

適応学習型汎用認識システム：ARGUS

— その理論的構成と応用 —

大津 展之

近年、映像の監視や目視検査等、さまざまな分野で視覚システムのニーズが高まっている。特に、簡便で高速な実用的な視覚システムの実現が望まれている。この論文では、その目標に向けて筆者がこれまで行ってきた理論研究とその応用について概説する。まずこれまでのアプローチの問題点を指摘し、基礎としてのパターン認識の基本的な枠組、特に特徴抽出理論について言及する。次にその実践として提案した高次局所自己相関と多変量解析手法の2段階の特徴抽出からなる適応学習型汎用認識方式と、その応用事例を示す。実験結果は本方式の柔軟で効果的な性能を示している。

キーワード：視覚システム、パターン認識、特徴抽出、適応学習

ARGUS: Adaptive Recognition for General Use System

– Its theoretical construction and applications –

Nobuyuki OTSU

In recent years, the need for computer vision systems is increasing in various fields, such as security monitoring and visual inspection. It is crucial to realize simple and high-speed vision systems especially for practical usage. This paper addresses the author's theoretical research and its applications developed thus far in working toward this goal. First, the problems of the conventional approach are pointed out, and the general framework of pattern recognition, in particular the feature extraction theory, is explained as the theoretical foundation of the present research. Then a scheme of adaptive vision system with learning capability is presented, which comprises two stages of feature extraction, namely, Higher-order Local Auto-Correlation and multivariate data analysis. Several applications are demonstrated, showing the flexible and effective performance of the proposed scheme.

Keywords : Vision system, pattern recognition, feature extraction, adaptive learning

1 はじめに

近年、視覚システム（コンピュータービジョン）への期待が大きいの。防犯分野での監視カメラ、生産分野での製品の外観検査、医療分野でのCT画像解析や組織検査、スポーツ分野での動作解析や評価、さらにはインターネットでの画像検索、ロボットの視覚等、多岐の分野にわたっている。その背景には、CCDカメラや各種センサー技術の発達、さらにはコンピューターや可視化技術の発達により、さまざまな画像の収集と処理が容易になっていることが挙げられる。

これらのニーズに呼応して、画像認識の研究が国際的にも盛んに行われているが、いまだ自動化や実用化は難しく、個別のアドホックな手法や専用の高価なシステムとなっていて、実際の場面ではいまだ人の能力に頼っているのが現状である。できればPCベースでの簡便で安価な、しかも高速で汎用性の高い柔軟な視覚システムの実現と普及が

強く望まれている。

この論文では、この目標に向けて筆者がこれまで行ってきたパターン認識の理論^[1]、特に特徴抽出理論^[2]と、それに基づく実践的なシステム構成方式として提案した適応学習型汎用認識システム^{[3][4]}、およびそのさまざまな応用展開について論ずるとともに^{[5][6]}、認識（一般に情報）システムの構成法においては特に理論的アプローチが有効かつ重要であることを示す。

2 従来方式とパターン認識

まず、パターン認識としての画像計測や画像認識の問題を考えてみよう。図1-a)は、大小2種類の粒子(円)があって、それぞれの個数を数え上げる画像計測(計数)の課題である。通常考える方式は、次のような逐次方式であろう。まず画面をスキャンして個々の粒子を切り出し、個別に円で近似してその半径を測り、半径の大小によって粒子

の大小を判定して数え上げる。しかし、これでは、明らかに計算時間は対象の個数に比例して増大してしまう。

一方、図1-b)は、それぞれの対象（動物）が何であるかを答える画像認識の問題である。通常、これら四つの対象を識別する（部分的な）特徴は何かと考えるであろう。それぞれのモデルに照らして、耳や尾、体型等を、部分画像として（あるいは数量化して）照合し、最終判断を行う。しかし、全体の認識が部分の認識に帰着されていて、部分の認識を誤れば全体の認識を誤る。

従来方式のほとんどは、このように、まず画像から個々の対象を切り出し、あらかじめ用意したモデルに照らして認識を行う「逐次手順型」の方式である。しかし、一般にパターンはさまざまな変形を伴うので、モデルもそれに応じてさまざまに複雑なものを用意しなければならない。また、逐次手順では各段での処理の誤差が累積するため全体として脆弱となり、計算量も多く、実用に足る認識性能を得ることは難しい。画像レベルでアドホックに論理的な手順として考えがちな所に問題がある。ある意味で、ノイマン型計算機のプログラミングのパラダイムに支配されたアプローチである。

これに対するアンチテーゼとして、1980年後半からニューラルネットによる「並列学習型」の方式が提唱され^[7]、その理論的性質の研究と共に、特にパターン認識や制御へのさまざまな応用が試みられてきた。しかし、素子の非線形や値域限定 [0,1] の制約から、一般には情報表現や特徴抽出の問題が曖昧となる嫌いがあり、さらにはモデルの恣意性や学習の速度と収束の問題等もあるため、近年はカーネル法^[8]等の非線形多変量解析へと流れが変わってきている。

視覚システム、一般に認識システムの新たな方式を考察するためには、その基礎となるパターン認識の一般的な枠組、特に情報表現と特徴抽出について、理論的枠組から再考する必要がある。

2.1 パターン認識の一般的な枠組

パターン認識では、一般に時空局在的な関数 f で表される信号としてのパターンから、認識に有効な何らかの特徴

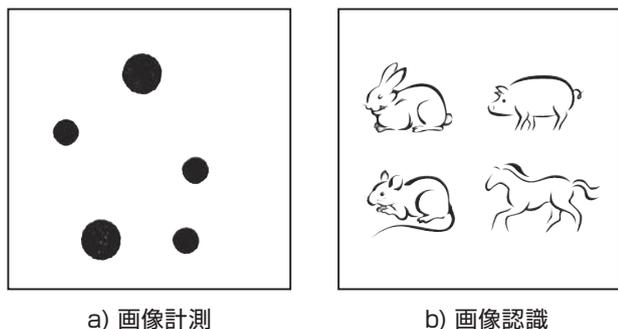


図1 視覚認識課題の例^[6]

値（一般に関数の関数としての汎関数 $x_i = \phi_i[f]$ ）を複数抽出して（したがってベクトル x で表現して）認識を行う。通常、これは図2に示すように、「特徴抽出」と「認識」の2段階処理の枠組として考えられている。認識には識別と類別とがある。識別は、入力パターンが既知の概念のいずれかを判定することであり、学習段階で答えが与えられる意味で教師有り学習と呼ばれる。類別は、教師無し学習と呼ばれ、入力パターンを幾つかの類（概念）に別けて区別することである。識別に関してはすでに多くの手法が提案されていて、誤識別率最小の識別方式は、事後確率 $P(C_j|x)$ が最大となる概念 C_j に決定するベイズ決定方式であることがすでに理論的に知られている。したがって、その意味では、前段の特徴抽出が認識システムの性能を左右する要件として重要であるが、これまで認識課題に応じたさまざまなアドホック、あるいは heuristic な手法が提案されてきた。

2.2 特徴抽出理論

筆者は、これら特徴抽出の理論的な研究を行ってきた^[2]。特徴抽出の一般的な枠組としては、幾何学的な側面としての「不変特徴抽出」と統計的な側面としての「判別特徴抽出」があり、この順序でこれら2段階からなる特徴抽出が原理的に重要である。したがって、この理論から帰結されるパターン認識の一般的な枠組は、図3のようになる。

2.2.1 不変特徴抽出（幾何学的側面）

パターンとしての観測像 f は、対象と認識主体との相対的な位置関係や運動により、平行移動、大小伸縮、回転

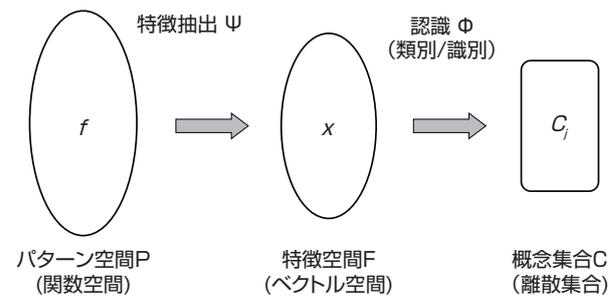


図2 パターン認識の一般的な枠組（通常）

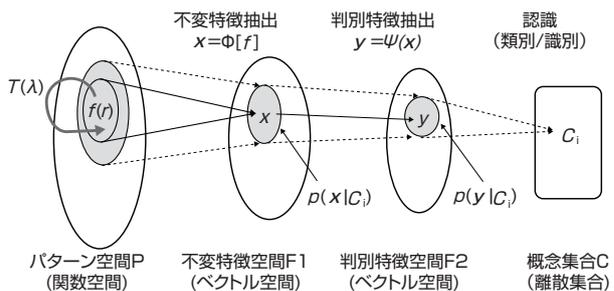


図3 パターン認識の一般的な枠組（詳細）

等さまざまな連続な幾何学的変換（一般には射影変換）を受けているが、認識結果はそれらによらず不変である。不変特徴抽出理論では、パターン関数 f に作用するこれらの概念対応を変えない幾何学的な変換（不変変換と呼ぶ）を作用素 $T(\lambda)$ で表し、その下で不変な特徴値、したがって不変汎関数 $x = \Phi[f]$ を追求する。

$$\Phi[T(\lambda)f] - \Phi[f] = 0 \quad (1)$$

Lie 群論に基づく作用素解析から、必要十分条件として導かれる偏微分方程式の基本解として、与えられた不変変換に対する不変特徴が求まる^{[11][2]}。これにより、パターンは、無関係な情報を捨象して認識に本質的な特徴として、理想的には、不変特徴ベクトル空間の1点 x として統一的に捉えられる。

2.2.2 判別特徴抽出（統計的側面）

しかし、実際のパターンはさまざまな変形やノイズを含み、概念のクラス C_j 毎に確率的な分布 $p(x|C_j)$ に従い分布する。次段の判別特徴抽出理論では、不変特徴ベクトル x から次元を縮小した新特徴ベクトル y への写像 $y = \Psi(x)$ を考え、概念クラスの判別等、 y に関する評価基準を最適化する最適写像を求める。線形写像の場合は、いわゆる多変量解析（例えば判別分析）となり、ある種の非線形写像の場合がニューラルネットやカーネル法である。

究極の最適非線形判別写像は、実は、変分法を用いて次式で陽に求められる^{[11][2]}。

$$y = \Psi_N(x) = \sum_{j=1}^K P(C_j|x) c_j \quad (2)$$

この結果は、パターン判別がベイズ事後確率 $P(C_j|x)$ と密接に関係し、パターン認識の背後のベイズ推定の本質的な枠組を示唆している。ここに c_j は写像先 Y での各概念を代表表現するベクトルであり、判別分析の場合、元の空間 X での概念クラス間の推移確率行列の固有ベクトルとして求まる。得られる最適判別空間の次元は本質的にクラス数から決まり、 $K-1$ 次元となることが分る。

実際の応用においては、これらの理論的な枠組を踏まえ

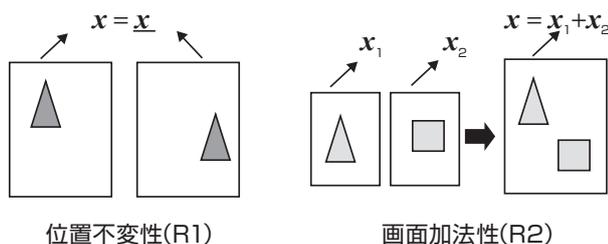


図4 位置不変性と画面加法性^[6]

て、実用ニーズに照らした適切な簡略化が必要である。

3 アプローチと構成法への条件

柔軟な視覚システムへのアプローチと構成法を考えるにあたり、まず、視覚システムに要請される基本的な条件として、次の3点を挙げた（図4）。

R1: 位置不変性、R2: 画面加法性、R3: 適応学習性。

認識 / 計測対象が画像枠のどこにあっても、その認識 / 計測の結果は変わらない。したがって、最初の条件 R1 は、パターンから抽出される特徴 x は対象の位置によらない（平行移動不変である）ことを要請している。不変変換としては、ほかにも大小スケールや回転等も考えられるが、平行移動が最も基本的であるので、これを要請条件とした。

次の条件 R2 は、画面全体に対する特徴が個々の対象の局所特徴の和になることを要請している。これは、R1 からの帰結でもあるが、特徴表現が認識（特に計数）にとって都合の良い表現（線形）となり、後の処理が簡単で高速になるための要請条件である。

最後の条件 R3 は、従来法のように特徴抽出がヒューリスティックな手順として与えられ、認識課題が変わると再度構成法も変わるというのではなく、例からの学習によって課題に適した新特徴 y が初期特徴 x から最適に自動構成（合成）され、課題の変化にも構造を変えることなく適応的に最適化される汎用的な方式であることを要請している。

さらに、これらの要請条件を満たすべく構成される特徴抽出法としては、もちろん計算量が少なく実時間処理が可能であることが望ましい。

4 適応学習型汎用認識システム

これらの基本要請条件を満たし、前記のパターン認識、特に特徴抽出理論の枠組を最も簡単な形で実践する「適応学習型汎用画像認識システム」^{注1}を考案した^{[3][4]}。本システムは、特徴抽出の理論的な枠組にしたがって2段階の特徴抽出からなっている（図5）。

4.1 不変特徴抽出 (HLAC/CHLAC)

初段の初期特徴すなわち幾何学的側面としての不変特徴抽出としては、最も基本的な平行移動（位置）不変な特徴を考えた。認識は基本的に時空間でのパターン $f(r)$ の位置 r によらないからである。

位置不変な特徴として、音声等時系列解析分野では古くから自己相関関数 $r(\tau) = \int f(t)f(t+\tau)dt$ が知られている。これは波形パターンの位置によらない相対的な関係を抽出している。これを高次へ拡張した N 次自己相関関数

$$x(a_1, \dots, a_N) = \int f(r) f(r+a_1) \cdots f(r+a_N) dr \quad (3)$$

が数学的に知られていて、その2次（一般に偶数次）の高次自己相関は位置不変の完全系となる等、パターン認識応用への幾つかの興味ある性質が論じられている^[9]。画像の場合は、 $f(r)$ は参照点（画素：pixel） r での濃淡値、 a_i は参照点 r 周りの相対的な変位である。しかし、 N 個の変位の組み合わせによって得られる特徴値の数は指数関数的に膨大となり、それらの計算はほとんど不可能となる。したがって実際の応用には、ある限られた次数、また変位の組み合わせを用いることになる。

実際、一般に実世界のパターンは時空局在化していて、局所的な相対的關係が本質的である。また、この局所限定は、画面加法性（R2）を満たすことにもなる。したがって、

（3）式の高次自己相関関数を局所変位に限定した高次（ N 次）局所自己相関（HLAC: Higher-order Local Auto-Correlation）特徴を、R1とR2を満たす非線形特徴として考案して採用した^{[3][4]}。

HLAC: 実際の2D画像（静止画像） $f(x, y)$ に対しては、次数を2次まで、局所近傍を 3×3 とすると、位置不変性を考慮して同値でない独立な積和の取り方に関する局所マスクのパターンは25通り（図6）となる。全画面（もしくは部分領域） XY に関して、図6で表される各局所マスクでス

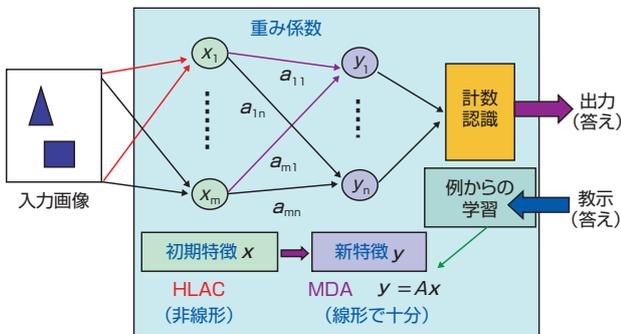


図5 適応学習型汎用認識システム（ARGUS）

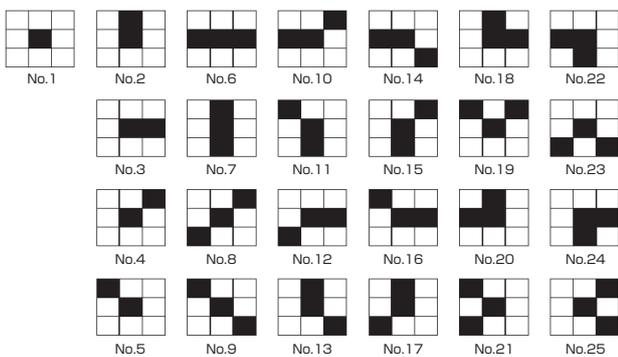


図6 2次までの局所 3×3 マスク^{[3][4]}

キャンして黒点に対応した画素値の積和を求め、HLAC特徴ベクトル x が得られる。その次元は、濃淡画像の場合には35次元（例えば、No.1のマスクに対しては、 f, f^2, f^3 が区別される）、2値（0/1、白/黒）画像の場合には、縮退して25次元となる（例えば、No.1のマスクに対しては、 $f = f^2 = f^3$ と冪等となる）^{注2}。

CHLAC: 動画（3D） $f(x, y, t)$ の場合は、2次元の静止画が時間軸に沿って並んだ3次元（立体） XYT の数値データであることから、HLACを時間軸を含めて自然に拡張したCHLAC（Cubic HLAC）特徴を抽出する^[11]。CHLACの局所 $3 \times 3 \times 3$ マスクの一例を図7に示す。独立な局所マスクのパターンは251通りとなる。HLACと同様に、立体枠 XYT にわたりそれらのマスクで積和を求め、CHLAC特徴ベクトル $x(t)$ が得られる。その次元は濃淡動画の場合は279次元、2値動画の場合は251次元となる。

積分特徴としてのHLAC（CHLAC）による特徴抽出方式は、要請条件のR1（位置不変性）とR2（画面加法性）を満たす基本的で汎用的な「対象の形（と動き）」の特徴抽出方式となっている。これらによって、認識対象は常に統一的に不変特徴空間における1点（ベクトル） x として捉えられ表現される。

4.2 判別特徴抽出（MDA）

次段の適応学習的な（R3を満たす）統計的判別特徴抽出としては、線形写像としての多種多様な多変量データ解析（MDA）手法を適用する（図8）。これは、HLACもしくはCHLAC特徴ベクトル x の要素の重み付き線形和として、与えられた認識課題に最適な新特徴 y を適応的に求めることであるが（R3：適応学習性）（図5）、線形写像であることから要請条件のR2（画面加法性）を確保している。

同様な方式としてニューラルネットがあるが、非線形ゆえにR2を保存しない。また、最適化に逐次解法を要し、計算時間も掛かる。これに対してMDAは、例からの学習によってタスクに最適な重みが解析的に陽な形で容易に求められる利点がある^{注3}。

4.3 認識システムARGUSの特長

これら2段の特徴抽出からなる本方式は、対象の切り出しや位置合わせが不要であり、対象に関する知識やモデ

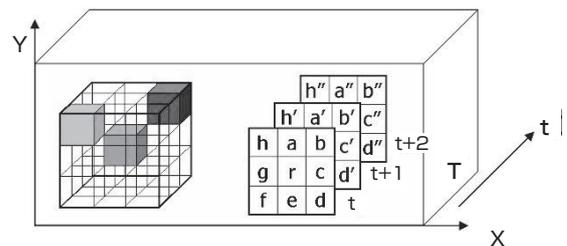


図7 CHLACマスクの例（ $hr'b''$ ）^{[5][11]}

ルを一切必要としない点で新しくユニークであり、静止画像や動画のさまざまな認識や計測・計数に応用できる汎用性をもっている。また、基本的に積和演算のみのため、CHLACでも通常のPCで非常に高速な処理（2 msec/frame）が可能である。

5 応用事例

5.1 複数対象の同時認識（計数）

静止画像の認識として、複数対象を同時認識して計数する課題への応用例を示す（図9）。HLAC特徴の持つ位置不変性と加法性から、因子分析（FA）を用いることで容易に実現することができる。図9の左図の各パターンを一度システムに提示すると、システムはテスト画像（右図）に対して瞬時にそれぞれの個数 y_i を $y = (F^T F)^{-1} F^T x$ として回答する。加法性により、右図全体の特徴ベクトル x は、各パターンの特徴（因子）ベクトル f_i の個数を係数とする線形和 $x = \sum_{i=1}^6 y_i f_i = [f_1, \dots, f_6] y = Fy$ に分解できるからである。

5.2 位相的特徴の認識（計数）

次に、形によらない認識の例として位相的特徴の計数結果を図10に示す。重回帰分析（MRA）を用いた例からの

学習により、システムは正しく対象の個数（a）あるいは穴の数（b）を答えている^[4]。興味深いことに、システムは例から位相幾何学の基礎となるオイラー数（点数-線数+面数）を学習し^{注4}、認識に用いていた。

5.3 顔認識と表情認識

2値画像に限らず、HLACは濃淡画像に対してもそのまま適用可能である。この応用として顔認識を行った^{[12][13]}。多解像度を表す画像ピラミッドの各層からのHLAC特徴を判別分析（DA）で統合することにより、簡単な識別法MDD^{注5}でも、119人に対して99%強の高認識率を達成した^[13]。さらに、困難とされる表情認識を行い、9名の7表情（JAFFE Dataset^[14]、図11）に対して、HLAC特徴の場所重みも考慮した判別分析とMDDを用いて、80%強の高い認識率が得られた^[15]。

5.4 動作と人の認識

HLAC特徴を動画の場合に自然に拡張したCHLAC特徴を用いて、動画からの対象と動きの認識が可能となる。5人の4動作（左/右方向への歩き/走り）の動画をフレーム差分後2値化し、CHLAC特徴を抽出し、動作と人それぞれに判別分析を適用した結果を図12に示す^[11]。それぞれの類（概念）が良く塊り分離されていて、CHLAC

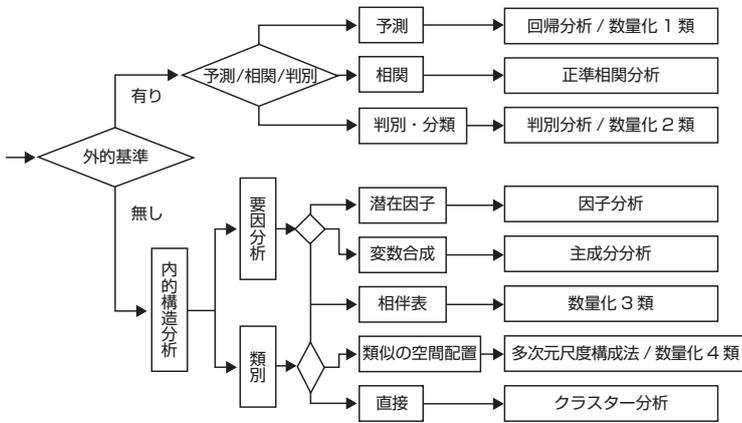
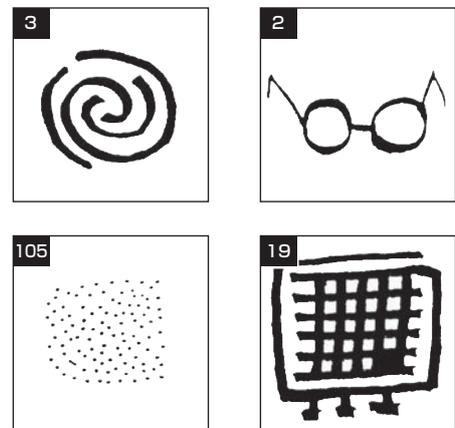


図8 多変量データ解析手法（目的別）

外的基準の有/無は教師有/無に対応する。数量化手法は質的データ（Yes/No, 1/0）の場合の手法である。



a) 対象の個数 b) 穴の数

図10 位相的特徴の認識（計数）^{[5][6]}

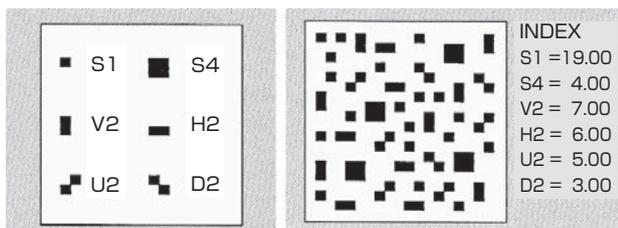


図9 複数対象の同時認識（計数）^{[5][6]}



図11 JAFFE 表情データの一部（3名）^[14]

特徴の有効性を示している。簡単な識別法 MDD でも、ほぼ 100 % の認識が得られた。

5.5 Gait認識

近年、遠方からの監視カメラによる個人同定（テロリスト等）のキーとして、gait（歩様）が注目されている。CHLAC + 判別分析 + k -NN 識別を、米国 NIST が取り纏めている 71 人の gait から個人を認識する Gait Challenge Dataset（図 13）に適用した結果、これまでの上位 5 位までの手法を大幅に上回る世界最高性能を達成した^[16]（図 14）。

5.6 異常検出

CHLAC は、画像中に複数の対象がある場合、それぞれの対象の特徴の和が全体の特徴となる加法性をもつので、通常（正常）動作の特徴ベクトルは、特徴空間（251 次元）のある線形部分空間（通常動作部分空間） S_N に分布することになる。したがって、常時学習（教師無し）で主成分分析（PCA）により S_N を求めておけば、異常行動は、あらかじめ定義する必要なく、 S_N からの逸脱（距離 / 角度）として直ちに高速かつ高精度に検出・認識される^[17]（図 15）。

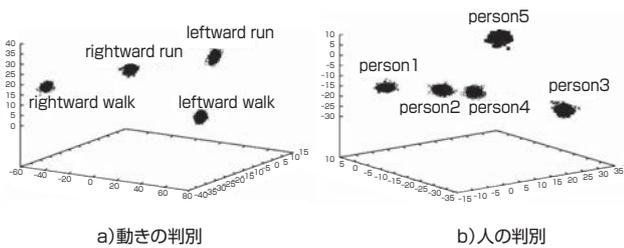


図 12 得られた判別特徴空間^[11]

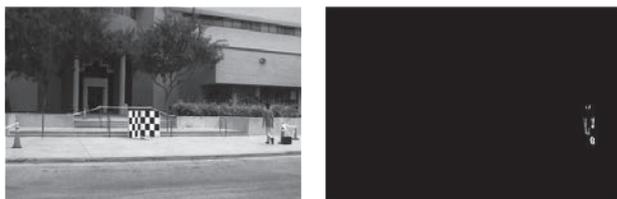


図 13 gait 動画とそのフレーム差分

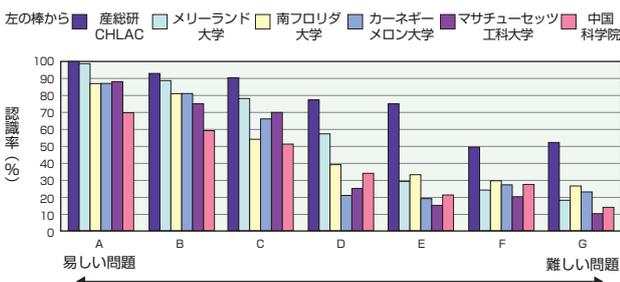


図 14 gait 認識の比較実験^{[5][6][16]}

加法性から複数人の場合でも異常検出力は同じである（図 16）。

この異常検出方式は、すでにエレベーター内の監視カメラ^{注6}に実用化されている。

通常例を CHLAC 特徴空間での統計的分布として学習し、そこからの逸脱（通常ではない）として異常を検知する本方式は、対象のモデルや知識を一切必要としない。したがって、監視カメラや車載カメラ等、映像からの異常検出のみならず、他のさまざまな異常検出に広く応用できる。例えば、静止画像の場合には HLAC 特徴空間を用いて、製造分野での半導体基板等の各種外観検査に応用できる（図 17）。

また、HLAC を用いた異常検出は医療分野での各種組織検査、特に癌の病理診断に同様に応用できる。癌は細胞の異常である。癌の病理診断は病理医が臓器組織の構造や細胞の変化の程度を顕微鏡で見て行うが、豊富な経験と知識が必要となる。しかし、経験豊かな病理医は不足しており、病理医の負担は増大している。したがって、そうしたスクリーニング検査の自動化による負担の軽減や、

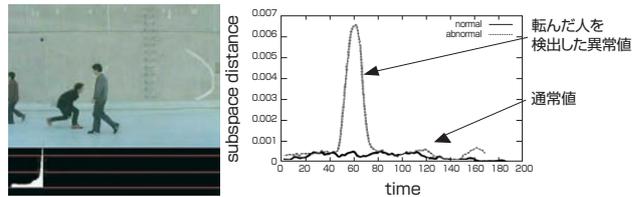


図 15 異常検出の例（ここでは「転ぶ」が異常）

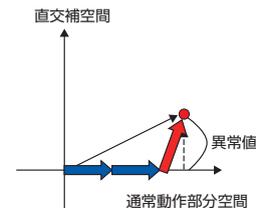


図 16 通常動作部分空間からの逸脱

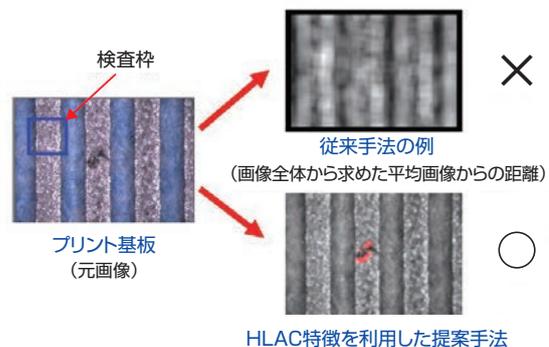


図 17 基板検査への応用例

ダブルチェックによる見落とし防止等、病理医を支援するシステム開発のニーズは大きい。実際の胃癌のリンパ節転移に対してこの方法を適用したところ、病理医の所見に近い解析結果を得ることができた^[18]（図18）。現在、癌病理診断支援システムの構築を目指し、大学病院や癌センターと共同研究を行っている。

5.7 時系列データ解析

画像に限らず、一般にセンシングデータは N 次元 (Ch.) の時系列データ $\{s_i(t)\}_{i=1}^N, t = 1, \dots, M$ で表される。これらを $N \times M$ の 2次元配列 (画像) とみなして HLAC 特徴を抽出することもできるが、一般に次元 (Ch.) の添え字 i の順序は随意である。そこで、例えば任意の 3 個の組み合わせをとると、 $K = {}_N C_3$ 個の 2次元配列 ($3 \times M$) が得られ、それぞれから 3×3 で HLAC 特徴をとると $K \times$ HLAC の次元の特徴ベクトルが得られる。これを多変量解析 (PCA, DA) することで、時系列データの解析 (異常検出や判別) ができる。この方法は心電図の異常検出^[19] や多自由度のロボット多指ハンドの動きの解析^[20] に応用されている。

また、時系列データ間の非対称な相互関係 (相関) として捉えられる因果関係は多くの分野で重要であり、線形自己回帰モデルを用いた解析指標として Granger Causality^[21] が提案されているが、これを多項式回帰モデル (したがって高次局所自己相関特徴が関係する) へ拡張した^[22]。さらに因果関係の有無を示す重み関数 $w(t)$ (Causality Marker) を導入し、因果関係の存在箇所を自動的に抽出する方式を提案した^[23]。

5.8 対応関係の学習

対応関係の学習は、一般的な幅広い応用に繋がる。パターン (静止画あるいは動画等) に対する人の判断や評価 (外的規準) の表現、例えばキーワードや印象 (感性) 語の質的表現 $y^{\text{質}}$ や評点 y と、パターンの特徴ベクトル表現 (HLAC/CHLAC) x との対応関係を学習的に近似 (正準相関分析 CCA や重回帰分析 MRA) することにより、印

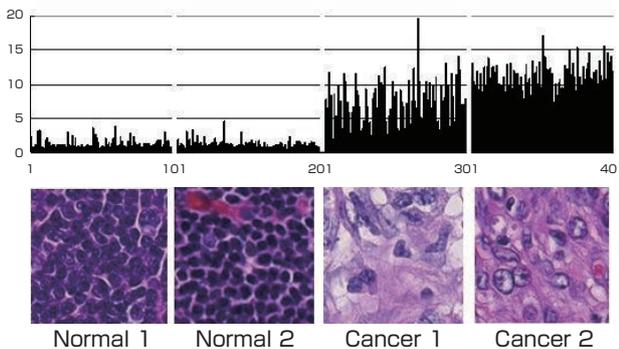


図18 癌検出への応用例 (上図は各標本の異常値)

象検索や双方向検索、さらに自動評価 (予測) が可能となる。図19に家紋の印象検索 (CCA)^[24] と運動の自動評価 (MRA)^[25] への応用を示す。

前者はさらに一般画像の検索 (annotation) に応用され^[26]、後者はスポーツ映像の自動インデキシング^[27] や、さらには超音波画像からの牛肉の肉質等級 (BMS) 判定にも応用されている^[28]。

6 理論的アプローチの有効性

以上、パターン認識における特徴抽出理論に基づき、基本要請条件を満たすべく構成された適応学習型汎用認識システム (ARGUS) とその応用、特に視覚システムとしての種々の実用について概説した。

物理や化学の科学的アプローチ (現象解明) と異なり、工学応用、とりわけ情報技術においては、機能実現に向けての構成法は自由度が高く、とかくアドホックで恣意的なものとなりがちであるが、応用ニーズの基本要請条件を押さえて理論的な視点から新規性のある本格的な解決を図ることが重要である。

本方式は、理論的な視点からパターン認識の基本的枠組を踏まえて、幾何学的不変特徴抽出としての高次局所自己相関 (HLAC/CHLAC) と統計的判別特徴抽出としての多変量解析手法の 2 段階からなり、後者によって課題に応じた例からの学習が可能となっている。対象に関するモデルや事前知識を一切必要とせず、対象パターンの形状とその動きを判別特徴空間の点として区別する。対象の切り出しも不要であり、計算量も一定の積和主体と少ないため、動画像に対しても通常の PC で実時間処理を遙かに超える高速処理が可能である。本方式の特長は次のとおりである。

- 非モデルベースな方式 → 高い汎用性
- 基本的な初期特徴 (HLAC/CHLAC) → 広範囲のデータ形式に適用可能
- 統計的な学習 (MDA) → 課題適応性と精度向上
- 並列積和演算 → 高速で大量データ処理が可能



図19 対応関係の学習^{[24][25]}

ほぼ想定どおり、種々の応用をとおしてこれまでの方式を上回る性能が得られている。これは、理論に裏付けされた方式、特に高次局所自己相関特徴の優位性とその本質性によるところが大と思われる。通常の自己相関が2点関係に留まるのに対し、3点関係へと高次化することにより、例えば静止画では輪郭の局所直線方向から曲率（凹凸）、動画像では速度から加速度へと、得られる特徴が詳細となっている。そして、これらの基本的で本質的な初期特徴が、個別の恣意的な逐次手順や論理判断（例えば閾値処理や条件分岐等）を用いず、多変量解析手法を用いて並列・総合的にタスクに有効な新特徴へと統合されていて、情報のロスが少なく頑健な方式となっている。

HLAC/CHLACは、時空間に局在する「パターン」の局所的形状パターンの統計量（相関や頻度）という基本的で汎用的な特徴である。その意味で、近年のHOGやSIFT特徴に代表される「モデル照合ベースから局所特徴の統計量へ」といった潮流の先鞭をなすものである。また、画像に限らず、音声や各種センサー情報等の多チャンネル時系列データ、さらには一般の3-wayデータへも広く応用可能である。今後の展開としては量的データから質的（カテゴリカル）データへの拡張が挙げられ、すでに手法^[29]の開発を行っている。

今後、ニーズの高い知的防犯カメラ等、セキュリティ分野における自動（無人）映像監視を始め、各種外観検査、画像のannotationと検索、ロボットの視覚、スポーツやリハビリ分野での動作解析や評価等、広くコンピュータービジョンへの応用が期待される。現在、医療応用として大病院や癌センターとの共同研究で、顕微鏡画像からの癌の自動検査システムへの応用を進めている。さらに、産総研認定ベンチャー（融合技術研究所）を中心に、半導体基板検査や各種外観検査への実用化や、地域コンソーシアムプロジェクトでの米の品質検査や乳牛の発情・分娩予知等、農業畜産分野への応用を展開している。

本方式の実用化においては、その他の細やかな事柄、例えば前処理やパラメータ（相関幅）の調整が必要であり、それらのノウハウの蓄積や自動化が今後の課題である。

謝辞

この研究は、併任先の東京大学知能機械情報専攻での卒論・修論指導のもと、大いに進展した。関係学生諸君に記して感謝する。特に小林匠氏（現：情報技術研究部門）には、その後の共同研究においても大いに貢献していただいた。ここに感謝したい。また、初期のHLACの実験で栗田多喜夫氏（現：広島大学）、近年の応用展開でポストドクの方々や坂上グループと樋口グループに貢献していただい

た。関係諸氏に感謝する。

注1) 当初、ギリシャ神話の百の目をもつ巨人にあやかって、ARGUS (Adaptive Recognition for General Use System) と呼ぶ予定であった。近年、本方式の略称にHLAC/CHLACが用いられる向きもあるが、それは初段の特徴抽出法を指すので適切ではない。したがって、今後はシステム/方式全体をARGUSと呼ぶことにする。

注2) 2値画像のHLACは、パーセプトロン^[10]における $N+1$ 位ベクトル、それによる図形のスペクトルと密接に関係している。ここでは、黒点(1)と白点(0)の組み合わせが考えられていて、我々の黒点のみを考える立場は一見不十分に見えるが、実は十分である。例えば $\blacksquare \square$ は、 $f_0 = f(\mathbf{r}) = 1$ かつ $f_1 = f(\mathbf{r} + \mathbf{a}_1) = 1$ として、論理的に $f_0 \cdot f_1 = f_0 \cdot (1 - f_1) = f_0 - f_0 \cdot f_1$ となり、マスク(No.1とNo.3)による特徴値の線形和の範囲で表される。

注3) この方式は、ニューラルネットの逆伝搬学習法^[7]より以前に提案されている^{[2][3]}。

注4) HLAC特徴は丁度それら位相幾何的要素の個数を数えあげていて、次段の重回帰でそれらの係数が学習的に決定されている。

注5) Minimum Distance Decision: 未知入力の特徴ベクトルからの各クラスの重心への距離を測り、最短のクラスへ識別する方式。

注6) ヘリオスウォッチャー (KK 日立ビルシステム)
http://www.hbs.co.jp/lineup/elevator/hw_outline.html

注7) 各該当語の有/無を1/0で表したベクトル。

参考文献

- [1] 大津展之, 栗田多喜夫, 関田巖: パターン認識-理論と応用, 行動計量学シリーズ12, 朝倉書店, 東京 (1996).
- [2] 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, 電総研究報告, 818, 210 (1981).
- [3] 大津展之, 島田俊之, 森俊二: N次自己相関マスクによる図形の特徴抽出, 電子通信学会技術報告, PRL-78 (31) (1978).
- [4] N. Otsu and T. Kurita: A new scheme for practical flexible and intelligent vision systems, *Proc. IAPR Workshop on Computer Vision (MVA1988)*, 431-435 (1988).
- [5] N. Otsu: Towards flexible and intelligent vision systems - from thresholding to CHLAC, *Proc. IAPR Conf. on Machine Vision Applications, Invited talk*, 430-439 (2005).
- [6] N. Otsu: CHLAC approach to flexible and intelligent vision systems, *Proc. ECSIS and IEEE Symposium on Bio-inspired Learning and Intelligent Systems for Security (BLISS2008), Invited talk*, 23-33 (2008).
- [7] D. Rumelhart, G. Hinton and R. Williams: Learning representations by back-propagating errors, *Nature*, 323 (9), 533-536 (1986).
- [8] J. Shawe-Taylor and N. Cristianini: *Kernel methods for pattern analysis*, Cambridge Univ. Press, Cambridge (2004).
- [9] J. McLaughlin and J. Raviv: N th-Order Autocorrelations in Pattern Recognition, *Information and Control*, 12, 121-142 (1968).
- [10] M. Minsky and S. Papert: *Perceptrons*, The MIT Press (1969).
- [11] T. Kobayashi and N. Otsu: Action and simultaneous multiple-person identification using cubic higher order local auto-correlation, *Proc. 17th Int. Conf. on Pattern Recognition (ICPR)*, 741-744 (2004).
- [12] T. Kurita and N. Otsu: Face recognition method using

- higher order local autocorrelation and multivariate analysis, *Proc. 11th ICPR*, 213-216 (1992).
- [13] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu: Face recognition system using local autocorrelations and multi-scale integration, *IEEE Trans. PAMI*, 18, 1024-1028 (1996).
- [14] M. Lyons and S. Akamatsu: Coding facial expressions with gabor wavelets, *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG1998)*, 200-205 (1998).
- [15] Y. Shinohara and N. Otsu: Facial expression recognition using Fisher weight maps, *Proc. 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG2004)*, 499-504 (2004).
- [16] T. Kobayashi and N. Otsu: A three-way auto-correlation based approach to human identification by gait, *6th IEEE Int. Workshop on Visual Surveillance*, 185-192 (2006).
- [17] T. Nanri and N. Otsu: Unsupervised abnormality detection in video surveillance, *Proc. IAPR Conf. on Machine Vision Applications*, 574-577 (2005).
- [18] 岩田健司他: 高次局所自己相関特徴法によるがん病理画像診断支援システム, *ViEW2009*, A-6H (I-15) (2009).
- [19] 荒木英人, 村川正宏, 小林匠, 樋口哲也, 久保田一, 大津展之: 高次局所自己相関特徴による多チャンネル時系列データからの異常検知, *電気学会論文誌C*, 129 (7), 1305-1310 (2009).
- [20] R. Fukano, Y. Kuniyoshi, T. Otani, T. Kobayashi and N. Otsu: Acquisition of unknown object property for manipulation by a compliant multi-fingered hand, *Journal of Robotics and Mechatronics*, 17 (6), 645-654 (2005).
- [21] C. W. J. Granger: Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424 (1969).
- [22] K. Ishiguro, N. Otsu, M. Lungarella and Y. Kuniyoshi: Comparison of nonlinear Granger causality extensions for low-dimensional systems, *Physical Review E*, 77 (issue 3), 036217 (1-9) (2008).
- [23] K. Ishiguro, N. Otsu, M. Lungarella and Y. Kuniyoshi: Detecting direction of causal interactions between dynamically coupled signals, *Physical Review E*, 77 (issue 2), 026216 (1-6) (2008).
- [24] 小林匠, 森崎巧一, 大津展之: 印象情報の付与による類似画像検索性能の評価, *電子情報通信学会論文誌*, J91-D (4), 1025-1032 (2008).
- [25] 森下雄介, 小林匠, 森崎巧一, 大津展之: 時間重みと外的規準を用いた動作評価手法, *電子情報通信学会技術研究報告*, PRMU-107 (539), 371-376 (2008).
- [26] H. Nakayama, T. Harada, Y. Kuniyoshi and N. Otsu: High-performance image annotation and retrieval for weakly labeled images, *Proc. Pacific-Rim Conf. on Multimedia*, 601-610 (2008).
- [27] F. Yoshikawa, T. Kobayashi, K. Watanabe and N. Otsu: Start and end point detection of weightlifting motion using CHLAC and MRA, *Proc. 1st Int. Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications*, 44-50 (2010).
- [28] T. Kobayashi, K. Watanabe, T. Higuchi, T. Miyajima and N. Otsu: Recognition of dynamic texture patterns using CHLAC features and linear regression, *International Journal of Database Theory and Application*, 2 (4), 13-26 (2009).
- [29] T. Kobayashi and N. Otsu: Image feature extraction using gradient local auto-correlations, *Proc. European Conf. on Computer Vision (ECCV)*, 346-358 (2008).

執筆者略歴

大津 展之 (おおつ のぶゆき)

1971年3月東京大学大学院工学系研究科計数工学数理コース専攻修士課程修了。同年4月電子技術総合研究所入所。パターン認識の理論と応用、特に特徴抽出理論とその実践としての画像認識の研究に従事。工学博士。1985年4月数理情報研究室長を経て、1990年4月首席研究官、1991年4月知能情報部長に就任。実世界情報処理(RWC)プロジェクト(1992-2001)の策定と特に実世界知能の研究推進に従事。2001年から産業技術総合研究所フェロー、現在に至る。1992年4月から2010年3月まで筑波大学連携大学院教授併任。2001年4月から2007年3月まで東京大学大学院情報理工学系研究科教授兼務。



査読者との議論

議論1 理論と応用と産業界への展開

質問 (赤松 幹之: 産業技術総合研究所ヒューマンライフテクノロジー研究部門)

ARGUSはしっかりとした理論に裏打ちされた手法であることから幅広く応用できた技術であると理解しています。こういった理論をベースにした研究のシンセシオロジーの論文として、理論ベースの研究のポイントや難しさ等を書いていただけませんか。また、種々の応用をしてみて、およそ理論どおりだったのか、理論どおりいなくて苦労したことはなかったのか等も記載していただけませんか。もし、前者のとおりだったのでしたら、なぜこの理論はうまくいったのか等を書いていただくと読者の大いなる参考になると思います。

回答 (大津 展之)

可能な範囲で対応いたしました。

議論2 要素技術の選択

質問 (上田 完次: 産業技術総合研究所)

この論文は、理論をベースとした技術の実問題への適用という第2種基礎研究であり、これまでにシンセシオロジーには掲載されていなかったタイプの研究の論文です。理論をベースとした構成的研究における要素技術の選択についてお伺いします。

実用化目標を達成するための構成要素をどのようにして選択したのでしょうか。既状態から演繹的に導かれる構成要素だけなのか、仮説的な構成要素があるのか等の説明をお願いします。

質問 (赤松 幹之)

3.1節において、R1からR3の要請条件を満たすシステムとして適応学習型汎用画像認識システムを開発したとあり、位置不変性を満たす特徴量を抽出する技術として、HLACとCHLACを採用したと述べられています。これを採用したプロセスにおいて、他にも候補となった技術があると思いますので、それらの技術と比較してHLACが優れていると判断した論拠を書いていただけませんか。

また、ここにおいてHLACを採用した理由として書かれているのは、パターンは局在していて局所的な相対関係が本質的であるという点です。局所の特徴だけを見ることが位置不変との関係等、専門外の読者にすぐに理解してもらえるかは分からないので、少し追記していただければと思います。

回答 (大津 展之)

認識システムでは対象パターンからの特徴抽出が性能を決める重要な構成要素となり、これまでアドホック (いわば仮説的、試

行錯誤的に種々選択されていたのに対して、理論的な基礎から幾何学的不変特徴抽出と統計的判別特徴抽出の2段階構成の枠組みを与え、実用化目標を達成するための基本要請3条件を満たすそれらの具体的な構成要素として、高次局所自己相関と多変量解析を採択しました。その意味では、理論から演繹的に導かれた構成要素ともいえますし、理論と条件を満たす仮説的しかし本質的な構成要素ともいえます。

基本要請条件（特にR1とR2）を同時に満たす特徴、しかもモデルベースではない汎用的な特徴としては、実は他にはあまり考えられません。

ご指摘のように、局所の特徴だけを見ているので位置不変となるわけではありません。むしろ、「相対的」な関係を自己相関として抽出するので位置不変となるわけです。

紙面の都合とも合わせて可能な範囲で説明を補足致しました。

議論3 視覚システムの要件

質問（赤松 幹之）

視覚システムに要請される基本的な条件として、「R1：位置不変性、R2：画面加法性、R3：適応学習性を掲げた」とありますが、これらを掲げた論拠が明記されていませんので、これらの技術開発を選択したシナリオを書いていただけませんかでしょうか。幾何学的不変性についても、例えば大きさに対する不変性、傾きに対する不変性、特徴間の位置関係の不変性等、他にも考えられると思います。また、不変特徴抽出において、幾何学的な変換に不変な特徴値が得られる汎関数を追求するとありますが、これは視覚システムを対象としていることから幾何学的な不変性が本質的に重要な性質になると理解してよろしいでしょうか。

また、画面加法性については、重なりがある場合には加法性が満たされないことになるとと思いますが、主に処理時間の観点からの選択でしょうか。

回答（大津 展之）

ここでの位置不変性とは、平行移動不変性のことです。「カメラと対象物の距離があまり変化しない」という意味ではなく、むしろカメラの向きの変化により対象物が画面枠内での平行移動として幾何学的な変換を受けて位置が変わりますが、そのような基本的な平行移動変換に不変な特徴が認識に本質的ということ。もちろんご指摘のように、他にもそのような不変変換としては大小（スケール）変換や回転等も考えられますが、平行移動（位置）不変性が最も基本的ということ。少し加筆して誤解のないようにいたしました。幾何学的変換に不変な特徴（汎関数）を追求する不変特徴抽出理論は、視覚に限定されるものではなく音声信号等も含めて、パターン一般にいえる普遍な理論です。

画面加法性は、重なりがある場合には、ご指摘のように厳密には成り立ちませんが、その場合も含めて、あえて要請しておくことが重要との主張です。これは、ご指摘のように、処理時間の観点からでもあります。特徴表現が認識（特に計数）にとって都合の良い表現（線形）となり、後の処理が簡単になるための要請条件です。少し説明を補足いたしました。

議論4 適応学習の意味

質問（上田 完次）

“適応学習”という言葉の用法はいくつかありますが、この論文での意味を明確にしていただけでないでしょうか。

回答（大津 展之）

そもそも、パターン認識では事前に必要な情報は完備されていません。有限個の例が学習サンプルとして与えられるのみで、そ

れに基づき未知サンプル（出来れば無限個）の認識を行います。ご指摘のように、確かに“適応学習”は用語的に多義性があります。まずパターン認識として、認識対象を限定しても、パターンの変動に対する適応ということがあります。これは特徴抽出と学習プロセスに関わるものです。また、この論文でいう適応学習は、さらにメタに、認識課題に対する適応学習という意味でも使っています。モデルベースの学習ですと課題（タスク）が代わるとモデルの入れ替えが必要なのに対して、この方式ではモデルを一切必要とせず構成要素もそのまま、後段の統計的な特徴抽出である多変量解析手法において、例からの学習によって最適に（重みが）構成され課題に適応します。このあたりは、少し分かり易く加筆いたしました。

議論5 パターン認識の正答率

質問（上田 完次）

正答率が100パーセントにならないのはなぜでしょうか。あるいは、どのような場合に100パーセントとなりうるのでしょうか。これまでの研究、他の研究者に比べて優れた結果が得られたということは十分に評価した上で、シンセシオロジー論文としての議論を深めるための質問です。

回答（大津 展之）

実世界のパターン、例えば音声の「a/i」や画像の「犬/猫」は、多様な変形やノイズを持ち、そこからの特徴（観測）値、例えば周波数や色は一般には確率統計的に分布して、概念は判別的としても裾野は限りなく接近して重複もします。したがって、学習サンプルに対しても正答率100パーセントにならないのが普通です。もちろん、有効な特徴を数多く抽出して統合すればするほど漸近的に100パーセントに近づきますが、コストの点で有限個の特徴抽出に押さえるのが現実的です。費用対効果の問題です。

もちろん、簡単な識別問題では100パーセントの正当率になりうる場合があります。例えば100円玉と10円玉の識別では、それらの特徴値（例えば直径や重さ）は、およそ確定的で異なるので、自動販売機が実用化されているわけです（たまたま誤認識もあるようですが）。この論文では、より困難で高度な認識問題への汎用的なアプローチの方式を示しています。

議論6 応用事例

質問（赤松 幹之）

第4章に応用事例が9例述べられていますが、これによって汎用システムであることを主張していると理解します。しかし、これらの例において、共通に使われているのはHLACとCHLACであり、多変量データ解析による判別特徴抽出には、因子分析、重回帰分析、判別分析、kNN識別、主成分分析、ARモデル、正準相関分析等、異なる手法が使われています。それぞれの課題に対してどの手法を使うのが最適なのか等、一部には説明がありますが、タスクに応じた手法の使い分けの基本的な考え方は理論について整理された記述があることを期待します。これによって、読者が持っている課題を解決するためには、どの手法を適用すれば良いかといった読者の理解が進むと思います。

回答（大津 展之）

ご指摘のように、HLAC/CHLAC特徴を基本初期特徴（不変特徴）として、タスクに応じたそれらの最適統合手法（線形重み和）として種々の多変量解析を用いています。多変量解析に不案内な読者の理解のためには、ご指摘のような配慮も必要と思われまので、紙面の都合から割愛しておりました対応表をつけることにしました。