 $j \in J$ 番目のカテゴリに帰属する代表パターン $\omega[j]$ を含むパターンの系列 $\phi[j; q]$ ($q=1, 2, \dots, r$)の全ての q についてBSC($\phi[j; q], j$)= 1 となるように、

$$b[j] = \min_{q=1, 2, \dots, r} W_{\text{mean}}(\phi[j; q], j) \quad (6.35)$$

と決める。このとき、少なくとも、カテゴリ間の相互排除性BSC($\omega[i], j$)= 0 ($i \neq j$)が満たされている。

6.3.4 学習による $b[j]$ の決定法

学習によって求められた $b[j_i, t]$ の値をそのまま $b[j]$ として、

$$b[j] = b[j_i, t] \quad (6.36)$$

と決める。このとき、カテゴリ間の相互排除性BSC($\omega[i], j$)= 0 ($i \neq j$)が満たされているとは限らない。

6.4 学習の計算機シミュレーション

大分類関数BSC内の助変数 $W[j]_{kl}$, $b[j]$, $\sigma[j]_{kl}$ はお互いに影響を与え合い、相互の関係の中で適切な値が定まってくるものである。ある1つのカテゴリについて学習を行う場合にも、他のカテゴリについての学習結果から影響を受けるため、ある1つのカテゴリの学習もまた、他のカテゴリとのお互いの関係の中で値が定まってくる。従って、単独の学習変数、単独のカテゴリについて学習するのではなく、 $W[j]_{kl}$, $b[j]$, $\sigma[j]_{kl}$, それぞれの変数をそれぞれのカテゴリについて、並列的かつ逐次的に学習を行えるアルゴリズムで処理手順を構成する必要がある。そこで、画像理解システムが実際に行っている学習処理の流れについて説明する。

学習を開始すると画像理解システムはまず代表パターンを決定する。代表パターンを学習によ

って決定する場合は、求まった代表パターンは各カテゴリに属する訓練パターンの最後のもの、つまり $\phi[j;r]$ (r は1つのカテゴリに属する訓練パターンの総数)の記憶場所に記憶される。代表パターンを視察から選ぶ場合は、代表的な訓練パターン $\phi[j;o]$ を代表パターン $\omega[j]$ として採用する。

代表パターンが決定した後、 $W[j, 0]_{ki}$, $b[j, 0]$, $\sigma[j, 0]_{ki}$ の初期化を行い、その値を一時的に $W[j]_{ki}$, $b[j]$, $\sigma[j]_{ki}$ として採用する。

次に、時刻 t を固定し、最も若いカテゴリ番号から時刻 $t-1$ において求めておいた、 $W[j, t+1]_{ki}$, $b[j, t+1]$, $\sigma[j, t+1]_{ki}$ の値を、今回の $W[j, t]_{ki}$, $b[j, t]$, $\sigma[j, t]_{ki}$ の値として採用し、その値から $W[j, t+1]_{ki}$, $b[j, t+1]$, $\sigma[j, t+1]_{ki}$ の順で値を計算し、カテゴリ番号を1つ進め同様の処理を行う。

全てのカテゴリについての $W[j, t+1]_{ki}$, $b[j, t+1]$, $\sigma[j, t+1]_{ki}$ を計算し終えたならば、その値を次の時刻までの一時的な $W[j]_{ki}$, $b[j]$, $\sigma[j]_{ki}$ として採用し、時刻 t を1つ進める。

このとき、もし $W[j, t]_{ki}$, $b[j, t]$, $\sigma[j, t]_{ki}$ がそれぞれの終了条件を満たしていれば、 $W[j]_{ki}$, $b[j]$, $\sigma[j]_{ki}$ の値の更新は行わなくて、次の時刻からその変数の学習は行わなくて、終了した時刻 t での値を学習結果として採用する。ただし、 $b[j]$ に関しては $b[j]$ を学習によって決定する場合以外のときは、全ての学習が終了したとき最適化を行うため、同じカテゴリについての $W[j]_{ki}$ と $\sigma[j]_{ki}$ の両方の学習が終了した時点で終了させてしまうものとする。

$\sigma[j, t]_{ki}$ の学習に関しては、もし時刻 t における訓練パターン $\phi[j, t]$ が $\omega[j]$ と同じものであるならば無条件に終了条件を満たしてしてしまうため、その時刻での学習は行わないものとする。

すべてのカテゴリについて $W[j]_{ki}$, $b[j]$, $\sigma[j]_{ki}$ のそれぞれの学習が終了したならば、 $b[j]$ の最適化を行い、学習処理全体が終了する。

6.5 非逐次学習

本研究では6.4で述べた逐次学習 (sequential learning)の他に、一括学習 (batch learning)を一部取り入れた非逐次学習 (non-sequential learning)をも取り入れている。その説明は次のとおりである。

$W[j]_{ki}$, $b[j]$, $\sigma[j]_{ki}$ の値を初期化した後、時刻 t を固定し、カテゴリ番号の若い番号から順に $W[j_{t+1}, t+1]_{ki}$ の値を求めていく。すべてのカテゴリについて $W[j_{t+1}, t+1]_{ki}$ を求めたならば、 $W[j_{t+1}, t+1]_{ki}$ の値を次の時刻までの一時的な $W[j]_{ki}$ として採用し時刻 t を1つ進める。次に、前段階での $W[j_{t+1}, t+1]_{ki}$ の値を、今回の $W[j_t, t]_{ki}$ として、今回の $W[j_{t+1}, t+1]_{ki}$ を求め、この処理を繰り返す。もし、終了条件を満たしているならば $W[j_t, t]_{ki}$ を $W[j]_{ki}$ として採用し、以後そのカテゴリについては時刻 t が進んでも学習を行わない。全てのカテゴリが終了条件を満たしたならば、 $W[j]$ の学習は終了し、次に $b[j]$ を $W[j]_{ki}$ と同様の方式で学習を行う。

$b[j]$ の学習が終了した後、 $\sigma[j]_{ki}$ を $W[j]_{ki}$ や $b[j]$ と同様の手順で学習を行う。

$\sigma[j]_{ki}$ の学習が終了した後、 $b[j]$ の最適化を行い、学習処理全体を終了する。

7. RECOGNITRONによる画素認識 (不動点連想形多段階認識)による画像内容の理解手法

RECOGNITRONはパターン(1枚の画像内の事物) ϕ のモデル $T\phi$ を多段階認識法 (multi-stage recognition)で、或るカテゴリ、 \mathcal{C}_j の代表パターン ω_j のモデル $T\omega_j$ へと連想変換し、パターン ϕ を認識する。第 t 認識段階における想起パターン ϕ' と、そのパターン ϕ' の帰属する可能性のあるカテゴリ番号リスト λ' との対 (pair)によって表現される“カテゴリ帰属知識” (categorical

membership knowledge) $\langle \varphi^t, \lambda^t \rangle$ を第 $t-1$ 段階でのカテゴリ帰属知識 $\langle \varphi^{t-1}, \lambda^{t-1} \rangle$ から

$$\begin{aligned} \langle \varphi^{t-1}, \lambda^{t-1} \rangle &\rightarrow \langle \varphi^t, \lambda^t \rangle \\ t &= 0, 1, 2, \dots \end{aligned} \quad (7.1)$$

と求め、つまり、段階的に連想していき、このカテゴリ帰属知識が不動点となる認識段階 t が発見されたとき、すなわち

$$\begin{aligned} \langle \varphi^{t+1}, \lambda^{t+1} \rangle &=_{\Delta} \langle \varphi^t, \lambda^t \rangle, \text{ つまり} \\ \varphi^{t+1} &= \varphi^{t+1} \text{ かつ } \lambda^{t+1} = \lambda^t \end{aligned} \quad (7.2)$$

が成立するとき、多段階認識が完了する。この方法がカテゴリ帰属知識を用いた不動点連想形多段階認識法である。 $\langle \varphi^t, \lambda^t \rangle$ が $\langle \varphi^0, \lambda^0 \rangle$ から式(7.7)の構造受精変換 $TA(\mu)T$ の不動点として連想されたカテゴリ帰属知識である。2元関係 $=_{\Delta}$ の意味は文献 [3] の定義3.1で説明されているが、カテゴリ帰属知識間の恒等関係である。

式(7.1)の変換については、7.1で説明される。

7.1 不動点の探索

RECOGNITRONの注目している入力画像の画素 $x = \langle x_1, x_2 \rangle$ における入力パターン

$$\varphi = \{ \varphi_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q \} \quad (7.3)$$

に対する、第 t ($t=0, 1, 2, \dots$) 認識段階における出力パターンを

$$\varphi^t = \{ \varphi^t_{kl} \mid k=0, \pm 1, \pm 2, \dots, \pm p, l=0, \pm 1, \pm 2, \dots, \pm q \} \quad (7.4)$$

と表現する。ここで、

$$\begin{aligned} \varphi^0 &= T\varphi \\ \varphi^1 &= TA(\mu)T\varphi^0 \\ &\vdots \\ \varphi^{t+1} &= TA(\mu)T\varphi^t \end{aligned}$$

と求めて行き、パターンモデルに関する不動点方程式

$$\varphi^{t+1} = \varphi^t \quad (7.5)$$

の成立するパターン φ^t を求める。このとき t は最終認識段階番号 (final recognition stage number) であり、 φ^t は画素 x において不動点として探索・想起されたパターンモデルである。

ただし本研究では、不等式

$$\| \varphi^{t+1} - \varphi^t \|^2 < \varepsilon^2 \quad (\varepsilon = 0.001) \quad (7.6)$$

が成立した時点で不動点とみなし、終了することになっている。

ここに φ^t は、モデル構成作用素 T を7.4の構造受精作用素 $A(\mu)$ の両側に配置して得られ、構造受精変換と呼ばれる写像

$$TA(\mu)T : \langle \Phi, 2^J \rangle \rightarrow \langle \Phi, 2^J \rangle \quad (7.7)$$

Φ : 処理の対象とする全てのパターン φ の集合

2^J : 集合 J の全ての部分集合の集まり (べき集合)

によって、

$$\varphi^t = TA(\mu \cap \lambda^{t-1})T\varphi^{t-1} \quad (7.8)$$

と求められるものである。なお、本研究では、カテゴリ番号リスト μ を

$$\mu = J \quad (J=1, 2, \dots, m) \quad (7.9)$$

としている。

このように、カテゴリ帰属知識に関する不動点方程式(7.2)が成立する認識段階番号 t が見つかったとき、入力画像の、処理の対象としている画素 $x = \langle x_1, x_2 \rangle$ は、

カテゴリ集合 \mathcal{C}_j , $j \in \lambda^t = \text{CSF}(\varphi^{t-1}, \mu \cap \lambda^{t-1})$ の内の一に帰属するという認識結果が得られるという。

画像内の個々の物体についてRECOGNITRONの持っている知識と学習能力とに応じ、

認識可能(該当するカテゴリが唯1個あること; $|\lambda^t| = 1$)

認識不定(該当するカテゴリが複数個あること; $|\lambda^t| \geq 2$)

認識不能(該当するカテゴリが1個もないこと; $|\lambda^t| = 0$)

という3解釈結果が得られ、何が何でも、認識可能、つまり、該当するカテゴリを唯1つ求めるという解釈結果を出力することはない。ここに、 $|B|$ は集合 B にふくまれる要素の個数の意である。

7.2 線形探索

式(7.1)のカテゴリ帰属知識 $\langle \varphi^t, \lambda^t \rangle$ における、カテゴリ番号リスト λ^t を

$$\lambda^t = \text{CSF}(\varphi^{t-1}, \mu \cap \lambda^{t-1}) \quad (7.10)$$

と選んだ探索である。登場しているカテゴリ選択関数CSFについては、7.5で説明されている。

7.3 修正線形探索

式(7.1)のカテゴリ帰属知識 $\langle \varphi^t, \lambda^t \rangle$ における、カテゴリ番号リスト λ^t を

$$\lambda^t = \text{CSF}(\varphi^{t-1}, \mu \cap \lambda^{t-1}) - \{i\} \quad (7.11)$$

と選んだ探索である。ここに、カテゴリ番号 i は

$$\textcircled{1} \quad |\text{CSF}(\varphi^{t-1}, \mu \cap \lambda^{t-1})| > 1 \quad \text{のとき}$$

$$i = \arg \min_k \text{SM}(\varphi^{t-1}, \omega_k) \quad (7.12)$$

$$k \in \text{CSF}(\varphi^{t-1}, \mu \cap \lambda^{t-1}) \quad (7.13)$$

と選ばれるが、

$$\textcircled{2} \quad |\text{CSF}(\varphi^{t-1}, \mu \cap \lambda^{t-1})| \leq 1 \quad \text{のときは}$$

$$|i| \equiv \phi, \quad \text{つまり } \lambda^t \text{ を式 (7.10) のように設定する} \quad (7.14)$$

としている。

7.4 構造受精作用素 $A(\mu)$

構造受精作用素

$$A(\mu) : \Phi \rightarrow \Phi \quad (\mu \in 2^I) \quad (7.15)$$

は、任意の $\varphi \in \Phi$ について、

$$\textcircled{1} \quad \varphi = 0 \quad \text{または} \quad \mu = \phi \quad (\text{空集合}) \quad \text{のとき}$$

$$A(\mu)\varphi \equiv 0 \quad (7.16)$$

$$\textcircled{2} \quad \varphi \neq 0 \quad \text{かつ} \quad \mu \neq \phi \quad (\text{空集合}) \quad \text{のとき}$$

$$\textcircled{2}-1 \quad \sum_{k \in \mu} \text{BSC}(\varphi, k) = 0 \quad \text{ならば}$$

$$A(\mu)\varphi \equiv \sum_{k \in \mu} \text{SM}(\varphi, \omega_k) \cdot T\omega_k \quad (7.17)$$

$$\textcircled{2}-2 \quad \sum_{k \in \mu} \text{BSC}(\varphi, k) > 0 \quad \text{ならば}$$

$$A(\mu)\varphi \equiv \sum_{k \in \mu} \text{SM}(\varphi, \omega_k) \cdot \text{BSC}(\varphi, k) \cdot T\omega_k \quad (7.18)$$

と定義される。

7.5 カテゴリ選択関数CSF

2文献 [3], [4] でのaxiom 4を満たすカテゴリ選択関数(category-selection function)

$$\text{CSF} : \Phi \times 2^J \rightarrow 2^J \quad (7.19)$$

は, 任意の $\varphi \in \Phi$, $\lambda \in 2^J$ について

- ① $\varphi = 0$ または $\lambda = \phi$ (空集合) のとき

$$\text{CSF}(\varphi, \lambda) \equiv \phi \text{ (空)} \quad (7.20)$$

- ② $\varphi \neq 0$ かつ $\lambda \neq \phi$ のとき

- ②-1 $\sum_{k \in \lambda} \text{BSC}(\varphi, k) = 0$ ならば

$$\text{CSF}(\varphi, \lambda) \equiv \{k \in \lambda \mid \text{SM}(\varphi, \omega_k) > 0\} \quad (7.21)$$

- ②-2 $\sum_{k \in \lambda} \text{BSC}(\varphi, k) > 0$ ならば

$$\text{CSF}(\varphi, \lambda) \equiv \{k \in \lambda \mid \text{SM}(\varphi, \omega_k) > 0 \text{ かつ } \text{BSC}(\varphi, k) = 1\} \quad (7.22)$$

と定義される.

7.6 任意の画素の名認識段階について詳細に検討する際に必要な諸量

任意の画素についての認識がどのような様相を経て行われているかを詳細に検討するために, 7.6.1~7.6.4での諸量を求めなければならない. これらの諸量は本シミュレーションで求められているが, その検討は紙面の都合上, 割愛される.

7.6.1 認識情報量RGetpy

認識情報量(recognition entropy)RGetpyは

$$\text{RGetpy}(\varphi^t) = \sum_{j \in J} \text{SM}(\varphi^t, \omega_j) \cdot \log_e [\text{SM}(\varphi^t, \omega_j) / P(\mathcal{C}_j)] \quad (7.23)$$

と定義される. ただし,

$\text{SM}(\varphi^t, \omega[j]) = 0$ のときは

$$\text{SM}(\varphi^t, \omega_j) \cdot \log_e [\text{SM}(\varphi^t, \omega_j) / P(\mathcal{C}_j)] = 0 \quad (7.24)$$

として計算する.

認識が正常に進んでいるときは通常,

$$\text{RGetpy}(\varphi^t) \leq \text{RGetpy}(\varphi^{t+1}) \quad (7.25)$$

が成立していると予想され, 値は増加していく.

7.6.2 認識情報量RAIN

φ^t を基準にしての φ^{t+1} の認識情報量RAINは

$$\text{RAIN}(\varphi^{t+1}, \varphi^t) = \sum_{j \in J} \text{SM}(\varphi^{t+1}, \omega_j) \cdot \log_e [\text{SM}(\varphi^{t+1}, \omega_j) / \text{SM}(\varphi^t, \omega_j)] \quad (7.26)$$

と定義される.

ただし, $\text{SM}(\varphi^{t+1}, \omega_j) = 0$ または $\text{SM}(\varphi^t, \omega_j) = 0$ のとき

$$\text{SM}(\varphi^{t+1}, \omega_j) \cdot \log_e [\text{SM}(\varphi^{t+1}, \omega_j) / \text{SM}(\varphi^t, \omega_j)] = 0 \quad (7.27)$$

として計算する.

認識が正常に進んでいるときは通常,

$$\text{RAIN}(\varphi^{t+1}, \varphi^t) \geq \text{RAIN}(\varphi^{t+2}, \varphi^{t+1}) \quad (7.28)$$

が成立していると予想される.

7.6.3 カテゴリ帰属知識のポテンシャルエネルギー $E(\varphi^t, \lambda^t)$

カテゴリ帰属知識 $\langle \varphi^t, \lambda^t \rangle$ のポテンシャルエネルギー(energy) $E(\varphi^t, \lambda^t)$ は

- ① $\varphi^t = 0$ あるいは $\lambda^t = \phi$ (空集合) のとき

$$E(\varphi^t, \lambda^t) = 0 \quad (7.29)$$

② $\varphi^t \neq 0$ かつ $\lambda^t \neq \phi$ (空集合) のとき

$$E(\varphi^t, \lambda^t) = |\varphi^t| - \sum_{j \in \lambda^t} SM(\varphi^t, \omega_j) \quad (7.30)$$

と定義される。

認識が正常に進んでいるときは通常、

$$E(\varphi^t, \lambda^t) \geq E(\varphi^{t+1}, \lambda^{t+1}) \quad (7.31)$$

が成立していると予想される。

7.6.4 カテゴリ帰属知識間の距離 SMdis

2つのカテゴリ帰属知識間の距離 SMdis は

$$SMdis(\langle \varphi^s, \lambda^s \rangle, \langle \varphi^{s+1}, \lambda^{s+1} \rangle) \equiv -\log_e \sum_j [SM(\varphi^s, \omega_j) \cdot SM(\varphi^{s+1}, \omega_j)]^{1/2} \quad (7.32)$$

$$j \in CSF(\varphi^s, \lambda^s) \cap CSF(\varphi^{s+1}, \lambda^{s+1}) \quad (7.33)$$

と定義される。

認識が正常に進んでいるときは通常、

$$SMdis(\langle \varphi^s, \lambda^s \rangle, \langle \varphi^{s+1}, \lambda^{s+1} \rangle) \geq SMdis(\langle \varphi^{s+1}, \lambda^{s+1} \rangle, \langle \varphi^{s+2}, \lambda^{s+2} \rangle) \quad (7.34)$$

が成立していると予想される。

7.7 認識過程における振動発生の可能性

式(7.1)の認識過程において、7.1で説明されている不動点探索を行っているとき、

$$\sum_j SM(\varphi^t, \omega_j) = 0 \quad (7.35)$$

ここに、式(7.35)においては、カテゴリ番号 j は集合

$$J - CSF(\varphi^{t-1}, \lambda^{t-1}) \quad (7.36)$$

に属する

が成立していれば、不動点方程式(7.5)は必ず成立し、認識はいずれ終了する。しかし上述の式(7.35)が成立していなければ、認識が振動し続け終了しない可能性がある。

本シミュレーションにおいては、式(7.35)は必ずしも成立しておらず、多段階認識過程において稀に振動が発生し認識が終了しない場合がある。なお画像理解システムはカテゴリ数の10倍認識段階が進んだならば振動しているとみなし、認識の中止を決定する。

8. シミュレーション結果とその検討

本研究で構築された画像理解システムには多種多様なモデル構成作用素 T 、類似度関数 SM 、学習方法、探索法などが用意されており、システムに与える各種パラメータや、訓練パターンや代表パターンの選び方、その大きさなど、少なくない選択肢が存在する。これらはマウスをクリックすることによってその1つが選択され、実行されるようになっている。従ってシミュレーションは、多数の選択肢の組み合わせの中から、1つ1つの選択がどのような影響を及ぼすかを、予測と実験を繰り返しその傾向を読み取りながら進められた。どの選択も一長一短で、常に最良の結果を出すということはほとんどなく、対象とする画像の性質によっても理解の性能が変動する。ある程度、認識率の向上を犠牲にしながらも、学習したパターンに近いものはなるべく正確に認識することができて、しかもできるだけ誤認識を少なくできると思われるものを採用し、シミュレーションを繰り返した。

本章ではその繰り返しをとおして読み取ることができた傾向を明らかにしつつ、検討を進めていく。

8.1 どのモデル構成作用素Tを採用すべきかについての検討

モデル構成作用素Tは3において説明されているとおり、本研究では3値、5値、Sを採用する3種類を構成した。

8.1.1 3値のTの傾向

3値のTはその量子化のあらっぼさから変形を吸収する機能が一番高いため、認識時のカテゴリ間の分離が最も良い。人工画像(artificial or man made scene)などの単純な画像では認識結果の良好性・不良性が明確に判明し、学習の働きや諸設定条件が適切に調整・整合されていれば良好な結果を期待することができる。しかし、自然画像(natural scene)はパターンの微妙な変化が重要である場合が多く、5値のTなどに比べると、認識がやや正確さに欠け大雑把である傾向がある。実際、風景画像で3値のTを採用した場合、大部分が電柱になってしまったり、空になってしまったりといった事態が生じた。

閾値 h_{kl}^{\pm} を学習によって2助変数 k , 1成分ごとに適当な値を求めることも可能であるが、本研究では一貫して ± 0.5 を採用している。今後、 k , 1成分ごとに適応的な値を求める学習研究をしなければならない。

8.1.2 5値のTの傾向

5値のTは変形を吸収する機能は3値のものに比べ劣るが、その反面認識が正確である。形状の抽出が精密であるため、別カテゴリとの分離があいまいになり、認識が進むにつれ例えば木から道路へ徐々に移ってしまったり、電柱が空と区別がつかなくなったりもする。

しかし、自然画像ではかなり優れた性能を示す。

8.1.3 Sを採用するTの傾向

Sを採用するモデル構成作用素Tは変形を吸収する機能が無いため、訓練パターンを用いたSMを使うかぎりでは、訓練パターン以外のものは認識不定になりやすい。このTを用いて不動点連想形多段階認識を行う場合、未知の画像に対する順応性は弱く、不動点への収束に多くの認識段階を要するが、学習が上首尾になされれば、誤認識が少ない精密で良好な認識結果を期待できる。

8.2 どの類似度関数SMを採用すべきかについての検討

本研究で採用した類似度関数SMは、例えばパターンモデルの値が全て1と全て-1のものなど正反対のパターンを区別できないため、空と道路といったような正反対の濃度値をもつ傾向のあるもの同士が区別できないときがあり、SMで混同するパターンを分離するのに大分類関数BSCのカテゴリ抽出機能を利用できた。白黒の度合いは風景画像ではパターンの帰属するカテゴリを特定するための重要な手がかりであり、パターン分離に関する“SMとBSCとの相互補完機能”を今後、詳細に、研究しなければならない。

4.2, 4.3~4.6には、「各カテゴリに帰属する複数個のパターンを使ったSM①~④」が説明されているが、訓練パターンの系列と称される式(4.8)の系列を使っていることに注意する。

訓練パターンを使うSMでは、1つのカテゴリのみが高い値になり、それ以外のカテゴリの値は全て同じになる。つまり帰属する可能性のあるカテゴリ以外は、すべて同じ類似度を示してしまう。ところが、式(4.4)の相互情報量MIを使ったSMでは、このような不都合な事態は生じないことが

判明している。従って、あるカテゴリのSMの値が1.0にならない限り、他のカテゴリについてのSMの値が0になることがないため、BSCで分離されない限り、SMのみではあるパターンが1つのカテゴリに帰属する可能性を排除することができない。4.2, 4.3~4.6のSMを使う場合はBSCのカテゴリ間の分離性能が非常に重要になってくる。

8.2.1 式(4.4)の相互情報量MIを使ったSMの傾向

代表パターンのみを考慮するためか、パターン間の違いに対する反応は最も大きいのが、4.1の、式(4.4)の相互情報量MIを使った式(4.6)のSMである。ただ、代表パターンのみを使用するため、同じカテゴリに帰属するパターンであっても、代表パターンに似ていないならば、SM値は小さくなってしまふ。それで、なるべく多くの変形を吸収させたいのであれば、代表パターンを入念に学習しなければならない。その場合はSMの値は視察で代表パターンを決めたときよりもカテゴリ間の分離性が失われるが、パターンの違いについてのSM値の変動幅が大きいので、大きく認識結果が間違ふことも少なくない。パターン間の違いに対しやや鈍感に5の代表パターン学習法で決定したほうが良い認識結果になる傾向がある。

8.2.2 訓練パターンを考慮したSM①の傾向

訓練パターンを考慮するSMの中では、2番目に全般的にSM値の変動が小さく、SM値を抑えるところは抑え、跳ね上げるところは跳ね上げられる。それで、使い勝手がよく、パターン間の変動について安定した印象を与える。誤認識をある程度抑え、認識すべきところもなるべく認識させる標準的な結果を望むときに使うと良いとの結論を得た。

8.2.3 量子化するSMの傾向

量子化するSMは、式(4.19)の Φ_j に、SMに用いる式(4.8)の Ψ_j を採用したためか本シミュレーションでは使用不可の事態を招いてしまった。

MIを使ったSMでは、2不等式(4.20), (4.21)を満たさなければならない2閾値 $s_0(j)$, $s_1(j)$ が不等式 $s_0(j) < s_1(j)$ を満たさず使用できなかったし、訓練パターンを考慮するSMでは、2式(4.24), (4.25)での $a(j)$, $b(j)$ がそれぞれ0, 1になり、 $s_0(j)=0$, $s_1(j)=0.83333\cdots$ になってしまった。これは、訓練パターンを考慮するSMを使う場合は、訓練パターンが帰属するカテゴリとのSMの値($SM(\phi[j; q], \omega[j])$)を1にするようになっているためである。あるカテゴリでSMの値が1を示すことで、帰属するカテゴリ以外とのSMの値($SM(\phi[j; q], \omega[i], (i \neq j))$)が全て0になってしまうためである。

閾値 $s_0(j)$ が0ではどんなSMの値も量子化されないし、閾値 $s_1(j)$ が約0.8では、量子化するまでもなく次の認識段階でSMの値は1.0になることは、実際にこの閾値で量子化したものとししないものの実行結果を比較したところ、まったく同じ結果になったことから、推測される。

SMに使っているパターンの系列とは別のパターンの系列を用いれば、SMの量子化が有効になる可能性がある。

8.2.4 訓練パターンを考慮したSM②の傾向

訓練パターンを取り入れ構成されるSMの中では、もっともSM値の変化が大きい。ほかのSMでは認識されず不定になってしまうものでも、このSMを使えば、認識できることがある。少しでも正認識可能性のあるカテゴリとのSMの値を大きくし、それ以外のSMの値を抑える働きが強いので扱いが難しい。認識されにくいものでもよく認識が進んで認識できてしまう反面、間違つたまま多段階認識が進んでしまうと修正ができなくて、そのまま誤認識されてしまう。

特定の画素について認識性能を確かめる場合やBSCの分離が悪いときは威力を発揮するが、正確な画像解釈を得たい場合使わないほうが無難である。

8.2.5 訓練パターンを考慮したSM③の傾向

訓練パターンを考慮するSMの中では、2番目にSM値の変動が大きい。正確な認識を求めるよりもむしろ、訓練パターンの妥当性を確かめる場合など試験的に画像解釈を行う場合に効果的である。

8.2.6 訓練パターンを考慮したSM④の傾向

訓練パターンを考慮するSMの中では、最もSM値の変動が小さい。かなり適切な画像解釈を行うことができるが、大分類関数BSCの備えているカテゴリ間の分離があまりよくないときはほとんど認識不定になることがあり、また、当然分離がよいと思われるときも、認識不定になってしまったりする。しかしながら、大抵の場合、適切な解釈結果を期待できる。

8.3 代表パターン $\omega[j]$ を学習で決定すべきか、視察で決定すべきか？

代表パターンを5の学習法を適用して決定すべきか、視察で選ぶべきか、そのどちらが適切であるかは一概には言えないが、学習によって決定した方が、全体的に良い結果になる場合が多かった。

唯、カテゴリ総数が大きくなり、代表パターン間の分離をもたらす特徴を視察で決定するのが困難になる場合、例えば5.の学習で決定するしか方法が残されていない。

8.4 修正探索について

画像の理解・解釈は2次元平面上で展開されるため、多数の同一のカテゴリラベルが集まれば、多少の別カテゴリが紛れ込んでいたとしても、すべて同じカテゴリとみなしてしまうこともできる。しかし誤認識が増えてくると正しく認識された部分が埋もれてしまい、その物体の帰属するカテゴリを特定することができなくなり、物体の存在する領域も不明確になってくる。誤認識が増えるなら、認識不定になったほうがよりその画像について正確に理解できるといえる。

修正線形探索は、その認識段階で最も可能性の少ないと思われるカテゴリを1つずつ削っていくため必ず1つのカテゴリを求めることができる。一見望まれる結果が出るように思われるが、訓練パターンに含まれないような微妙な部分があるパターンや、視点の移動によって訓練パターンの中心から外れたパターンなど、元々認識不定になってよいパターンさえカテゴリを確定してしまうため、誤認識が非常に多くなる。本シミュレーションでは振動を回避したい場合や、訓練パターンに近いパターンの中心付近などの認識しやすい特定の部分を確実に認識させたい場合以外は、あまり使用する価値が見られなかった。

8.5 重み $W[j]_k$ の学習についての検討

2式(6.1)、(6.2)の大分類関数BSC内の式(6.5)の $W[j]_k$ の、6.2.1で説明済み学習においては、式(6.5)の更新分 $\Delta W[j_i, t]_k$ 内の係数 ϵ_2 の与え方次第で、収束するかどうかが決まる。 ϵ_2 の初期値として0.1を採用し、それ以降学習時刻 t が進むごとに減少していくように、つまり、式(8.3)のように ϵ_2 を変化させて行くと、上首尾に収束することが判明した。

式(8.1)のように、 ϵ_2 の初期値を1になるように与えた場合を説明する。この場合、 $\Delta W[j; t]_k$ の値が訓練パターンを2順する位まで、例えば10、-30、100、-200のように、時刻が進むたびに正と負の値を繰り返しながらその絶対値が増加しつづけ、やがて変動がおさまって来ると、今度は変動が大きかったときに $W[j; t]_k$ にもたらされた変動分を修正することができないまま学習

が終了してしまう。 $W[j; t]_{kl}$ の値が大きくなると、 $\sigma[j; t]_{kl}$ の値も大きくなり、 そうなると、 異常な学習結果を得てしまう。 そこで、 訓練パターンを12から6に減らして見ると、 正常と思われる学習結果が得られた。 7個でも正常と思われる学習結果が得られたが、 8個以上に訓練パターンを増加させると異常な学習結果を得てしまう。

そこで、 ϵ_2 の値が0.1以下になるように、

$$\epsilon_2 [j, t]_{kl} = 1 / (1 + q) \quad (8.1)$$

$$\text{ここに、 } q = t \div r \text{ の商} \quad (8.2)$$

という ϵ_2 の値に0.1をかけ、 改めて

$$\epsilon_2 [j, t]_{kl} = 0.1 \cdot \{1 / (1 + q)\} \quad (8.3)$$

にすると、 正常と思われる学習結果が得られるようになった。 ちなみに、 0.1の代りに0.5や0.2をかけてみたが、 0.2や0.5では $\Delta W[j; t]_{kl}$ の、 異常な変動を抑えきれなかった。 ϵ_2 の組織的選定方法を研究する必要がある。

8.6 閾値 $b[j]$ の学習についての検討

式(6.4)内の閾値 $b[j]$ を6.2.2の最急降下学習で求めると、 訓練パターンと同じパターンしか正しく認識できないことが判明した。 訓練パターンを含まない画像について画像解釈を行ってみたが、 訓練パターンと同じパターンを持つ空だけがSMによって1段階で正認識されただけで、 それ以外のカテゴリを表す画素については全て認識不定になった。 大分類関数BSCがほとんどの場合1を示してしまい、 式(7.17)からわかるように、 類似度関数SMによってしか帰属する可能性能あるカテゴリを絞り込めなくなるからである。

代表パターンのみを使用し、 $b[j]$ を決定する方法では訓練パターンを必ずしも認識できない。 訓練パターンの数が多くした場合で、 訓練パターンを使用する $b[j]$ の決定法ではカテゴリ間の分離が悪くなりすぎるときに、 採用するのがよい。 本シミュレーションでは、 代表パターンのみを使用する $b[j]$ の決定法と、 訓練パターンを使用する $b[j]$ の決定法とを比較・検討し、 全ての訓練パターンについてBSCを1にする $b[j]$ の決定法を標準的に採用したが、 画像の性質によってこの両者の方法を使い分けるべきである。

8.7 標準偏差 $\sigma[j]_{kl}$ の学習についての検討

式(6.7)内の標準偏差(の1/2) $\sigma[j]_{kl}$ の、 6.2.3で説明されている最急降下学習では、 初期値の選び方が学習過程の収束に大きな影響を与える。 初期値は、 1.0から10.0位までの間にした方がよい。 あまり大きく選ぶと、 学習時の $\Delta W[j; t]_{kl}$ の値が k, l 成分ごとに差がでなくなってしまう。 そのため $W[j; t]_{kl}$ があまり変化せず、 結果として $\sigma[j; t]_{kl}$ の値もあまり変化しないまま学習が進んでしまう。 $\sigma[j]_{kl}$ の初期値を100くらいに大きく選んでしまうと、 式(6.28)の更新分 $\Delta \sigma[j; t]_{kl}$ の値が非常に小さくなって、 例えば10の6乗などをかけないとほとんど変わらないまま学習が終了し、 $W[j]_{kl}$ などもほとんど変わらなくなってしまうといったことが起こる。 逆に初期値を0.1などに小さくとり過ぎると、 $\Delta \sigma[j; t]_{kl}$ が負の値になったり、 $W[j; t]_{kl}$ の値が非常に大きくなってしまったりもするので、 1.0位に選ぶのが適当であった。

8.8 理想出力 $y^*(j)$ についての考察

3式(6.15)、 (6.22)、 (6.28)内に登場している理想出力 $y^*(j)$ は、 本シミュレーションでは標準的

に5/8を採用したが、1/8とした場合との違いは次のように説明される。条件を全て同じにして、理想出力だけを5/8と1/8にした2つの実行結果を比べてみると、1/8にした方の学習結果が、5/8にした方の学習結果にくらべ、全体的に $W[j]_k$, $b[j]$, $\sigma[j]_k$ などの値がやや小さかった。

解釈結果では、5/8の場合で誤認識した画素が、1/8のものでは認識不定になり、誤認識が少なくなったが、逆に5/8が正しく認識できた画素も1/8の場合では認識不定になってしまうことがあった。理想出力を小さくとると、学習の結果得られる諸量が全体的に小さくなるため、パターン間の差異が抑制され、各カテゴリ間についてBSC間の距離が縮まり、認識不定になりやすくなった。

8.9 訓練パターンにの選定について

訓練パターンが少なければ代表パターンに比べて変形した未知の画像に対する適応性が弱まる。しかし訓練パターンが多くなれば、それだけ幅広くパターンを同一カテゴリとして扱わなければならなくなるため、カテゴリ間の分離が悪くなり、認識不定が増えたり、認識段階が進むにつれ似ている別カテゴリへ移行し誤認識されたりする。そこで、本シミュレーションでは1つのカテゴリにつき、カテゴリ数の2倍~3倍くらいの数の訓練パターンを大体同じような傾向をもたせて作っていくことにした。例えば車といっても全体像が視野に収まらなければ、視点によってフロントガラスであったり、タイヤであったりする。本研究ではフロントグリルやボンネット、前照灯付近の車の「顔」の部分を一貫して車として学習させた。そのため、車の「顔」以外の部分は空や家になってしまったりもしたが、車の「顔」に関しては未知の画像であってもかなり良く認識することができた。

8.10 非逐次学習について

6.5の逐次学習では学習係数 ϵ 2の値を式(8.3)のようにおさえないで学習する式(8.1)のような場合、重み $W[j;t]$ の値が増加し続ける。学習がある程度進むと次第におさまってくるが、値の増加が落ち着いてきてからでは増加し続けていたときに受けた影響を修正できないときがあるため、重み $W[j;t]$ を先に決定してから、標準偏差 $\sigma[j, t]$ を求めてみようという考えから、6.5の非逐次学習方式を考案したのであるが、実際実行してみても、逐次学習と比べ、格別によい画像理解結果は得られなかった。

8.11 実行時間について

パソコン処理系を6台、用意し、本シミュレーションを実行している。

処理系が同じであっても使用するTやSM、カテゴリ数や訓練パターンの数や、視野の大きさなど、諸条件によってその各処理系による実行時間は大きく変わってくる。学習の相では、カテゴリごとに終了する時刻がまちまちで、あるカテゴリについては、時刻41で終わっているのに、最後に終了したカテゴリは時刻13720まで単独で進んでいたりすることもあり、時刻 t が1進むのにどれくらいの時間がかかるかは不明である。認識の相では、用いたTやSMや、BSCの分離の度合いによってもその各処理系による実行時間まったく違って来る。

そこで、学習の相についてはカテゴリ数を6、視野の大きさを 11×9 ($p=5, q=4$)として、全てのカテゴリについての全ての $W[j]$, $b[j]$, $\sigma[j]$ の学習が時刻1000になれば同時に全て終了するようにして、クロック周波数が1.2GHzのパソコンで実行させてみたところ、24分55秒かかった。従って、時刻 t が1

進むのに約1.495秒かかることがわかった。さらにカテゴリ数6で割ると、1つのカテゴリの学習につき、約0.249秒かかっている事がわかった。この間の事情を表6.1にまとめている。

表8.1 学習に要する時間

①計算機のクロック周波数	1.2GHz
②カテゴリの数	m=6
③視野の大きさ	p=5, q=4(11×9=99画素)
④モデル構成作用素T	5値のT
⑤学習時刻が1進むためにかかる時間	約1.495秒
⑥1つのカテゴリの学習にかかる時間(⑤/②)	約0.249秒

画像の解釈にかかる時間については、本研究で用いた比較的標準的な条件で行った解釈結果では、クロック周波数1.2GHzの計算機上で、画像1枚の解釈にかかった時間が2時間18分38秒となっている。画像サイズが101×81であるから、1画素あたりの認識にかかる時間は、平均約1.017秒であることがわかる(表8.2を参照)。

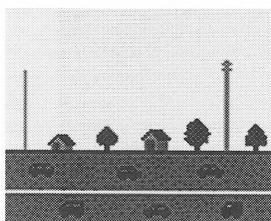
表8.2 画像解釈に要する時間

①計算機のクロック周波数	1.2GHz
②カテゴリの数	m=6
③視野の大きさ	p=5, q=4(11×9=99画素)
④画像の大きさ	101×81(=8181画素)
⑤モデル構成作用素T	3値のT
⑥類似度関数SM	訓練パターンを考慮するSM①
⑦カテゴリごとの訓練パターンの数	r=12
⑧画像一枚の解釈にかかる時間	2時間18分38秒
⑨1画素あたりの認識にかかる時間(⑧/④)	平均約1.017秒

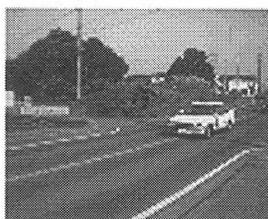
8.12 シミュレーション結果の実例

空、木、車、家、道路、電柱の計6種類のカテゴリを採用し計算機シミュレーションを行った。シミュレーション結果の内、以下に、幾つかを説明しよう。

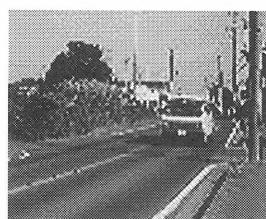
処理を行った画像の内、図8.1には、訓練パターンを含む二つの画像である人工風景画像A.bmpと実際の風景自然画像035.bmp、並びに、訓練パターンを全く含まない画像である実際の風景自然画像002.bmpが示されている。風景画像002.bmpはシステムにとっては未知の画像である。



人工風景画像A.bmp



002.bmp



035.bmp

図8.1 処理画像の3例

解釈結果の色とカテゴリの対応関係は図8.2のとおりである。

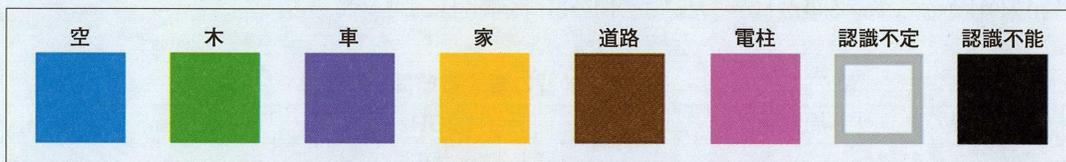


図8.2 カテゴリラベルと色の対応関係

処理画像とその結果の下に書かれている各種の「Selector」は、学習・認識・解釈を行うときに与えた条件を示すものである。各「Selector」の説明は次のとおりである：

- ・ Tselector：1=3値のT，2=5値のT，3=Sを採用するT，4=移動，回転，伸縮を修正するT，5=カテゴリごとに修正するT
- ・ SMSelector：1=MIを使ったSM，2=訓練パターンを考慮したSM①，3=なし，4=訓練パターンを考慮したSM②，5=訓練パターンを考慮したSM③，6=訓練パターンを考慮したSM④
- ・ QUANTIZATION：true=SMを量子化する，false=SMを量子化しない
- ・ omegaSelector：1=学習による $\omega[j]$ の決定，2=視察による $\omega[j]$ の決定
- ・ bSelector：1=代表パターンのみを考慮した $b[j]$ の決定法，2=訓練パターンを考慮した $b[j]$ の決定法，3=全ての訓練パターンについてBSCを1にする $b[j]$ の決定法，4=学習による $b[j]$ の決定法
- ・ linerSearchSelector：1=線形探索，2=修正線形探索
- ・ learningSelector：1=逐次学習，2=非逐次学習

以下に，7結果1～7が示されている

[結果1]



結果03-01-1-modelA

TSelector = 2

SMSSelector = 2 (QUANTIZATION = false)

omegaSelector = 2

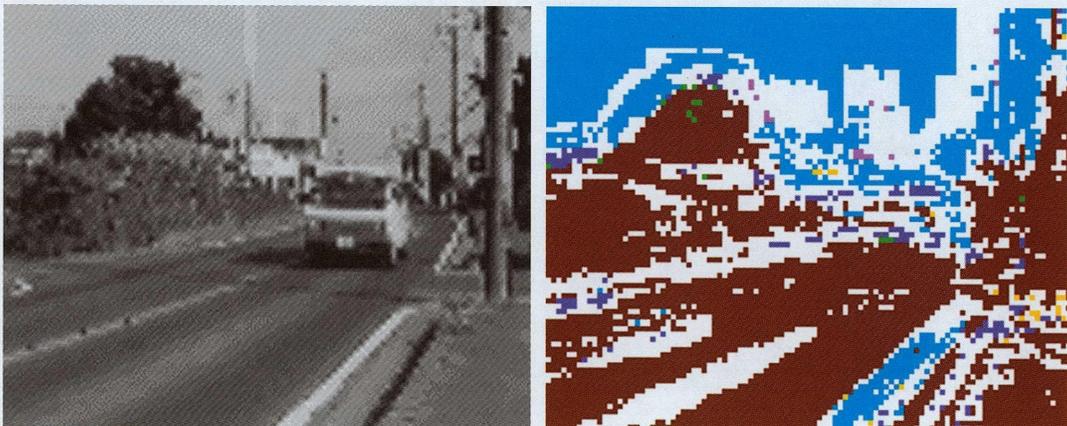
bSelector = 3

linerSearchSelector = 1

learningSelector = 1

dispSelector = 1

[結果2]



結果03-09-5-035

TSelector = 2

SMSSelector = 2 (QUANTIZATION = false)

omegaSelector = 1

bSelector = 3

linerSearchSelector = 1

learningSelector = 1

[結果3]



結果03-05-3-002

TSelector = 1

SMSSelector = 5 (QUANTIZATION = false)

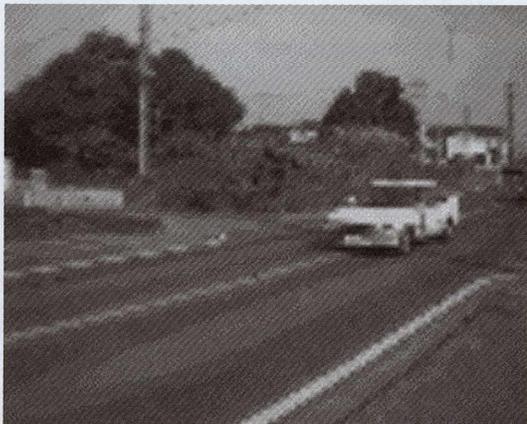
omegaSelector = 2

bSelector = 3

linerSearchSelector = 1

learningSelector = 1

[結果4]



結果03-08-3-002

TSelector = 2

SMSSelector = 5 (QUANTIZATION = false)

omegaSelector = 1

bSelector = 3

linerSearchSelector = 1

learningSelector = 1

[結果5]



結果03-08-5-002

TSelector = 2

SMSSelector = 1 (QUANTIZATION = false)

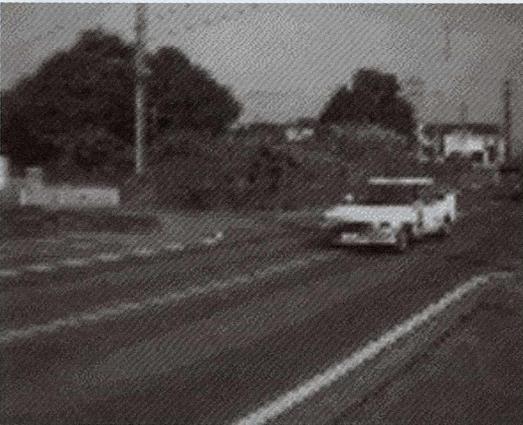
omegaSelector = 1

bSelector = 3

linerSearchSelector = 1

learningSelector = 1

[結果6]



結果03-09-2-002

TSelector = 2

SMSSelector = 6 (QUANTIZATION = false)

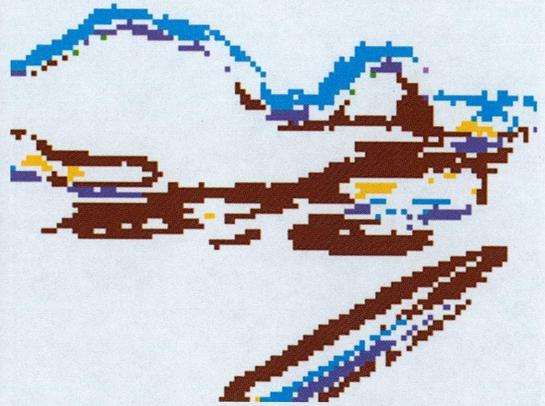
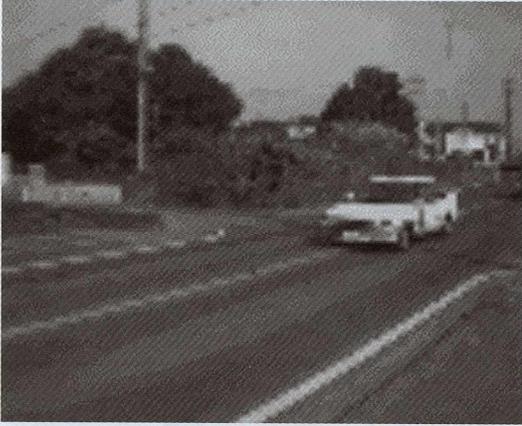
omegaSelector = 1

bSelector = 2

linerSearchSelector = 1

learningSelector = 1

[結果7]



結果03-10-6-002

TSelector = 3

SMSSelector = 6 (QUANTIZATION = false)

omegaSelector = 1

bSelector = 3

linerSearchSelector = 1

learningSelector = 1

9. おわりに

S.Suzukiは万能性認識システムRECOGNITRONに関する数理 [3], [4] を展開済みである. 実は, 7.1の不動点探索の, RECOGNITRONによる認識過程はS.Suzukiが考案したある連想形認識方程式を解く過程である. この点については, 文献 [3] の付録GのG7, G8両節に説明されている. RECOGNITRONによる認識の, 求解過程による意味付けは, 人工知能言語Prologを使って書かれたプログラムの実行が論理(導出原理)によって問題を解決する反駁証明になっていることと対応している. このような対応付けはパターン情報処理分野では, S.Suzuki以外の研究では全く見られない.

本論文では, 文献 [6] で提案されている“画素単位に画像内容を理解するシステムの一設計理論を基盤として, JAVA言語で計算機シミュレーションされた結果が報告された. 文献 [6] で解析された数式のささいな間違いも訂正され実行された. 式(6.11)の大分類関数BSCが学習的に形成されることに, 画像内容理解能力の源泉があることが確認された.

事物の存在する領域を大きさ $(2p+1)(2q+1)$ の画像成分と仮定し, 領域抽出(セグメンテーション)が自動的になされる画素単位の認識方式を内蔵している認識システムRECOGNITRONについてその性能を確かめる計算機シミュレーションをJAVA言語で実行し, その結果を検討した. モデル構成要素T, 類似度関数SM, 大分類関数BSC, カテゴリ選択関数CSFはそれぞれ, axiom 1~4を満たすように構成されているので, 2文献 [3], [4] で解析されているRECOGNITRONの基本的な認識性能は保証されている.

JAVA言語でRECOGNITRONの動作を記述し、**graphic user interface (GUI)**を設けたことによって、マウスをクリックすることによってその殆どの動作をコントロールするようになってきた。例えば、画像内の任意の一つの画素をクリックすることによってこの画素が帰属するカテゴリを7.1の不動点探索で決定できるようになっている。

3.1での3値のモデル構成作用素Tについてその助変数である閾値 $h_{\pm k}$ を学習によってk, l成分ごとに適当な値を求めることも可能であるが、本研究では一貫して ± 0.5 を採用している。今後の研究では、適応的な設定を考えなければならない。

8.1.3におけるSを採用するTの場合、学習が適切に機能していれば誤認識が少ない精密な認識結果を期待できる(結果03-10-6-002を参照)。しかし、自然画像はパターン of の微妙な変化が重要である場合が多く、5値のTなどに比べると、認識がやや正確さに欠け大雑把である傾向がある。実際、風景画像で3値のTを採用した場合、大部分が電柱になってしまったり、空になってしまったりといった事が起こった。なお、処理対象パターンの変形・回転・伸縮を修正するTを採用すれば、画像理解に関する性能が良くなる。このTについては文献7)の2付録A, Bで研究されているが、本研究では処理時間が多大であることが判明し、途中でその処理を中止している。

本シミュレーションにおいては、7. の認識方法による認識結果に、

認識可能(該当するカテゴリが唯1個あること)・認識不定(該当するカテゴリが複数個あること)・認識不能(該当するカテゴリが1個もないこと)という3結果

が生じたが、これは、採用した類似度関数SMがいずれもミックスチュア条件を満たすことが保証されていないためである(文献[4]の定理6.6を参照)。

また、7.7で説明されている認識過程の振動は、実際生じたが、これは次のように説明される。採用した類似度関数SMがいずれも直交条件を満たすことが保証されていないため、7.6.3のカテゴリ帰属知識 $\langle \varphi^i, \lambda^i \rangle$ のポテンシャルエネルギー $E(\varphi^i, \lambda^i)$ が認識段階番号tの非増加関数であるとは限らないからである(文献[4]の3定理8.1, 8.3, 8.5を参照)。

今回構築された画像理解システムRECOGNITRONは、2次元平面に投影された物体の濃淡値から得られる形状情報のみを手がかりに認識・解釈を行うが、代表パターンに対する変形をある程度おさえた画像であればかなり高い認識率を得られることが今回のシミュレーションにおいて確認された。本研究では認識対象を自動的に縮小・拡大しマッチングをとる処理や、2次元の投影像からもとの3次元形状を脳内で復元し新しい視点からの像を作り上げる感覚運動的な知能処理も取り入れていないし、遠近・高低・大小・明暗・色・カテゴリ同士の関係(雲は空にある等)などの人間の常識もRECOGNITRONに反映させていない。これらの空間的・知能的・知識工学的操作を扱えるようにするためには、モデル構成作用素、類似度関数、大分類関数等を改良することと、ある程度の時間内で処理できる計算機の処理速度や記憶容量などの向上が求められるが、この様な操作を取り入れることができれば今回の結果よりもさらに優れた認識を期待できることは容易に予想できる。また本研究で提案されている処理方式の特徴のひとつに、処理画像を限定しない一般性を挙げることができるが、これらの処理を取り入れることができれば、この一般性を更に高めることもできる。

本研究はSS理論を風景画像の理解に適用した計算機シミュレーション研究の初期の段階のものであり、認識率の改善よりもむしろ画像解釈における限界点を模索し、問題点を露呈させることに、主力が注がれた。本シミュレーションの段階では無作為に撮影した変形の大きい自然画像では満足のいく解釈結果は得られなかった。しかし、物体の大きさや向きをある程度規格化した画

像であれば、ある程度の成果を得る事が可能であるといえる。風景画像は一般に顔画像などに比べ遙かに自由度が高く、撮影者と対象物との距離により生ずる大きさの変動や、対象物に向かう撮影者の位置による像の回転や左右反転、青空と曇り空・白い車と黒い車などの同一カテゴリ内での色(明暗)の逆転や、物体同士の重なり合いなどが常に起こり得る。むしろそれらの変形が含まれない風景画を撮影することの方が稀である。従って、これらの変形を如何に処理し、RECOGNITRONに代表パターンのモデルを想起させるかが今後の課題であろう。

無造作に撮影された自然画像を正しく認識させるためには、まずある程度規格化された画像を認識させることができるようになってはじめてその可能性が見えてくる。よって、本シミュレーションでは、ある程度規格化された風景画像の解釈がRECOGNITRONに可能かどうかの検討に重点を置いた。今回のシミュレーション結果を参考とし、改良を重ねなければならない。

尚、本研究で扱った入力風景画像のみであるが、画像理解システムの従来構成法とは異なり、学習に用いる訓練パターンと、あらかじめ用意するカテゴリラベルを変更するだけで、顔画像の理解や、指紋解析、文字認識などに適用することも可能である。今後もパソコン能力の向上にあわせて改良を重ね、より高度な画像理解システムを実現していく努力をしていかねばならない。

文 献

- [1] 鈴木昇一：“認識工学”，柏書房，Feb.1975
- [2] 鈴木昇一：“ニューラルネットの新数理”，近代文芸社，Sept.1996
- [3] 鈴木昇一：“パターン認識の数理的な一般解決”，近代文芸社，Jun.1997
- [4] 鈴木昇一：“認識知能情報論の新展開”，近代文芸社，Aug.1998
- [5] 松山隆司：“画像理解における推論方式”，人工知能学会誌，vol.4, no.1, pp.21-29, Jan.1989
- [6] 鈴木昇一：“風景画像から知識を抽出し、解釈するシステムの、ファジィ推論ニューラルによる構成”，文教大学情報学部「情報研究」，no.23, pp.183-265, 2000
- [7] 鈴木昇一：“SS大分類関数BSCの適応的構成への，計算論的学習理論の適用”，文教大学情報学部「情報研究」，no.25, pp.187-238, 2000
- [8] Sameer Singh, John Haddon, Markos Markou “Nearest-neighbour classifiers in natural scene analysis”，Pattern Recognition, vol.34, pp.1601-1612, 2001
- [9] 彌富仁，萩原将文：“ファジィ推論ニューラルネットワークを用いた風景画像からの知識抽出と認識”，電子情報通信学会論文誌D-II，vol.J82-D-II, no.4, pp.685-693, Apr.1999
- [10] Takatoshi Nishina, Masafumi Hagiwara：“Fuzzy inference neural network”，Neurocomputing, vol.14, pp.223-239, 1997
- [11] Nikhil R.Pal, Sankar K.Pal：“A review on image segmentation techniques”，Pattern Recognition, vol.26, no.9, pp.1277-1294, 1993
- [12] Jia-Ping Wang：“Stochastic Relaxation on Partitions with Connected Components and Its Application to Image Segmentation”，IEEE, Transactions on Pattern Analysis and Machine Intelligence vol.20, no.6, pp.619-636, June 1998
- [13] 和田俊和，佐藤正行，松山隆司：“選択的注視に基づく複数対象の動作認識”，電子情報通信学会論文誌D-II，vol.J82-D-II, no.6, pp.1031-104, June 1999

- [14] Toshikazu Wada, Takashi Matsuyama : “Multiobject behavior recognition by event driven selective attention method”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.22, no.8, pp.873-887, Aug.2000
- [15] 加藤ジェーン, 渡邊豊英, 米田政明 : “HMMに基づく交通監視映像の背景・物体・影の分離手法”, 情報処理学会論文誌, vol.42, no.1, pp.1-14, Jan.2001

(鈴木昇一, 川俣博司, 大槻善樹, 風景画の理解に関するJAVA言語によるRECOGNITRONの計算機シミュレーション, 文教大学・情報学部・情報研究, no.27への投稿論文, 投稿年月日 2002年01月28日)

(すずき しょういち 文教大学情報学部, かわまた ひろし 富士ソフトABC株式会社, おおつき ぜんじゅ 文教大学情報学部非常勤講師)