

である多層パーセプトロン (MLP) と Gaussian GRBF の間の関係を明らかにし、近似能力の比較を行なう。

7.2 種々の学習・近似方式の数値比較

7.2.1 種々の学習・近似方式

本章では、学習を与えられた入出力例からの元の写像を近似、復元することであると見なす立場に立ち議論を進める。いま、 \mathbf{R}^d で定義された関数族 G を考え、 G に属する関数 $g: \mathbf{R}^d \rightarrow \mathbf{R}$ から、入出力データ (例題) $S = \{(x_i, y_i) \in \mathbf{R}^d \times \mathbf{R}\}_{i=1}^N$ が得られたものとする。この場合の学習・近似の目的は、例題集合 S から元の関数 g を復元、近似することである。このような問題に対しては、従来から関数近似論や統計学の分野で数多くの研究がなされ、なかでも parametric regression と呼ばれる手法がよく用いられてきている。これは、データの元となる関数の近似を有限次元のパラメータで決定される関数族 $H(\subset G)$ の中で探索しようとするものである。このとき、関数族 H の要素による g の近似誤差の測定尺度が与えられれば、学習・近似問題の解は、近似誤差を最小化するようなパラメータを探索することにより得ることができる。したがって、parametric regression は以下の2つの要素から構成されているものと見なすことができる。

- 有限次元のパラメータ集合 M から G の部分集合 H への写像 A 。これは、パラメータ p から H の要素であるパラメトリック関数 $h_p(x)$ を決定するものであり、次のように表すことができる。

$$\begin{aligned} A: M &\longrightarrow H \subset G \\ A(p) &= h_p(x), \quad p \in M, \quad h_p \in H \end{aligned} \quad (7.1)$$

- 近似誤差評価の尺度

$$E: G \times G \longrightarrow \mathbf{R} \quad (7.2)$$

いま $g \in G$ が与えられ、(2) より、 M から \mathbf{R} への写像

$$E_g(p) = E(A(p), g) = E(h_p, g) \quad (7.3)$$

が導かれる。この場合、学習・近似の解は、 M 上の最小化問題

$$\min_{p \in M} E_g(p) \quad (7.4)$$

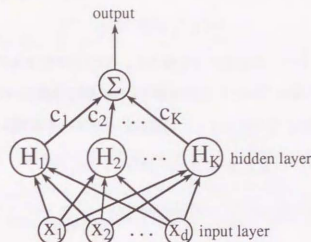


図 7.1: 隠れ層 1 層を持つ 3 層構造ネットワーク

を解くことによって得られる。誤差尺度 E として一般に最もよく用いられるものは、データ S 上での最小自乗誤差

$$E = \sum_{i=1}^N (y_i - h_p(x_i))^2 \quad (7.5)$$

である。この他にも、学習の頑健性に着目した種々の評価の提案や [30]、その理論的定式化 [31] もなされているが、本章では最小自乗誤差尺度を用いて議論を進めるものとする。また、parametric regression を決定する写像 A 、即ちパラメトリック関数 $h_p(x)$ の形式としても数多くのものが考えられるが、本章ではパラメトリックな基底関数の線形結合による以下のような方式を考える。

$$h_p(x) = \sum_{\alpha=1}^K c_{\alpha} H_{\alpha}(x) \quad (7.6)$$

ここに H_{α} は基底関数である。この場合の全体のパラメータ p は $\{c_{\alpha}\}$ 及び H_{α} を決定するパラメータの和によって構成されることになる。式 (7.6) の形式で表される学習・近似方式は、基底関数 $\{H_{\alpha}\}$ から成る隠れ層 1 層を持つ 3 層ネットワークによって実現することができる (図 7.1)。以下では本章で用いる種々の学習・近似手法について簡単に述べる。

- HBF (Hyper Basis Function)

HBF (Hyper Basis Function) は Poggio and Girosi[89] によって提案された学習方式で、パラメトリック関数 $h_p(x)$ は次のように表すことができる。

$$h_p(x) = \sum_{\alpha=1}^K c_{\alpha} G(\|x - t_{\alpha}\|_W^2) \quad (7.7)$$

ここに G は (conditionally) positive definite function であり、 W は $d \times d$ 行列で、重み付き

ノルム $\|\cdot\|_W$ は、

$$\|x\|_W^2 = x^T W^T W x \quad (7.8)$$

で定義される。本章では (7.7) の形式のみを扱うが、(7.7) にはさらに低次多項式を付加してもよい。HBF におけるパラメータは K 個の係数 $\{c_\alpha\}$, K 個の d 次元ベクトル $\{t_\alpha\}$, 及び $d \times d$ 行列 W によって構成される。したがって、この場合、パラメータ p は

$$p \in M = \mathbf{R}^K \times (\mathbf{R}^d)^K \times GL(d) \quad (7.9)$$

となる。本章では Gaussian $G = e^{-r^2}$ と Multiquadrics $G = \sqrt{\alpha^2 + r^2}$ [53] を基底関数として用いる。

Gaussian HBF の場合、 $h_p(x)$ 全体の作る集合 H は、積に関して閉じていることを示すことができる (Appendix C)。

• Ridge 関数による学習・近似方式

非線形基底関数の線形結合による学習・近似方式 (7.6) において、基底関数としては、入力ベクトルとパラメータの内積によって値が決まる ridge 関数がよく用いられる。このとき、パラメトリック関数 $h_p(x)$ は

$$h_p(x) = \sum_{\alpha=1}^K c_\alpha h(w_\alpha \cdot x + \theta_\alpha) \quad (7.10)$$

と表せる。ここに h は基底関数である。この場合、全体のパラメータ p は、 K 個の展開の係数 $\{c_\alpha\}$ 、 K 個の d 次元ベクトル $\{w_\alpha\}$ 、 K 個のバイアスパラメータ $\{\theta_\alpha\}$ によって構成され、したがって、 p は、次のような $K(d+2)$ 次元集合 M 上に存在することになる。

$$M = \mathbf{R}^K \times (\mathbf{R}^d)^K \times \mathbf{R}^K \quad (7.11)$$

ridge 関数全体を特徴付ける性質としては、ridge 方向への微係数が 0 になるという点が挙げられる。したがって、ridge 関数の線形和から成る関数は、適当な有限回の方向微分によって 0 とすることができ、これを ridge 基底による近似方式を特徴付ける性質の 1 つであると見なすことができる。

ここでは、以下のような基底関数を用いて比較計算機実験を行う。

1. 3 角関数による近似 (Flexible Fourier)

$$h_p(x) = \sum_{\alpha=1}^K c_\alpha \cos(w_\alpha \cdot x + \theta_\alpha) \quad (7.12)$$

このとき、 w_α は周波数として解釈できる。但し、ここに、 w_α は通常の Fourier 展開の場合と異なり、任意のベクトルであってよいものとする。

2. Sigmoid 関数による近似

$$h_p(x) = \sum_{\alpha=1}^K c_\alpha \sigma(w_\alpha \cdot x + \theta_\alpha), \quad \sigma(t) = \frac{1}{1 + \exp(-t)} \quad (7.13)$$

基底関数として Sigmoid 関数を用いる学習・近似方式は、隠れ層 1 層を持つ多層 (3 層) パーセプトロン的一种であると解釈できる (図 7.1)。この場合のパラメータ w_α の解釈は (Flexible) Fourier の場合のように明確ではないが、近年盛んに研究が行われている Projection Pursuit (Regression) [27, 41, 103] の観点からすると、 w_α は、近似対象関数の情報、特徴を示す "interesting direction" を表しているとも見なすことができる。

3. 指数関数 (Exponential Sums) による近似

1 次元関数の近似理論においては、指数関数による近似手法が研究されてきている [16]。

多次元においては、その自然な拡張として

$$h_p(x) = \sum_{\alpha=1}^K c_\alpha \exp(w_\alpha \cdot x) \quad (7.14)$$

なる近似形式を考えることができる。いま、バイアスパラメータは無視できるから、パラメータ p は

$$p \in M = \mathbf{R}^K \times (\mathbf{R}^d)^K \quad (7.15)$$

となる。

7.2.2 学習アルゴリズム

本章では "学習=写像の近似" の立場に立ち、学習の目的は例題として与えられた集合 S 上での自乗誤差 (7.5) を最小化するようなパラメータ p を求める。このとき、例題 S は一括して与えられるような off-line 型の学習のみを考える。この場合、学習アルゴリズムは式 (7.5) で与えられる E を最小化する最適化アルゴリズムに他ならない。本論文では、学習アルゴリズムとして、標準的な最適化アルゴリズムの 1 つである Levenberg-Marquardt 法 [92] を主として用いる。但し、最急降下法をはじめとし、Levenberg-Marquardt 法などの勾配法によって得られる解は、その初期状態 (初期パラメータ設定) に依存し、最適解 (global minimum) が得られる保証はない。本論文では、同一の近似実験について、数回の初期パラメータ設定を行ない、得られた結果のうち、最良の結果を採用するものとした。

HBF/GRBF に関しては、その中心点とノルム行列の設定に関して、すべてのパラメータを同時に Levenberg-Marquardt 法によって求める手法と併用して、以下のような制限付きの手法を用いた。

● frozen centers (FRZ)

HBF の最適パラメータを求める際には、すべてのパラメータを同時に探索するのではなく、より重要と見なせるパラメータについてのみ探索を行ない、その他のパラメータに関しては適当な値に初期設定した後は、その値に固定したままにしておくことが考えられる。Poggio らは、HBF においてはノルム行列 W の設定が重要であると考え、中心点位置は、学習データ中から選択された点に固定したままにしておく（固定中心点、frozen centers）手法を提案している [87]。本節においては、すべてのパラメータを同時に探索する手法（movable centers）と共に固定中心点（frozen centers）も用いて実験を行なった。このとき、固定中心点位置の決定手法としては、与えられた学習データ中からランダムに選択する手法と LVQ (Learning Vector Quantization) [56] による手法とを試みた。これらの手法は、固定中心点以外の場合（movable centers）の中心パラメータの設定の際にも適用した。

● ノルム行列への制限 (Diag)

HBF においては、ノルム行列 W の設定が重要であると考えられる。本論文では、 W として一般の行列 (HBF) と共に W に対角行列

$$W = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_d \end{pmatrix}$$

であるという制限を加えた場合の HBF に関しても実験を行なった。

7.2.3 計算機実験結果

上で述べた種々の基底関数に基づく学習・近似方式を用いて、多変数関数の近似実験を行った。実験においては、誤差を含まない例題集合（学習データ） S を用い、学習・近似のための評価関数は式 (7.5) に示したような最小自乗誤差評価

$$E_p(p) = \sum_{i=1}^N (h_p(x_i) - y_i)^2 \quad (7.16)$$

を用いた。近似実験結果を評価するためには、学習データ $S = \{(x_i, y_i)\}_{i=1}^N$ の他に評価データ $S' = \{(x'_i, y'_i)\}_{i=1}^{N'}$ を用い、 S, S' 双方について下に示すような評価式によって近似結果の誤差評

価を行なった。

$$L_2 = \frac{\sum_{i=1}^N (h_p(x_i) - y_i)^2}{\sum_{i=1}^N y_i^2}, \quad L_\infty = \frac{\max_{i=1}^N |h_p(x_i) - y_i|}{\max_{i=1}^N |y_i|}$$

$$L'_2 = \frac{\sum_{j=1}^{N'} (h_p(x'_j) - y'_j)^2}{\sum_{j=1}^{N'} (y'_j)^2}, \quad L'_\infty = \frac{\max_{j=1}^{N'} |h_p(x'_j) - y'_j|}{\max_{j=1}^{N'} |y'_j|}$$

学習の目的は、例題集合（学習データ）を用いて、未知入力に対しても適当な出力を与えるような関数を生成すること（即ち、汎化）にある。上の評価式において、ある方式によって L_2, L_∞ と共に L'_2, L'_∞ も十分小さい値を達成できた場合は、その問題に関して、その方式が十分な汎化能力を示したことを表すものである。逆に、 L_2, L_∞ が十分小さくなったにも関わらず、 L'_2, L'_∞ が大きな値を持つ場合には、汎化を行うことができず、学習データ S への overfitting が起こったことを示すものである。ここで行った学習・近似実験の対象とした関数を表 7.1 にまとめる。実験に用いた学習データ S のサイズ N は、2 次元関数の場合には $N = 100$ 、5 次元の場合には $N = 1000$ を用いた。また、すべての実験において評価用データ S' のサイズは $N' = 100$ とした。なお、データは S, S' とともに、一様乱数によって発生させたものを用いている。また、実験において、最適化アルゴリズムは $L_2 < 10^{-3}$ となった時点で停止させた。

表 7.1: 近似対象関数

Experiment No.	Target Function	Dimensionality
1	$\exp(-0.8x_1 - 1.2x_2) \cos(5x_1) \sin(3x_2)$	2
2	$x_1 x_2$	2
3	$x_1 x_2 \cdots x_5$	5
4	$\exp(-\ x\ ^2) \cos(\frac{3}{4}\pi \sum_{i=1}^2 x_i)$	2
5	$\exp(-\ x\ ^2) \cos(\frac{3}{4}\pi \sum_{i=1}^5 x_i)$	5

以下に実験 1 ～ 実験 5 の結果を示す。

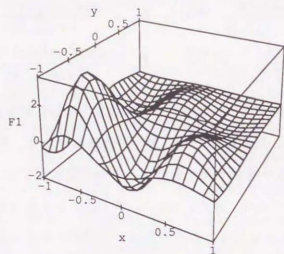


図 7.2: 近似対象関数 (Experiment1)

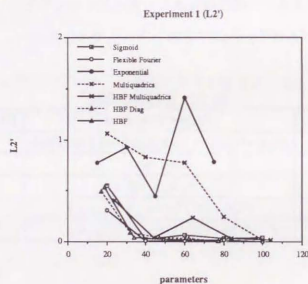


図 7.3: 実験結果 (Experiment1)

● Experiment 1

図 7.2 に近似対象関数、図 7.3 に近似実験結果のうち L_2 の挙動をパラメータ数に関して表示した図を示す。図から明らかなように、用いた方式の中では、指数関数が他方式と比べて悪い結果しか得られていない。その他の方式の中では Flexible Fourier が最も良く、Sigmoid, (Gaussian)HBF, Multiquadrics HBF は、ほぼ同様の結果を得た。凹凸を持つ近似対象関数の形状と、ランダムに与えられた 100 点というデータ数を考えると、これらの結果は Sigmoid, HBFs, Flexible Fourier

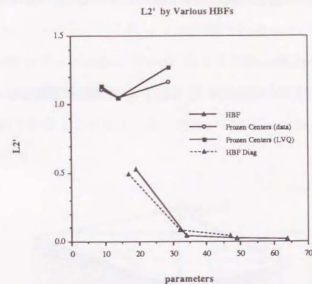


図 7.4: 種々の HBF による実験結果 (Experiment1)

の高い関数近似能力を示すものと言えよう。なお、図 7.3 中の指数関数による L_2 の挙動は、最適解を発見できなかったこと (local minima の存在) を示すものであると考えられる。

図 7.4 には、種々の HBF による近似実験結果を示す。固定中心点 (FRZ: frozen centers) を用いた手法よりも、中心点をパラメータとして最適化を行なった方式の方が良好な近似結果を得た。これは、この近似対象関数に関しては、中心点の適正配置が重要であることを示すものである。例えば、図 7.5 に 10 個の Gaussian (HBF) を用いた場合に得られた中心点のうち、近似対象の定義域 $[-1, 1] \times [-1, 1]$ のものと、近似対象の等高線を示す。比較的少数のユニットを用いた

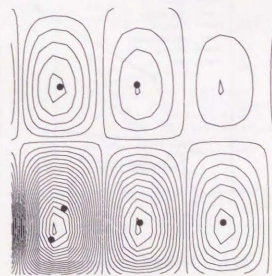


図 7.5: 近似対象の等高線と Gaussian HBF の中心点配置

この場合には、中心点がほぼ近似対象の構造を反映していると思われる位置に移動していることが分る。

● Experiment 2, Experiment 3

図 7.6 に近似対象関数、図 7.7 に実験結果 L_2 を示す。すべての方式について良好な結果が得られたことが分る。

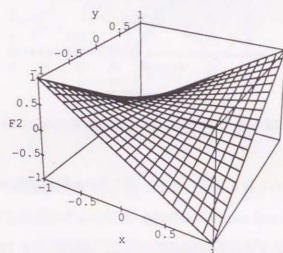


図 7.6: 近似対象関数 (Experiment2)

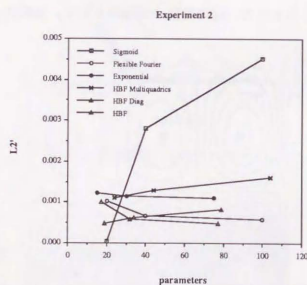


図 7.7: 実験結果 (Experiment2)

Experiment 3 においては、良好な結果が得られた Experiment2 の近似対象関数 $F = x_1 x_2$ を 5 次元に拡張し ($F = x_1 \cdots x_5$, $|x_i| < 1.5$)、近似実験を行なった。実験の結果得られた L_2 及び L_2 の挙動を図 7.8, 7.9 に示す。Flexible Fourier などを除き、ほとんどの場合について使用パラメータ数の増加と共に近似評価結果 (L_2) は良くなっているものの、Sigmoid、指数関数などを除き十分良い近似能力を発揮しているとは言えない。特に Flexible Fourier においては $L_2 \ll L_2$ となる現象が顕著である。

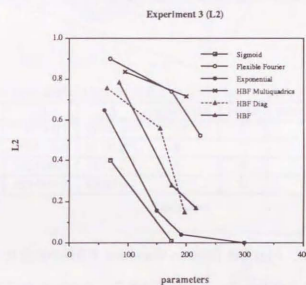


図 7.8: 実験結果 L_2 (Experiment3)

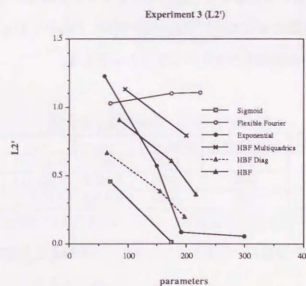


図 7.9: 実験結果 L_2 (Experiment3)

しかしながら、これらの結果は、必ずしも、各方式の近似能力自体の不足を示すものではない。例えば、指数関数、Flexible Fourier、Gaussian HBF は積に関して閉じているから (Appendix C)、各基底によって $F = x_i, F = x_i x_j$ 等を少数のユニット数で近似できれば、これらの積を考えることにより、比較的少数のユニット数 (パラメータ数) で $F = x_1 \cdots x_5$ を近似できる解を合成できることが考えられる。実際、これらの基底を用いて、 $F = x, F = xy$ ($x, y \in [-1.5, 1.5]$) を少数のユニットで近似できることが、数値実験によって確認できる。この実験結果を表 7.2 に示す。また、これらの低次元の解の積によるパラメータを初期条件とした場合の実験結果を表 7.3 に示す。

表 7.2: $F = x_1 \cdots x_d$ ($|x_i| < 1.5, d = 1, 2$) (data size = 1000)

scheme	dimension	units	parameters	L_∞	L_2	L'_∞	L'_2
Flexible Fourier	1	2	6	0.0209	9.0×10^{-5}	0.0183	7.8×10^{-5}
	2	2	8	0.0370	0.000280	0.0295	0.000228
Exponential	1	2	4	0.0247	0.000887	0.02505	0.000770
HBF-Diagonal	1	2	5	0.0265	0.000426	0.0261	0.000430

上の考察により、指数関数、Flexible Fourier、Gaussian HBF の場合については、良好な近似解を合成する能力は有することがわかる。この結果はまた、これらの方式が

$$F = \prod_{i=1}^k \sum_{j=1}^d a_{ij} x_j \quad (7.17)$$

を 2^k 個以下のユニットを用いて近似できることを示すものでもある。したがって、図 7.8, 7.9 の結果は、(この近似対象関数にとっては) Gaussian HBF, Flexible Fourier などについてのより有効な学習アルゴリズムの必要性を示すものであるとも言える。

表 7.3: $F = x_1 \cdots x_5$ ($|x_i| < 1.5$)

scheme	units	parameters	data size	L_∞	L_2	L'_∞	L'_2
Flexible Fourier	32	224	1000	0.0511	0.000881	0.0417	0.00128
Exponential	32	192	1000	0.0678	0.000943	0.0424	0.000773
HBF-Diagonal	32	197	1000	0.0484	0.00132	0.0451	0.00158

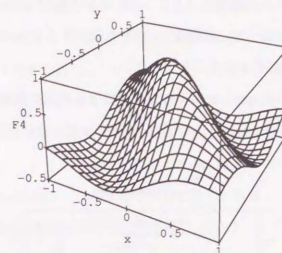


図 7.10: 近似対象関数 (Experiment 4)

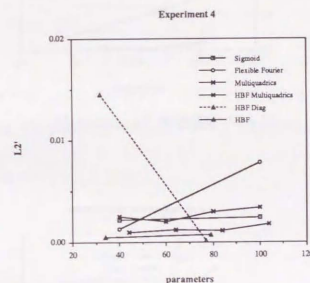


図 7.11: 実験結果 (Experiment 4)

• Experiment 4, Experiment 5

Experiment 4 の近似対象関数の形状を図 7.10 に、指数関数を除く他の方式による近似実験結果を図 7.11 に示す。指数関数による結果は、50 ユニットを用いた場合でも $L_2 = 0.330$, $L'_2 = 0.474$ と良好な結果を得ることはできなかった。

近似対象関数の形状から、HBF の場合については、制限なしのノルム行列 W を用いる場合が近似に適していることが想像される。実験の結果も、図 7.11 に示す通り、Gaussian HBF に

いて、ノルム行列に対する制限がない場合の方が、ユニット数が少ない場合には特に明らかに良好な近似結果を得ている。これらの結果により、HBFにおける W の重要性がわかる。特に、対角行列などを使用した場合と比べて、制限なしの W を使用することにより、少ないパラメータ数で同等の近似性能を示す可能性があると考えられる。

Experiment 5 においては、Experiment 4 の対象関数を 5 次元に拡張し、近似実験を行なった。図 7.12 に実験結果を示す。Experiment 4 同様、HBF が最も良い近似結果を得た。

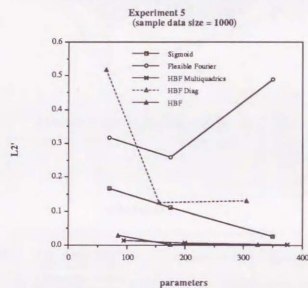


図 7.12: 実験結果 (Experiment5)

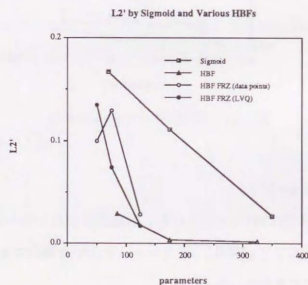


図 7.13: Sigmoid と Gaussian HBF(Frozen Centers) の比較 (Experiment5)

図 7.13 には固定中心点を用いた種々の HBF による近似結果を示す。図 7.13 より、固定中心点を用いた場合でも、 W を最適化することにより、Sigmoid などと比べ、1/3 程度のパラメータ数で、同程度の近似性能を示せたことが分る。これらは、HBF におけるパラメータ W の重要性を示すものであると考えられる。

● Flexible Fourier への Structure Removal アルゴリズムの適用

Flexible Fourier に関しては、図 7.8, 7.9 に見られるように、十分な近似能力を持つ場合でも、単純な最適化アルゴリズムだけでは不十分な結果しか得られていない。これらは、データの次元の増大に伴うデータ密度の減少が 1 つの原因であると考えられる。

与えられたデータからの関数近似問題において、近似すべきデータの次元が高くなると、データの密度が減少し、これに起因する種々の難点が生じることが良く知られている (curse of dimensionality)。これに対する 1 つの方策として、データを適当な低次元空間に射影して、この低次元空間上で近似を行うことが考えられる。PPR (Projection Pursuit Regression) [41] は、このための有力な手法の 1 つであると考えられる。PPR における関数の近似は、ridge 関数の線形結合を用い、

$$g \sim \sum_{j=1}^m g_j(a_j \cdot x) \quad (\|a_j\| = 1) \quad (7.18)$$

なる形式で行われる。ここに、 g_j はノンパラメトリックに決定される 1 次元関数である。通常、PPR においては、次のような structure removal アルゴリズムが用いられる。

1. 第 (k-1) 項までが得られていると仮定し、

$$r_i = y_i - \sum_{j=1}^{k-1} g_j(a_j \cdot x_i)$$

とおく。

2. 第 k 項を

$$\sum_{i=1}^N (r_i - g_k(a_k \cdot x_i))^2$$

を最小化するような方向ベクトル a_k 及びノンパラメトリック関数 g_k を探索することによって決定する。

PPR においてノンパラメトリックに決定される関数 g_j を式 (7.10) における ridge 基底に対応させると、PPR は、その特別な場合として本章で取り上げた近似方式を含むことが分る。

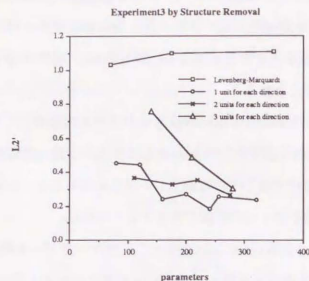


図 7.14: Structure Removal による実験結果 (Flexible Fourier, Experiment3)

そこで、ここでは、PPR においてはノンパラメトリックに決定される関数 g_j を少数のパラメトリックな基底関数で

$$g_j(\mathbf{a}_j \cdot \mathbf{x}) = \sum_{m=1}^{M(j)} c_{jm} h(\mu_{jm} \mathbf{a}_j \cdot \mathbf{x} + \theta_{jm}) \quad (7.19)$$

と近似することを考え、上の structure removal アルゴリズムを適用した。なお、この方式の適用に当たっては、各方向への基底数 $M(k)$ は一定であるとした。また、射影を行う方向 \mathbf{a}_j の総数もあらかじめ決定し、structure removal アルゴリズムを収束するまで繰り返し適用した。なお、structure removal アルゴリズムにおける最適化手法としては、Levenberg - Marquardt 法を用いたが、初期パラメータの設定に際しては Genetic Algorithm [33] も援用した。

この structure removal アルゴリズムを Flexible Fourier に適用した場合の、Experiment 3, 5 に関する結果を図 7.14, 7.15 に示す。Experiment 3 に対する結果は未だ十分満足できるものとは言えないものの、Levenberg-Marquardt のみの場合に比べ近似結果は大幅に改善されていることが分かる。また、Experiment 5 に関しては Structure Removal アルゴリズムでかなり良好な近似結果が得られた。実験で用いた Structure Removal アルゴリズムはまだ多くの改善の余地はあるものの、これらの結果は Ridge 基底を用いた場合の、このアルゴリズムの有効性を示すものであると思われる。なお、指数関数に関しては、Experiment 5 に関して Structure Removal アルゴリズムを適用した場合でも良好な結果は得られなかった。

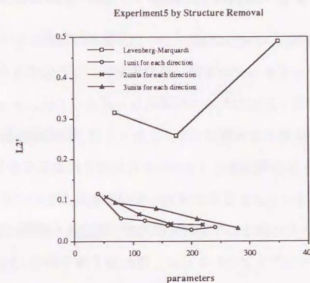


図 7.15: Structure Removal による実験結果 (Flexible Fourier, Experiment5)

7.2.4 考察

本節においては、“学習=写像の近似”の立場から、種々の学習・近似方式について、その近似能力を数値的に比較することを試みた。ここで行なった比較実験はごく初歩的なものであり、一般的な性質について考察を加える際には注意を要することは言うまでもないものの、今回比較を行なった基底の中では、指数関数の近似能力が、他と比較して劣っていると考えられる。実験を通して、他より常に優れた近似性能を持つ方式 (clear winner) は見い出されなかった。ridge 基底に基づく近似方式に関しては、Sigmoid が全体を通して、良好な近似能力を発揮したもの、Flexible Fourier も、アルゴリズムの工夫などにより、同等の近似性能を発揮し得ると考えられる。また、HBF についても同様に良好な近似能力を示す実験結果を得た。Lippmann らは、種々のパターン識別問題に関して多層パーセプトロンや RBF の能力を比較し、問題により必要とされるユニット数や学習時間に大きな差が見られるものの、識別能力自体はどの方式もほぼ同等であることを結論づけている [57, 58, 76]。本節における結果は、関数近似問題においても同様のことが成立することを示すものであると言える。したがって、実際の問題適用する学習手法を選択する際には、近似能力のみならず、初期状態の設定のしやすさ、見積られる必要ユニット数や必要学習時間を考慮することが重要であると考えられる。

7.3 正規化入力に対する GRBF と多層パーセプトロンの能力比較

“学習=関数(写像)近似”という立場に立つと、種々の学習ネットワーク・学習方式は、例題として与えられた疎データから元の写像(多変数関数)を近似するための方式であると見なすことができる。この、学習・近似のための代表的な方式としては、ネットワーク構成要素(基底)としてシグモイド関数を用いる多層パーセプトロン(MLP=Multilayer Perceptron)や、正則化理論(Regularization)から導かれ、Gaussianに代表されるようなRBF(Radial Basis Function)を用いるGRBFネットワーク、さらにその一般化であるHyper BFネットワークなどが提案されている[86, 87]。これらの異なる学習方式の比較、関連性の解明は重要な課題である。

いま、MLP(多層パーセプトロン)として、隠れ層1層を持ち、出力層に線形出力素子を持つような3層パーセプトロンを考えると、MLP、GRBFとも(非線形)基底関数の線形結合に基づく学習方式であると見なすことができる。この場合、MLPはシグモイド関数を、GRBFはRBF(Radial Basis Function)を基底として用いていることになる。このように、MLP、GRBFは形式的に同様の構造を持ち、理論的にも、また、前節で示した通り実験的[13, 68]にも、両者とも良好な近似能力を有している。しかしながら、両者の比較、関係の解明[22]は十分になされていない。

MLP、GRBFの両者の明らかな相違点は、MLPで基底として用いられているシグモイド関数が入力ベクトルと荷重パラメータの内積で値が決まるridge関数であるのに対し、GRBFで用いられるRBFは、入力と“中心点”パラメータとの距離によって値が決まるradial関数であるという点である。本節では、入力ベクトル \mathbf{x} が正規化条件 $\|\mathbf{x}\| = 1$ を満足する場合を想定し、正規化入力の下でradial関数を基底とする学習方式とridge関数を基底とする学習方式を比較する。特に、正規化入力の場合の、シグモイドを基底として用いるMLPとGaussianを基底として用いるGaussian GRBFを比較し、両者の関連を明らかにする。これにより、MLPは同数のユニットから成るGaussian GRBFを十分良く近似できることを示すと共に、MLPとGaussian GRBFを同一視できるための条件を与える。さらに、Multiscale Gaussian GRBFを用いた場合には、同様なパラメータ数を持つMLPを十分良く近似し得ることを計算機実験により明らかにする。

7.3.1 正規化入力に対する ridge 関数と radial 関数の比較

学習・近似方式として、非線形基底関数の線形結合に基づく方式

$$f(\mathbf{x}) = \sum_{\alpha=1}^K c_{\alpha} H_{\alpha}(\mathbf{x}) \quad (7.20)$$

を考える。ここに H_{α} は基底関数である。式(7.20)に示された学習方式は、隠れ層1層を持ち、線形出力ユニットを持つ3層構造のネットワークとして表現できることが分かる(図7.1)。式(7.20)の学習方式における基底関数 H_{α} としてはさまざまなものが考えられるが、代表的な基底関数としては、radial関数とridge関数の2つが挙げられる。

radial関数は、一般に

$$H_{\alpha}(\mathbf{x}) = h(\|\mathbf{x} - \mathbf{t}_{\alpha}\|_W^2), \quad \mathbf{x} \in \mathbf{R}^d, \mathbf{t}_{\alpha} \in \mathbf{R}^d \quad (7.21)$$

なる形式で表される。ここに、 $\{\mathbf{t}_{\alpha}\}$ は“中心点”、 W は $d \times d$ 行列であり、 $\|\mathbf{x}\|_W$ は重み付きノルム

$$\|\mathbf{x}\|_W^2 = \mathbf{x}^T W^T W \mathbf{x} \quad (7.22)$$

を表す。GRBFの場合には、 h は(conditionally) positive definiteであり、 W は単位行列に固定されたものが用いられる[87]。 h の例としては、Gaussian, (Hardy's) Multiquadrics[53]などが挙げられる。これらの関数はRBF(Radial Basis Function)と呼ばれる。Gaussian GRBF等は、正則化理論(Regularization Theory)から導くことができ、確固たる理論的背景を有している。

一方、ridge関数は、

$$H_{\alpha}(\mathbf{x}) = h_{\alpha}(\mathbf{w}_{\alpha} \cdot \mathbf{x} + \theta_{\alpha}), \quad \mathbf{x} \in \mathbf{R}^d, \mathbf{w}_{\alpha} \in \mathbf{R}^d, \theta_{\alpha} \in \mathbf{R} \quad (7.23)$$

なる形式で表される。ここに \mathbf{w}_{α} は荷重ベクトルであり、 θ_{α} はバイアスパラメータである。この学習・近似方式の例としては、次のような手法が含まれる。

- MLP(3層パーセプトロン)

$$H_{\alpha}(\mathbf{x}) = \sigma(\mathbf{w}_{\alpha} \cdot \mathbf{x} + \theta_{\alpha}), \quad \sigma(t) = \frac{1}{1 + \exp(-t)}$$

- Flexible Fourier Series $H_{\alpha}(\mathbf{x}) = \cos(\mathbf{w}_{\alpha} \cdot \mathbf{x} + \theta_{\alpha})$

- Exponential Sums $H_{\alpha}(\mathbf{x}) = \exp(\mathbf{w}_{\alpha} \cdot \mathbf{x})$

radial 基底と ridge 基底の明らかな違いは、各関数の値が、パラメータと入力ベクトルとの距離によって決まるか、あるいは、パラメータと入力ベクトルの内積によって決まるかという点である。ここでは、2種類の基底の比較を行ない、関連を調べるために、入力ベクトルとして、正規化入力、すなわち $\|x\| = 1$ を満たすものを考える。正規化されていない一般の入力集合についても、入力の次元を1つ増やし、例えば、次のような変換を施すことにより、正規化入力として取り扱うことができる。

$$\begin{aligned} x &= (x_1, \dots, x_d)^T \in \mathbf{R}^d, |x_i| \leq \ell \\ x &\rightarrow x' = (x'_1, \dots, x'_{d+1})^T \in S^d \subset \mathbf{R}^{d+1} \\ x'_i &= \frac{x_i}{\ell\sqrt{d}} \quad (i = 1, \dots, d) \quad x'_{d+1} = \sqrt{1 - \sum_{i=1}^d x_i'^2} \end{aligned} \quad (7.24)$$

ここに S^d は d 次元単位球面を表す。以下では、正規化入力を仮定して議論を進めるものとする。

入力が正規化されている場合には、

$$\|x - t\|^2 = \|x\|^2 + \|t\|^2 - 2x \cdot t = 1 + \|t\|^2 - 2x \cdot t \quad (7.25)$$

が成立する。この関係式を用いて、正規化入力に対する一般の ridge 関数と radial 関数の関係について考察する。いま、radial 関数を用いた学習・近似方式として、

$$f(x) = \sum_{\alpha=1}^K c_{\alpha} h(\|x - t_{\alpha}\|^2) \quad (7.26)$$

を考えると、式 (7.25) より、次の関係式を容易に導くことができる。

$$f(x) = \sum_{\alpha=1}^K c_{\alpha} h(\|x - t_{\alpha}\|^2) = \sum_{\alpha=1}^K c_{\alpha} \tilde{h}(t_{\alpha} \cdot x + \theta_{\alpha}) \quad (7.27)$$

ここに、

$$\tilde{h}(x) = h(-2x), \quad \theta_{\alpha} = -\frac{1}{2}(1 + \|t_{\alpha}\|^2) \quad (7.28)$$

である。上式は、正規化入力の下では、radial 関数は常に ridge 関数と見なせること、但し、この場合、バイアスパラメータ θ_{α} は任意に値を設定することができず、 $\theta_{\alpha} \in (-\infty, -\frac{1}{2}]$ なる拘束を受けることを示している。この場合、radial 関数における“中心点”は ridge 関数における結合荷重に相当する。上の関係は、正規化入力の下では、radial 関数 $h(\|x - t_{\alpha}\|^2)$ を用いた3層構

造のネットワーク (図 7.1) は、常に同数のユニットを用いた ridge 関数 $\tilde{h}(t_{\alpha} \cdot x + \theta_{\alpha})$ を用いたネットワークで実現できることを示すものである。

次いで、正規化入力の下での ridge 関数による学習方式 (7.23) を考えると、式 (7.25) より次の関係が成立する。

$$\begin{aligned} f &= \sum_{\alpha=1}^K c_{\alpha} h(w_{\alpha} \cdot x + \theta_{\alpha}) \\ &= \sum_{\alpha=1}^K c_{\alpha} h(-\lambda \|x - \frac{w_{\alpha}}{2\lambda}\|^2 + d_{\alpha}) \quad (\lambda \in \mathbf{R}, \lambda \neq 0) \\ d_{\alpha} &= \theta_{\alpha} + \lambda(1 + \frac{\|w_{\alpha}\|^2}{4\lambda^2}) \end{aligned} \quad (7.29)$$

ここに λ は任意の実数パラメータである。したがって、正規化入力の下で、一般に ridge 関数による学習方式 (7.23) は、次の条件のいずれかが成立した場合に radial 関数による学習方式 (7.26) と同一視できることが分かる。

$$\theta_{\alpha} = -\lambda(1 + \frac{\|w_{\alpha}\|^2}{4\lambda^2}) \quad (7.30)$$

$$h(-\lambda x + y) = g_1(x)g_2(y) \quad (7.31)$$

いずれの場合も、ridge 関数の荷重ベクトルは、radial 関数の中心点に対応する。(7.30) の場合、解が存在するのは $\{\theta_{\alpha}\} (\alpha = 1, \dots, K)$ の符号がすべて等しい場合に限られることが分かる。

一方、(7.31) が成立する場合、学習方式は

$$f(x) = \sum_{\alpha=1}^K c'_{\alpha} g_1(\|x - \frac{w_{\alpha}}{2\lambda}\|^2) \quad (7.32)$$

となる。ここに、

$$c'_{\alpha} = c_{\alpha} g_2(\theta_{\alpha} + \lambda(1 + \frac{\|w_{\alpha}\|^2}{4\lambda^2})) \quad (7.33)$$

である。(7.31) の場合については、 $x = 0, y = 0$ を代入して、

$$g_1(x) = \frac{h(-\lambda x)}{g_2(0)}, \quad g_2(y) = \frac{h(y)}{g_1(0)} \quad (7.34)$$

となるから、結局、 h は、

$$h(x+y) = \frac{h(x)h(y)}{h(0)} \quad (7.35)$$

を満たす。これを満たす関数族は $h = ce^x$ である。

以下では、これらの考察に基づき、特に、radial 関数として Gaussian を用いる Gaussian GRBF と、ridge 関数としてシグモイド関数を用いる3層パーセプトロンを正規化入力の下で比較する。

7.3.2 正規化入力に対するシグモイド関数と Gaussian

ここでは、正規化入力の下で、radial 関数として Gaussian、ridge 関数としてシグモイド関数を考え、これらを基底とする Gaussian GRBF

$$f(\mathbf{x}) = \sum_{\alpha=1}^K c_{\alpha} \exp\{-\mu^2 \|\mathbf{x} - \mathbf{t}_{\alpha}\|^2\}, \mathbf{t}_{\alpha} \in \mathbf{R}^d, \mu \in \mathbf{R} \quad (7.36)$$

及び、3層パーセプトロン（出力素子は、線形であるものとする）

$$f(\mathbf{x}) = \sum_{\alpha=1}^K c_{\alpha} \sigma(\mathbf{w}_{\alpha} \cdot \mathbf{x} + \theta_{\alpha}), \sigma(t) = \frac{1}{1 + \exp(-t)} \quad (7.37)$$

を考える。これらは、いずれも、図 7.1 に示すような、隠れ層 1 つを持つ 3 層構造のネットワークで実現することができる。このとき、入力ベクトル $\mathbf{x} \in \mathbf{R}^d$ は、 $\|\mathbf{x}\| = 1$ を満たすから、パラメータ $\lambda \in \mathbf{R}$ を導入して、

$$\mathbf{w} \cdot \mathbf{x} + \theta = -\lambda \|\mathbf{x} - \frac{\mathbf{w}}{2\lambda}\|^2 + \lambda(1 + \frac{\|\mathbf{w}\|^2}{4\lambda^2}) + \theta \quad (7.38)$$

が成立する。したがって、

$$\begin{aligned} \sigma(\mathbf{w} \cdot \mathbf{x} + \theta) &= \frac{s}{1 + \exp\{-(-\mathbf{w} \cdot \mathbf{x} + \theta)\}} \\ &= \frac{s}{1 + C(\lambda) \exp\{\lambda \|\mathbf{x} - \mathbf{t}\|^2\}} \end{aligned} \quad (7.39)$$

ここに、

$$\mathbf{t} = \frac{\mathbf{w}}{2\lambda}, C(\lambda) = \exp\{-(\theta + \lambda(1 + \frac{\|\mathbf{w}\|^2}{4\lambda^2}))\} \quad (7.40)$$

である。これは、シグモイド関数は正規化入力に関して radial 関数 (7.39) と等価であることを表している。また、 λ は任意の値を持つパラメータであり、 λ の存在は、どんな λ をとっても、式 (7.39) で表される radial 関数は、正規化入力に対して等価であることを示している。ここでは、正規化入力に対する Gaussian と、式 (7.39) で表されるようなシグモイド関数に対応する radial 関数を比較するために、式 (7.39) で与えられる 1 パラメータ関数族のうちで、Gaussian に最も近いと思われるものを選択し、比較を行なう。式 (7.39) より明らかなように、この radial 関数は、

$$C(\lambda) \gg 1 \quad (7.41)$$

が成立する場合には、Gaussian の良い近似を与えることになる。ここでは、式 (7.39) の radial 関数と Gaussian の近さの尺度として $C(\lambda)$ をとり、 λ で与えられる 1 パラメータ関数族の中で、 $C(\lambda)$ を最大にするような $\lambda^* > 0$ を持つものを考える。

$$C(\lambda^*) = \max_{\lambda > 0} C(\lambda) \quad (7.42)$$

λ^* は、 $\partial C / \partial \lambda = 0$ を解いて、

$$\lambda^* = \frac{\|\mathbf{w}\|}{2} \quad (7.43)$$

となることがわかる。この値を式 (7.39) に代入すると、中心点を単位球面上に持つような、次の radial 関数が得られる。

$$\begin{aligned} \sigma(\mathbf{w} \cdot \mathbf{x} + \theta) &= \frac{s}{1 + C(\theta, \|\mathbf{w}\|) \exp\{\frac{\|\mathbf{w}\|}{2} \|\mathbf{x} - \frac{\mathbf{w}}{\|\mathbf{w}\|}\|^2\}} \\ C(\theta, \|\mathbf{w}\|) &= \exp\{-(\theta + \|\mathbf{w}\|)\} \end{aligned} \quad (7.44)$$

図 7.16 に示すように、上の radial 関数は、 $C \gg 1$ の場合、Gaussian にきわめて近い形状となる。条件 $C \gg 1$ は、バイアスパラメータに関する条件式

$$\theta \ll -\|\mathbf{w}\| \quad (7.45)$$

と等価であることは明らかである。この関係は、正規化入力の下では、シグモイド関数はバイアスパラメータ θ を制御することにより、Gaussian を十分良く近似できることを示している。このとき、1 つのシグモイド関数のパラメータ数は $d+2$ となる。

一方、正規化入力 $\|\mathbf{x}\| = 1$ に対する Gaussian の出力を考えると、次のような関係が成立する。

$$\text{cexp}\{-\lambda^2 \|\mathbf{x} - \mathbf{t}\|^2\} = c'(\lambda) \exp\{-\mu'(\lambda)^2 \|\mathbf{x} - \mathbf{t}'(\lambda)\|^2\} \quad (7.46)$$

ここに、パラメータ $\lambda \in \mathbf{R}$ は任意の正実数であり、

$$c'(\lambda) = \text{cexp}\{-\mu^2(1-\lambda)(1 - \frac{\|\mathbf{t}\|^2}{\lambda})\}, \mu'^2(\lambda) = \lambda\mu^2, \mathbf{t}'(\lambda) = \frac{\mathbf{t}}{\lambda} \quad (7.47)$$

である。上の式は、Gaussian の表現 $\text{cexp}\{-\mu^2 \|\mathbf{x} - \mathbf{t}\|^2\}$ は正規化入力の下では冗長であること示している。例えば、任意の Gaussian $\text{cexp}\{-\mu^2 \|\mathbf{x} - \mathbf{t}\|^2\}$ に対して、以下のような Gaussian も、それと等価であることを示すことができる。

$$\begin{aligned} \text{cexp}\{-\mu^2 \|\mathbf{x} - \mathbf{t}\|^2\} &= c' e^{-\mu'^2 \|\mathbf{x} - \frac{\mathbf{t}}{\lambda}\|^2} (\lambda = \|\mathbf{t}\|) \\ &= c'' e^{-\|\mathbf{x} - \mu'^2 \mathbf{t}\|^2} (\lambda = \frac{1}{\mu'^2}) \end{aligned} \quad (7.48)$$

したがって、正規化入力に対する Gaussian 基底としては、例えば、

$$\text{cexp}\{-\mu^2 \|\mathbf{x} - \mathbf{t}\|^2\}, \|\mathbf{t}\| = 1 \quad (7.49)$$

を用いればよく、1 つの Gaussian ユニットのパラメータ数は $d+1$ である。式 (7.44) 及び (7.49) を比較すると、正規化入力の下では、3 層パーセプトロン (MLP) は条件 (7.45) を満足するよ

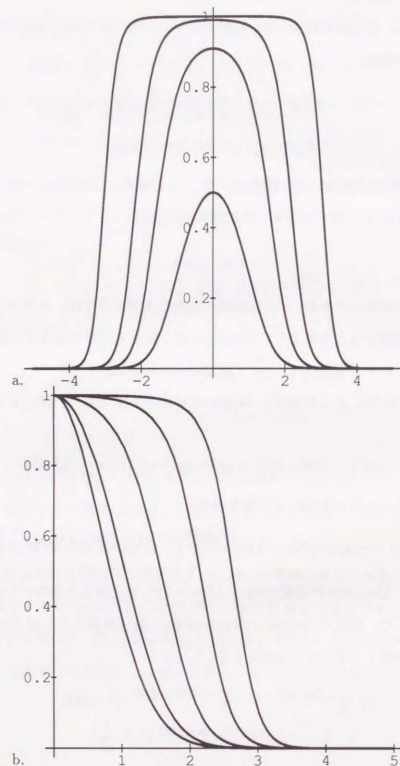


図 7.16: シグモイド関数に対応する radial 関数と Gaussian : (a) 種々の C に関する radial 関数 $f(r) = 1/(1+Ce^{r^2})$ (内側より $C = 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$)、(b) Gaussian $f(r) = e^{-r^2}$ と種々の C に関する radial 関数 $f(r) = (1+C)/(1+Ce^{r^2})$ (内側より Gaussian, $C = 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$)

らにバイアスパラメータを調整することにより、同数のユニットを持つどんな Gaussian GRBF も十分良く近似できることが分かる。但し、その場合、MLP のパラメータ総数は、ユニット数を K とすると $K(d+2)$ であるのに対し、Gaussian GRBF のパラメータ総数は $K(d+1)$ である。

7.3.3 計算機実験結果

前節において、正規化入力の下で、MLP は同数のユニットから成る任意の Gaussian GRBF を近似できることを明らかにした。これは、MLP のバイアスパラメータが条件 (7.45) を満たす場合に成立し、この場合には MLP と Gaussian GRBF を同一視できる。そこで、ここでは、実際の正規化入力の場合の学習・近似において条件 (7.45) が成立しているかを確認するために、MLP による簡単な関数近似の計算機実験を行なった。実験においては、次のような関数を学習・近似対象として用いた。

$$F(x, y) = F(x), |x| \leq 1, y = \sqrt{1-x^2} \quad (7.50)$$

$$\begin{aligned} & \bullet F_1 = x & \bullet F_2 = e^{-x^2} \cos\left(\frac{3}{4}\pi x\right) & \bullet F_3 = e^{-x^2} \cos\left(\frac{3}{2}\pi x\right) \\ & \bullet F_4 = \frac{\sin 3\pi x}{3\pi x} & \bullet F_5 = e^{-x} \sin 5x \end{aligned}$$

実験においては (7.37) の方式を用い、学習データは一括して与えられるものとして、評価関数

$$L = \sum_{i=1}^N \{F(x_i) - f(x_i)\}^2 \quad (7.51)$$

の最小化を行なった。また、得られた実験結果は、以下のような尺度を用いて評価した。

$$\begin{aligned} L_2 &= \frac{\sum_{i=1}^N \{F(x_i) - f(x_i)\}^2}{\sum_{i=1}^N F(x_i)^2}, & L_\infty &= \frac{\max_{i=1}^N |F(x_i) - f(x_i)|}{\max_{i=1}^N |F(x_i)|} \\ L'_2 &= \frac{\sum_{p=1}^M \{F(x_p) - f(x_p)\}^2}{\sum_{p=1}^M F(x_p)^2}, & L'_\infty &= \frac{\max_{p=1}^M |F(x_p) - f(x_p)|}{\max_{p=1}^M |F(x_p)|} \end{aligned}$$

ここに、 $\{x_i\}_{i=1}^N$, $\{x_p\}_{p=1}^M$ は、それぞれ、一様乱数を用いて定められた学習データ、評価用データであり、 $N = 20, M = 100$ である。なお、実験においては学習 (最適化) アルゴリズムとして gradient-descent 法的一种である Levenberg-Marquardt 法を用いた [92]

表 7.4 に実験結果、表 7.5 にパラメータ ($C_0, \|w_0\|$) の値を示す。これらは、条件 (7.45) が常に成立する訳ではないことを示すものである。したがって、MLP は実際上も同数ユニットの Gaussian GRBF と同一視できる訳ではないことが分かる。

表 7.4: MLP による関数近似結果

Function	Units	L_∞	L_2	L'_∞	L'_2
F_1	2	0.00632	2.0×10^{-5}	0.00853	2.2×10^{-5}
F_2	3	0.00786	2.6×10^{-5}	0.00895	5.3×10^{-5}
F_3	3	0.0106	5.4×10^{-5}	0.0236	7.6×10^{-5}
F_4	3	0.0130	0.000130	0.0153	0.000312
F_5	3	0.00874	9.9×10^{-5}	0.0579	0.000719

表 7.5: 学習により得られた $C(\theta_\alpha, \|\mathbf{w}_\alpha\|, \|\mathbf{w}_\alpha\|)$ の値

	$C(\theta_\alpha, \ \mathbf{w}_\alpha\ , \ \mathbf{w}_\alpha\)$	
F_1	(0.0893, 2.03)	(0.0509, 2.14)
F_2	(7.67, 6.30)	(0.829, 0.492)
F_3	(11.4, 10.5)	(0.00371, 0.504)
F_4	(7.48, 8.01)	(1.60, 8.63)
F_5	(1.25, 2.37)	(0.0271, 7.70)
		(1.5×10^{-4} , 4.17)
		(1.2×10^{-5} , 7.64)

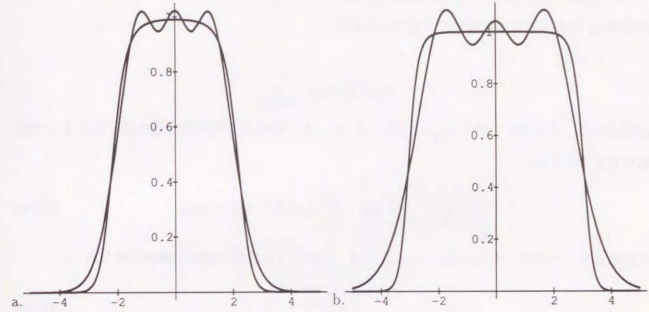
7.3.4 Multiscale-GRBF による 3 層パーセプトロンの近似

これまでの議論により、正規化入力の下では、もしユニット数が等しい場合には、3 層パーセプトロンはどんな Gaussian GRBF ネットワークも十分良く近似できることが示された。これは、シグモイド関数と Gaussian のユニット同士の 1 対 1 の比較に基づくものである。しかしながら、この場合には、パラメータ数が MLP の方が多くなり、パラメータ数が等しい場合のネットワーク全体としての能力比較は、上の議論では明らかではない。そこで、ここでは、複数の Gaussian を用いて、式 (7.44) の radial 関数を近似することを考える。その際、パラメータ数をできるだけ少なくするために、式 (7.44) を同一の中心点を持つ Gaussian を用いて次のように近似することを考える。

$$\frac{s}{1 + Ce^{\theta r^2}} \sim \sum_{j=1}^m a_j e^{-b_j r^2} \quad (b_j > 0) \quad (7.52)$$

図 7.17 に、 $C = 0.01, 0.0001$ の場合について 3 個の Gaussian で式 (7.44) の radial 関数を近似した計算機実験結果を示す。図において $\frac{1}{1+0.01 \exp\{r^2\}}$ は

$$g_1 = 8.146 \exp\{-0.5072r^2\} - 7.831 \exp\{-0.6836r^2\} + 0.7069 \exp\{-2.531r^2\}$$

図 7.17: $\frac{1}{1+0.01 \exp\{r^2\}}$ の 3 Gaussians による近似: (a) $1/(1+0.01 \exp\{r^2\})$ の近似 (b) $1/(1+0.0001 \exp\{r^2\})$ の近似

で、また $\frac{1}{1+0.0001 \exp\{r^2\}}$ は

$$g_2 = 5.094 \exp\{-0.2422r^2\} - 5.670 \exp\{-0.4547r^2\} + 1.617 \exp\{-1.048r^2\}$$

で近似されている。この実験結果は、複数の Gaussian で、式 (7.44) の radial 関数をかなり良く近似できる可能性があることを示すものである。しかしながら、式 (7.52) の近似方式を用いると、シグモイド関数のパラメータ数が $d+2$ であるのに対し、必要とされる Gaussian の全パラメータ数は、 $d+2m-1$ になってしまうという欠点がある。以下では、シグモイド関数を、それと同程度のパラメータ数の Gaussian で近似する可能性について考察する。

同程度のパラメータ数で MLP を近似するために、次のような複数次定スケール $\{B_j\}_{j=1}^m$ を持つ基底関数

$$H_\alpha = \sum_{j=1}^m a_{\alpha j} e^{-B_j \|\mathbf{x} - \mathbf{t}_\alpha\|^2}, \quad (\|\mathbf{t}_\alpha\| = 1) \quad (7.53)$$

に基づく Multiscale GRBF を考え、これにより、シグモイド関数に対応する radial 関数 (7.44) を近似することを試みる。多重スケールの radial 基底による学習方式は正則化理論に基づいて導くことができることが明らかにされている [87]。

このとき、上の Multiscale Gaussian 基底 H_0 のパラメータ数は、 $d+m-1$ である。いま、中心点 t_0 および入力 x は正規化されているから、

$$0 \leq \|x - t_0\| \leq 2$$

が成立する。そこで、ここでは、パラメータ $\{a_j\}$ は、次のような評価関数の最小化によって得られるものとする。

$$L = \int_0^2 \left\{ \frac{1}{1 + Ce^{\frac{B_j}{2}r^2}} - \sum_{j=1}^m a_j e^{-B_j r^2} \right\}^2 dr \rightarrow \min \quad (7.54)$$

これにより、パラメータ $a = (a_1, \dots, a_m)^T$ は、次のように線形方程式を解いて得られる。

$$Ua = v \quad (7.55)$$

$$U_{ij} = \int_0^2 e^{-(B_i + B_j)r^2} dr, \quad v_i = \int_0^2 \frac{e^{-B_i r^2}}{1 + Ce^{\frac{B_i}{2}r^2}} dr$$

図 7.18 に $m = 3, 4, 5$ (即ち、パラメータ数は $d+2, d+3, d+4$) の場合の近似実験結果を示す。

図 7.18 において、近似誤差 ϵ は、

$$\epsilon = L / \int_0^2 \frac{dr}{(1 + Ce^{\frac{B}{2}r^2})^2} \quad (7.56)$$

を用いて評価している。これにより、Multiscale Gaussian GRBF は MLP をかなり良く近似し得ることが分かる。なお、実験において、スケールパラメータ $\{B_i\}_{i=1}^m$ は図 7.17 の実験の結果得られたものを使用している。この他、固定スケール比を持つ Multiscale Gaussian などによる MLP の近似も考えられる [69]。

7.3.5 考察

本節では、MLP と Gaussian GRBF の能力比較を、入力ベクトルが正規化条件 $\|x\| = 1$ を満足する場合について行ってきた。正規化入力に関しては、従来から MLP の有効性を示すものとして広く知られている研究の中には、正規化入力と見なすことができる入力表現を用いているものが存在する。例えば NETtalk[96] がこの一例である。また、2 値表現も正規化入力の一種と考えることができる (Appendix D)。

正規化入力の条件が成立している場合には、ridge 関数を radial 関数として扱うことができる。この場合、ridge 関数における荷重パラメータは radial 関数における中心点パラメータに相当することになる。特に、シグモイド関数の場合にはバイアスパラメータを調整することにより、

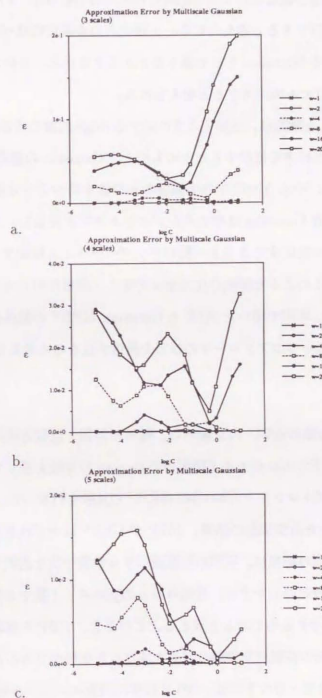


図 7.18: Multiscale Gaussian によるシグモイドの近似結果: (a) 3 scales (b) 4 scales (c) 5 scales

これを Gaussian と見なすことができる。このとき、ユニット数が十分多く与えられていれば、中心点を与えられた学習データと一致するように配置するという、RBF で用いられる手法 [87] を MLP にも適用することができる。また、Gaussian を用いる学習ネットワークについては、その局所性を生かした高速学習法がいくつか提案されている [74, 82]。また LVQ[56] など中心点パラメータの設定に利用できると考えられる。正規化入力の場合にはバイアスパラメータの調整によりシグモイド関数を Gaussian として扱うことができるから、これらの手法は正規化入力に関する MLP の初期設計にも利用できると考えられる。

さらに、本節で展開した議論は、正規化入力に対する ridge 関数の荷重パラメータを radial 関数の中心点と見なす解釈の仕方を提供することにもなる。Gaussian の場合、いま、学習の結果、分散パラメータが小さく、中心点の周りの小領域のみで値を持つような関数が生成されたとする。この場合、このような Gaussian は中心点をプロトタイプと見なし、プロトタイプに近い入力が与えられたときのみに反応するような一種のテンプレートとも解釈することができる [90]。正規化入力の場合には、このような解釈の仕方をシグモイド関数に対しても適用してみることが可能になる。これにより、本章で導いた MLP と Gaussian GRBF の関係は、学習の結果得られた MLP を解釈するための一つのアプローチの仕方を提供するものと考えられる。

7.4 結言

本章では、“学習=写像の近似”の立場から、種々の学習・近似方式の比較を数値的に行なうと共に、正規化入力の下でのシグモイド関数と Gaussian の比較を通して、代表的学習方式である MLP (3層パーセプトロン) と Gaussian GRBF の比較を行なった。

各種の基底関数を用いた近似実験の結果、MLP や GRBF はいずれも良好な近似能力を持つことを明らかにした。本章の結果は、実際の問題適用する学習手法を選択する際には、近似能力のみならず、初期状態の設定のしやすさ、見積られる必要ユニット数や必要学習時間を考慮することが重要であることを示すものであると言える。このとき、RBF を基底として用いる学習方式の持つ、基底関数の意味の直観的な理解の容易さ等は大きな利点であると考えられる。

本章では、また、正規化入力の下では、MLP は常に同数ユニットから成る任意の Gaussian GRBF を近似できることを明らかにすると共に、MLP と Gaussian GRBF を同一視できるための条件を与えた。次いで、MLP と Gaussian GRBF をそのパラメータ数に基づいて比較するという観点から Multiscale Gaussian GRBF と MLP の比較を行ない、Multiscale Gaussian GRBF は同様なパラメータ数で MLP をかなり良く近似し得ることを計算機実験により示した。本章で

導いた MLP と Gaussian GRBF の関係は、また、正規化入力に関する MLP の初期設定やネットワークの解釈に対する1つのアプローチの仕方を与えるものと考えられる。

8 例題からの学習に基づく3次元変形可能物体の認識

8.1 緒言

3次元物体の認識を行なうには、通常、モデルに基づく手法が用いられる。物体を記述するモデルとしては、3次元モデルがまず考えられるが[11, 17]、一般に未知物体の3次元モデルを自動構築するのは非常に困難である。これに対して、3次元モデルの代りに、複数視点からの物体の見え方(view)に基づく2次元モデルを用いる方法[100]、さらにモデル構築の困難さを回避するために、2次元画像からの3次元物体の認識手法を学習によって獲得する方法が提案されている[18, 23, 91]。Poggioらは対応付けの済んだ特徴点を入力とし、Gaussianを基本要素とする学習ネットワークGRBF (Generalized Radial Basis Function) [87]を用いることにより、物体毎に、任意視点からの像を標準像(典型的な視点からの像)に変換する画像を学習できることを示した[91]。その他、GRBFと種々の手法の比較[23]及び、実際の画像に対してこの方式を適用する試みについても発表されている[18, 129]。

Poggioらの認識手法は、“未知姿勢の既知剛体物体の認識”であると考えられる。したがって、この手法では複数の物体を認識するためには個々の物体毎に学習ネットワークが必要になってしまう。また、1つの物体であっても、関節を含むような変形する物体の認識は考慮していない。

例題からの学習は、画像の近似・補間であると見なすことができる。Poggioらの結果は、少数の例題-像-を補間することにより、任意視点の像から標準像への画像を近似できることを示したものであると言える。このとき、学習方式が十分な近似・補間能力を有していれば、彼らの手法をさらに拡張して、物体の形状があるルールに従って変形する(例えば蝶の羽根が閉じたり開いたりする)ような場合にも、少数例題からの補間によって、物体の構造に関する知識なしに、任意視点からの変形物体の像をその形状の標準像に変換する画像が1つのネットワークで獲得できると考えられる。

本章ではGRBFを用いて変形する物体、すなわち単一の形状のみならず同一クラスに属す形状全体を例題からの学習に基づいて認識するための手法について述べる。変形可能な物体としては、関節部分を持つ3次元構造物(butterfly, dragonfly)を考え、まず、透視射影の下で、これらの変形する物体の任意視点からの像を標準像に変換する画像がGRBFを用いた学習によって少数の例題から近似できることを示す。さらに、像の拡大・縮小を含むような場合でも、このような画像が近似できることを示す。次いで、物体の標準像への変換画像のみならず、変形可能物体について、任意視点からの像を一定角度変換させたときの像に変換する視点変換画像も同様に近

似的に学習できることを示す。

次に、物体の像からその物体の種類を識別する問題を考える。変形する物体の場合は、剛体物体の場合とは異なり、照合すべきモデルとして、すべての形状について画像の望ましい出力(標準像など)が学習データとして与えられるわけではなく、形状によってはモデルが未知である。したがってこの場合は、画像の出力像から物体を単純に識別することはできない。本章ではこの問題に対して、物体を識別する関数(characteristic function)を、標準像へ変換する画像と同時に学習し、この関数値によって識別する手法、及び、多視点を利用した手法として、視点変換画像を学習し、これを用いることにより認識を行なう手法について述べる。

さらに、本章では、Ullman and Basriによって提案された、物体の複数視点からの像を用いて3次元物体を認識する手法[100]を変形可能物体等の場合にも拡張することを考える。物体が剛体であり、その複数視点からの像が与えられているときには、これらをモデルとして用いて3次元物体の認識ができることが知られている[100]が、これは変形可能物体等の場合のように入力に未知形状が含まれるような場合には適用することはできない。このような場合の手法として、対象物体の像が与えられたとき、この物体と同じクラスに属す既知物体の例から、与えられた未知対象物体の複数視点からの像を推定し、これらをモデルとして用いて認識を行なうことが考えられる。ここでは、標準像への変換画像、及び、視点変換画像を用いて、入力像に対して複数視点の像を推定し、変形可能物体(未知形状)の3次元物体認識を行なう手法について述べる。

さらに、画像とモデルの間で特徴点の対応が未知の場合の識別手法について述べる。画像とモデルとの可能な対応の組み合わせの中から正しい対応を選択するためには、各対応の評価関数を与える必要がある。本章では学習によって得られた識別関数と、入力像と標準像の2枚の画像間に成立すべき拘束条件による整合性の評価を組み合わせる手法を考え、これにより対応付けが未知の場合でも識別が可能であることを計算機実験によって示す。

8.2 物体の標準像の学習

8.2.1 変形する物体の標準像の学習

Poggioらは、隠れがないようなワイヤフレーム状の物体を対象とし、物体の任意視点からの2次元投影像を標準投影像に変換する画像を学習によって近似できることを示した。このとき物体の表現として、各頂点

$$\{X_i\} (i = 1, \dots, N), X_i = (X_i, Y_i, Z_i) \in R^3$$

の座標値を並べたベクトル表現

$$\tilde{X} = (X_1, \dots, X_N) \in \mathbb{R}^{3N} \quad (8.1)$$

を用い、各視点の像の表現としては、各頂点の2次元座標を並べたベクトル

$$\tilde{x} = (x_1, \dots, x_N), x_i = (x_i, y_i) \in \mathbb{R}^2, \tilde{x} \in \mathbb{R}^{2N} \quad (8.2)$$

を用いている。以下では、これと同様の物体表現、像の表現を用いるものとし、3次元空間から画像面上への射影は、基本的には透視射影を考えるものとする。なお、以下では2次元像の表現式(8.2)において、重心と原点とが一致するように像の平行移動が行なわれている、即ち、2次元像の表現において

$$\sum_{i=1}^N x_i = \sum_{i=1}^N y_i = 0 \quad (8.3)$$

が成立するものとする。

いま、物体が3次元空間中で受ける変換として線形変換を考え、これを 3×3 行列 L で表す。

L により物体は

$$\tilde{X}' = (LX_1, \dots, LX_N) \quad (8.4)$$

に変換される。視点の変更、すなわち物体の回転等はこのような線形変換 L により表現できる。

ここで、物体の任意視点からの像を入力として、その物体の典型的視点からの像(標準像)へ変換する写像 T を考えると、剛体の場合 T は、

$$T: \tilde{x}(\phi) \mapsto \tilde{x}_s = \tilde{x}(\phi_s) \quad (8.5)$$

と表わすことができる。ここに、 ϕ は物体の姿勢を表すパラメータ、 ϕ_s は標準像の姿勢を表すパラメータである。

物体が剛体で3次元空間から2次元画面への射影が直交射影である場合には、標準像への変換を実現する写像は例題から構成できることが明らかにされている[100][付録E]。すなわち、直交射影の場合、剛体物体の任意の像 \tilde{x} は複数個の視点からの像 \tilde{x}_i ($i = 1, \dots, k$)の線形結合で、

$$\tilde{x} = \sum_{i=1}^k a_i \tilde{x}_i \quad (8.6)$$

と表せるから、式(8.5)に示す写像 T としては $\tilde{L}\tilde{x}_i = 0$ ($i = 1, \dots, k$)を満足する行列 \tilde{L} と標準像 \tilde{x}_s を用いて、

$$T[\tilde{x}] = \tilde{x}_s - \tilde{L}\tilde{x} \quad (8.7)$$

とすればよい。Poggioらは直交射影だけでなく透視射影の場合でも、この標準像への変換がGRBFを用いた学習によって獲得できることを示した[23, 91]。

ここでは、Poggioらのこの手法を変形する物体の認識へ拡張して用いることを考える。いま、対象物体として剛体を拡張し、形状がパラメータ α で表現されるような変形する物体を考える。このような変形する物体としては、例えば、関節を持つようなワイヤフレーム状物体が考えられる。この場合、2次元画像上での特徴ベクトルは変形パラメータ(関節角度) α を用いて $\tilde{x}(\alpha; \phi)$ と表せる。剛体の場合の変換を表す式(8.5)を拡張して、変形する物体の標準像への変換を

$$T: \tilde{x}(\alpha; \phi) \mapsto \tilde{x}_s(\alpha) = \tilde{x}(\alpha; \phi_s) \quad (8.8)$$

で定義する。このとき、十分な汎化能力を有する学習方式を用いれば、少数の例題から未知形状パラメータ(関節角) α 、未知姿勢 ϕ のパターンに関しても標準像へ変換する写像が、例題からの補間によって近似的に実現できると考えられる。これは変形する物体の一つのクラスに属する物体群と見なした、物体クラスの認識とも捉えることができる。

例題からの学習の方式としては、多層パーセプトロン等さまざまな方法が考えられるが、本章では、Gaussianを基本要素とするGRBF[87]を用いる。いま、学習すべき写像の入力を n 次元実数、出力を m 次元実数とすると、Gaussian GRBFによる学習方式は次式で表される。

$$f(x) = \sum_{k=1}^K c_k G_k(x) = \sum_{k=1}^K c_k \exp\{-\|x - t_k\|_W^2\} c_k \in \mathbb{R}^m, t_k \in \mathbb{R}^n \quad (8.9)$$

ここに、 x は入力ベクトル、 $c_k \in \mathbb{R}^m$ は係数ベクトル、 $t_k \in \mathbb{R}^n$ はGaussianの中心点、 W は重み付きノルムを表す $n \times n$ 行列であり、

$$\|x\|_W^2 = x^T W^T W x \quad (8.10)$$

である。 c_k, t_k, W は学習によって決定されるパラメータである。

いま、学習データとして与えられる入出力例を $\{(x_i, y_i)\}$ ($i = 1, \dots, N_0$), $x_i \in \mathbb{R}^n, y_i \in \mathbb{R}^m$ とすると、学習では、評価関数 E (残差平方和)

$$E = \sum_{i=1}^{N_0} \|f(x_i) - y_i\|^2 \quad (8.11)$$

を最小化するようにパラメータ c_k, t_k, W の最適化を行なうものとする。このとき、最適化アルゴリズムとしてはLevenberg - Marquardt法[92]を用いる。

8.2.2 Linear Deformation の場合の標準像への変換写像

ここでは、簡単な1パラメータ形状変形の例として、形状が2種類の物体A,Bの間で直線的に変形する場合を考える。いま、物体A,BのN個の特徴点を並べた物体表現を $\tilde{x}_A, \tilde{x}_B \in \mathbb{R}^{2N}$ とおく。このとき、変形パラメータ α を持つ物体の像 $\tilde{x}(\alpha)$ は、

$$\tilde{x}(\alpha) = (1 - \alpha)\tilde{x}_A + \alpha\tilde{x}_B \quad (0 \leq \alpha \leq 1) \quad (8.12)$$

と書ける。Ullman and Basriの結果[100],[付録E]より、 \tilde{x}_A, \tilde{x}_B は、物体A,Bを同一方向から見た複数個の像 $\{\tilde{p}_i\}, \{\tilde{q}_i\}$ を用いて、次のように表せる。

$$\tilde{x}_A = \sum_{i=1}^k \kappa_i \tilde{p}_i, \quad \tilde{x}_B = \sum_{i=1}^k \kappa_i \tilde{q}_i \quad (8.13)$$

但し、 $\{\tilde{p}_i\}, \{\tilde{q}_i\}$ は互いに独立であるとする。これを式(8.12)に代入して、

$$\begin{aligned} \tilde{x}(\alpha) &= (1 - \alpha)\tilde{x}_A + \alpha\tilde{x}_B \\ &= \sum_{i=1}^k \kappa_i \tilde{p}_i + \alpha \sum_{i=1}^k \kappa_i (\tilde{q}_i - \tilde{p}_i) \\ &= \sum_{i=1}^k \kappa_i \tilde{p}_i + \alpha \sum_{i=1}^k \kappa_i \tilde{u}_i \end{aligned} \quad (8.14)$$

を得る。この変形可能物体の標準像 $\tilde{x}_s(\alpha)$ を

$$\tilde{x}_s(\alpha) = (1 - \alpha)\tilde{p}_0 + \alpha\tilde{q}_0 = \tilde{p}_0 + \alpha\tilde{u}_0 \quad (8.15)$$

で定義する。いま、 $\{\tilde{p}_i\}, \{\tilde{q}_i\}$ と直交するベクトル $\{\tilde{v}_i\}$ ($i = 1, \dots, 2N - 2k$) を考え、線形変換Aとして以下の関係式を満足するものとする。

$$\begin{cases} A\tilde{p}_i = \tilde{p}_0 & (i = 1, \dots, k) \\ A\tilde{u}_i = \tilde{u}_0 & (i = 1, \dots, k) \\ A\tilde{v}_i = \tilde{v}_i & (i = 1, \dots, 2N - 2k) \end{cases} \quad (8.16)$$

このとき、式(8.14)(8.15)より、

$$\begin{aligned} A\tilde{x}(\alpha) &= \sum_{i=1}^k \kappa_i A\tilde{p}_i + \alpha \sum_{i=1}^k \kappa_i A\tilde{u}_i \\ &= \left(\sum_{i=1}^k \kappa_i \right) \tilde{p}_0 + \alpha \left(\sum_{i=1}^k \kappa_i \right) \tilde{u}_0 \\ &= \left(\sum_{i=1}^k \kappa_i \right) \tilde{x}_s(\alpha) \end{aligned} \quad (8.17)$$

となる。標準像への変換を得るために、さらに、長さ1の適当なベクトル \tilde{s}_0 をとり、次式で定義される線形変換 A_P, A_Q を考える。

$$\begin{cases} A_P \tilde{p}_i = \tilde{s}_0 & (i = 1, \dots, k) \quad \|\tilde{s}_0\| = 1 \\ A_P \tilde{q}_i = \mathbf{0} & (i = 1, \dots, k) \\ A_P \tilde{v}_i = \tilde{v}_i & (i = 1, \dots, 2N - 2k) \end{cases} \quad (8.18)$$

$$\begin{cases} A_Q \tilde{p}_i = \mathbf{0} & (i = 1, \dots, k) \\ A_Q \tilde{u}_i = \tilde{s}_0 & (i = 1, \dots, k) \\ A_Q \tilde{v}_i = \tilde{v}_i & (i = 1, \dots, 2N - 2k) \end{cases} \quad (8.19)$$

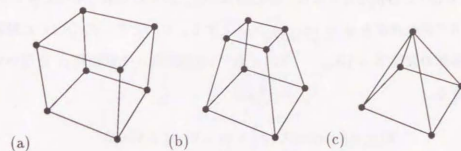


図 8.1: linear deformation の場合の変形物体の例 “cube to pyramid”: 変形パラメータ (a) $\alpha = 0.0$ (b) $\alpha = 0.5$ (c) $\alpha = 1.0$

このとき、式(8.14)より、

$$A_P \tilde{x}(\alpha) = (1 - \alpha) \left(\sum_{i=1}^k \kappa_i \right) \tilde{s}_0, \quad A_Q \tilde{x}(\alpha) = \alpha \left(\sum_{i=1}^k \kappa_i \right) \tilde{s}_0 \quad (8.20)$$

となり、

$$\|A_P \tilde{x}(\alpha) + A_Q \tilde{x}(\alpha)\| = \left| \sum_{i=1}^k \kappa_i \right| \quad (8.21)$$

が成立する。したがって、式(8.17)より、

$$T\tilde{x}(\alpha) = \frac{A\tilde{x}(\alpha)}{\|A_P \tilde{x}(\alpha) + A_Q \tilde{x}(\alpha)\|} \quad (8.22)$$

により、一つの標準像への変換写像が得られる。この写像は明らかに非線形写像であり、これにより、変形可能物体の場合、標準像への変換写像を得るためには、GRBFのような非線形性を有する学習・近似方式が必要とされることがわかる。

8.2.3 計算機実験

GRBF による変形物体の標準像への変換写像の学習能力を検証するために計算機実験を行なった。

実験では主として図8.2に示すように6個の頂点 $\{X_i\}$ ($i = 1, \dots, 6$) から成る物体を用いた。

この物体 butterfly は、 $\overline{X_2 X_5}$ を軸として、関節によって“羽根”を上下して形状が変化する。関

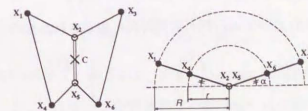


図 8.2: 対象物体 butterfly

節の回転角を α で表す。各頂点はさらに、中心軸 $\overline{X_2 X_5}$ 上の中点に取った中心点 C の回りで回転する。このときの回転角度を $\phi = (\phi_X, \phi_Y, \phi_Z)$ とする。このとき、式 (8.1) に対応する 3 次元空間における物体の表現 $\tilde{X} = (X_1, \dots, X_6) \in \mathbb{R}^{6 \times 3}$ は関節角 α と回転角 ϕ を用いて次のように表すことができる。

$$\tilde{X}(\alpha, \phi) = \cos \alpha \tilde{U}(\phi) + \sin \alpha \tilde{V}(\phi) + \tilde{W}(\phi) \quad (8.23)$$

したがって、画面上への射影として直交射影を考えると、物体の像は

$$\tilde{x}(\alpha; \phi) = \cos \alpha \tilde{u}(\phi) + \sin \alpha \tilde{v}(\phi) + \tilde{w}(\phi) \quad (8.24)$$

と表すことができる。明らかに、これは式 (8.14) で表される linear deformation の場合の像とは異なっている。しかしながら、関節角 α の取り得る範囲が十分小さい場合には、近似的に *butterfly* を linear deformation の場合の変形可能物体として扱うことができる。実際、いま、 \tilde{x}_A, \tilde{x}_B を式 (8.24) において関節角がそれぞれ θ_0 と $\theta_0 + \Delta$ の形状の像であるとする。このとき次式が成立する。

$$\tilde{x}_A = \cos \theta_0 \tilde{u} + \sin \theta_0 \tilde{v} + \tilde{w}, \quad \tilde{x}_B = \cos(\theta_0 + \Delta) \tilde{u} + \sin(\theta_0 + \Delta) \tilde{v} + \tilde{w} \quad (8.25)$$

関節角の変位 Δ が十分小さければ

$$\cos(\theta_0 + \Delta) \approx \cos \theta_0 - \Delta \sin \theta_0, \quad \sin(\theta_0 + \Delta) \approx \sin \theta_0 + \Delta \cos \theta_0 \quad (8.26)$$

が成立するから、

$$\begin{aligned} (1 - \xi) \tilde{x}_A + \xi \tilde{x}_B &\approx (\cos \theta_0 - \xi \Delta \sin \theta_0) \tilde{u} + (\sin \theta_0 + \xi \Delta \cos \theta_0) \tilde{v} + \tilde{w} \\ &\approx \tilde{x}(\theta_0 + \xi \Delta, \phi) \end{aligned} \quad (8.27)$$

を得る。これは十分小さな関節範囲であれば、linear deformation で *butterfly* のような関節物体の変形を近似できることを示すものである。関節範囲が大きくなった場合も、いくつかの、形状に固有な標準像への変換写像や linear deformation の場合の変換写像を接続・補間することにより、適当な変換写像が学習によって得られることが考えられる。

以下の実験では、式 (8.10) のノルム行列が対角行列に限定されたもの

$$W = \begin{pmatrix} w_{11} & & \\ & \ddots & \\ & & w_{2N} \end{pmatrix} \quad (8.28)$$

を用い、物体の回転角度 ϕ は、 $-30^\circ \leq \phi_X, \phi_Y, \phi_Z \leq 30^\circ$ とした。このときの、学習 (GRBF) ネットワークの構成を図 8.3 に示す。

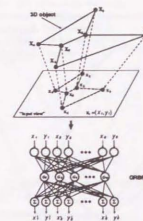


図 8.3: 3 次元物体認識のための学習ネットワーク

いま、画面上に投影された対象物体 *butterfly* の像は $\tilde{x}(\alpha; \phi) \in \mathbb{R}^{2 \times 6}$ と表すことができる。実験においては透視射影を用いた。学習データとして、入出力例の組 $S = \{(\tilde{x}(\alpha_i; \phi_{ij}), \tilde{x}_s(\alpha_i))\}$ を与え、学習データに含まれない未知関節角度 α に関しても、標準像へ変換する写像が GRBF を用いた学習によって得られるか実験を行なった。この場合、未知関節角度に対する変換は、学習データとして与えられる関節角の場合の変換を補間することによって得られると考えられる。実験は、次の 3 つの場合の関節角による学習データについて行なった。

$$(a) \alpha = 30^\circ, (b) \alpha_i \in \{0^\circ, 60^\circ\}, (c) \alpha_i \in \{0^\circ, 30^\circ, 60^\circ\}$$

それぞれの場合、各関節角 α_i について回転角 ϕ_{ij} を乱数によって同数ずつ、学習データ総数が 3 つの場合について等しくなるように与えた。ここでは、学習データ総数は 60 とし、GRBF のユニット数は 10 とした。学習後の写像は、各関節角ごとに回転角 100 例を乱数によって発生させ、次のような誤差関数でその性能を評価した。

$$\text{Error} = E(\alpha) = \frac{1}{M} \sum_{k=1}^M \frac{\|f(\tilde{x}_k(\alpha; \phi_k)) - \tilde{x}_s(\alpha)\|}{\|\tilde{x}_s(\alpha)\|}, (M = 100) \quad (8.29)$$

実験結果を図 8.4, 図 8.5 に示す。図 8.4 は、学習入力と異なる関節角度 α の像を与えたときの出力誤差である。このとき、(a) は一つの剛体物体の標準像への変換を学習する Poggio らの実験 [91] に対応することになる。関節角度が学習した入力 ($\alpha = 30^\circ$) からずれるにつれて、出力像と関節角 α の標準像との誤差は大きくなる。(b), (c) のように複数の角度の形状を学習するこ

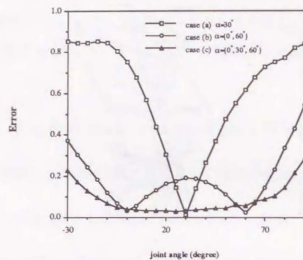


図 8.4: 標準像への変換誤差の評価 (butterfly)

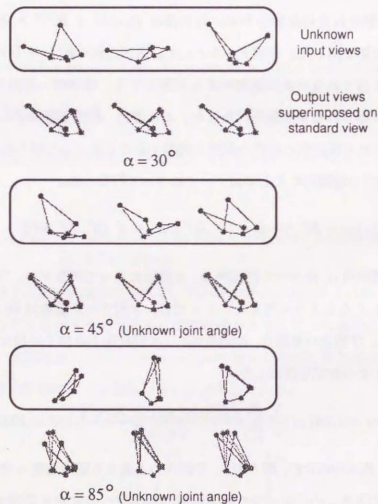


図 8.5: 標準像への変換例

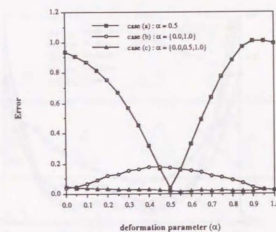


図 8.6: 標準像への変換誤差の評価 (cube to pyramid)

とにより、学習データの間の α に対して、補間によって標準像への変換写像を近似できることがわかる。図 8.4(c) のネットワークに対して、学習入力と異なる回転角の入力像を与えたときの出力像を図 8.5 に示す。図において、四角形内は学習していない回転角の入力像、その下はそれらの入力像に対するネットワークの出力像と標準像を重ね書きしたものである。学習データとして与えられた形状 ($\alpha = 30^\circ$) のみならず、未知形状 ($\alpha = 45^\circ$) に対しても、精度良く標準像への変換ができることがわかる。

同様の計算機実験を図 8.1 に示す変形可能物体 ("cube to pyramid") に対しても行なった。実験は次の 3 つの場合の変形パラメータによる学習データについて行なった。

$$(a) \alpha = 0.5, (b) \alpha_i \in \{0.0, 1.0\}, (c) \alpha_i \in \{0.0, 0.5, 1.0\}$$

それぞれの場合について、butterfly の場合と同様に回転角を決定して学習データを生成し、8 個の Gaussian から成る GRBF により学習を行なった。学習によって得られた写像を式 (8.29) で評価した結果を図 8.6 に示す。butterfly の場合と同様、補間により標準像への変換写像が近似できることが分かる。

8.2.4 拡大・縮小に関して不変な写像の学習

任意視点の像から標準像への変換は、姿勢の変化に関して不変であることが要求される。剛体かつ直交射影の場合は、明らかに式 (8.7) に示す変換はこの性質を満たす。変形可能形状の場合でも、例題として与えられた入出力例を補間することによって、拡大・縮小に関して不変な写像が近似的に実現できると考えられる。画面上に投影された像がさらに拡大・縮小された場合で

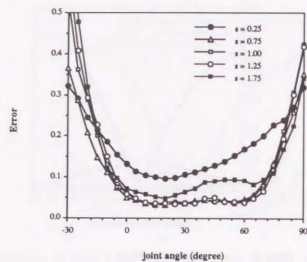


図 8.7: 拡大・縮小を含む場合の標準像への変換誤差評価

も、標準像に変換する写像が得られるかどうかの検証を行なった。この場合、拡大・縮小のパラメータを s とすると、各視点からの像は

$$\tilde{x}(\alpha; s, \phi) = s\tilde{x}(\alpha; \phi) \quad (8.30)$$

となる。

拡大・縮小を含む場合の標準像への変換写像の学習可能性を *butterfly* (図 8.2) を用いて計算機実験によって検証した。

実験において学習データは、関節角度 $\alpha_i \in \{0^\circ, 30^\circ, 60^\circ\}$ 、拡大率 $s_{ij} \in \{0.5, 1.0, 1.5\}$ とし、 $\{\alpha_i, s_{ij}\}$ それぞれについて回転角 ϕ を乱数により 20 例与えた。学習データ総数は 180、GRBF のユニット数は 15 とした。図 8.7 に得られた写像を

$$E(\alpha) = \frac{1}{M} \sum_{k=1}^M \frac{\|f(\tilde{x}(\alpha; s_k, \phi_k)) - \tilde{x}_s(\alpha)\|}{\|\tilde{x}_s(\alpha)\|}, \quad (M = 100)$$

によって評価した結果を示し、図 8.8 に出力例を示す。

拡大・縮小を含む場合も、図 8.7 に示すように、学習した $s = 0.5, 1.0, 1.5$ の間である $s = 0.75, 1.25$ に対する出力誤差が小さいことから、学習データの補間によって、像の拡大・縮小に関しても不変な写像が近似的に得られることがわかる。学習入力と異なる回転角の人力像を与えたときの出力像を図 8.8 に示す。図において、四角形内は学習していない回転角で拡大率 $s = 0.5, 0.75, 1.0, 1.25, 1.5$ の入力像、その下はそれらの入力像に対するネットワークの出力像と標準像を重ね書きしたものである。学習データとして与えられていない、拡大率 $s = 0.75, 1.25$ 、および関節角度 $\alpha = 45^\circ$ についても望ましい標準像が得られたことがわかる。

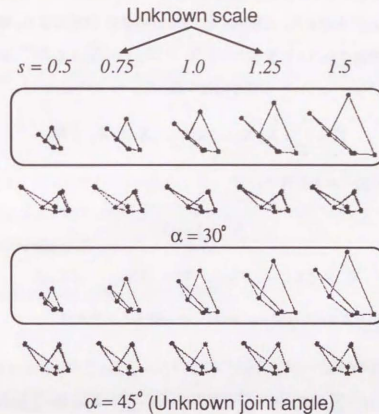


図 8.8: 拡大・縮小を含む場合の標準像への変換例

8.3 視点変換写像の学習

前節では入力像に対して、その物体の標準像を推定する写像の学習について述べた。人間は、与えられた物体の認識ができれば、このような標準像の推定のみならず、その物体を視点在一定角度変換させたときの像をも推定できると考えられる。ここでは、このような視点変換写像について考え、物体が剛体かつ像が直交射影によって生成される場合には、このような視点変換写像を、与えられた例題から構成できることを示す。さらに、これを変形可能物体、即ち、ある形状のクラス全体に関する視点変換写像に拡張して考え、このような写像の学習可能性を計算機実験によって検証する。

8.3.1 剛体、直交射影の場合の視点変換写像の獲得

物体が剛体かつ 3 次元空間から画面への射影が直交射影の場合について考える。このとき、任意視点からの物体の像を \tilde{x} 、それに対してある一定の変換 (回転、視点の一定角度の変化) を施した後の像を \tilde{x}' とおく。この視点変化の予測を表す写像

$$\tau: \tilde{x} \mapsto \tilde{x}' \quad (8.31)$$

を学習によって獲得することを考える。視点変化を表す3次元空間中での変換として、線形変換(行列) $L = [\ell_1 \ \ell_2 \ \ell_3]^T$ を考える。Ullman and Basriの結果[100]により、剛体かつ直交射影の場合には、3次元座標値に基づく特徴ベクトル $\tilde{X} = (X_1, \dots, X_N) \in \mathbb{R}^{3N}$ はいくつかの視点からの特徴ベクトル \tilde{X}_i ($i = 1, \dots, k$) の線形結合によって

$$\tilde{X} = \sum_{i=1}^k a_i \tilde{X}_i, \tilde{X}_i = (X_{i1}, \dots, X_{iN}), X_{ij} \in \mathbb{R}^3 \quad (8.32)$$

と書ける。変換後の特徴ベクトル \tilde{X}' は

$$\tilde{X}' = \sum_{i=1}^k a_i \tilde{X}'_i \quad (8.33)$$

$$\tilde{X}' = (LX_1, \dots, LX_N), \tilde{X}'_i = (LX_{i1}, \dots, LX_{iN})$$

となる。いま直交射影を考えているから、画面上の特徴ベクトル \tilde{x} は

$$\begin{aligned} \tilde{x} &= (X_1, Y_1, \dots, X_N, Y_N), \\ \tilde{x}' &= (\ell_1 \cdot \sum a_i X_{i1}, \ell_2 \cdot \sum a_i X_{i1}, \dots, \ell_1 \cdot \sum a_i X_{iN}, \ell_2 \cdot \sum a_i X_{iN}), \\ \tilde{x}'_i &= (\ell_1 \cdot X_{i1}, \ell_2 \cdot X_{i1}, \dots, \ell_1 \cdot X_{iN}, \ell_2 \cdot X_{iN}) \end{aligned} \quad (8.34)$$

となる。したがって、

$$\tilde{x} = \sum_{i=1}^k a_i \tilde{x}_i, \tilde{x}' = \sum_{i=1}^k a_i \tilde{x}'_i \quad (8.35)$$

が得られる。このとき、変換として並進運動も含まれる場合は、

$$\begin{aligned} \tilde{x} &= \sum_{i=1}^k a_i \tilde{x}_i + \tilde{t}, \tilde{t} = (t_x, t_y, t_x, t_y, \dots)^T \\ \tilde{x}' &= \sum_{i=1}^k a_i \tilde{x}'_i + \tilde{t}', \tilde{t}' = (t'_x, t'_y, t'_x, t'_y, \dots)^T \end{aligned} \quad (8.36)$$

となる。いま、重心と原点は一致しているものとする、

$$\sum_{i=1}^N X_i = \sum_{i=1}^N Y_i = N t_x = N t_y = \sum_{i=1}^N X'_i = \sum_{i=1}^N Y'_i = N t'_x = N t'_y = 0 \quad (8.37)$$

となるから、並進運動を含む場合も式(8.35)のみを考えればよいことがわかる。

式(8.35)は

$$\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_k], \tilde{X}' = [\tilde{x}'_1, \dots, \tilde{x}'_k], \mathbf{a} = (a_1, \dots, a_k) \quad (8.38)$$

を用いて

$$\tilde{x} = \tilde{X} \mathbf{a}, \tilde{x}' = \tilde{X}' \mathbf{a} \quad (8.39)$$

と表すことができる。したがって係数 $\mathbf{a} = (a_1, \dots, a_k)^T$ は

$$\mathbf{a} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{x} \quad (8.40)$$

であるから、式(8.39)より、 $\{(\tilde{x}_i, \tilde{x}'_i)\}$ が学習データとして与えられれば、式(8.31)に示す視点変換の写像 τ は

$$\tau = \tilde{X}' (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \quad (8.41)$$

と求めることができる。明らかに、上の議論は、単一の剛体の場合にとどまらず、式(8.32)で表されるような物体のクラス、即ち、ある物体の像の集合 $\{\tilde{X}_i\}$ の線形結合によって表されるクラスの物体すべてについて成立する。

8.3.2 Linear Deformation の場合の視点変換写像

直交射影の下で、簡単な変形可能物体として、第8.2.2節で扱った2種類の物体A,Bの間で形状が直線的に変形するものを考える。このとき、この物体A,Bの像は、物体A,Bを同一方向から見た複数個の像 $\{\tilde{p}_i\}, \{\tilde{q}_i\}$ を用いて、式(8.13)のように表せる。いま、 $\{\tilde{p}_i\}, \{\tilde{q}_i\}$ を一定角度視点を変換させたときの像を $\{\tilde{p}'_i\}, \{\tilde{q}'_i\}$ とすると、式(8.35)より、物体A,Bの視点変換後の像 $\tilde{x}'_A, \tilde{x}'_B$ は次式ようになる。

$$\tilde{x}'_A = \sum_{i=1}^k \kappa_i \tilde{p}'_i, \tilde{x}'_B = \sum_{i=1}^k \kappa_i \tilde{q}'_i \quad (8.42)$$

このとき、変形可能物体の像 \tilde{x}' に対して、視点変換像を与える写像 τ は

$$\tau: \tilde{x}(\alpha) \longrightarrow \tilde{x}'(\alpha)$$

$$\tilde{x}(\alpha) = (1 - \alpha) \tilde{x}_A + \alpha \tilde{x}_B, \tilde{x}'(\alpha) = (1 - \alpha) \tilde{x}'_A + \alpha \tilde{x}'_B \quad (8.43)$$

となる。式(8.14)(8.43)より、

$$\tilde{x} = \sum_{i=1}^k \kappa_i ((1 - \alpha) \tilde{p}_i + \alpha \tilde{q}_i), \tilde{x}' = \sum_{i=1}^k \kappa_i ((1 - \alpha) \tilde{p}'_i + \alpha \tilde{q}'_i) \quad (8.44)$$

であるから、式(8.41)より、

$$P = [\tilde{p}_1 \dots \tilde{p}_k], P' = [\tilde{p}'_1 \dots \tilde{p}'_k], Q = [\tilde{q}_1 \dots \tilde{q}_k], Q' = [\tilde{q}'_1 \dots \tilde{q}'_k] \quad (8.45)$$

とおくと、入力像 $\tilde{x}(\alpha)$ から α を求めることができれば、

$$\tau[\tilde{x}(\alpha)] = D'(D^T D)^{-1} D^T \tilde{x}(\alpha)$$

$$D = (1 - \alpha)P + \alpha Q, D' = (1 - \alpha)P' + \alpha Q' \quad (8.46)$$

によって、視点変換画像を定義することができる。このとき、式(8.18)(8.19)で定義される線形変換 A_P, A_Q をとると、

$$1 - \alpha = \frac{\|A_P \tilde{x}\|}{\|A_P \tilde{x} + A_Q \tilde{x}\|}, \alpha = \frac{\|A_Q \tilde{x}\|}{\|A_P \tilde{x} + A_Q \tilde{x}\|} \quad (8.47)$$

となるから、結局、この場合、視点変換画像 τ は次式で求まる事がわかる。

$$\begin{aligned} D &= \frac{\|A_P \tilde{x}\|}{\|A_P \tilde{x} + A_Q \tilde{x}\|} P + \frac{\|A_Q \tilde{x}\|}{\|A_P \tilde{x} + A_Q \tilde{x}\|} Q \\ D' &= \frac{\|A_P \tilde{x}\|}{\|A_P \tilde{x} + A_Q \tilde{x}\|} P' + \frac{\|A_Q \tilde{x}\|}{\|A_P \tilde{x} + A_Q \tilde{x}\|} Q' \\ \tau &= D'(D^T D)^{-1} D^T \end{aligned} \quad (8.48)$$

8.3.3 変形可能物体の視点変換画像の獲得

物体が変形可能である場合も、標準像への変換画像の場合と同様に、剛体の場合の画像を補間することにより、視点変換画像を近似的に実現できると考えられる。入力像に対し、一定角度視点を変換した像を出力する画像が、変形する物体の場合にも学習できるかどうか butterfly を用いて、透視射影の下で計算機実験を行なった。

視点変換画像としては、与えられた物体の像を、3次元空間中でY軸の回りに ϕ_Y 回転した像に変換することを考え、 $\phi_Y = 15^\circ$ として実験を行なった。学習データとしては、入出力例の組として、関節角度 $\alpha_i \in \{0^\circ, 30^\circ, 60^\circ\}$ を用いて、

$$S = \{\tilde{x}(\alpha_i; \phi_{ij}), \tilde{x}(\alpha_i; \phi'_{ij})\} \quad (8.49)$$

を与えた。ここに、姿勢 ϕ'_{ij} は ϕ_{ij} からさらに物体をY軸回りに 15° 回転した姿勢を示す。各 ϕ_{ij} は各 α_i について乱数を用いて30例与えた。学習方式としては、一般のノルム行列を持つ Gaussian GRBF を用いた。なお、実験においては、GRBFのユニット数は20とした。学習によって得られた画像は各関節角度 α について100例発生させ

$$E(\alpha; \phi) = \frac{\|f(\tilde{x}(\alpha; \phi) - \tilde{x}(\alpha; \phi))\|}{\|\tilde{x}(\alpha; \phi)\|} \quad (8.50)$$

の平均値、最大値、最小値を用いて評価した。結果を図8.9に示す。標準像への変換画像の場合と同様に、学習データとして与えられた形状のみならず、未知関節角度による形状に関しても補間によって視点変換画像が近似的に獲得できることがわかる。

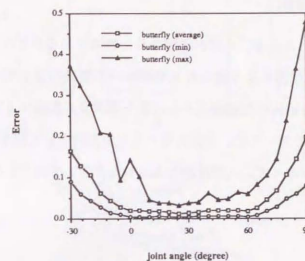


図 8.9: 学習によって得られた視点変換画像の誤差評価 (butterfly)

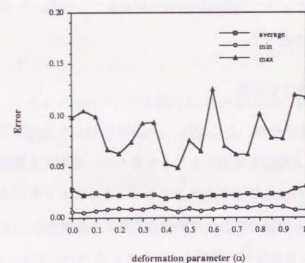


図 8.10: 学習によって得られた視点変換画像の誤差評価 (cube to pyramid)

同様の実験を図8.1に示した変形可能物体について行なった結果を図8.10に示す。なお、この場合、学習データとしては変形パラメータ $\alpha_i \in \{0.0, 0.5, 1.0\}$ を用いて学習データを生成している。関節を持つ変形可能物体 butterfly の場合と同様、この場合も視点変換画像が近似的に獲得できていることが分かる。

8.4 変形する物体の識別

ある物体の像が与えられたとき、それが学習された物体であるかどうかを認識・識別する問題を考える。このために、学習により得られた標準像への変換写像を用いることが考えられる[18, 91]。いま、学習により物体の任意視点からの像を標準像に変換する写像 T が得られたものとする。通常、剛体の場合においては、学習に用いる入出力例として望ましい出力、すなわち標準像 \tilde{x}_s が与えられる。したがって、入力 \tilde{x} が与えられたとき、出力 $T\tilde{x}$ と \tilde{x}_s を比較して、例えば

$$E = \|T\tilde{x} - \tilde{x}_s\|$$

を評価することにより、識別を行うことができると考えられる。

しかしながら、変形する物体の場合、学習パターンとして提示されていない未知形状に対しては、写像の望ましい出力値が不明であり、上の手法をそのまま用いることはできない。本節ではこの問題に対して、(1) 物体クラスの識別関数の学習に基づく認識、(2) 視点変換の学習に基づく認識、の2つの手法について述べる。

8.4.1 識別関数の学習に基づく手法

変形する物体を識別するために、ここでは、入力像に対してその形状の標準像に変換する写像と同時に、物体を識別する関数を学習することを考える。変形する物体は、それぞれの形状に対応する剛体物体の集合から成る物体のクラスであるとみなすことができるから、この場合の識別関数は物体のクラス、即ち形状の集合を識別する関数とみなせる。

いま、学習対象の変形する物体が k 種類与えられているものとする。このとき、出力例として

$$y(\tilde{x}; \alpha; \phi) = (\tilde{x}_s, F_1, F_2, \dots, F_k) \in \mathbb{R}^{2N+k} \quad (8.51)$$

を与え、それに対するネットワークの出力を、

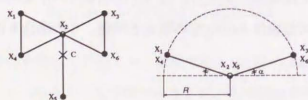


図 8.11: 対象物体 dragonfly

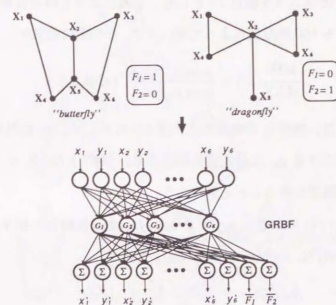


図 8.12: 物体の識別のための学習ネットワーク

$$f(\tilde{x}; \alpha; \phi) = (\tilde{x}_s, \bar{F}_1, \bar{F}_2, \dots, \bar{F}_k) \in \mathbb{R}^{2N+k} \quad (8.52)$$

とする。ここに、 F_i は、入力 \tilde{x} が物体 i である場合に 1、それ以外の場合には 0 を出力するような関数である。したがって、物体 j の像を $\tilde{x}^j(\alpha; \phi)$ とおくと、 F_i はデルタ関数を用いて

$$F_i(\tilde{x}^j) = \delta_{ij} \quad (8.53)$$

と表すことができる。このとき変形物体の識別関数を

$$\rho_j = \sum_{i=1}^k |\bar{F}_i(\tilde{x}) - \delta_{ij}|, \quad (j = 1, \dots, k) \quad (8.54)$$

とし、最小値をとる ρ_j がしきい値 ε 以下である場合、すなわち、

$$\rho_0 = \min_j \rho_j, \quad \rho_0 < \varepsilon, \quad (j = 1, \dots, k) \quad (8.55)$$

のとき、与えられた像は物体 0 であるものと判定する。但し、このとき $\varepsilon < 0.5$ とする。

本方式の有効性を調べるために、計算機実験を行なった。実験においては認識対象として、butterfly(図 8.2) の他に同様に関節で変形する物体 dragonfly(図 8.11) を用いた。学習ネットワークの構造を図 8.12 に示す。学習データは、 $\alpha_i \in \{0^\circ, 30^\circ, 60^\circ\}$ とし、それぞれの物体の各関節角度 α_i について、回転角 ϕ を乱数によって同数ずつ与えた。さらに、学習データとしてランダムパターンを与え、ランダムパターンに対する出力は $\tilde{x}_s = 0$, $F_i = 0$ ($i = 1, 2$) とした。学習データ

の総数は 140 個、GRBF のユニット数は 15 とした。学習によって得られた写像は、物体 o_j の各関節角度に対して回転角を 100 例乱数によって発生させ、その識別性能を、

$$\frac{\text{MIN}}{\text{MAX}}(o_j) = \frac{\min \rho_{o_i}(\tilde{x}^{o_j})}{\max \rho_{o_j}(\tilde{x}^{o_j})}, (o_i \neq o_j) \quad (8.56)$$

によって評価した。これは、物体 o_j の複数の入力像 \tilde{x}^{o_j} に対して、 o_j 以外の物体に対応する ρ_{o_i} ($o_i \neq o_j$) の最小値と、 o_j に対応する ρ_{o_j} の最大値の比である。MIN / MAX > 1 の場合、物体 o_j とそれ以外の物体が完全に識別できることを意味する。

図 8.13 に $o_j = \text{butterfly}$ および、 $o_j = \text{dragonfly}$ とした場合の結果を示す。この場合、 $\{o_i\} = \{\text{butterfly}, \text{dragonfly}, \text{unknown}(\text{random})\}$ であり、

$$\begin{aligned} \rho_{\text{butterfly}} &= |F_1 - 1| + |F_2| \\ \rho_{\text{dragonfly}} &= |F_1| + |F_2 - 1| \\ \rho_{\text{unknown}} &= |F_1| + |F_2| \end{aligned} \quad (8.57)$$

とした。この場合も、学習した形状 $\alpha = \{0^\circ, 30^\circ, 60^\circ\}$ による識別関数の学習によって、未知形状に対しても適正に識別する関数が得られていることがわかる。

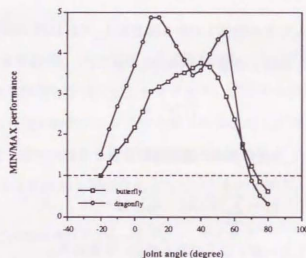


図 8.13: butterfly と dragonfly の識別実験結果

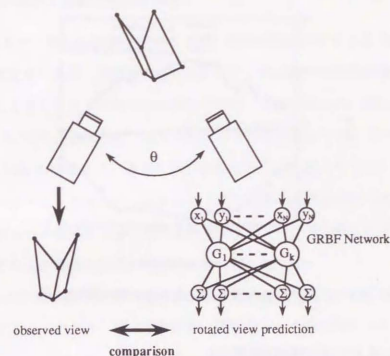


図 8.14: 視点変換写像を用いた物体の識別

8.4.2 視点変換写像を用いた手法

第 8.3 節において、変形可能物体の場合でも、学習によって物体の像をその物体を一定角度視点を変換した場合の像に変換する視点変換写像が獲得できることを示した。ここでは、物体の識別にこの視点変換写像を用いることを考える。

いま、観測者が物体に対して能動的な観測を行なうことができ、例えば、物体を一定角度別の視点から観測することができる場合を想定する。ある物体の像から、それを一定角度視点を変えた場合の像を予測、生成する写像が学習できれば、実際に対象物体をその視点から観測した像と、予測された像とを比較することにより物体認識が可能であると考えられる (図 8.14)。

第 8.3 節で得られた視点変換写像に、学習した物体 *butterfly* を入力した場合と、学習していない物体 *dragonfly* を入力した場合の誤差

$$E(\alpha, \phi_k) = \frac{\|f(\tilde{x}(\alpha; \phi_k) - \tilde{x}(\alpha; \phi'_k))\|}{\|\tilde{x}(\alpha; \phi'_k)\|} \quad (8.58)$$

の平均値、最大値、最小値の比較を行なった。図 8.15 に *butterfly*、*dragonfly* の誤差平均、及び *butterfly* の誤差最大値と *dragonfly* の誤差最小値の比を示す。なお、各関節角について姿勢 ϕ_k は乱数を用いて 100 例発生させた。この結果は、複数視点の画像が得られる場合には、視点変換写像の学習により、物体の認識・識別が可能であることを示すものであると言える。

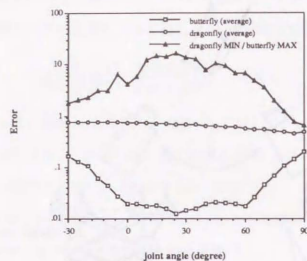


図 8.15: 視点変換画像を用いた物体の識別実験結果

8.5 像推定に基づく3次元物体の認識

物体が剛体かつ像が直交射影によって生成される場合には、物体の任意視点からの像は、その物体の複数視点からの像をモデルとして保持して用いることにより認識を行なうことができる。しかしながら、物体が変形可能である場合等、入力像として与えられた形状が（既知の形状と同一クラスには属するものの）未知形状である場合には、この手法をそのまま適用することはできない。このような場合の手法として、対象物体の像が与えられたとき、この物体と同じクラスに属す既知物体の例から、与えられた未知対象物体の複数視点からの像を推定し、これらをモデルとして用いて認識を行なうことが考えられる。このとき、本章でこれまで述べてきた標準像への変換画像、視点変換画像は、一枚の画像から、その物体の別視点からの像を推定する画像であるとみなせるから、これらの画像を用いて、入力像に対して複数視点の像を推定し、変形可能物体（未知形状）の3次元物体認識を行なうことが考えられる。

本節では、変形する3次元物体を対象として、画像からその物体の複数視点からの像を推定する関数を例題からの学習によって獲得し、この関数を用いて得られた像をモデルとして用いて3次元物体を認識する手法について述べる。なお、本節では、3次元空間から2次元画面への射影は直交射影で表されるものとし、また画像上の特徴点 $\{x_i\}_{i=1}^N$ はその重心が原点と一致するように平行移動されているものとして議論を進める。

8.5.1 像の整合性に基づく3次元物体の認識

本章では、これまで、物体が関節を持つような、変形可能物体である場合でも、十分な汎化能力を持つ学習方式を用いれば、例題からの補間により、標準像への変換画像、視点変換画像が近似的に得られることを示した。これらの画像を用いて、未知形状の入力像に対して得られる標準像、視点変換像は、この入力形状の、異なる視点からの像を推定したものであると見なせる。Ullman and Basri[100]の結果より、ある物体の複数視点からの像が与えられれば、これらの像（の線形結合）をモデルとして用いることにより3次元物体の認識が可能である。したがって、いま、学習の結果得られた画像によって推定された像を用いて、入力像と推定された像の間の整合性を検証することにより物体の認識が可能であると考えられる。

いま、例題からの学習によってY軸回りの回転角度 ϕ の視点変換画像、標準像への変換画像が得られたものとする。このとき、視点変換画像、標準像への変換画像より入力像中の点 $x_i = (x_i, y_i)$ はそれぞれ $x_i^v = (x_i^v, y_i^v)$, $x_i^s = (x_i^s, y_i^s)$ に変換されるものとする。これらの変換が整合性を持つ場合には、これらの点は3次元空間中の1点 P_i を用いて次のように表すことができる。

$$x_i = e_1 \cdot P_i + T_X, \quad y_i = e_2 \cdot P_i + T_Y$$

$$x_i^v = e_1 \cdot R_Y(\phi) \cdot P_i + T_X^v = u_1 \cdot P_i + T_X^v, \quad y_i^v = e_2 \cdot R_Y(\phi) \cdot P_i + T_Y^v = e_2 \cdot P_i + T_Y^v$$

$$x_i^s = e_1 \cdot R \cdot P_i + T_X^s = r_1 \cdot P_i + T_X^s, \quad y_i^s = e_2 \cdot R \cdot P_i + T_Y^s = r_2 \cdot P_i + T_Y^s \quad (8.59)$$

ここに、

$$R_Y(\phi) = [u_1 \ u_2 \ u_3]^T = \begin{pmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{pmatrix}, \quad R = [r_1 \ r_2 \ r_3]^T = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (8.60)$$

である。各点は、特徴点の重心が原点と一致するように平行移動されていることを仮定しているから、式(8.60)より

$$\sum_{i=1}^N y_i = e_2 \cdot \sum_{i=1}^N P_i + N T_Y = 0$$

$$\sum_{i=1}^N y_i^v = e_2 \cdot \sum_{i=1}^N P_i + N T_Y^v = 0 \quad (8.61)$$

が成立し、したがって、 $T_Y = T_Y^v$ すなわち

$$y_i = y_i^v \quad (i = 1, \dots, N) \quad (8.62)$$

が成立することがわかる。また、 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{u}_1 = (\cos \phi, 0, \sin \phi)^T$ は独立であるから、係数 $\{\alpha_i\}, \{\beta_i\}$ が存在して、

$$\begin{aligned} \mathbf{r}_1 &= \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{u}_1, \\ \mathbf{r}_2 &= \beta_1 \mathbf{e}_1 + \beta_2 \mathbf{e}_2 + \beta_3 \mathbf{u}_1 \end{aligned} \quad (8.63)$$

とできる。したがって式(8.60)より、 (x_i^s, y_i^s) は (x_i, y_i) 、 (x_i^v, y_i^v) を用いて、

$$\begin{aligned} x_i^s &= \alpha_1 x_i + \alpha_2 y_i + \alpha_3 x_i^v + (T_X^v - \alpha_1 T_X - \alpha_2 T_Y - \alpha_3 T_X^v) \\ y_i^s &= \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^v + (T_Y^v - \beta_1 T_X - \beta_2 T_Y - \beta_3 T_X^v) \end{aligned} \quad (8.64)$$

と表すことができ、いま、重心は原点と一致しているから

$$\begin{aligned} x_i^s &= \alpha_1 x_i + \alpha_2 y_i + \alpha_3 x_i^v \\ y_i^s &= \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^v \end{aligned} \quad (8.65)$$

が成立する。したがって、

$$\bar{\mathbf{x}}^s = \begin{pmatrix} x_1^s \\ \vdots \\ x_N^s \end{pmatrix}, \bar{\mathbf{y}}^s = \begin{pmatrix} y_1^s \\ \vdots \\ y_N^s \end{pmatrix}, \Xi = \begin{pmatrix} x_1 & y_1 & x_1^v \\ \vdots & \vdots & \vdots \\ x_N & y_N & x_N^v \end{pmatrix}$$

とおくと、 $\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$ 、 $\beta = (\beta_1, \beta_2, \beta_3)^T$ は $\text{rank}(\Xi) = 3$ のとき、

$$\hat{\alpha} = (\Xi^T \Xi)^{-1} \Xi^T \bar{\mathbf{x}}^s, \hat{\beta} = (\Xi^T \Xi)^{-1} \Xi^T \bar{\mathbf{y}}^s \quad (8.66)$$

により得ることができる。得られた α, β を用いることにより、式(8.63)に基づいて $\mathbf{r}_1, \mathbf{r}_2$ を求めることができる。入力像、標準像、視点変換像が整合性を持つ場合には、次の拘束条件が(近似的に)成立すると考えられる。

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = \alpha_1 \beta_1 + \alpha_2 \beta_2 + \alpha_3 \beta_3 + (\alpha_1 \beta_3 + \alpha_3 \beta_1) \cos \phi = 0 \quad (8.67)$$

$$\|\mathbf{r}_1\|^2 = \sum_{i=1}^3 \alpha_i^2 + 2\alpha_1 \alpha_3 \cos \phi = \|\mathbf{r}_2\|^2 = \sum_{i=1}^3 \beta_i^2 + 2\beta_1 \beta_3 \cos \phi \quad (8.68)$$

学習方式の汎化能力を利用して、学習時に与えられた例題の補間により、学習データと同一クラスの未知形状に対しても標準像、視点変換像を精度良く近似する写像が得られれば、式(8.62)(8.68)等に基づいて整合性の検証を行なうことにより、物体(形状)の認識ができると考えられる。

なお、このとき、対象物体が平面形状の場合には、式(8.59)で用いた3次元空間中の点 $\mathbf{P}_i = (X_i, Y_i, Z_i)^T$ に関して

$$Z_i = n_X X_i + n_Y Y_i + d = n_X x_i + n_Y y_i + Z_0 \quad (8.69)$$

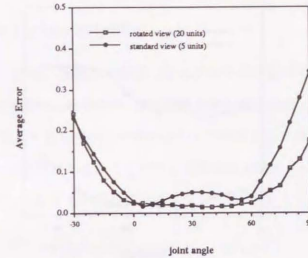


図 8.16: 標準像、視点変換像の学習結果 (butterfly)

と表せる。これを式(8.59)に代入し、いま重心と原点が一致することから、

$$\begin{aligned} x_i^v &= (\cos \phi + n_X \sin \phi) x_i + n_Y \sin \phi y_i \\ x_i^s &= (r_{11} + n_X r_{13}) x_i + (r_{12} + n_Y r_{13}) y_i \\ y_i^s &= (r_{21} + n_X r_{23}) x_i + (r_{22} + n_Y r_{23}) y_i \end{aligned} \quad (8.70)$$

となり、 $\text{rank}(\Xi) \leq 2$ となる。 $\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}_v$ の一次従属性より物体が平面状と判定される場合には平面形状の整合性の検証を行えばよい。

8.5.2 計算機実験

対象物体として、図 8.2 に示すような関節角 α によって形状が決まるような物体を考え、GRBF を用いて $\alpha = \{0^\circ, 30^\circ, 60^\circ\}$ の例題から、標準像と Y 軸回りに 15° 回転した視点変換像を推定する写像の学習を試みた。実験結果を望ましい出力 \mathbf{y} とネットワーク出力との差の平均値

$$E(\alpha) = 1/M \sum_{i=1}^M \|\mathbf{f}(\mathbf{x}_i(\alpha)) - \mathbf{y}_i\|^2 / \|\mathbf{y}_i\|^2$$

で評価したものを図 8.16 に示す。標準像の学習には 5 個の Gaussian から成る GRBF を、視点変換像の学習には 20 個の Gaussian から成る GRBF を用いた。

学習の結果得られた写像を用いて、推定像及び入力像の整合性の検証シミュレーションを行なった。整合性の検証のために、次の評価関数を用いた。

$$\begin{aligned} E &= (\|\bar{\mathbf{x}}_s - \Xi \hat{\alpha}\|^2 + \|\bar{\mathbf{y}}_s - \Xi \hat{\beta}\|^2) / (\|\bar{\mathbf{x}}_s\|^2 + \|\bar{\mathbf{y}}_s\|^2) + \|\bar{\mathbf{y}} - \bar{\mathbf{y}}_v\|^2 / \|\bar{\mathbf{y}}\|^2 \\ &+ \|\mathbf{r}_1\|^2 - \|\mathbf{r}_2\|^2 / (\|\mathbf{r}_1\|^2 + \|\mathbf{r}_2\|^2) + \mathbf{r}_1 \cdot \mathbf{r}_2 / \|\mathbf{r}_1\| \|\mathbf{r}_2\| \end{aligned} \quad (8.71)$$

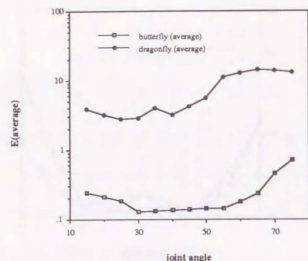


図 8.17: 推定像の整合性の評価

図 8.17に $E(\alpha)$ の平均値を示す。比較のために図 8.2(butterfly) と類似の対象物体 (dragonfly, 図 8.11) を入力した場合の値も示す。また、図 8.18には、図 8.1に示した変形可能物体を用いた場合の実験結果を示す。なお、この場合、図 8.18(a) に示す学習において、学習データは変形パラメータ $\alpha_i \in \{0.0, 0.5, 1.0\}$ を用いて生成した。図 8.17、図 8.18(b) に示すように、いずれの場合も、整合性の検証により物体の識別・認識が可能であると考えられる。ここで述べた手法は、3次元空間から2次元画面への射影が直交射影である場合のものであるが、透視射影である場合でも、直交射影が近似的に成立する場合には本手法を用いることが可能であると考えられる。

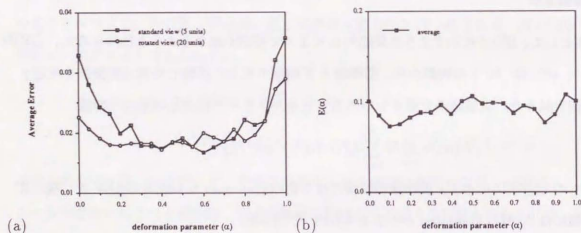


図 8.18: 推定像の整合性の検証による linear deformation の場合の変形可能物体の認識 (cube to pyramid) : (a) 標準像、視点変換像の学習結果 (b) 推定像の整合性の評価

8.6 特徴点の対応付け

8.6.1 識別関数を用いた対応付け候補の限定

これまで述べてきた手法は、画像上の特徴点とモデルとの対応は既知としており、対応が未知である場合には適用できない。物体が N 個の特徴点から成る場合、モデルと画像との対応の可能性は $N!$ 通り存在する。これらすべての組み合わせの中から正しい対応を検出するためには、各対応の評価関数を与える必要がある。ここでは、各組み合わせによって得られる入力ベクトル $\bar{x}_\kappa \in \mathbf{R}^{2N}$ ($\kappa = 1, \dots, N!$) に対する学習後のネットワークの出力

$$f(\bar{x}_\kappa) = (\bar{x}_\kappa(\kappa), \bar{F}_1(\kappa), \dots, \bar{F}_k(\kappa)) \quad (8.72)$$

に基づき識別関数 $\rho_j(\kappa) = \sum_{i=1}^k |\bar{F}_i(\kappa) - \delta_{ij}|$ を求め、

$$\rho_0(\kappa) = \min_j \rho_j(\kappa) < \varepsilon \quad (8.73)$$

となるものを正しい対応の候補と考える。

8.6.2 標準像と入力像の整合性に基づく対応付け

一般に、異なる視点からの2つの像の間には、対応点間に幾何学的拘束条件が存在する。これまでの結果より、我々の学習ネットワークを用いれば、入力像に対して標準像が得られる。従って、正しい対応付けのなされた入力像に対しては、この幾何学的拘束条件を満たすような標準像が得られると考えられる。逆に、対応が正しくない場合には、この拘束条件が満たされないと考えられる。

いま、3次元空間から画像への投影は直交射影であるとし、3次元空間中の点 $\mathbf{X} = (X, Y, Z)$ は剛体変換を受けるものとする。このとき、変換後の点 \mathbf{X}' は回転行列 $R = [r_1 \ r_2 \ r_3]^T$ 、並進ベクトル $\mathbf{T} = (T_X, T_Y, T_Z)^T$ を用いて

$$\mathbf{X}' = R\mathbf{X} + \mathbf{T} \quad (8.74)$$

と表せる。 \mathbf{X}, \mathbf{X}' を画面上に投影した点をそれぞれ $\mathbf{x} = (x, y)$, $\mathbf{x}' = (x', y')$ とすると、いま直交射影を考えているから、

$$x = X, y = Y$$

$$\begin{aligned} x' &= \mathbf{r}_1 \cdot \mathbf{X} + T_X = r_{11}x + r_{12}y + r_{13}Z + T_X \\ y' &= \mathbf{r}_2 \cdot \mathbf{X} + T_Y = r_{21}x + r_{22}y + r_{23}Z + T_Y \end{aligned} \quad (8.75)$$

したがって、

$$\begin{aligned} & (r_{11}r_{23} - r_{13}r_{21})x + (r_{12}r_{23} - r_{13}r_{22})y - r_{23}x' + r_{13}y' \\ & = -r_{23}Tx + r_{13}Ty \end{aligned} \quad (8.76)$$

となり、変換前後の画像上の特徴点の組 (x, y, x', y') はすべて同一の超平面上に存在することがわかる。この超平面を次式で表す。

$$a_1x_i + a_2y_i + a_3x'_i + a_4y'_i = a_5, \quad \left(\sum_{i=1}^4 a_i^2 = 1\right) \quad (8.77)$$

近似的に直交射影が成立する場合には、対応仮定により得られる \tilde{x}_κ とそれに対応する他視点からの像 \tilde{x}'_κ が与えられれば、例えば N 個の特徴点 ($N \geq 5$) から 1 個の点 $\{(x_\ell, y_\ell), (x'_\ell, y'_\ell)\}$ を除いて $\{a_i\}$ を計算し、除いた 1 点について得られた超平面上に存在するかを評価することにより、次のような対応付けの評価関数 E_d を定義できる。

$$E_d = \frac{1}{N} \sum_{\ell=1}^N |a_1x_\ell + a_2y_\ell + a_3x'_\ell + a_4y'_\ell - a_5| \quad (8.78)$$

いま、学習によって得られた GRBF ネットワークは、正しい対応付けにより得られた入力ベクトルに対しては、その形状の標準像を精度良く近似できると考えられる。また、正しくない対応に基づく入力から得られた出力 (標準像) を用いた場合には式 (8.77) の拘束は成立せず、式 (8.78) によって対応の正しさを評価ができると考えられる。そこで、ここでは他視点からの像 \tilde{x}'_κ として、入力ベクトル \tilde{x}_κ に対応するネットワーク出力 (8.72) 中の標準像 $\tilde{x}_s(\kappa)$ を用い、式 (8.73) の条件を満たすものについて式 (8.78) に示す E_d を計算し、各評価関数の荷重和

$$E = \xi\rho_0 + \eta E_d \quad (8.79)$$

が最小となるものを正しい対応付けとみなす。

8.6.3 計算機実験

上に述べた手法の有効性を計算機実験によって検証した。実験において、対象は butterfly, dragonfly の像、およびランダムベクトルとした。なお、butterfly, dragonfly の像は透視投影によって得られるものとした。学習においては入出力像の表現として、特徴点集合の重心が原点と一致するように平行移動したものを用いた。入出力例データは、8.4.1 で述べたものと同一のものを用い、20 ユニットから成る GRBF によって学習を行なった。なお、学習データの特徴点の対応付けは正しいとする。学習後のネットワークを用いて、特徴点の対応付け及び物体の識別を行なっ

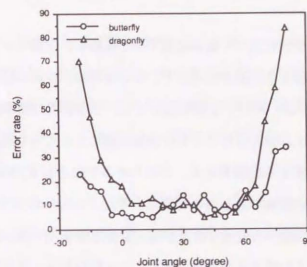


図 8.19: 頂点の対応付けにおける誤答率

た場合の誤答率を図 8.19 に示す。なお、ここでは $\rho_{\text{unknown}} < \varepsilon$ を満たす入力ベクトルは認識対象外であるとしている。実験においては $\xi = \eta = 1.0, \varepsilon = 0.5$ とした。本手法を用いることにより、モデルとの対応未知かつ学習データに含まれない未知形状の特徴点が与えられた場合でも 90% 近い精度で識別ができることがわかる。

いま、正しい対応付けから得られる入力ベクトルを \tilde{x}_κ 、式 (8.79) を最小にし、正しい対応であると判定された入力ベクトルを \tilde{x}_λ とすると、ネットワークは 2 次元の画像を $2N$ 次元空間上の点として評価しているため、 $\|\tilde{x}_\kappa - \tilde{x}_\lambda\| < \varepsilon$ である場合には、これは許容できる誤りであると考えられる。

実験において発生したこのような誤認定の例を図 8.20 に示す。このような誤差を許容した場合、頂点座標だけの情報を用いても、関節角 $0^\circ \leq \alpha \leq 60^\circ$ に関してはほぼ 90 ~ 95% の精度で識別ができる。さらに識別能力を向上させるためには、特徴点の隣接情報等を用いることが必要になると考えられる。

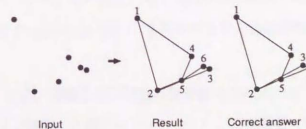


図 8.20: 頂点の対応付けにおける誤答例

8.7 結言

本章では、Poggioらの学習に基づく物体認識手法を拡張し、学習ネットワーク GRBF によって変形する3次元物体を認識する手法を提案した。このために、少数の姿勢・形状の像を例題とした学習によって、“未知形状”に対する像に関しても、その形状の標準像や、視点を一定角度変換した像に変換する写像が近似できることを計算機実験によって示した。また、変形する物体の像を他種類の物体から識別する問題と考え、このための手法として、物体の識別関数を標準像と共に学習する手法及び視点変換写像を用いる手法を考え、これらの手法の有効性を計算機実験によって示した。また、標準像への変換写像及び視点変換写像を用いた、像の推定に基づく3次元物体認識手法を提案し、その能力を計算機実験によって検証した。さらに、画像とモデルの間で特徴点の対応が未知の場合の識別手法として、学習によって得られた識別関数と、ネットワークによって生成される標準像の整合性を用いる手法を考え、その能力を計算機実験によって検証し、有効性を示した。今後の課題としては、対応付けのためのより効率のよいアルゴリズムや、隠れ (occlusions) にも対処する認識手法の確立などが挙げられる。

9 結論

本論文では、Marrの提案した視覚情報処理の流れに沿って、3次元情報の抽出手法と3次元物体の学習について考察し、対象物体に強い拘束条件が成立する場合の3次元情報抽出手法、及び、例題として与えられた2次元像からの3次元物体の学習・認識手法について述べた。

第2章においては従来コンピュータビジョン研究の中心課題の一つとして研究が進められて来ている3次元情報の抽出手法について、光パターンの投影などを行わず、自然界に存在する拘束条件を用いて3次元情報の抽出を行なう受動的手法、及び、光パターン等を投影して非接触で3次元計測を行なう能動的手法について、その代表的なものを概説し、これらの手法の持つ問題点について論じた。

第3章においては、単眼視の条件の下での3次元運動物体の認識問題と考え、対象物体が平面形状である場合には、従来手法では必要とされていた特徴点の対応付けを必要とせずに物体の立体構造と運動が線形特徴量の計算に基づいて求めることができることを明らかにした。さらに、線形特徴量の抽出に基づく、平面の運動と3次元構造の具体的計算手法を示した。

第4章では、形状が多面体である場合を考え、対象物体の稜線の隣接構造を用いた弛緩法によって立体視を行ない、3次元構造を復元する手法を示した。さらに、立体視によって得られた結果を、面の前後関係や面内パターンを用いて検証、修正する手法を示し、実験例によってその有効性を示した。

第5章においては能動的3次元計測手法について考え、パターン投影に基づく3次元計測手法として、特殊な計測機器を必要とせず、1回のパターン投影で対象物体全体に関する3次元情報を取得する、ランダムな切断を持つ多重スリットパターン投影法を提案し、その有効性を実験例によって示した。

第6章においては、学習に基づく3次元物体認識を考える際に重要となる種々の学習方式について、従来の研究を概観した。特に、第7章、第8章において用いるGRBF (Generalized Radial Basis Function) について、その背景、考え方について概説した。また、3次元物体認識に学習を適用した例としてPoggioらのGRBFを用いた3次元物体認識の研究[86, 87, 88, 89, 90, 91]について概説した。

第7章においては、まず、非線形基底関数の線形結合の形式で表される種々の学習方式として、基底としてシグモイド関数(3層パーセプトロン)、3角関数(Flexible Fourier)、指数関数、Gaussian等を用いた場合の関数近似能力比較を行ない、その結果、シグモイド関数、Gaussianは良好な近似性能を持つことを明らかにした。続いて、入力が正規化されている場合を取り

上げ、この場合に成立する ridge 関数と radial 関数の関係を導くと共に、この関係に基づいて、正規化入力の下で MLP (3 層パーセプトロン) と Gaussian GRBF を比較し、MLP は常に同数のユニットから成る Gaussian GRBF を十分に良く近似できること、及び、MLP と Gaussian GRBF を同一視できるための条件を明らかにした。また、Multiscale Gaussian GRBF を用いた場合には、同様なパラメータ数を持つ MLP を十分に良く近似できることを計算機実験により明らかにした。

第 8 章においては、個々の物体ではなく、物体のクラスを学習するために、対象物体として変形可能物体を考え、GRBF を用いた例題からの学習により、任意視点の像を標準像や一定角度回転した像に変換する写像が近似的に獲得できることを示した。さらに、これらの写像を用いた 3 次元変形可能物体の認識手法として、識別関数の学習による手法、視点変換写像を利用する手法、さらには像推定に基づいて 2 次元像に基づくモデルを合成し、これにより認識を行なう手法を示し、計算機実験によってその能力を検証した。また、学習により得られた標準像への変換写像を用いた特徴点の対応付け手法についても検討を行なった。

今後の研究課題としては、2 次元画像からの 3 次元情報の取得に関しては、より頑健な手法を確立することがまず挙げられる。各種の視覚モジュールに関しては多くの手法が提案されて来ている。今後は、これら複数、異種の視覚モジュールを統合し、相互の結果を参照し合うことによって信頼度の高い結果を得るための方式の研究が重要であると考えられる。さらに、視覚システムの置かれた環境に応じ、その環境に最適な視覚アルゴリズムの調整を行なうなどの、視覚アルゴリズムの学習、適応型視覚アルゴリズムの研究も、柔軟な視覚システムの実現のためには重要な課題である。3 次元物体の学習に関しては、本論文では、2 次元例題のみに基づく手法について考察したが、3 次元情報をも用いた物体の学習方式の検討が必要になると考えられる。さらには、効率の良い 3 次元物体の内部表現、即ち、3 次元物体の記憶・管理方式の研究が今後の重要な研究課題として挙げられる。

謝辞

本研究は、筆者が三菱電機中央研究所で行なった研究、及び東京大学工学部計数工学科に受託研究員として滞在中に行なった研究、マサチューセッツ工科大学人工知能研究所に客員研究員として滞在中に行なった研究を、東京大学工学部計数工学科 甘利俊一教授の御指導のもとにまとめたものである。学生時代、及び受託研究員として東京大学工学部計数工学科滞在中に、筆者を学習並びに視覚情報処理の研究に導いて下さると共に、本研究をまとめるに当たり、終始懇切なる御指導を賜りました甘利俊一教授に深謝いたします。

マサチューセッツ工科大学人工知能研究所滞在中、Tomaso Poggio 教授には、学習およびその視覚情報処理への応用に関する研究の御指導を頂き、研究面で多くの有益な御示唆、御助言を頂くと共に、慣れぬ異国の地にあった筆者を暖かく遇して頂いた。ここに厚く御礼申し上げます。

三菱電機中央研究所システム基礎研究部部長 武田捷一博士、同研究部前部長 田中千代治博士、同研究部元部長 上村勝彦博士には、良好なる研究環境を与えて頂くと共に、本研究を遂行する上で多くの御理解を頂いた。同研究部第 3 G 前グループマネージャ 阿部茂 博士には、研究の御指導を頂くと共に、多くの貴重な御助言を頂いた。同研究部第 4 G グループマネージャ 西田正吾博士には多大な御便宜と貴重な御助言を頂いた。三菱電機中央研究所システム基礎研究部の先輩、同僚諸氏からは、本研究を遂行する上で多大な御協力を頂いた。特に、中村泰明博士からは多くの有益な御助言を頂いた。寺岡照彦氏には、学習に基づく変形可能物体の認識の研究に関して御助力頂いた。また、マサチューセッツ工科大学人工知能研究所研究員の Federico Girosi 博士からは、学習の研究に関して御討議頂き、多くの有益な御示唆を頂いた。以上の方々に衷心より感謝の意を表します。

参考文献

- [1] G.Agin and T.Binford, "Computer description of curved objects", Proc. 3rd International Joint Conference on Artificial Intelligence, pp. 629 - 640 (1973).
- [2] S.Amari, "Invariant structure of signal and feature spaces in pattern recognition problems", RAAG Memoirs Vol.4, pp.553 - 566 (1968).
- [3] S.Amari, "Feature spaces which admit and detect invariant signal transformations", Proc. 4th International Joint Conference on Pattern Recognition, pp. 452 - 456 (1978).
- [4] S.Amari and M.Maruyama, "A theory on the determination of 3D motion and 3D structure from features", Spatial Vision, Vol.2, No.2, pp.151 - 168 (1987).
- [5] S.Amari, "Mathematical foundations of neurocomputing", Proc. IEEE, Vol. 78, No.9, pp. 1443 - 1463 (1990).
- [6] S.Amari, "Dualistic geometry of the manifold of higher-order neurons", Neural Networks, Vol.4, No.4, pp.443 - 451 (1991).
- [7] D.Ballard and C.Brown, "Computer Vision", Prentice-Hall (1982).
- [8] S.T.Barnard and W.B.Thompson, "Disparity analysis of images", IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 2, pp. 333 - 340 (1980).
- [9] A.R.Barron, "Approximation and estimation bounds for artificial neural networks", Proc. 4th Workshop on Computational Learning Theory, pp. 243 - 249 (1991).
- [10] A.R.Barron, "Universal approximation bounds for superpositions of a sigmoidal function", IEEE Trans. Information Theory (to appear).
- [11] R.Bolle, P.Horand and J.Hannah, "3DPO: A three - dimensional part orientation system", Proc. 1st International Symposium on Robotics Research, pp.413 - 424, MIT Press(1984).
- [12] B.Baumgart, "Winged edge polyhedron representation", STAN-CS-320, AIM-179, Stanford Univ. AI Lab. (1972).
- [13] S.Botros and C.Atkeson, "Generalization properties of radial basis functions", Advances in Neural Information Processing 3, pp. 707 - 713 Morgan Kaufmann Publishers(1991).
- [14] K.L.Boyer and A.C.Kak, "Color-encoded structured light for rapid active ranging", IEEE Trans. Pattern Analysis Machine Intelligence, Vol.9, pp. 14 - 28 (1987).
- [15] M.Brady, J.Ponce, A.Yuille and H.Asada, "Describing surfaces", Human and Machine Vision II, pp. 58 - 85, Academic Press (1985).
- [16] D.Braess, "Nonlinear approximation theory", Springer-Verlag (1986).
- [17] R.Brooks, "Symbolic reasoning among 3-D models and 2-D images", Artificial Intelligence, Vol.17, Nos. 1 - 3, pp.285 - 348 (1981).
- [18] R.Brunelli and T.Poggio, "Hyper BF networks for real object recognition", Proc. International Joint Conference on Artificial Intelligence-91, pp.1278 - 1284 (1991).
- [19] R.Brunelli and T.Poggio, "Hyper BF networks for gender classification", Proc. Image Understanding Workshop, pp.311 - 314 (1992).
- [20] P.Cohen and E.A.Figenbaum, "The handbook of artificial intelligence", Pittman (1982).
- [21] G.W.Cottrell and J.Matecalfe, "EMPATH: Face, emotion and gender recognition using holons", Advances in Neural Information Processing Systems 3, pp. 564 - 571 (1990).
- [22] D.Donoho and I.Johnstone, "Projection-based approximation and a duality with kernel methods", The Annals of Statistics, Vol.17, No.1, pp. 58 - 106 (1989).
- [23] S.Edelman and T.Poggio, "Bringing the grandmother back to the picture: a memory-based view of object recognition", MIT Artificial Intelligence Lab. Memo No.1181 (1990).
- [24] S.Edelman and H.Bülthoff, "Viewpoint-Specific Representations in Three-Dimensional Object Recognition", MIT Artificial Intelligence Lab. Memo No. 1239 (1990).
- [25] J.Q.Fang, T.S.Huang, "Solving three-dimensional small-rotation motion equations: uniqueness, algorithms, and numerical results", Computer Vision Graphics and Image Processing, Vol.26, pp.183 - 206 (1984).

- [26] J.D.Farmer and J.Sidorowich,"Predicting chaotic time series", Physical Review Letters, Vol.59, No. 8 pp.845 - 848 (1987).
- [27] J.Friedman and W.Stuetzel,"Projection pursuit regression", Journal of the American Statistical Association, Vol. 76, pp. 817 - 823 (1981).
- [28] K.Funahashi,"On the approximate realization of continuous mappings by neural networks", Neural Networks, Vol. 2, pp. 183 - 192 (1989).
- [29] F.Girosi and T.Poggio,"Networks and the best approximation property", Biological Cybernetics, Vol.63, pp. 169 - 176 (1990).
- [30] F.Girosi, T.Poggio and B.Capriole,"Extensions of a theory of networks for approximation and learning : outliers and negative examples", Advances in Neural Information Processing Systems 3, pp. 750 - 756, Morgan-Kaufman Publishers (1991).
- [31] F.Girosi,"Models of noise and robust estimates", MIT Artificial Intelligence Lab. Memo No. 1287 (1991).
- [32] F.Girosi and G.Anzellotti , " Rates of convergence for radial basis functions and neural networks " , in Artificial Neural Networks with Applications in Speech and Vision, Chapman & Hall (to appear).
- [33] D.Goldberg,"Genetic algorithms in search, optimization and machine learning", Addison-Wesley (1989).
- [34] W.E.Grimson,"From images to surfaces",MIT Press (1981).
- [35] M.Herman,T.Kanade and S.Kuroe,"Incremental acquisition of a three-dimensional scene model from images", IEEE Trans. Pattern Analysis Machine Intelligence, Vol.6, pp.331 - 340 (1984).
- [36] M.Herman and T.Kanade,"Incremental reconstruction of 3D scenes from multiple complex images", Artificial Intelligence, Vol.30, pp.289 - 341 (1986).
- [37] E.Hildreth,"The computation of the velocity field", Proc. Royal Society of London, Vol.B221, pp. 189 - 220 (1984).

- [38] B.K.P.Horn and B.G.Schunck,"Determining optical flow",Artificial Intelligence, Vol.17 Nos. 1-3 pp.183 - 206 (1981).
- [39] B.K.P.Horn,"Robot Vision",MIT Press (1986).
- [40] M.K.Hu,"Visual pattern recognition by moment invariant", IRE Trans. IT-8, pp.179 - 187 (1962).
- [41] P.Huber,"Projection Pursuit", The Annals of Statistics, Vol.13, No.2, pp. 435 - 475 (1985).
- [42] P.Huber,"Data analysis and projection pursuit", MIT Dept. Mathematics, Technical Report PJH-90-1 (1990).
- [43] A.Hurlbert and T.Poggio,"Synthesizing a color algorithm from examples", Science, Vol. 239, pp. 482 - 485 (1988).
- [44] J.Hutchinson, C.Koch and C.Mead,"Computing motion using analog and binary resistive networks",Computer, Vol.21,No.3,pp.52-63 (1988).
- [45] B.Irie and S.Miyake,"Capabilities of three-layered perceptrons",Proc. International Conference on Neural Networks, Vol.1, pp. 641 - 648 (1988).
- [46] R.Jain and H.Nagel,"On the analysis of accumulative difference pictures from image sequences of real world scene",IEEE Trans. Pattern Analysis Machine Intelligence, Vol.1, No.2, pp.206 - 213 (1979).
- [47] R.A.Jarvis,"A perspective on range finding techniques for computer vision", IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 5, pp. 122 - 139 (1983).
- [48] L.K.Jones , "A simple lemma on greedy approximation in Hilbert space and convergence rate for projection pursuit regression and neural network training" , The Annals of Statistics, Vol.20, No.1, pp. 608 - 613 (1992).
- [49] T.Kanade,"Geometrical aspects of interpreting images as a three-dimensional scene", Proc. IEEE, Vol.71, pp.789 - 802 (1983).

- [50] K.Kanatani, "Detection of surface orientation and motion from texture by a stereological technique", *Artificial Intelligence*, Vol.23 , pp. 217 - 237 (1984).
- [51] K.Kanatani, "Tracing planar surface motion from projection without knowing correspondence", *Computer Vision Graphics and Image Processing* Vol.29, pp 1 - 12 (1985).
- [52] K.Kanatani, "Detecting the motion of a planar surface by line and surface integrals", *Computer Vision Graphics and Image Processing* Vol.29, pp 13 - 22 (1985).
- [53] E.J.Kansa, "Multiquadrics - a scattered data approximation scheme with application to computational fluid dynamics", *Computers Math. Applic.* Vol.19, pp. 127 - 145 (1990).
- [54] J.J.Koenderink and A.J. van Doorn, "Invariant properties of the motion parallax field due to the rigid bodies relative to the observer", *Optica Acta*, Vol.22, pp.773 - 791 (1975).
- [55] J.J.Koenderink and A.J. van Doorn, "Local structure of movement parallax of the plane", *J. Opt. Society of America*, Vol. 66, pp. 717 - 723 (1976).
- [56] T.Kohonen, "An introduction to neural computing", *Neural Networks*, Vol.1, No.1, pp.3-16 (1988).
- [57] Y.Lee and R.Lippmann, "Practical characteristics of neural networks and conventional pattern classifiers on artificial and speech problems", *Advances in Neural Information Processing Systems*, Vol.2, pp. 168 - 177 (1990).
- [58] R.Lippmann and Y.Lee, "A critical overview of neural network pattern classifiers", *Proc. Neural Networks for Computing Conference*, Snowbird (1991).
- [59] H.C.Longuet-Higgins and K.Prazdny, "The interpretation of a moving retinal image", *Proc. Royal Society of London* Vol. B208, pp. 385 - 397 (1980).
- [60] H.C.Longuet-Higgins, "The visual ambiguity of a moving plane", *Proc. Royal Society of London* Vol. B223, pp. 165 - 175 (1984).
- [61] D.Marr and T.Poggio, "Cooperative computation of stereo disparity", *Science*, Vol. 194, pp. 283 - 287 (1976).

- [62] D.Marr and T.Poggio, "A computational theory of human stereo vision", *Proc. Royal Society of London*, Vol. B204, pp.301 - 328 (1979).
- [63] D.Marr and E.Hildreth, "Theory of edge detection", *Proc. Royal Society of London*, Vol.B207, pp. 187 - 217 (1980).
- [64] D.Marr, "Vision", *Freeman* (1982).
- [65] M.Maruyama and S.Abe, "Acquiring a polyhedral structure through face extraction and verification", *Proc. 9th International Conference on Pattern Recognition* pp. 579 - 581 (1988).
- [66] M.Maruyama and S.Abe, "Range sensing by projecting multi-slits with random cuts", *Proc. International Workshop on Industrial Applications of Machine Intelligence and Vision* , pp. 163 - 168 (1989).
- [67] M.Maruyama and S.Abe, "Range sensing by projecting multiple slits with random cuts", *SPIE Vol.1194, Optics, Illumination, and Image Sensing for Machine Vision IV* , pp. 216 - 224 (1989).
- [68] M.Maruyama, F.Girosi and T.Poggio, "Techniques for learning from examples : numerical comparisons and approximation power", *MIT Artificial Intelligence Lab. Memo No.1290* (1992).
- [69] M.Maruyama, F.Girosi and T.Poggio, "A connection between GRBF and MLP" , *MIT Artificial Intelligence Lab. Memo No. 1291* (1992).
- [70] M.Maruyama and S.Abe, "Range sensing by projecting multiple slits with random cuts", *IEEE Trans. Pattern Analysis Machine Intelligence*, Vol. 15, No.6, pp. 647 - 651 (1993).
- [71] M.Maruyama, T.Teraoka and S.Abe, "Recognition of 3D flexible objects by GRBF", *Biological Cybernetics* (to appear).
- [72] M.Maruyama and T.Teraoka, "Recognition of 3D nonrigid objects by learning view change transformations", *Proc. International Joint Conference on Neural Networks '93* (to appear).

- [73] G. Medioni and R. Nevatia, "Segment-based stereo matching", Computer Vision, Graphics and Image Processing, Vol. 31, No 1, pp. 2 - 18 (1985).
- [74] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units", Neural Computation, Vol.1 pp.281-294 (1989).
- [75] J. Moody, "The *Efficient* number of parameters : An analysis of generalization and regularization in nonlinear learning systems", Advances in Neural Information Processing 4, pp. 847 - 854, Morgan Kaufmann Publishers (1992).
- [76] K. Ng and R. Lippmann, "A comparative study of the practical characteristics of neural networks and conventional pattern classifiers", Advances in Neural Information Processing Systems, Vol.3, pp. 970 - 976 (1991).
- [77] K. Nishihara, "Practical real-time imaging stereo matcher", Optical Engineering, Vol 23, No. 5, pp. 536 - 545 (1984).
- [78] M. Oshima and Y. Shirai, "Object recognition using three dimensional information", IEEE Trans. Pattern Analysis Machine Intelligence, Vol.5, pp. 353 - 361 (1983).
- [79] S. Omohundro, "Efficient algorithms with neural network behavior", Complex Systems, Vol.1, pp. 273 - 347 (1987).
- [80] S. Omohundro, "Geometric learning algorithms", Physica D, Vol. 42, pp.307 - 321 (1990).
- [81] A. Pentland, "Perceptual organization and the representation of natural form", Artificial Intelligence, Vol.28, pp. 293 - 331 (1986).
- [82] J. Platt, "A resource-allocating network for function interpolation", Neural Computation, Vol.3 pp.213-225 (1991).
- [83] J. Tajima, "Rainbow range finder principle for range data acquisition", Proc. International Workshop on Industrial Applications of Machine Vision and Machine Intelligence, pp. 381 - 386 (1987).
- [84] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory", Nature, Vol.317, pp.314 - 319 (1985).

- [85] T. Poggio, E. Gamble and J. J. Little, "Parallel integration of vision modules", Science, Vol.242, pp.436 - 440 (1988).
- [86] T. Poggio and F. Girosi, "Regularization algorithms for learning that are equivalent to multilayer networks", Science, Vol. 247, pp. 978 - 982 (1990).
- [87] T. Poggio and F. Girosi, "Networks for approximation and learning", Proc. IEEE, Vol. 78, No.9, pp. 1481 - 1497 (1990).
- [88] T. Poggio and F. Girosi, "Extension of a theory of networks for approximation and learning : dimensionality reduction and clustering", MIT Artificial Intelligence Lab. Memo No. 1167 (1990).
- [89] T. Poggio and F. Girosi, "Hyper BF : A powerful approximation technique for learning", In Artificial Intelligence at MIT : Expanding Frontiers, Vol. 1, pp. 271 - 285, MIT Press (1990).
- [90] T. Poggio, "A theory of how the brain might work", MIT Artificial Intelligence Lab. Memo No.1253 (1990).
- [91] T. Poggio and S. Edelman, "A network that learns to recognize 3D objects", Nature, Vol.343, pp.263 - 266 (1990).
- [92] W. H. Press et al., "Numerical recipes in C", Cambridge Univ. Press (1987).
- [93] J. W. Roach and J. K. Aggarwal, "Determining the movements of objects from a sequence of images", IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 2, pp. 554 - 562 (1980).
- [94] K. Sato, H. Yamamoto and S. Inokuchi, "Tuned range finder for high precision 3D data", Proc. 8th International Conference Pattern Recognition, pp. 1168 - 1171 (1986).
- [95] K. Sato and S. Inokuchi, "Range-imaging system utilizing nematic liquid crystal mask", Proc. 1st International Conference on Computer Vision, pp. 657 - 661 (1987).
- [96] T. J. Sejnowski and C. R. Rosenberg, "Parallel networks that learn to pronounce English text", Complex Systems, Vol. 1, pp. 145-168 (1987).

- [97] R.Y.Tsai and T.S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces", IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 6, pp. 13 - 27 (1984).
- [98] S.Ullman, "The interpretation of visual motion", MIT Press (1979).
- [99] S.Ullman, "The optical flow of planar surfaces", Spatial Vision, Vol.1, pp. 263 - 276 (1986).
- [100] S.Ullman and R.Basri, "Recognition by linear combinations of models", IEEE Trans. Pattern Analysis Machine Intelligence, Vol.13, No.10, pp.992 - 1006 (1991).
- [101] V.Vapnik, "Principles of risk minimization for learning theory", Advances in Neural Information Processing 4, pp. 831 - 838, Morgan Kaufmann Publishers (1992).
- [102] A.Yuille, "Generalized deformable models, statistical physics, and matching problems", Neural Computation, Vol.2, No.1, pp.1-24 (1990).
- [103] Y.Zhao and C.Atkeson, "Projection pursuit learning", Proc. International Joint Conference on Neural Networks - '91, Vol.1, pp. 869 - 874 (1991).
- [104] 甘利俊一, "パターン信号の特徴空間における正規化理論", 信学誌, Vol.49, No.7 pp.1342 - 1350 (1966).
- [105] 甘利俊一, "情報理論 II", 共立出版 (1968).
- [106] 甘利俊一, "神経回路網の数理", 産業図書 (1978).
- [107] 甘利俊一, "神経モデルとコネクションイズム", 東大出版会 (1989).
- [108] 福島邦彦, "コグニトロンのパターン分離能力の向上", 信学論 (A), Vol. J62-A, No.10, pp.650 - 657(1979).
- [109] 福島邦彦, "位置ずれに影響されないパターン認識の神経回路モデル - ネオコグニロン", 信学論 (A), Vol. J62-A, No.10, pp.658 - 665(1979).
- [110] 鴻, 杉原, 杉江, "パターン投影像とカメラの前進移動を利用した3次元物体計測", 日本ロボット学会誌, Vol.5, No. 6, pp. 14 - 23 (昭 62)

- [111] 飯島泰蔵, "図形の構造認識に関する基礎理論", 信学論 (D), Vol J55-D, No.8, pp.499 - 505 (1974).
- [112] 井口, 佐藤, "三次元画像計測", 昭晃堂 (1990).
- [113] 北村, 来海, 谷内田, "エッジの連続性と局所情報を利用した3眼視", 信学技報 PRU 86 - 2 (1986)
- [114] 松島与三, "多様体入門", 掌華房 (昭 40) (1965).
- [115] 丸山, 甘利, "物体の運動と立体構造の単眼視による認識の基礎", 信学論 (A), Vol.J67-A, No.12, pp. 1107 - 1114 (1984).
- [116] 丸山, 阿部, "立体視に基づく面の抽出と面内ベクトルの認定による多面体構造の抽出", 信学論 (D), Vol.J70-D, No.12, pp.2581 - 2589 (1987).
- [117] 丸山, 阿部, "ランダムな切断を持つスリットパターン投影による形状計測", 信学論 (D), Vol.J71-D, No.12, pp.2596 - 2603 (1988).
- [118] 丸山, Girosi, Poggio, "正規化入力に対する GRBF と多層パーセプトロンの能力比較", 信学技報 NC91-73 (1992).
- [119] 丸山, Girosi, Poggio, "学習における種々の基底関数の能力比較", 信学技報 NC91-124 (1992).
- [120] 丸山稔, "Radial Basis Functions を用いた学習ネットワーク - ニューロコンピュータインクに対する新しいアプローチ -", システム/制御/情報, Vol.36, No. 5, pp. 38 - 45 (1992).
- [121] 丸山, 寺岡, 阿部, "学習ネットワーク GRBF の変形する物体の認識能力", 信学技報 NC92-76 (1992).
- [122] 丸山, 寺岡, 阿部, "学習ネットワーク GRBF の変形する3次元物体の認識能力", 信学論 (DII) (論文採録)
- [123] 中村, 阿部, 大沢, 坂内, "多次元データの平衡木による管理 - MD 木の提案 -", 信学論 (D), Vol.J71-D, No.9, pp.1745 - 1752 (1988).
- [124] 西村, 徐, 辻, "1枚の画像をモデルとした物体認識", 情処研資 CV79-6 (1992).

- [125] 大田, 正井, 池田, "動的計画法によるステレオ画像の区間対応法", 信学論 (D), Vol. J68-D, No.4, pp.554 - 561 (昭60) (1985).
- [126] 大田, 金出, "走査線間の整合性を考慮した2段の動的計画法によるステレオ対応探索", 情報処理, Vol.26 No.11, pp.1356 - 1363 (昭60)(1985)
- [127] 大津展之, "パターン認識における特徴抽出に関する数値的研究", 電総研研究報告 818 (1981).
- [128] 佐藤, 井口, "空間コード化による距離画像入力", 信学論 (D), J68-D, No.3, pp. 369 - 375 (昭60) (1985).
- [129] 高橋, 徐, 辻, "並列 GBF ネットワークの自動生成による物体認識", 画像の認識・理解シンポジウム (MIRU-92), Vol.2, pp. 231 - 238 (1992).
- [130] 田村秀行監修, "コンピュータ画像処理入門", 総研出版 (1984).
- [131] 富田文明, "ステレオ画像の領域の輪郭線セグメントに基づく対応", 情報処理学会コンピュータビジョン研究 38-1 (1986).
- [132] 米沢, 玉邑, "符号化格子を用いた物体形状の計測", 信学論 (D), Vol. J61-D, No. 6, pp. 411 - 418 (昭53) (1978).
- [133] ニューラルネットワークの応用総合特集号, システム / 制御 / 情報, Vol.35, No.1 (1991).
- [134] 日本写真測量学会編, "写真による3次元測定", 共立出版 (昭58) (1983).

付録

A 点パターンの場合の可解性

いま、平面要素上に n 個の点から成るパターンが存在するものとする。このとき、2次元画面上のパターン $s(x, y, t)$ は

$$s(x, y, t) = \sum_{i=1}^n \delta(x - x_i(t), y - y_i(t)) \quad (A.1)$$

と書ける。したがって、このパターンから測定される特徴量は

$$f_j = \langle m_j, s \rangle = \sum_{i=1}^n m_j(x_i, y_i) \quad (A.2)$$

と表せる。式 (3.63) より、特徴量の時間変化率は

$$\dot{f}_j = \sum_{a=1}^8 \dot{u}_a \bar{r}_a(\bar{n}) m_j(x_i, y_i) = \sum_{a=1}^8 \left(\sum_{i=1}^n h'_{ja}(x_i, y_i) \right) \dot{q}_a \quad (A.3)$$

ここに、 h'_{ja} は式 (3.48) を用いて次のようになる。

$$h'_{ja} = t_a m_j(a = 1, \dots, 6), \quad h'_{j7} = -t_7 m_j, \quad h'_{j8} = -t_8 m_j \quad (A.4)$$

式 (A.3) は行列表現を用いて

$$\dot{f} = H \dot{q}, \quad \dot{q} = (q_1, q_2, q_3, q_4, q_5, q_6, -q_7, -q_8) \quad (A.5)$$

と表される。ここに、 H は $k \times 8$ 行列であり、

$$H = \sum_{i=1}^n \begin{pmatrix} \vdots & \vdots & \vdots \\ \bar{h}_{j1}(i) & \bar{h}_{j2}(i) & \dots & \bar{h}_{j8}(i) \\ \vdots & \vdots & \vdots \end{pmatrix} \quad (A.6)$$

$$\bar{h}_{j1}(i) = \frac{\partial m_j}{\partial x}(x_i, y_i), \quad \bar{h}_{j2}(i) = x_i \frac{\partial m_j}{\partial x}(x_i, y_i), \quad \bar{h}_{j3}(i) = y_i \frac{\partial m_j}{\partial x}(x_i, y_i),$$

$$\bar{h}_{j4}(i) = \frac{\partial m_j}{\partial y}(x_i, y_i), \quad \bar{h}_{j5}(i) = x_i \frac{\partial m_j}{\partial y}(x_i, y_i), \quad \bar{h}_{j6}(i) = y_i \frac{\partial m_j}{\partial y}(x_i, y_i)$$

$$\bar{h}_{j7}(i) = x_i(x_i \frac{\partial m_j}{\partial x}(x_i, y_i) + y_i \frac{\partial m_j}{\partial y}(x_i, y_i)), \quad \bar{h}_{j8}(i) = y_i(x_i \frac{\partial m_j}{\partial x}(x_i, y_i) + y_i \frac{\partial m_j}{\partial y}(x_i, y_i))$$

である。 $k \times 8$ 行列 H は $k \times 2n$ 行列 H_0 と $2n \times 8$ 行列 H_1 を用いて

$$H = H_0 H_1 \quad (A.7)$$

と書ける。ここに

$$H_0 = \begin{pmatrix} \vdots & \vdots & \vdots \\ \bar{h}_{j1}(1) & \dots & \bar{h}_{j1}(n) & \bar{h}_{j4}(1) & \dots & \bar{h}_{j4}(n) \\ \vdots & \vdots & \vdots \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 1 & x_1 & y_1 & 0 & x_1^2 & x_1 y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & 0 & x_n^2 & x_n y_n \\ 0 & 1 & x_1 & y_1 & x_1 y_1 & y_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & x_n y_n & y_n^2 & y_n^2 \end{pmatrix} \quad (\text{A.8})$$

である H がフルランクである場合 ($\text{rank}(H) = 8$) には、 q は線形方程式を解いて得ることができる。このとき、次の定理が成立する。

定理

行列 H_0 がフルランクであると仮定する。このとき行列 H がフルランクとなるための必要条件は、点パターンは 4 点以上の点を含み、かつ、そのうちのどの 3 点をとっても一直線上に並ばないことである。 $k > 2n$ を満足する場合は、この条件は必要十分条件を与える

証明

行列 H がフルランクであると $\text{rank}(H) = 8$ となる。このためには $\text{rank}(H_1) = 8$ を満たさなければならない。したがって、パターン s は少なくとも 4 点を含まなければならないことがわかる。いまパターンが 4 点を含むとし、その座標を (x_i, y_i) ($i = 1, \dots, 4$) とする。このとき行列 H_1 に関して次の関係式が成立する。

$$\det \begin{pmatrix} 1 & x_1 & y_1 & 0 & x_1^2 & x_1 y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_4 & y_4 & 0 & x_4^2 & x_4 y_4 \\ 0 & 1 & x_1 & y_1 & x_1 y_1 & y_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_4 & y_4 & x_4 y_4 & y_4^2 & y_4^2 \end{pmatrix} = d_1 d_2 d_3 d_4 \quad (\text{A.9})$$

但し、ここに

$$d_1 = \begin{vmatrix} 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{vmatrix}, d_2 = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{vmatrix}, d_3 = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_4 & y_4 \end{vmatrix}, d_4 = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix} \quad (\text{A.10})$$

明らかに、 $\begin{vmatrix} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_k & y_k \end{vmatrix} = 0$ を満足するのは 3 点 $(x_i, y_i), (x_j, y_j), (x_k, y_k)$ が一直線上に存在するときで、また、そのときに限られる。したがって明らかに、 H_1 が正則となるのはパターン中のどの 3 点をとっても一直線上に並ばないときである。

B 水平エビボーラ線生成のための画像変換

いま図 B.1 のようなカメラモデルを考える。このカメラに付随した座標系として視線方向の Z 軸をとり、3 次元空間中の 1 点 $X = (X, Y, Z)$ から 2 次元画面上の点 $x = (x, y)$ への変換を透視射影

$$x = f \frac{X}{Z}, y = f \frac{Y}{Z} \quad (\text{B.1})$$

によって表すものとする。ここに f は射影の中心点から画面までの距離である。このとき画面上に 1 点 $p = (p, q)$ をとると、これに対応する 3 次元空間中の点は、このカメラ系の原点 O (射影の中心点) と (p, q, f) を通る直線上に存在することが分かる。

逆に、いま 3 次元空間中の直線

$$X(t) = at + b, \quad a = (a_1, a_2, a_3), \quad b = (b_1, b_2, b_3) \quad (\text{B.2})$$

を考えると、この直線による画面上の像は、式 (B.1) より、次式で与えられる。

$$x = f \frac{a_1 t + b_1}{a_3 t + b_3}, \quad y = f \frac{a_2 t + b_2}{a_3 t + b_3} \quad (\text{B.3})$$

したがって、2 次元画面上の像はパラメータ t を消去して

$$(a_2 b_3 - a_3 b_2)x + (a_3 b_1 - a_1 b_3)y + (a_1 b_2 - a_2 b_1)f = 0 \quad (\text{B.4})$$

となる。このとき、 $\tilde{x} = (x, y, f)$ とおくと、式 (B.4) は次のように表すことができる。

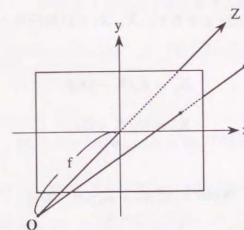


図 B.1: 透視射影

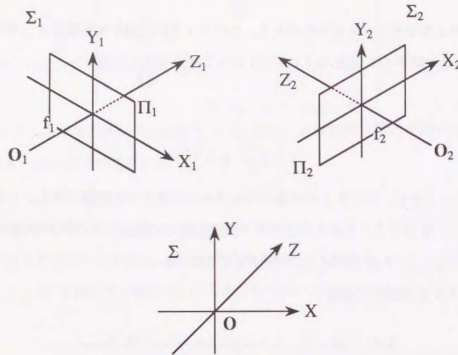


図 B.2: 複数カメラの固有座標系

$$(\mathbf{a} \times \mathbf{b}) \cdot \hat{\mathbf{x}} = 0 \quad (\text{B.5})$$

ここで、図 B.2 に示すように、2つのカメラ C_1, C_2 を考え、それぞれの画面を Π_1, Π_2 、得られた画像を I_1, I_2 とおく。また、それぞれのカメラに固有の座標系を Σ_1, Σ_2 とおく。このとき、3次元空間中のある座標系 Σ を考え、3次元空間中の1点をこの座標系で表したものを X 、同じ点を座標系 Σ_1, Σ_2 で表現したものをそれぞれ X_1, X_2 とする。いま座標系 Σ を用いて表した座標系 Σ_1, Σ_2 の原点を O_1, O_2 とすると、 X_1, X_2 は回転行列 R_1, R_2 を用いて次のように表すことができる。

$$X_1 = R_1(X - O_1) \quad (\text{B.6})$$

$$X_2 = R_2(X - O_2) \quad (\text{B.7})$$

したがって

$$X_1 = R_1 R_2^T X_2 + R_1(O_2 - O_1) \quad (\text{B.8})$$

$$X_2 = R_2 R_1^T X_1 + R_2(O_1 - O_2) \quad (\text{B.9})$$

が成立する。いま3次元空間中のある点 P の I_2 上の座標値が (p_2, q_2) であったとする。このとき $P_2 = (p_2, q_2, f_2)$ とおくとき P は座標系 Σ_2 では $X_2(t) = P_2 t$ と表される直線上に存在する。

この直線は座標系 Σ_1 では、式 (B.8) より

$$X_1(t) = R_1 R_2^T P_2 t + R_1(O_2 - O_1) \quad (\text{B.10})$$

と表すことができる。いま

$$\mathbf{a}_1 = R_1 R_2^T P_2, \quad \mathbf{b}_1 = R_1(O_2 - O_1) \quad (\text{B.11})$$

とおくと、この直線の I_1 上での像は式 (B.5) より

$$(\mathbf{a}_1 \times \mathbf{b}_1) \cdot \hat{\mathbf{x}}_1 = 0 \quad (\text{B.12})$$

と表すことができる。したがって、カメラ C_2 で撮影した座標系 Σ_2 による点 (p_2, q_2) に対応するカメラ C_1 で撮影した画像上の点を探索するためには、 I_1 上で、式 (B.12) で示される直線上を探索すればよいことがわかる。この直線はエビボア線と呼ばれる。

このとき、エビボア線が画面上に導入された座標軸 (x 軸) と平行すなわち水平になる場合には、対応点探索のためのアルゴリズムを単純にできる。いま、 $\mathbf{c} = \mathbf{a} \times \mathbf{b} = (c_1, c_2, c_3)$ とおき、画面上の直線 $\mathbf{c} \cdot \hat{\mathbf{x}} = c_1 x + c_2 y + c_3 f = 0$ を考える。この直線が x 軸と平行となるのは、 $c_1 = 0$ のとき、すなわちベクトル $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ と $\mathbf{e}_1 = (1, 0, 0)^T$ が直交する場合である。 $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ であるから、 $\mathbf{c} \cdot \mathbf{a} = \mathbf{c} \cdot \mathbf{b} = 0$ が成立する。したがって、 \mathbf{a}, \mathbf{b} いずれかの方向ベクトルが $\mathbf{e}_1 = (1, 0, 0)^T$ と一致する場合には、この直線は画面上で水平になることがわかる。いまエビボア線の式 (B.12) を考えると、式 (B.11) より $\mathbf{a}_1 = R_1 R_2^T P_2, P_2 = (p_2, q_2, f_2)$ であり、 (p_2, q_2) はさまざまな値を取り得るから、結局エビボア線が水平となるのは、 $\mathbf{b}_1 = R_1(O_1 - O_2)$ が \mathbf{e}_1 と平行な場合である。したがって、画面 I_1, I_2 についてエビボア線が共に水平となるのは、

$$R_1(O_1 - O_2) // \mathbf{e}_1, \quad R_2(O_1 - O_2) // \mathbf{e}_1 \quad (\text{B.13})$$

が成立する場合、すなわち

$$O_1 - O_2 // R_1^T \mathbf{e}_1, \quad O_1 - O_2 // R_2^T \mathbf{e}_1 \quad (\text{B.14})$$

となる場合である。式 (B.6) - (B.9) より、 $R_1^T \mathbf{e}_1, R_2^T \mathbf{e}_1$ はそれぞれ座標系 Σ で見たカメラ1、カメラ2に付随した x 軸 (x_1 軸、 x_2 軸) 方向の方向ベクトルに等しい。したがって、エビボア線が水平となるのは、2つの座標系 Σ_1, Σ_2 の x 軸の方向ベクトルが等しく、かつ、2つのカメラ系の射影の中心点を結んでできるベクトル $O_1 - O_2$ と、各カメラの座標系の x 軸の方向が等しいときである。

カメラ位置を任意に設定した場合には、上の条件は撮影された画像について常に成立する訳ではない。しかしながら、画像変換によってエピソード線が水平になるようにすることができる。画面 I_1 上にある特徴点 $x_1 = (x_1, y_1)$ は 3 次元空間中では 1 本の直線に相当する。したがって、画面 I_1 を点 Q_1 の回りで回転した仮想画面 I'_1 を考えると、特徴点 x_1 に相当する I'_1 上の点は、画面 I'_1 と直線の交点によって求めることができ、これにより I_1 上の原画像を仮想画面 I'_1 上の画像に変換することができる。いま、仮想画面 I'_1 として画面 Π_1 を Q_1 の回りで回転し、仮想画面上の x 軸がベクトル $O_1 - O_2$ と平行になるようにとり、同様の変換を I_2 についても行えば、変換後の画像ではエピソード線は共に水平とすることができる。

C 積に関する Gaussian HBF, Flexible Fourier, 指数関数の性質

第7章で用いた近似方式のうち、Gaussian HBF, Flexible Fourier, 指数関数については、これらの方式によって生成される関数全体の集合は積に関して閉じていることを示すことができる。Flexible Fourier, 指数関数については、

$$\cos \theta \cos \phi = \frac{1}{2} \{ \cos(\theta + \phi) + \cos(\theta - \phi) \} \quad (C.1)$$

$$e^{ax} e^{bx} = e^{(a+b)x} \quad (C.2)$$

が成立するから、積に関して閉じていることは明らかである。Gaussian HBF に関して、この性質を示すために、いま、任意の Gaussian HBF $h_1, h_2 \in H$,

$h_1 = \sum_{\alpha=1}^{K_1} c_{1\alpha} \exp(-\|W_1(x - t_{1\alpha})\|^2)$ を考える。 $h_1 \cdot h_2 \in H$ を示すためには

$$e^{-\|W_1(x-t_1)\|^2} e^{-\|W_2(x-t_2)\|^2} = ce^{-\|W_0(x-t_0)\|^2} \quad (C.3)$$

とできること、即ち、

$$\|W_1(x - t_1)\|^2 + \|W_2(x - t_2)\|^2 = \|W_0(x - t_0)\|^2 + c' \quad (C.4)$$

とできることを示せばよい。いま、

$$W = W_1^T W_1 + W_2^T W_2, \quad W > 0 \quad (C.5)$$

とすると、 W はユニタリ行列 U で対角化でき、かつ、 W は正定値であるから

$$W = U^T \Lambda U, \quad \Lambda = \text{diag}(\lambda_1^2, \dots, \lambda_d^2) = \tilde{\Lambda}^2 \quad (C.6)$$

とできる。したがって、

$$W_0 = \tilde{\Lambda} U, \quad t_0 = W^{-1}(W_1^T W_1 t_1 + W_2^T W_2 t_2) \quad (C.7)$$

とすることにより、Gaussian HBF は式 (C.3) を満足すること、即ち、積に関して閉じていることが分かる。

D 正規化入力と等価な各種入力表現

MLP の有効性を示すものとして、広く知られている研究の中には、入力表現として、正規化入力とも見なすことができる方式を用いているものが存在する。例えば、文字列からの音声合成システムに MLP を適用した NETalk[96] が 1 つの例である。NETalk においては、入力された 7 つの文字列から、文字列の中央の文字の発音を出力する写像が学習すべき対象となる。NETalk においては、入力層は、29 個のユニットから成る 7 つのグループによって構成されている。各グループはそれぞれ入力された文字に対応しており、各グループのユニットは 1 つのアルファベットに対応している。したがって、入力文字列が与えられると、それぞれのグループ内のユニットのうち、ただ 1 つだけが値 1 を持ち、他のユニットは 0 となる。これをベクトル表示すると、

$$x = (\underbrace{0, \dots, 1, \dots, 0}_{\text{group 1}}, \dots, \underbrace{0, \dots, 1, \dots, 0}_{\text{group 7}}) \quad (D.1)$$

となり、常に $\|x\|^2 = 7 = \text{const.}$ が成立するから、これは正規化入力の 1 種であると思えることができる。

さらに、次のような 2 値入力表現

$$x = (x_1, \dots, x_d), \quad x_i = \begin{cases} \alpha \\ \beta \end{cases} \quad (D.2)$$

を考える。このとき、

$$\tilde{x} = \frac{a}{\sqrt{d}} x + \frac{b}{\sqrt{d}} \mathbf{1} \quad (D.3)$$

を導入する。ここに、

$$a = \frac{2}{\alpha - \beta}, \quad b = -\frac{\alpha + \beta}{\alpha - \beta} \quad (D.4)$$

である。 \tilde{x} は $\|\tilde{x}\| = 1$ を満足する。これは、2 値表現 x は正規化表現 \tilde{x} と等価と見なせることを示している。このとき、以下の関係式が成立する。

$$w \cdot x + \theta = \tilde{w} \cdot \tilde{x} + \tilde{\theta}, \quad \mu^2 \|x - t\|^2 = \tilde{\mu}^2 \|\tilde{x} - \tilde{t}\|^2 \quad (D.5)$$

ここに、

$$\tilde{w} = \frac{\sqrt{d}}{a} w, \quad \tilde{\theta} = \theta - \frac{b}{a} \sum_{i=1}^d w_i, \quad \tilde{\mu} = \frac{\sqrt{d}}{a} \mu, \quad \tilde{t} = \frac{a}{\sqrt{d}} (t + \frac{b}{a} \mathbf{1}) \quad (D.6)$$

である。

E 2次元像の線形結合による3次元物体の認識 - Ullman&Basriの結果 -

Ullman and Basri は、直交射影の下で、3次元物体は3次元モデルを持たなくても、その物体の複数視点からの2次元像の線形結合をモデルとして用いることにより認識できることを示した[100]。ここでは、彼らの結果を[23]に従ってまとめる。

E.1 モデル画像を用いた任意視点からの像の記述

3次元物体上の特徴点を $\tilde{\mathbf{X}}^{obj} = (x_1, y_1, z_1, \dots, x_n, y_n, z_n)^T \in \mathbf{R}^{3n}$ とする。この物体に対しては、3次元空間中で(回転等)線形変換が施されるものとする。物体に対する線形変換は、

$$L^{3n} = \begin{pmatrix} L & & \\ & \ddots & \\ & & L \end{pmatrix}, L = \begin{pmatrix} \ell_{11} & \ell_{12} & \ell_{13} \\ \ell_{21} & \ell_{22} & \ell_{23} \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix} \quad (\text{E.1})$$

で表わすことができる。 L は $3 \times 3 = 9$ だから、 L の作るベクトル空間は9個の基底で張られる。したがって、

$$L = \sum_{i=1}^9 a_i L_i^{3n} \\ \tilde{\mathbf{X}}^{obj} = L^{3n} \mathbf{X}_0 = \sum_{i=1}^9 a_i L_i^{3n} \mathbf{X}_0 = \sum_{i=1}^9 a_i \tilde{\mathbf{X}}_i \quad (\text{E.2})$$

となり、 $\tilde{\mathbf{X}}^{obj}$ の作る空間は9次元ベクトル空間になることが分かる。即ち、任意の3次元特徴点は9-viewsで合成可能であることが分かる。

3次元空間中の点は直交射影によって画面上に投影されるものとし、この射影を $P: \mathbf{R}^{3n} \rightarrow \mathbf{R}^{2n}$ で表す。

$P\tilde{\mathbf{X}} = \tilde{\mathbf{x}}: P: 3n \times 2n$ matrix

$$\tilde{\mathbf{X}}^{obj} = L^{3n} \tilde{\mathbf{X}}_0^{obj}, P\tilde{\mathbf{X}}^{obj} = PL^{3n} \tilde{\mathbf{X}}_0^{obj} \quad (\text{E.3})$$

であるから、

$$PL^{3n} = \begin{pmatrix} \tilde{L} & & \\ & \ddots & \\ & & \tilde{L} \end{pmatrix}, \tilde{L} = \begin{pmatrix} \ell_{11} & \ell_{12} & \ell_{13} \\ \ell_{21} & \ell_{22} & \ell_{23} \end{pmatrix} \quad (\text{E.4})$$

となり、 PL^{3n} は6次元ベクトル空間をなすことが分かる。これにより、前と同様にして、

$$\tilde{\mathbf{x}}^{obj} = \sum_{i=1}^6 a_i \tilde{\mathbf{x}}_i^{obj} \quad (\text{E.5})$$

とできる。平行移動を含む場合には次の2つの基底を追加すればよい。

$$\tilde{\mathbf{t}}_x = (1 \ 0 \ 1 \ 0 \dots)^T, \tilde{\mathbf{t}}_y = (0 \ 1 \ 0 \ 1 \dots)^T \quad (\text{E.6})$$

上の結果より、任意の PL^{3n} は、適当な6つのオペレータ(基底) $\{PL_i^{3n}\} (i=1, \dots, 6)$ を用いて

$$PL^{3n} = \sum_{i=1}^6 a_i PL_i^{3n} \quad (\text{E.7})$$

このとき、係数 $\{a_i\}$ は \tilde{L} のみによって決まる。いま、異なる2つの物体の特徴点 $\tilde{\mathbf{X}}^{obj}, \tilde{\mathbf{Y}}^{obj}$ を考える。 $\tilde{\mathbf{X}}^{obj} = L^{3n} \tilde{\mathbf{X}}_0^{obj}, \tilde{\mathbf{Y}}^{obj} = L^{3n} \tilde{\mathbf{Y}}_0^{obj}$ とすると、式(E.7)より、 $\tilde{\mathbf{X}}_0^{obj}, \tilde{\mathbf{Y}}_0^{obj}$ に同じオペレータ PL_i^{3n} を作用させて得られる像、即ち、 $\tilde{\mathbf{X}}_0^{obj}, \tilde{\mathbf{Y}}_0^{obj}$ を同じ方向から見た像 $\tilde{\mathbf{x}}_i^{obj}, \tilde{\mathbf{y}}_i^{obj}$ を用いれば、同じ係数 $\{a_i\}$ を用いて、

$$\tilde{\mathbf{x}}^{obj} = \sum_{i=1}^6 a_i \tilde{\mathbf{x}}_i^{obj}, \tilde{\mathbf{y}}^{obj} = \sum_{i=1}^6 a_i \tilde{\mathbf{y}}_i^{obj} \quad (\text{E.8})$$

と表現することができる。

さらに、次のような線形変換 T を考える。

$$T\tilde{\mathbf{X}} = (x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n)^T \quad (\text{E.9})$$

$$\begin{array}{ccc} \tilde{\mathbf{X}} & \xrightarrow{L^{3n}} & L^{3n} \tilde{\mathbf{X}} \\ T \downarrow & & \downarrow T \\ T\tilde{\mathbf{X}} & \xrightarrow{L'} & TL^{3n} \tilde{\mathbf{X}} \end{array}$$

このとき、

$$L' = TL^{3n}T^{-1}$$

$$L' = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}, I_{ij} = \ell_{ij}I \quad (\text{E.10})$$

である。いま、次のような3次元(線形)変換を考える

$$\begin{cases} x'_i = \ell_{11}x_i + \ell_{12}y_i + \ell_{13}z_i \\ y'_i = \ell_{21}x_i + \ell_{22}y_i + \ell_{23}z_i \\ z'_i = \ell_{31}x_i + \ell_{32}y_i + \ell_{33}z_i \end{cases} \quad (\text{E.11})$$

このとき $\xi = (x_1, \dots, x_n)^T, \eta = (y_1, \dots, y_n)^T$ とおくと、標準位置からの変換の形で、

$$\begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix} \begin{pmatrix} \xi_0 \\ \eta_0 \\ \zeta_0 \end{pmatrix} \quad (\text{E.12})$$

と表すことができる。したがって、画面上の像に関して、

$$\xi = I_{11}\xi_0 + I_{12}\eta_0 + I_{13}\zeta_0 = \sum_{i=1}^3 \ell_{1i}\xi_i$$

$$\eta = I_{21}\xi_0 + I_{22}\eta_0 + I_{23}\zeta_0 = \sum_{i=1}^3 \ell_{2i} a_i \quad (\text{E.13})$$

が成立する。ここに、 $a_1 = \xi_0$, $a_2 = \eta_0$, $a_3 = \zeta_0$ である。これにより、それぞれ ξ , η で張られるベクトル空間 V_x, V_y を考えると、

$$\dim V_x = \dim V_y = 3, V_x = V_y \quad (\text{E.14})$$

となることが分かる。したがって、式(E.5)と同様にして、各特徴点を x 座標値、 y 座標値に分けて物体を表現した場合、各特徴ベクトル(像の表現) ξ , η は、同じ物体を異なる視点から見た3つの像に相当する特徴ベクトルの線形結合により

$$\xi = \sum_{i=1}^3 a_i \xi_i, \quad \eta = \sum_{i=1}^3 b_i \eta_i \quad (\text{E.15})$$

と表現できる。また、また、式(E.8)と同様にして、異なる物体A, Bの像の x 座標値を並べた表現 ξ_A, ξ_B を考えると、これらは、A, Bを同じ方向から見たときの像 ξ_{Ai}, ξ_{Bi} を考えると、このような3つの像をとれば、同じ係数 $\{a_i\}$ を用いて、

$$\xi_A = \sum_{i=1}^3 a_i \xi_{Ai}, \quad \xi_B = \sum_{i=1}^3 a_i \xi_{Bi} \quad (\text{E.16})$$

とできる。明らかに、 η に関しても同じ結果が成立する。

E.2 剛体、直交射影の場合の標準像への変換写像

物体が剛体かつ、3次元空間から2次元画面への射影が直交射影である場合には、任意視点からの像 \tilde{x} は、いくつかの視点からの像 \tilde{x}_i ($i = 1, \dots, k$)の線形結合を用いて

$$\tilde{x} = \sum_{i=1}^k a_i \tilde{x}_i, \quad \tilde{x}_i \in \mathbf{R}^{2N} \quad (\text{E.17})$$

と表すことができる。いま、すべての \tilde{x}_i ($i = 1, \dots, k$)と直交するベクトル集合を $\{\tilde{u}_j\}$ ($j = 1, \dots, 2N - k$)とする。このとき、

$$\begin{aligned} \tilde{L} \tilde{x}_i &= 0 \quad (i = 1, \dots, k) \\ \tilde{L} \tilde{u}_j &= \tilde{u}_j \quad (j = 1, \dots, 2N - k) \end{aligned} \quad (\text{E.18})$$

なる線形変換 \tilde{L} を考える。この \tilde{L} により、式(E.17)より、この物体の任意の像 \tilde{x} について、

$$\tilde{L} \tilde{x} = \sum_{i=1}^k a_i \tilde{L} \tilde{x}_i = 0 \quad (\text{E.19})$$

が成立する。したがって、この物体の標準像を \tilde{x}_s とすると、

$$T[\tilde{x}] = \tilde{x}_s - \tilde{L} \tilde{x} \quad (\text{E.20})$$

で定義される変換 T は、この剛体物体の任意の像を標準像に変換する。 \tilde{L} は、与えられた物体の像 $\{\tilde{x}_i\}$ から式(E.18)を解いて求めることができるから、この写像 T は与えられた例題から求めることができる。このことがわかる。

研究発表一覧

●寄稿論文、解説論文

- [1] 丸山, 甘利, "物体の運動と立体構造の単眼視による認識の基礎", 信学論 (A), Vol.J67-A, No.12, pp. 1107 - 1114 (1984).
- [2] 丸山, 阿部, "立体視に基づく面の抽出と面内ベクトルの認定による多面体構造の抽出", 信学論 (D), Vol.J70-D, No.12, pp.2581 - 2589 (1987).
- [3] 丸山, 阿部, "ランダムな切断を持つスリットパターン投影による形状計測", 信学論 (D), Vol.J71-D, No.12, pp.2596 - 2603 (1988).
- [4] 丸山, 寺岡, 阿部, "学習ネットワーク GRBF の変形する 3 次元物体の認識能力", 信学論 (DII) (論文採録).
- [5] M.Maruyama and S.Abe, "Range sensing by projecting multiple slits with random cuts", IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 15, No.6, pp. 647 - 651 (1993).
- [6] M.Maruyama, T.Teraoka, S.Abe, "Recognition of 3D flexible objects by GRBF", Biological Cybernetics (to appear).
- [7] 丸山稔, "Radial Basis Functions を用いた学習ネットワーク - ニューロコンピューティングに対する新しいアプローチ -", システム/制御/情報, Vol.36, No. 5, pp. 38 - 45 (1992).

●国際会議

- [8] M.Maruyama and S.Abe, "Acquiring a polyhedral structure through face extraction and verification", Proc. 9th International Conference on Pattern Recognition pp. 579 - 581 (1988).
- [9] M.Maruyama and S.Abe, "Range sensing by projecting multi-slits with random cuts", Proc. International Workshop on Industrial Applications of Machine Intelligence and Vision, pp. 163 - 168 (1989).
- [10] M.Maruyama and S.Abe, "Range sensing by projecting multiple slits with random cuts", SPIE Vol.1194 Optics, Illumination, and Image Sensing for Machine Vision IV, pp. 216 - 224 (1989).
- [11] M.Maruyama and T.Teraoka, "Recognition of 3D nonrigid objects by learning view change transformations", Proc. International Joint Conference on Neural Networks '93 (1993) (to appear).

●研究会

- [12] 丸山, 甘利, "動的物体の認識についての基礎的考察", 信学技報 PRL83-64 (1984).
- [13] 丸山, 阿部, "立体視による建築物の多面体構造の抽出", 信学技報 PRU86-82 (1987).
- [14] 丸山, Girosi, Poggio, "正規化入力に対する GRBF と多層パーセプトロンの能力比較", 信学技報 NC91-73 (1992).
- [15] 丸山, Girosi, Poggio, "学習における種々の基底関数の能力比較", 信学技報 NC91-124 (1992).
- [16] 丸山, 寺岡, 阿部, "学習ネットワーク GRBF の変形する物体の認識能力", 信学技報 NC92-76 (1992).

●全国大会

- [17] 丸山, 阿部, "複数透視画像からの 3 D モデルの復元", 信学総全大, 1638, 分冊 6, p.280 (昭 61) (1986).
- [18] 丸山, 阿部, "法線ベクトル分布からの基本立体の抽出", 信学情報・システム部門全大, 分冊 1 p.120 (昭 62) (1987).
- [19] 丸山, 阿部, "距離画像からの物体の基本要素への分割", 信学春全大, D-509, 分冊 D-1, p.252 (昭 63) (1988).
- [20] 丸山, 阿部, "ランダムな切断を持つスリットパターン投影による距離計測", 信学秋全大, D-220 分冊 D-1, p.221 (昭 63) (1988).
- [21] 丸山, 寺岡, 阿部, "GRBF を用いた像推定に基づく 3 次元物体認識", 信学春全大, D-59 分冊 D-6, p.59 (平 5) (1993).

●その他

- [22] 丸山, 甘利, "神経場における情報表現細胞のトポロジカルな配置について", 信学技報 MBES4-116 (1985).
- [23] 丸山, 亀井, 阿部, "建築 CAD における MD 木を用いた 3 次元表示 - 座標変換による隠線消去の高速化 -", 信学技報 IES9-67 (1989).
- [24] 丸山, 阿部, "MD 木を用いた建築 CAD のための 3 次元表示", 情報全大 5K-7, Vol.II, pp. 917-918 (平元) (1989).
- [25] 丸山, 玉田, 中村, 阿部, "Memory Based Learning を用いた配水量予測", 電学情報処理研究 IP-92-4 (1992).

[26]S.Amari and M.Maruyama,"A theory on the determination of 3D motion and 3D structure from features", Spatial Vision, Vol.2, No.2, pp. 151 - 168 (1987).

[27]S.Amari and M.Maruyama,"On the topological representation of signals in self-organizing nerve fields", In E.Teramoto and M.Yamaguti Eds. Mathematical topics in population biology, morphogenesis and neurosciences, (Springer Lecture Notes in Biomathematics) pp.282 - 291, Springer (1987).

[28]M.Maruyama, F.Girosi and T.Poggio,"Techniques for learning from examples : numerical comparisons and approximation power", MIT Artificial Intelligence Lab. Memo No.1290 (1991).

[29]M.Maruyama, F.Girosi and T.Poggio,"A connection between GRBF and MLP", MIT Artificial Intelligence Lab. Memo No. 1291 (1991).

