

ファジィ推論ニューラルネットワーク を用いた画像認識に関する研究

慶應義塾大学大学院 理工学研究科 開放環境科学専攻

平成 15 年度

彌富 仁

目次

論文要旨	1
1 序論	2
1.1 はじめに	3
1.2 画像認識研究の歴史と概要	5
1.2.1 初期の画像認識研究	5
1.2.2 画像認識問題における知識処理	7
1.2.3 画像認識研究の現状	8
1.2.4 ニューラルネットワークの利用	9
1.2.5 ファジィニューラルネットワーク	10
1.3 ファジィニューラルネットワークの研究の歴史	11
1.3.1 ファジィ理論研究の経緯	11
1.3.2 ニューラルネットワーク研究の経緯	12
1.3.3 ファジィニューラルネットワーク研究の経緯	14
1.4 本研究の目的と位置づけ	16
1.5 本論文の構成	18
2 ファジィ推論ニューラルネットワークを用いた風景画像の認識とルール抽出	19
2.1 はじめに	20
2.2 画素を処理単位とした画像認識システム	22
2.2.1 システム概要	22
2.2.2 画像の認識結果	23
2.2.3 抽出されたルール	27
2.2.4 画素単位認識システムのまとめ	28
2.3 領域を処理単位とした画像認識システム	29
2.3.1 システム構成	29
2.3.2 領域分割過程	31
2.3.3 画像認識過程	33

2.3.4	画像解釈過程	36
2.3.5	画像認識、解釈に関するルール抽出	39
2.3.6	計算機実験	41
2.3.7	考察	51
2.4	結論	55
3	適応ファジィ推論ニューラルネットワーク (AFINN)	56
3.1	はじめに	57
3.2	AFINN の構造と動作	58
3.2.1	AFINN の構造	58
3.2.2	AFINN の動作	58
3.3	AFINN のモデリングと学習	59
3.3.1	初期ルール仮決定段階	59
3.3.2	入力要素選択段階	61
3.3.3	ネットワーク構造決定段階	62
3.3.4	システムパラメータの決定	62
3.4	計算機実験	64
3.4.1	数値モデルの近似とモデル構築	64
3.4.2	音声認識問題への適応	68
3.4.3	車の評価問題	70
3.5	結論	72
4	適応ファジィ推論ニューラルネットワークとアクティブ探索法を用いた画像認識	73
4.1	はじめに	74
4.2	複合型画像認識システム	75
4.2.1	システム概要	75
4.2.2	探索過程	77
4.2.3	検証過程	78
4.3	計算機実験	79
4.3.1	概要と準備	79
4.3.2	実験結果	81
4.4	考察と課題	85
4.4.1	物体ごとの認識結果について	85
4.4.2	認識誤りについて	86
4.4.3	検証過程とシステム構成について	87

4.5 結論	89
5 結論	90
謝辞	92
参考文献	93
付録	101
A ファジィ推論ニューラルネットワーク (FINN)	101
A.1 FINN の構造	101
A.2 FINN の動作	102
A.3 FINN の学習	103
A.3.1 自己組織化過程	103
A.3.2 LMS 学習過程	105
A.3.3 画像認識システム適用時の FINN の変更点	106
B 画素単位画像認識システムの詳細	107
B.1 システム概要	107
B.1.1 知識抽出ネットワークの構成	107
B.1.2 風景画の認識	108
B.1.3 ルール抽出過程	109
B.2 計算機実験	111
B.2.1 実験条件	111
B.2.2 学習結果	113
B.3 知識抽出と画像認識結果に対する考察	115
B.3.1 知識抽出に関して	115
B.3.2 画像認識に関して	115
B.3.3 画素単位認識システムの課題	116
C 領域分割とテクスチャ	117
C.1 K 平均アルゴリズム	117
C.2 テクスチャ	118
関連文献	119

論文要旨

画像の認識には、対象に関する適切な知識の獲得と利用が必要であり、人間は高度な判断をそれまでの経験や学習に基づく多くの知識を利用して行っている。従来の画像認識システムの多くは、あらかじめ登録されているルールやデータとのパターンマッチングにより処理を行う。そのため、限定された場面では優れた性能を発揮する一方、想定されていない状況では正しい処理は困難である。本論文ではこうした知識獲得、統合、利用に起因する画像認識システムの汎用性の限界を解決するために、ニューラルネットワークの学習能力を利用する研究について述べる。

まず、適度な複雑さをもつ風景画像を対象に、画像認識および認識のためのルールの自動生成を行った。この研究ではニューラルネットワークの学習能力に、論理的処理の記述が可能なファジィの長所を融合したファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network) を用いることによって、画像に対する明示的な知識を必要とせずに高精度の画像認識を実現した。生成されたルールはファジィ if-then 形式で可読性も高く、ルールベースとの親和性も高い。次に、精度を高めるためにより多くの特徴量を用い、また FINN を並列に複数用いることでスケラブルに認識精度を向上できる事を確認した。また、この研究では認識から一步踏み込み、画像の内容を複数の単語の組み合わせで説明する結果を得た。

一方、認識する対象を風景画像のみではなく一般的な画像も含めた場合、認識すべき物体の種類が大幅に増加するために、扱わなければならない情報量も増加する。従来の基本枠組みであった FINN には、推論の汎用性向上に必要な推論のモデリング機能や、高次元のデータへの対応がなされておらず、課題として残されていた。そこで FINN のメリットを残しつつ、これらの機能を追加し、大量のデータを効率よく、かつ高度な知識処理が可能な新しいニューラルネットワークである適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference neural network) を提案した。

学習システムは画像認識問題に効果的ではあるが、様々な状況に適切に対応するためには、ルールベースもしくは、マッチングの手法の併用も必要であると考えられる。本論文の最後には、優れたパターンマッチングの手法で知られるアクティブ探索法に、AFINN を組み合わせた画像認識手法を示した。この研究では、コンピュータ等が設置された研究室内部の画像を例題に扱い、学習システムにパターンマッチングの手法を効果的に組み込むことで、高速で高精度の認識実現を確認した。

第 1 章

序論

本論文はコンピュータの学習機能および、知識処理に基づく画像認識、画像理解実現を目指した研究である。本章では、本論文に関連する研究の歴史と背景について概観する。

1.1節では、画像認識問題について概要を述べる。1.2節では、画像認識問題に関して、従来行われてきた研究の歴史と現状、課題について系統立てて簡単に紹介する。1.3節では、本論文の重要事項である学習と知識処理の双方を可能にしたファジィニューラルネットワーク研究について、研究背景と歴史について述べる。1.4節では、従来の画像認識の研究を踏まえた本研究の位置づけと目的について述べる。1.5節で、本論文の構成を述べる。

1.1 はじめに

画像認識問題は、入力画像に対して何らかの判断を与える処理である。これらの処理は目的別に以下の様に分類することができる [1]。

- 画像全体を分類する。(例：X線写真を「健康」と「要精査」に分ける)
- 画像の中から必要なものを探す。(例：画像中から特定の人を探す)
- 画素単位で分類する。(例：衛星写真からの土地利用分類)
- 画像シーンの状況を説明する。(例：ロボットの視覚)

これらの結果は文字や記号、あるいは結果を記録した画像であったりする。一般的に画像認識のプロセスは以下の様なステップに分けて行われることが多い。

1. 前処理

画像に含まれる望ましくない成分の除去、補正 (明るさ、歪みの補正や、ノイズの除去)、注目している物体や範囲の抽出 (segmentation)、後の処理に必要な特徴の強調などが行われる。幾何学変換や画像強調、画質改善などの多くの処理もここに含まれる。

2. 特徴抽出

画像認識の目的を達成するために用いる特徴量を、前処理が終わった画像データから抽出する。これまでパターン認識問題で広く使われている代表的な特徴量および、表現方法を表 1.1 にまとめた。これらの特徴量、データ表現手法は一般的なものの例であり、用途によって効果的なものを選択する必要がある。また、特徴抽出を行った後では入力データは画像ではなく、特徴量のデータ組として扱われる。

3. 決定

この段階では、抽出された特徴量を用いて認識結果を得る。主な方法に、論理の組み合わせ (枝分かれ論理、決定木などのルールベースを含む)、パターンマッチング、統計的決定法 (Bayes 推定など)、識別関数法 (回帰分析、判別分析など)、学習アルゴリズムを用いるもの (Neural networks など) がある。

前処理と、特徴抽出は、画質改善や、画像計測など単体で用いられ役に立つ場合が多い。画像認識を行うには、これらの処理に加え、最後の決定部分が非常に重要となる。

決定部分を構成する手法にはそれぞれ長所、短所があるが、どれを扱うにしても画像認識を行わせる機械には何らかの知識を与える必要がある。この知識の適切な獲得、利用は非常に困難な問題である。対象に関する知識を多く持てば、対象物に対する認識精度の向上が期待できる一方で、その知識がない場合の処理が難しくなり汎用性は

低下する。知識の獲得および付与の処理は自動化が難しく、人間が行っているような非常に多くの経験と知識の上に成り立っている画像認識、画像理解を実現するには多くの課題が残されている。

表 1.1 パターン認識の代表的な特徴量とデータ表現法の例

代表的な特徴量	
色情報	RGB, YCrCb, xyz, Lab HSV(色相、彩度、輝度)、HSL(色相、彩度、明度)
エッジ検出	1 階差分に基づく方法 (Sobel フィルタ等) 2 階差分に基づく方法 (Laplacian フィルタ, zero-crossing 法等)
領域分割	エッジ追跡法、エッジ予測法、ハフ変換 分割、統合法 (K 平均法など)
texture 解析	構造解析法、統計的解析法
動きの検出	オプティカルフロー
代表的なデータ表現方法	
境界線の表現	折れ線近似、chain coding, フーリエ級数展開法 など
領域表現	リスト表現、4 進木表現、骨格表現 など
立体表現	境界線による表現 境界面による表現 (面パッチ法、翼エッジ法など) 立体の組み合わせ (一般化円筒法など)

1.2 画像認識研究の歴史と概要

机の上のコップを認識するためには、それぞれの物体を境界線で囲まれた個々の物体に区別し、かつ物体に関する適切な知識を利用する必要がある。画像認識問題は、対象とする画像から何らかの特徴を抽出し、知識に基づいて対象を概念と結びつけることに相当する [2]。

前節で述べたように画像認識は、前処理、特徴抽出、決定の3段階に分けられ段階的に処理が行われることが多い。この画像認識問題は、画像から抽出されたパターンを元にした、パターン認識問題に帰着でき、決定段階に相当する知識処理をより高度にしたものにしたものが画像理解といえる。なお、本論文では、画像理解を画像認識よりも、高度な知識処理によってなされている処理と定義する。

パターン認識の基本は、与えられたパターンを特徴を反映した別の表現で記述することである。与えられた記号列もしくは画像、音声などの特徴データの組を分類する場合、画像認識問題で要求されるような高度な知識処理は必ずしも必要とされないことも多い。

1.2.1 初期の画像認識研究

画像認識および理解研究の歴史は、基礎となるパターン認識研究から始まった。パターン認識の研究は1950年ごろ、米国での文字認識の研究に始まり、1955年には世界発の数字のOCR(optical character reader)が開発された。日本でも1960年頃から研究が始まり、手書き文字認識もこの時代から研究が始まった [3]。表 1.2 に初期のコンピュータとパターン認識研究、実用化の歴史をまとめた。当初は、計算量の問題などから文字や、簡単な幾何学図形、単純な物体の検出などが中心であった。しかし1965年のFFTの開発 [4] やコンピュータの進歩などから計算量の課題は改善され、また1968年に採用された郵便番号制度の導入などの要因から、パターン認識を含む画像認識関連の研究は急速に発展した。

画像認識に必要な基礎技術としてのデジタル画像処理は、これらの関連技術の進歩により、1960年代後半から始まった。この当時の画像処理の主な対象は、宇宙探査用の人工衛星画像であり、画質改善や画像復元、補正などが中心であった。その他の分野として、医用画像の解析や分類といったパターン認識が試みられた [1]。3次元物体の認識は、線画の解釈の研究に始まり [5]、個々の積木の位置関係を理解するいわゆる”block world”の問題として、後のコンピュータビジョン、ロボットビジョンの基礎となった。

表 1.2 初期のコンピュータと、パターン認識研究、実用化の歴史
(文献 [1][3] より抜粋)

年	内容
1946	ENIAC(世界初のコンピュータ) 完成
1948	トランジスタの発明 (Shockley) 情報理論提唱 (Shanon)
1950	文字読取り研究開始 (アメリカ)
1955	数字 OCR 発売 (アメリカ)
1956	ダートマス会議。人工知能 (パターン認識含む) の提唱と研究開始
1960	文字読取り研究開始 (日本) 光学的画像処理研究開始 染色体写真解析
1964	レンジャー 7 号 月面写真撮影に成功
1965	FFT アルゴリズム開発 [4] 航空写真解析
1966	IBM OCR 開発
1968	郵便番号の自動読み取り開始 (日本) 漢字、かな読み取り研究開始 (日本) 線画の解釈による 3 次元物体認識研究の提案 [5]
1969	画像処理アルゴリズムの体系化 [6]
1970	LSI 国内開発成功。集積回路の時代へ

1.2.2 画像認識問題における知識処理

画像認識、理解をコンピュータで行うには、対象に関する適切な知識が推論機構 (inference mechanism) によって適切に使われる必要がある。一般的に、コンピュータ上で画像認識や画像理解のような知的機能を実現するためには、以下の3つの条件が必要であり、これらを表現する枠組のことを知識表現 (knowledge representation) という [2]。

1. 人間のもつ知識を素直に表現できる。
2. 推論の効率が高い。
3. 知識を記述しやすく、修正しやすい。また記述が分かりやすい。

人間の知識をコンピュータ上で表現する方法としては、リスト表現、意味ネットワーク表現、ルール表現といった情報をシンボルで表す記号表現が多く用いられ、これらの知識は2つに分類できる。一つは「りんごは赤い」などの宣言的な表現による宣言的知識 (declarative knowledge) であり、もう一方は「もし円筒状で上部が空いているガラスの物体ならコップである」のような手続き的な表現であらわされる知識で手続き的知識 (procedural knowledge) と呼ばれる。2章以降、本論文で扱うファジィif-thenルールは手続き的知識に分類できる [2]。

一般的に、人間の知識をコンピュータ上に実装するにあたっては、その獲得と表現が困難であった。これらの知識を必要とする画像認識において、ルールベースのシステムは表現された知識をそのまま数値化するなどしてシステムを実現化する。Bayes推定に代表される統計的処理では、与えられた知識から事前確率などを求めることで知識を反映させている。判別分析やニューラルネットワークのように、識別境界を学習などによって求め、それを元に推論を行う手法は、最終的な入出力のみの関係を与えるだけで、明示的な知識を必要とせずにその関係を近似することができる。しかしながら効果的な入力データの特徴を選択する必要があり、また知識の埋め込み、特に手続き的知識を表現するのが非常に難しい。また、これらの議論はいかに知識を表現するかに着目したものであって、知識の獲得はまたこれとは別の大変困難な課題である。

1.2.3 画像認識研究の現状

画像認識技術は文書画像処理から始まり、産業用画像処理、ロボティクス、軍事、医療、セキュリティー、物体認識など様々に利用されている。近年では顔認識や生体情報認証技術 [7]-[9]、自動運転のための車両、道路認識技術 [10]、大容量データベースでのメタデータ自動作成や画像検索 [11]-[16] 等の研究が実社会からの要望も高く注目されている。

これまでの多くの画像認識システムは、画像から対象とする物体の何らかの特徴を抽出し、予めシステムに登録されたルール [17]-[22] や参照モデルとの照合によって認識を行うもの [9][10][23]-[31] が多い。

ルールベースのシステムは限られた条件下では優れた認識結果を行える反面、汎用性に乏しい。従ってより一般的な認識問題を扱おうとした場合、その対象に与える適切な知識の定義や表現など、より困難な問題を解決する必要がある。3次元参照モデルを利用して照合を行うシステム [7][23]-[25] は、光源や遮蔽などによって多様に变化する対象物を認識できるというメリットを持つ。しかし、モデルの照合に時間がかかり、また認識対象が変わる場合新たに登録する3次元モデルを構築しなければならない。一方、2次元参照モデルを用いるシステム [20][26][27] は、適切に認識を行うためには非常に多くの参照画像が必要であり、やはり計算量や記憶容量などの問題があった。村瀬らは、2次元の参照画像を複数用いて画像の中から所望の物体を高速に検出するアルゴリズムを開発している [29][30]。特にアクティブ探索法 [30] は、高速で汎用性が高いアルゴリズムであるため、多くの応用がなされて来た [31][32]。

一方、画像認識問題の1つの応用例である、画像検索システム (CBIR system : content-based image retrieval system) の研究が1990年初頭より盛んに行われるようになってきた [12]。初期の研究では、画像毎にあらかじめ付与されているテキスト情報を元に画像を検索するものがほとんどであり、これを応用したものがすでにwwwの検索エンジンであるgoogle[94]等で実現されている。これらの画像1枚1枚に予め解釈の情報を付与する方法は、高い検索精度を実現するがインターネットや大規模データベースでの利用を考えると、データ作成などのコストが問題になる。googleの画像検索システムは、画像ファイル名やコンテンツ内のテキスト情報などを用いてweb上の莫大なデータから検索を行うため、システム作成コストは低く抑えられ、検索範囲が広い実用に耐えうるシステムになっている。しかしながら、予め与えられたキーワード以外の検索が適切に行えない等の課題もある。

こうしたキーワードを元にした検索ではなく、画像のみから内容を自動認識し、検索が行えるシステムも研究されている。これらは画像が何から構成されているかを自動的に識別する必要があるため、知識に基づいた処理が必要である。これらの研究の初

期のものは画像の類似性のみに着目し、言語的、概念的な類似を考慮していないもの多かったが、今日では物体を表す概念も考慮された研究が行われている [14]-[16]。こうした画像データの言語による自動変換には、近年主に統計的手法が用いられる [14][16]。

Wang の SIMPLIcity[14] は、統計手法を用いて画像をカテゴリ毎に分類し、データベースの検索範囲を削減している。しかしながらこれらの手法は問題に特化したシステムであり汎用性、拡張性に課題が残っていた。Li ら [16] は、階層構造を持つ 2 次元隠れマルコフモデルを用いて、6 万枚にのぼる大規模な画像データを用いて、600 ものコンセプトの自動付与を実現する画像のインデキシングシステムを提案している。

こうした画像検索システムは、パターン情報である画像を、テキスト情報へ変換する必要があるため、画像の認識、理解問題を含む困難な問題である。また、対象となる画像範囲が広いために、汎用性と精度の両立には多くの課題が残っている。

コンピュータの処理能力の向上や、数多くの有効なアルゴリズムの開発によって、個々の場面での画像認識能力は人間をはるかに凌ぐレベルに達しているものの、様々な場面に対応する能力、様々な知識を統合することが要求される画像理解は人間に遠く及ばないのが現状である。

1.2.4 ニューラルネットワークの利用

一方で、システム毎に専用の設計が必要であったルールベースシステムの限界から、ニューラルネットワークの学習機能を利用した様々な用途の画像認識研究も盛んに行われて来ている [11][34]-[38]。

誤差逆伝播法 (BP:Back propagation)[58] に代表される典型的な階層型ニューラルネットワークは、十分な数の中間層ニューロンを準備すれば学習によって任意の入出力関係を近似することが可能である。このため画像認識問題に適切に適応させた場合、認識対象に関する明示的な知識や情報を与えなくとも優れた認識結果を得ることができ適用例は数多い。

Minka ら [11] は、競合学習のニューラルネットワークを用い、クラスタリングを階層的に行う画像検索システムを提案しているが、汎用性、精度に課題があった。Jing ら [37] は領域分割からパターンマッチングまでの階層構造を持つ手法を提案している。この手法は、学習によって良好な領域分割のパラメータを獲得し、例題として背景が複雑な一般的な画像から車両の検出を行っている。しかしながら認識の部分でパターンマッチングを行っており、ルールベースのシステムとなっている。

椋木ら [35] は、風景画像検索問題にニューラルネットワークを適用し、その過程で優れた認識結果を実現している。しかしながら、この研究の様にバックプロパゲーション学習アルゴリズムを用いた階層型ニューラルネットワークを使用しているシステムは、中間層ニューロンの役割が不明であり入出力の関係の解析が困難である。したがっ

て必要に応じた機能の追加や、変更、ルールベースとの親和性などに難点があり、拡張性、汎用性に課題が残る。

これらの研究の多くは、ニューラルネットワークを単なる判別器、あるいはクラスタリングのための道具として用いているに過ぎず、画像認識で本当に必要とされる知識処理と呼べる処理を行っていない。つまり、与えられた学習データから誤差が最小、あるいはそれに準じる状態になるような識別境界を作るといった従来の判別分析や重回帰分析などで行っていたものと、根本的には差異がない。

1.2.5 ファジィニューラルネットワーク

我々人間は様々な経験や学習を通して知識を獲得し、それを基に様々な決定を行っている。コンピュータによる画像認識や理解といった高度情報処理実現のためには、知識の自動獲得は非常に重要な課題である。そのような中で、優れた学習能力を持つニューラルネットワークと論理的な処理を得意とするファジィの融合により、双方の長所を活かしたファジィニューラルネットワークが提案されている。ファジィニューラルネットワークは、動作を決定するメンバシップ関数等のパラメータを、ニューラルネットワークの学習機能を利用して決定するもので、1974年にS.C.Leeらによって最初に提案された[43]。そしてBPが提案された後の1990年ごろから多くの研究がなされるようになった[44][45]。

ニューラルネットワークとファジィの融合には、形態や融合の度合によって系統的に分類されている[46]。本論文で扱うファジィ推論ニューラルネットワークは、この分類に従うとニューラルネットワークの一部がファジィ化した「ファジィ化ニューロ」に属する。ファジィニューラルネットワーク研究の詳細については続く第1.3節で紹介する。本論文は、このファジィニューラルネットワークの知識獲得機能に着目し、外部からの明示的な知識の付与を必要としない画像認識システムの実現を目指したものである。

1.3 ファジィニューラルネットワークの研究の歴史

ファジィニューラルネットワークは、その名が示すようにファジィの持つ言語記述能力と、ニューラルネットワークの学習能力を融合することで、効果的な情報処理の実現を目指したネットワークである。以下、その構成要素であるファジィと、ニューラルネットワークそれぞれ研究の経緯から順に述べる。

1.3.1 ファジィ理論研究の経緯

曖昧な情報や言語情報をそのまま扱う「ファジィ理論」は California 大学の Zadeh によって 1965 年に提言された [39]。当初はあまり注目されることも無かったが、1970 年代ごろから実システムへの応用が報告され工学分野で注目をされ始め [40]、1984 年には国際ファジィシステム学会が結成されるに至った [41]。そして 1985 年に高木、菅野 [42] らによってファジィの数値的解釈が行われて以来、多くの実用的な応用研究がなされてきた。

それまでの厳密さを求めていた情報処理と一線を画し、曖昧さを積極的に扱うファジィは以下の様なメリットがあり、制御分野を中心に研究が進み、1990 年代初期にはそれらを実応用した数多くの家電などが実用化された。

- 扱う対象のモデリングを行う際に厳密な数学的な解析が不要。
- 人間のノウハウをそのままシステムに記述でき、また動作の解釈が容易。
- 部分的なシステム変更が容易である。

ファジィの実応用を考えた場合、制御あるいは注目する対象に対する処理を知識としてシステムに与える必要がある。人間の知識をコンピュータで表現するには、前節で触れたようにシンボルで表された記号を用いることが多い。ファジィ制御で用いられるルールは、記号化された知的表現であり、手続き的知識を表現するのに適している。ファジィルールが他の知的表現と大きく異なるのは曖昧さをもつ点で、これをファジィ理論では Fuzziness と呼ぶ。これまでの工学、理学など一般の学問分野において不確かさというのは、ノイズであったり確率的な不確かさを指すことがほとんどであった。ファジィ理論の Fuzziness は、これらとは異なり、「大きい」とか「若い」などに見られる言葉の意味や概念の定義の曖昧さを指す。

この曖昧さを効果的に利用することによって、制御する対象の複雑さや、データの不完全さ、様々な状況に対する我々の判断や思考を手続きとして表すことが可能になる。しかしながら、ファジィはその動作を決定する最も重要なメンバーシップ関数の形状や、パラメータの決定などを開発者自身の手で行わなければならない。ファジィ技術

の一部はすでに実用段階を過ぎ成熟期を迎えており、基礎技術の一つとして他の分野と融合した研究が行われている。

1.3.2 ニューラルネットワーク研究の経緯

人間の脳で行っている情報処理機能をコンピュータ上で模倣、あるいは実現することで、新しい情報処理手法の確立と従来の問題点の解決を目指し、ニューラルネットワーク (neural network) の研究は始まった [47]。ニューラルネットワークは、コンピュータ上のソフトウェア、あるいはハードウェアそのものであるため、人工ニューラルネットワーク (artificial neural networks) と呼び区別することもある。ここでは、ニューラルネットワークの研究の経緯について簡単に述べる。

1937 年の Turing による Turing 機械の基本的構想の発表、1945 年の Neumann によるプログラム内蔵方式の提唱による今日のコンピュータの歴史が始まった頃とほぼ時を同じくして、ニューラルネットワーク研究も始まった [47]。

1943 年に McCulloch と Pitts が人間の脳細胞 (neuron) の特性を単純化し、数理的にとりあつかいやすい形で発表した [48]。この形式ニューロンの提案は、脳機能の構成要素としてニューロンを位置付けたことで、それらを複数接続させることにより情報処理を行うニューラルネットワーク開発にとって極めて重要な成果であった。1949 年には Hebb が、ニューラルネットワークの可塑性に関する仮説を発表した [49]。この仮説はニューロンの興奮 (発火) に関与したシナプスの伝達効率が増加するというものであり、学習や記憶をシナプスの伝達効率の変化と捉えた画期的な考え方であった。1958 年には Rosenblatt がこうした学習機能をニューラルネットワークモデルに組み込んだ Perceptron を発表した [50]。Perceptron は、学習、認識のできる識別器として大いに注目を集め、学習収束定理など様々な理論整備が進んだ。ところが、1969 年 Minsky らが著書 "Perceptron" の中で、Perceptron が線形分離不可能な問題に対して学習が行えない限界 [51] を指摘することによりブームは後退し、ニューロコンピュータの研究は次第に人工知能 (AI) にとって代わられるようになった。

多くの研究者が AI に関心を移していったが、地道な研究は続けられ、連想記憶モデルとして 1972 年に中野ら [52] により Associatron が提案され、同年に Kohonen [53]、Anderson [54] によって相関行列を元にした連想記憶モデルが提案された。1973 年には Malsburg によって競合学習モデルが提案された [55]。1980 年には福島らによって生体の視覚構造を模擬した Neocognitron が提案された [56]。

そして 1982 年に Hopfield が Hopfield model [57] を発表したことにより、組み合わせ最適化問題など新しいニューラルネットワークの利用法が示され、再び研究者の注目

を集めるようになった。Hopfield の発表から 4 年遅れた 1986 年、Rumelhart らによって、誤差逆伝播法 (BP:Back Propagation) 法が発表された [58]。

BP は、Perceptron では不可能であった中間層の学習を可能にし、非線形問題に対応できる汎用性の高い強力なアルゴリズムである。Sejnowski[59] らが BP を用いた、アルファベットの発音を出力する実応用例”NETtalk”を示したことによって急激に広まった。

BP はその学習能力と汎用性を活かして、様々なパターン識別問題や制御分野などに応用されている。また BP の様な階層構造を持つニューラルネットワーク以外で重要なモデルとして、Hopfield network に代表される相互結合型のネットワークとして、Kosko の BAM(Bidirectional Associative Memory)[60], 確率的に動作する Ackley の Boltzmann machine[61] や、Kohonen の自己組織化マップ [62] に代表される競合学習型ネットワークなどがある。

また BP に対して、学習の高速化や汎用性向上の様々な改良研究や、モデル提案がなされてきたが、現在になってもそれらは大きな枠組で見ると BP の派生研究であり根本的に新しいアルゴリズムは提案されていない。

近年、BP の様にニューロンが発火した信号を重み付け加算し、閾値処理をする現在のニューラルネットワークの根本的なアルゴリズムに代わり、発火信号を実際のニューロンと同じようにパルスで行うモデルであるパルスニューラルネットワークの研究も行われて来ている [63]。しかしながら、BP の様な確立された学習アルゴリズムは提案されておらず、また多くの信号処理は従来の BP で可能であり、計算量が極端に大きくなるパルスニューラルネットワークの明らかなメリットは現在までのところ発見されていない。パルスニューラルネットワークの汎用的学習則、および適切な応用例が発見されれば、次のニューラルネットワーク研究の中心になるとも考えられる。

1.3.3 ファジィニューラルネットワーク研究の経緯

BP は優れた学習能力を持ち、様々な分野に応用されているが、学習後のネットワーク状態を解析するのは非常に困難である。つまり、学習後のネットワークはブラックボックスであり、必要に応じて機能を追加したり変更するといった作業は困難である。また人間が通常行っている論理的な知識処理も行えない。

一方、この言語処理を得意とするファジィに、学習能力をもつニューラルネットワークの長所を組み合わせたファジィニューラルネットワークが提案されている。これはファジィの弱点であるモデル構築を、あらかじめ与えられた入出力関係を学習して獲得したニューラルネットワークによって行うもので、1974年に S.C.Lee らによって最初に提案され [43]、1990年頃から盛んに研究され始めた [44][45]。

ファジィニューラルネットワークは、BP のような単なるパターンの分類器としての利用だけではなく、学習後のネットワークから、学習データ中にあるルールを抽出することも可能であり、さまざまな応用が期待できる。

初期のファジィニューラルネットワークの中で、Lin[45] は 1991年、Kohonen の自己組織化アルゴリズム [62] を基にネットワーク構成を仮決定し、BP によってパラメータを調整する現在のファジィニューラルネットワークモデルに良く見られる構造を提案している。また代表的なものに、1993年 Jang らによって提案された ANFIS[64] がある。ANFIS ではメンバーシップ関数の形状やファジールールの抽出にニューラルネットワークが用いられている。しかし事前の入力空間の分割が必要であり、システムの特性は前処理に大きく依存する。これに対し Wang らによって、入力空間を自動分割しファジールールを抽出する方法が提案されている [65]。しかしこれらの手法は出力空間が考慮されていないため、得られたルールが必ずしも最適とは言いがたい。

また、より精度が高く、適切なルールが得られるファジィシステム構築を目指し、学習データの中から有力な情報や生成するルールの効果的な選択をするシステムも提案されてきた [66]-[73]。不要なルールおよび、入力の適切な削除は、計算量の削減の他に、ファジィの大きなメリットの一つである獲得されるルールの解釈を容易にする。

これらの中で、Linken ら [72] は、有効な入力選択、ルール選択を実現し、鉄の強度を題材にした実問題でも優れた精度と、適切なルール抽出が可能であることを示した。しかしながらこの手法は、最終的にルール数を決定する場面で菅野 [66] や Pal[68] 等のファジィニューラルネットワークと同様に、FCM(fuzzy c-means) をベースにしたアルゴリズムを使用している。このために、これらのシステムではあらかじめ適切なルール数が既知である必要がある [71]。またアルゴリズムが複雑な上に、明示されたアルゴリズムならびに結果は出力が 1次元のものに限られている。

一方で仁科ら [74] は、入出力空間を自動的に分割し、ファジールールを自動抽出する

ことができ、しかも構成が簡易なファジー推論ニューラルネットワーク (FINN: fuzzy inference neural network) を提案している。FINN は、他の多くのファジィニューラルネットワークと同様、自己組織化によって初期ルールを決定した後、ルール間のユークリッド距離を基に不要なルールの削除を行う。その後、誤差降下法である LMS (least mean square) アルゴリズムを用いてパラメータの決定ならびに、精度の向上を達成している。FINN は、重要な入力を選択やルール数の調整などのモデリング機能が十分でない点もあったが、様々な用途に応用しやすい簡易な構成のため、推論機能の拡張 [75] や画像認識などの問題に対して基礎コンポーネントとして用いられて来た [76]-[78]。本論文において基礎コンポーネントとして用いたこの FINN の構造や動作、学習については、付録 A で詳しく述べる。

1.4 本研究の目的と位置づけ

本研究は、ファジィ推論ニューラルネットワークの学習能力、ルール生成能力に着目した画像認識の研究である。

図 1.1 に、ルールベースのシステムと学習によるシステムとの比較を示す。

多くの画像認識システムは、あらかじめ登録されているルールやデータとのパターンマッチングにより処理を行うため、限定された場面では優れた性能を発揮する反面、想定されていない場面の対応が困難である。つまり様々な場面に柔軟に対応するには、膨大な数のルールを登録する必要があり、与える適切な知識の定義や表現など、より困難な問題を解決する必要がある。

学習によるシステムは、特定の場面の認識において、専用のルールが用意された知識ベースのシステムに対して精度の面で不利である。しかしながら、データから必要なルールを自動的に構築し利用することができるために、事前のルールの獲得や組み込みが不要という大きなメリットを持つ。様々な状況、場面に対応できるコンピュータビジョンの実現のために、無数のルールを逐次入力するのはコスト、実用性を考慮すると現実的でないため、学習機能が必須であると考えられる。こうした背景を踏まえ本研究は、認識に必要な特徴成分や知識をニューラルネットワークの学習能力で獲得して利用する、新しい流れの研究である。

本研究と、これまでの画像認識、ニューラルネットワーク関連の研究との関係を図 1.2 に示した。ニューラルネットワークの学習能力を利用した画像認識の研究は過去にも多くあったが、データからの知識獲得を念頭においた研究は見られなかった。本研究では、コンピュータの学習と、知識獲得に重点を置き、従来のルールベースのシステムでは必須であった前提の知識の付与がない状態で、精度良い画像認識、および認識を発展させた画像理解実現を目指している。

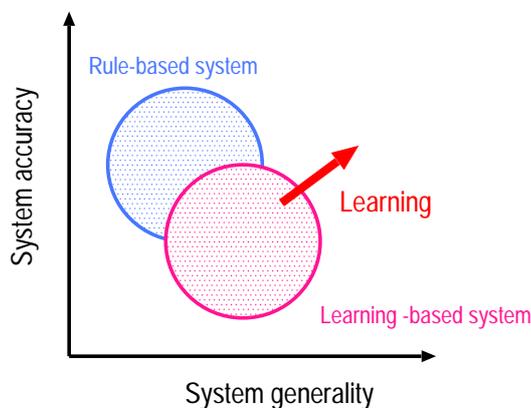


図 1.1 ルールベースシステムと学習ベースシステム

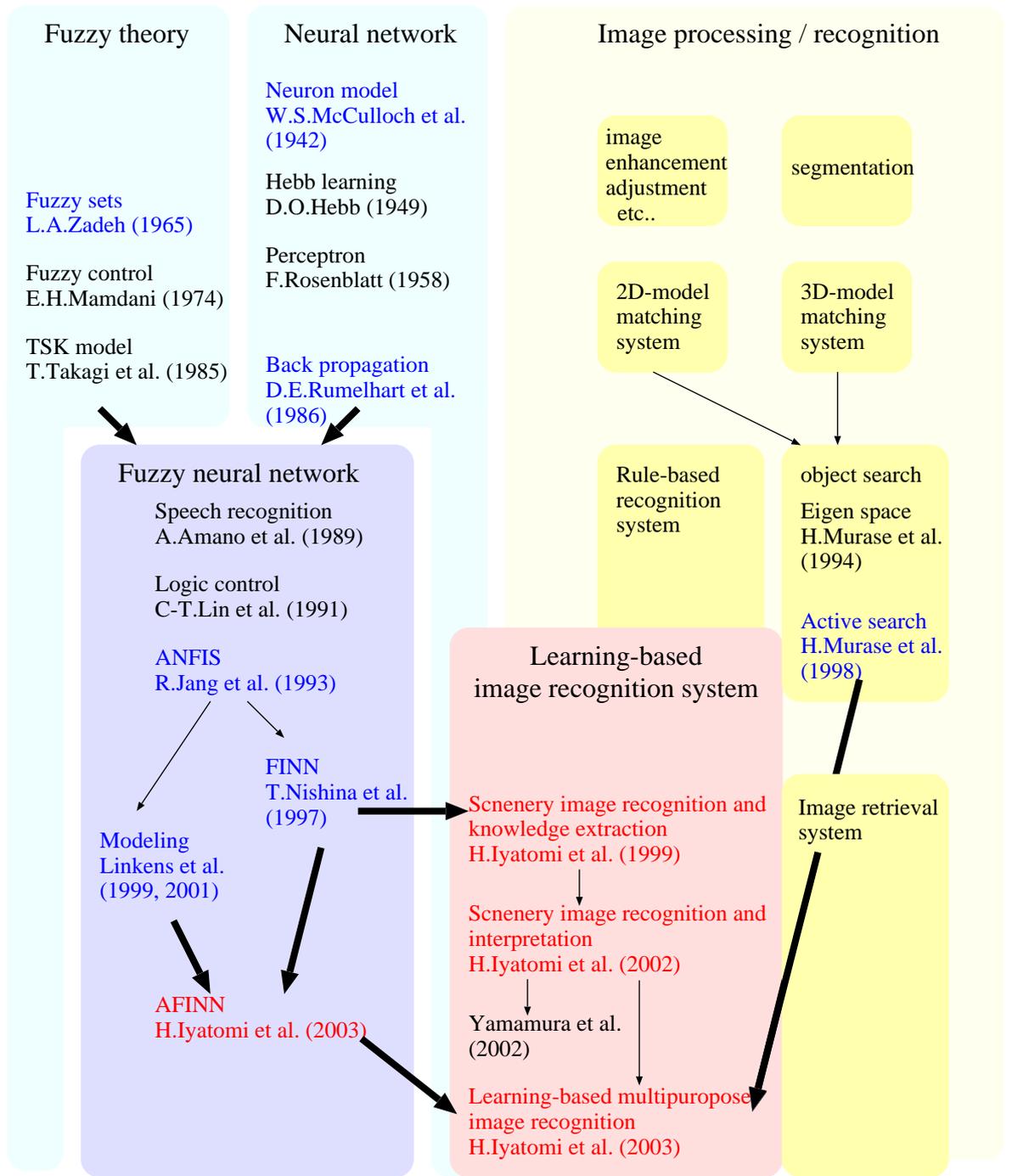


図 1.2 本研究の位置付け

1.5 本論文の構成

本論文は全 5 章から構成される。

第 2 章では、風景画像を例題に認識と、学習によって自動生成される知識の抽出に関する研究について述べる [76][77]。本研究はファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network)[74] を基本要素として用いることによって、認識対象に関する明示的な知識などを必要としない画像認識と、認識に関するルール抽出が行えることを示す。本研究では認識のアプローチとして、前処理の影響を受けにくく、認識物体に対する事前の知識が不要な画素単位の認識手法と、画像中の物体毎など、意味のある領域を処理の単位とした認識手法の双方について述べる。まず、基本的な画素単位の認識問題において、学習システムの有効性を検証する。次にその結果を踏まえ、画像を複数の領域の集まりとして扱い、また利用する特徴量を大幅に増やすことによって、認識精度の向上と、複数の単語の組み合わせによる画像の説明を行っている。また、推論ネットワークを複数並列に利用することによってスケラブルに性能が向上することを確認している。

第 3 章では、これまでの学習システムによる画像認識問題で基本要素となる FINN の弱点を補う新しい処理モデルである適応推論ファジィニューラルネットワーク (AFINN: adaptive fuzzy inference neural network)[79] について述べる。AFINN は、学習データから適切な入力要素の選択やルール数、パラメータの調整等のモデリングを自動的に行い、より一般性の高いルールを生成することのできるファジィニューラルネットワークである。本章では、California 大学 Irvine 校 Machine Learning Repository[91] で公開されているいくつかの例題を用いて有効性を確認している。

第 4 章では、AFINN と有効なパターンマッチング手法であるアクティブ探索法 [30] の長所を融合した画像認識法 [80] について述べる。本研究は、学習システムと、従来のパターンマッチング手法の双方の長所を利用することで、汎用性と精度の両立を目指している。本章では、コンピュータ等が多数設置された情報系の研究室内部の認識を例題として扱い、優れた認識結果を確認している。

第 5 章で、本論文のまとめを行う。

第 2 章

ファジィ推論ニューラルネットワークを用いた風景画像の認識とルール抽出

本章では、適度な複雑さを持つ風景画像を例題にコンピュータの学習機能を利用した、画像認識および認識に関するルールを生成することのできる新しいシステム [76][77] について述べる。

これまでの画像の認識手法の多くは、あらかじめ与えられたルールやパターンを元にマッチングを行うものがほとんどであった。こうしたルールベースや、パターンマッチングを利用した手法は、想定された場面では優れた効果を発揮するが、状況ごとに適切なルールや知識の付与が必要であり、汎化性に非常に大きな課題があった。

本研究はこうした知識獲得の問題に、学習能力を有するニューラルネットワークと、ルールや知識などの言語を表現するのに優れたファジィの長所を融合し解決を目指した研究である。

本章ではまず、画像の画素単位での認識および、認識に関するルールの自動生成、抽出を行う [76]。画素単位の認識手法は、領域分割などの前処理の結果の影響を受けにくく、また対象となる物体に対する明確な知識などを必要としないというメリットをもつ。各画素が何であるかの推論には、ファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network)[74] を用いて、学習システムを画像認識問題に適用した場合の基本的な効果について検証する。

続いて、一般的な画像認識手法として広く行われている前処理、特徴量抽出、決定の手法に基づいた画像中の物体を処理の単位とした学習ベースの認識システムの有効性を検証する [77]。このシステムは、1) 領域分割過程 2) 画像認識過程 3) 画像解釈過程の 3 過程から構成される。前処理である領域分割過程では、領域分割アルゴリズムとして広く知られている K 平均アルゴリズム [81] を基に画像を適当な領域に分割する。画像認識、解釈各過程では FINN を基にしたネットワークを利用している。画像認識過程では、より高い認識精度を実現するために、複数のネットワークを並列的に動作させる。それらの結果を用いて、最終的に画像解釈段階において、画像の言語的な解釈が行われる。

本章では、これらのシステムについて内容と実験結果を評価し、学習によって獲得

されたルールのみを利用して汎用性の高い良好な認識、解釈結果が得られる事を確認した。またシステムは追加学習が可能であり、汎用性の向上も確認している。

2.1 はじめに

画像理解は高度情報化社会において非常に重要なテーマであり、従来から多くの研究が行われている。しかし、それらの多くは、認識のためにあらかじめ画像中の対象領域に関するヒューリスティックな知識を必要としている [17]-[22]。また前処理として、領域分割という別の問題を解決しなければならない [81]-[88]。そのため限られた条件下では優れた認識を行える反面、汎用性に乏しい。従ってより一般的な認識問題を扱おうとした場合、その対象に与える適切な知識の定義や表現など、より困難な問題を解決する必要がある。

その様な考えのもとに、階層型ニューラルネットワークの学習能力を利用して、画素単位で対象物の認識を行う方法が提案されている [33]-[35]。画像中の画素を単位とする認識法は、画像中の対象の領域について明確な知識や定義が必要なく、処理を自動化しやすい。Ghosh[33] は、自己組織的に領域を分割し物体の抽出を行い、ファジィ測度を用いることにより、耐ノイズ性を高めている。しかしながら扱われているのは形状が比較的単純な物体のみである。野上 [34] は、複数のニューラルネットワークを組合せた画像の領域認識の手法が提案されている。しかし、対象画像は衛星写真に限られている。椋木 [35] は、対象に関する知識を必要とせず風景画像の優れた認識結果を報告している。しかしこれらは、バックプロパゲーション学習アルゴリズムを用いた階層型ニューラルネットワークを使用しているため、中間層ニューロンの役割が不明である。従って、学習後のネットワークの解析は困難であり、人間に有益なルールや知識の抽出は難しい。

一方、コンピュータによる知識の自動獲得は大変困難ではあるが、さまざまな応用が期待され多くの研究がなされている。そのような中で、知識獲得の問題解決のために、優れた学習能力を持つニューラルネットワークと、言語的処理を得意とするファジィの融合により、双方の長所を活かした手法が提案されてきている。

本章では、これらを融合したファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network)[74] を基本要素として用いることにより、適度な複雑さを持つ風景画像を例題に、認識および、認識に必要な言語形式によるルールの抽出法について述べる。

まず、画像を画素単位で扱う最も基本的な認識手法について述べ、知識処理が可能な学習システムの効果を検証する [76]。その後、領域分割を行い、対象物の形状情報

等を扱うことによって、より高いレベルのルール抽出および画像解釈を目指した研究 [77] について述べる。この研究において領域分割には、多くの特徴量を一括して扱え、広く一般的に知られている K 平均法 [81] を拡張して、パラメータ依存性の低い [82] の手法を用いた。

本章で述べる学習ベースの画像認識システムは、

1. 学習によって画像認識、解釈を自動的に行える。
2. 追加学習により汎用性向上が可能。
3. システムから画像認識に関するルールを抽出することが可能。
4. システムにルールを組み込むことができ、ルールベースとの統合が可能。

という特徴を有している。

これらの研究は画像理解実現のために、自動学習の側面からアプローチしたものであり、例題として適度に複雑さを持ち実画像応用への一段階として考えられる風景画像を扱った。画像認識問題は、知識ベースのシステムのみでは限界があり、高い汎用性実現のためにはシステムの自動学習が必要不可欠であると考えられる。本章では学習機能によって獲得されたルールのみを使用して画像認識を行うシステムについて述べ、成果と課題について検討する。

2.2 画素を処理単位とした画像認識システム

画像の認識には、航空写真の分析などの画素単位で行われる処理と、領域分割などの処理によって画像中の注目している物体を切り出してから認識を行う方法がある [1]。画素単位の処理は、周囲の状況や物体の形状情報などが扱いにくい半面、前処理の影響を受けにくく、また対象とする明示的な知識を必要としない。

本節では、コンピュータの学習機能と言語処理能力を融合したファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network)[74] を用いて、画素単位での最も基本的な画像認識を行い、また学習によって獲得された画像認識に関するルールの抽出を行い、学習システムの有効性を検証する。本論文では、この画素単位で処理を行う画像認識システムを画素単位認識システムと呼ぶ。

2.2.1 システム概要

図 2.1 に画素単位認識システムの概要を示す。図中の KEN と書かれた各要素が、FINN を拡張して構成された知識抽出ネットワーク (KEN: knowledge extraction network) である。画素単位認識システムは複数の KEN が学習により認識に関するルールを獲得し、それらをもとに画像の認識を行う。ルールの抽出は、各ネットワークの学習が終了した時点で生成されたファジィルールが言語形式に変換されることで行われる。画像の認識は、認識したい画像を学習の終わった各 KEN に入力し、それぞれの推論結果を統合し行われる。最終的なルールの抽出は、学習段階によって各 KEN が獲得したルールを統合して得られる。FINN の動作アルゴリズムについては付録 A において詳述する。また、このシステム構成の詳細については付録 B で詳述する。

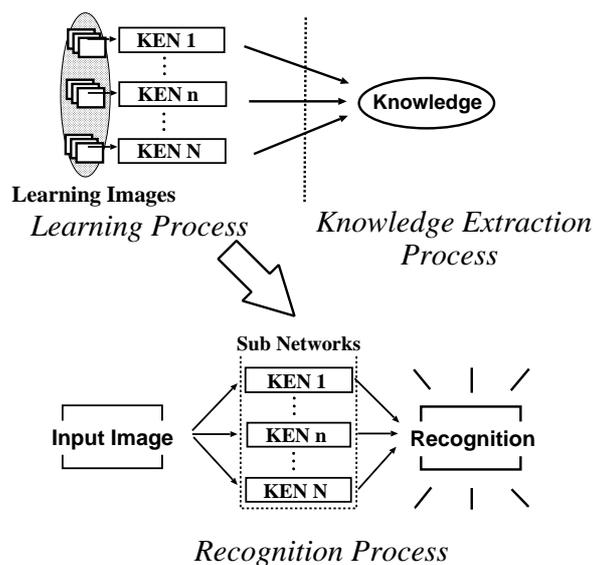


図 2.1 画素単位認識システムの概要

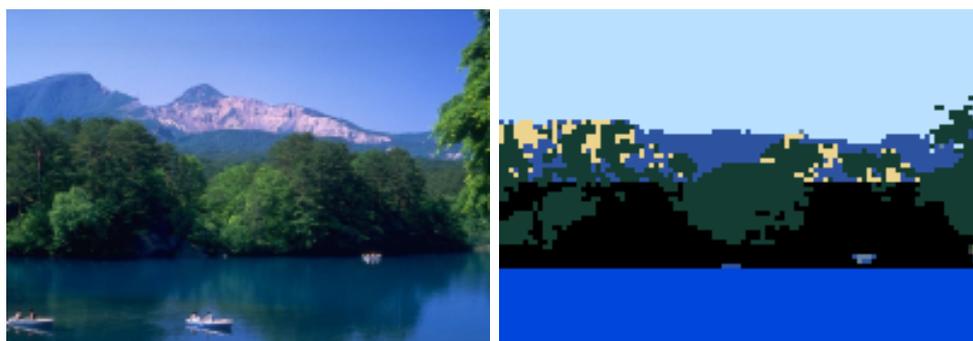
2.2.2 画像の認識結果

システムは、KEN を用いて入力された画像を画素ごとに空、山などの 10 種類の認識ラベルの内 1 つを決定する。全画素についてこの処理を行い、認識結果として出力する。未学習画像に対するシステムの認識結果と、人間による認識結果を表 2.1 にまとめた。表中の数値は対応する各要素と判定されたピクセル数で 20 枚の画像の平均値である。風景画像での適切な 10 種類のラベルつけは人間が行っても難しい部分が多く、評価の基準として必ずしも絶対的なものではないが、参考のため記載している。

また、画素単位認識システムによって得られた風景画像の認識結果例を図 2.2-2.4 に示す。それぞれの画像対において、左側が入力画像で、右側が認識結果である。図中の数値は評価用に与えた人間による認識結果との一致ピクセル数の割合を示している。全未学習データの画素単位での認識において人間との一致度は 59.0%であった。

表 2.1 人間とシステムの認識の違い (20 枚の画像の平均)

		システムによる認識									
		空	雲	水	山	岩	木	草	影	雪	他
人間による認識	空	1778	234	0	13	91	8	8	43	27	12
	雲	424	93	0	2	13	5	0	0	17	0
	水	1	10	406	20	3	21	0	4	3	0
	山	185	25	1	161	12	103	0	4	17	14
	岩	31	9	8	10	290	102	6	59	20	14
	木	59	20	14	101	95	514	40	69	4	10
	草	5	2	3	7	177	236	304	12	10	9
	影	12	3	20	24	108	86	53	410	4	0
	雪	45	14	4	3	3	4	0	0	108	0
	他	14	0	3	1	6	9	0	0	7	1



(a) 79.3%



(b) 71.6%



(c) 85.1%

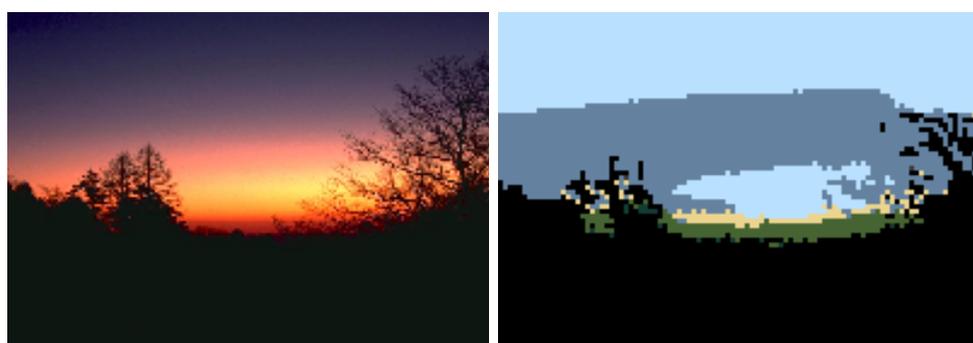
Sky	Woods
Cloud	Grass
Mountains	Shadow
Water	Snow
Rocks	Unknown

(d) 凡例

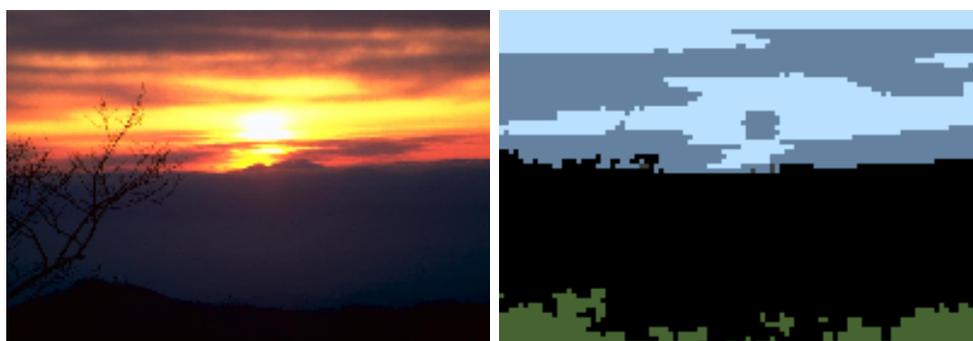
図 2.2 認識結果例 1



(a) 57.8%



(b) 67.8%



(c) 75.2%

Sky	Woods
Cloud	Grass
Mountains	Shadow
Water	Snow
Rocks	Unknown

(d) 凡例

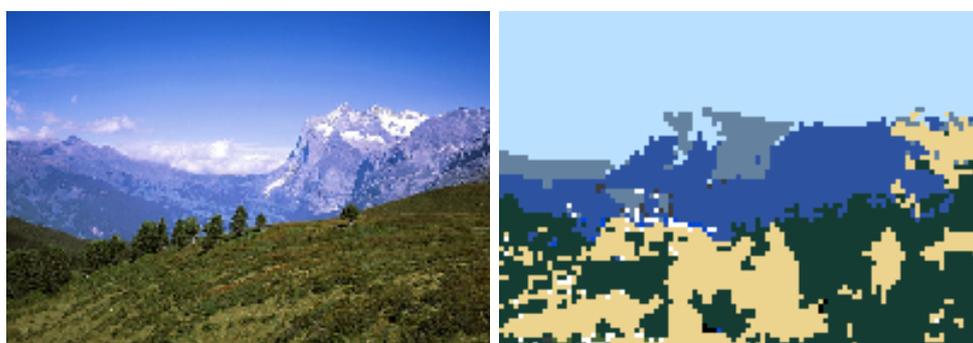
図 2.3 認識結果例 2



(a) 75.3%



(b) 40.6%



(c) 60.8%

Sky	Woods
Cloud	Grass
Mountains	Shadow
Water	Snow
Rocks	Unknown

(d) 凡例

図 2.4 認識結果例 3

2.2.3 抽出されたルール

画素単位での認識手法で獲得された、画像認識に関するルールの例を表 2.2 にまとめる。表中の [] 内の値は、そのファジィルールを出力したニューロンの数で、この値が大きい程一般的なルールであると言える。

表 2.2 獲得したルールの例

空は、
右で、上のほうにあり、白系統の色です。[3]
左右中央で、上のほうにあり、白系統の色です。[3]
左で、上下中央にあり、白系統の色です。[4]
左はじで、上にあり、青系統の色です。[2]
雲は、
左はじで、上のほうにあり、灰色系統の色です。[5]
左で、上のほうにあり、にぶい緑系統の色です。[1]
左右中央で、上にあり、白系統の色です。[1]
水(海、川等)は、
右はじで、下にあり、青系統の色です。[22]
左右には、ほぼ関係なく下にあり、青系統の色です。[45]
左で、下にあり、青系統の色です。[21]
遠くの山(等)は、
左右中央で、上のほうにあり、暗く青系統の色です。[17]
右で、上のほうにあり、暗く青系統の色です。[13]
左で、上下中央にあり、暗く青系統の色です。[12]
土砂類、岩山は、
左右中央で、下にあり、にぶいあまり色に依存しないものです。[3]
左はじで、上下中央にあり、明るくオレンジ(黄色)系統の色です。[2]
右はじで、下にあり、にぶい黄色系統の色です。[2]
左右には、ほぼ関係なく下にあり、明るく黄色系統の色です。[2]
林/樹木等は、
右で、上下中央にあり、暗く灰色系統の色です。[2]
右はじで、上下中央にあり、黒系統の色です。[11]
右はじで、下のほうにあり、暗く灰色系統の色です。[3]
草/草原は、
左はじで、下にあり、暗く灰色系統の色です。[2]
左で、下にあり、にぶい黄色系統の色です。[2]
右で、下にあり、暗くにぶいあまり色に依存しないものです。[1]
かげの部分は、
左右中央で、下のほうにあり、黒系統の色です。[2]
右はじで、下のほうにあり、暗く灰色系統の色です。[2]
右はじで、上下中央にあり、灰色系統の色です。[1]

2.2.4 画素単位認識システムのまとめ

本システムは、画素ごとの処理を行うため、従来の前処理での領域分割などの前処理を必要としない。また、FINN の学習能力とルール記述能力によってデータから認識に関するルールを自動生成するために、明示的な知識を必要としないで、未知の画像の認識を行うことができた。画素単位での認識において人間との一致度が 59.0% という精度は、数値だけとりあげると十分ではない。しかし多くの領域において空と雲、木と草原など人間が行う認識でも曖昧な部分が多く、それらを考慮すると学習のみのシステムとしては良好であると考えられる。

ただしこの手法は、物体の形状に考慮していないため、人工物などの対応に大きな課題が残る。また、色情報と位置情報という最も初歩的な情報のみを用いて処理を行っているため、より高い精度で認識を行うためには、周囲の情報の考慮や、効果的な特徴量を利用し、洗練されたルールの獲得が必要である。

画素単位認識システムにおけるルール抽出は、学習画像というパターン情報のみから、ルールという形でテキスト形式の情報を抽出する、パターンからシンボルへの自動変換の初歩であると言える。

2.3 領域を処理単位とした画像認識システム

本節では、多くの画像認識システムで行われている、前処理、特徴抽出、決定の手法に基づく手法に、学習システムを適用した結果について述べる。本論文ではこのシステムを領域単位認識システムと呼ぶ。領域単位処理システムは分割された各領域を処理単位として認識を行い、また画像が何から構成されているかの情報を複数の単語の組み合わせとして出力する。

「広大な空」「雄大な山」「沢山の木がある」等の言語による画像の構成要素の提示は、画像理解実現に向けた大きな成果であり、応用例として画像検索システムなどへも応用できる。近年の画像検索システムにおいて、画像の内容を自動的に識別しキャラクタ情報に変換する技術は重要であり、本システムが出力する結果は、画像を説明するメタデータに相当する。本論文では、キャラクタ情報である文字列による画像の構成情報表示結果を解釈結果と定義する。

領域単位認識システムでも、前節で利用した FINN を基礎ネットワークとして利用しているが、認識精度向上のため、より多くの特徴情報を利用し、また複数のネットワークを並列に動作させている。基本となる FINN には、適切なルール抽出に利用するために若干の変更を加えた。これについては、付録の A.3.3 に詳述する。

以降、まずシステム構成について述べ、続いて結果と考察を述べる。

2.3.1 システム構成

図 2.5 に領域単位認識システムの概要を示す。システムは 3 つの過程からなる。

1. 領域分割過程 (regional segmentation process)
2. 画像認識過程 (image recognition process)
3. 画像解釈過程 (image interpretation process)

システムはこれらの処理を行うことによって、風景画像を認識しその解釈を行う。

システムはまず、領域分割段階において入力画像をいくつかの領域に分割する。この過程で分割された各領域を、本システムでは処理の単位として扱う。そのため、画像を画素単位で扱うよりも物体の形状や大きさ、その他の様々な特徴の利用が可能となっている。

画像認識過程では、各領域ごとに並列化されたネットワークによる推論が行われ、画像の認識結果が得られる。画像解釈段階においては、認識過程の結果を基に複数の解釈語からなる解釈文を出力する。さらに、ネットワーク構造から風景画像認識、解釈に関するルール抽出もなされる。

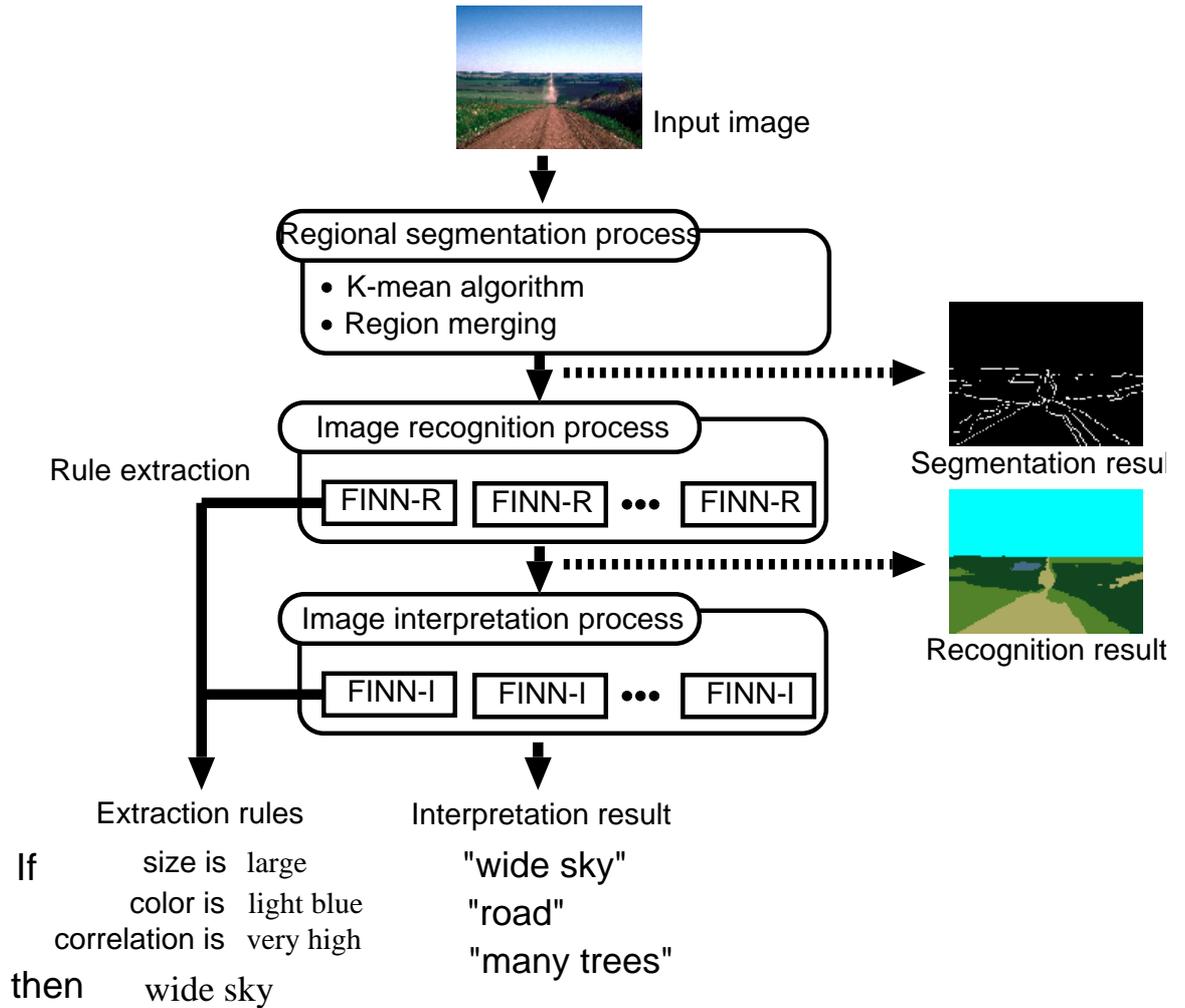


図 2.5 領域単位認識システムの概要

2.3.2 領域分割過程

本システムでは領域を処理の単位として扱うために、前処理として領域分割を行う。分割された各領域から特徴量を抽出し、それを元に推論が行うため、領域分割は非常に重要な前処理である。

入力された画像は適当な数の領域に分割される。この領域分割過程は

- (A) K 平均法による領域分割
- (B) 多数決フィルタによる領域統合
- (C) 小領域の統合
- (D) 原画像とのエッジ比較による領域統合

の4段階で構成される。

この領域分割手法は、最初に細かく分割しそれらを統合していくボトムアップ的なものである。

K 平均法 [81] は、アルゴリズムが簡単で複数の特徴量を1次元的に扱える長所を持っているが、領域を細かく分割しすぎたり、分割結果が初期クラスタやパラメータに大きく影響する等という欠点もある。そこで堺田ら [82] の手法を参考に (B),(C) を利用し、最後にさらに (D) の原画像のエッジと比較して不要な分割を統合するという処理方式を採用した。以下、領域分割過程のそれぞれの処理について述べる。

A. K 平均法による領域分割

K 平均アルゴリズムについての詳しい説明は付録Cに譲り、ここでは簡単に説明する。

本システムでは K 平均アルゴリズムでの領域分割に色情報として (I_1, I_2, I_3) [17][83]、位置情報である (x, y) の計5次元を用いる。 I_1, I_2, I_3 は、各々に輝度情報を持つ RGB 色表示系と比べ、互いに独立性が高く、テクスチャを含む自然画や人物画などの領域分割で優れた結果が報告されている [17]。

I_1, I_2, I_3 は、それぞれ以下の式 (2.1) で求められる。

$$\begin{aligned} I_1 &= (R + G + B)/3. \\ I_2 &= R - B. \\ I_3 &= (2G - R - B)/2. \end{aligned} \quad (2.1)$$

これらの値を用いて、画素とクラスタの距離 L は次式によって計算される。

$$L = \sqrt{k_l \{(I_1 - \bar{I}_1)^2 + (I_2 - \bar{I}_2)^2 + (I_3 - \bar{I}_3)^2\} + k_p \{(x - \bar{x})^2 + (y - \bar{y})^2\}}. \quad (2.2)$$

ここで、 k_l, k_p は特徴量への重み付け係数であり、予備実験の結果より、 $k_l : k_p = 100 : 1$ をパラメータとして採用し、また K 平均法における初期クラスタ数は 3×3 とした。

B. 多数決フィルタによる領域統合

K 平均アルゴリズムは、大きな領域の周囲などに粒状の小さい領域が多数発生するという欠点がある [82]。その不必要に多くなりすぎた小領域を統合する手段としてまず、多数決フィルタと呼ばれる処理を行う。多数決フィルタの概念図を、図 2.6 に示す。この処理では注目しているピクセルを中心とした正方形を用い、中心のラベルの属性をその周囲の属性の中で最も多いものに変化させる。これにより、領域の境界線上にあった細かい多数の小領域が統合できる。フィルタのサイズを大きくすることにより、その効果を調整できるが、大きくしすぎると境界が崩れてしまう。本システムでは予備実験の結果、フィルタサイズを 5×5 とした。

C. 小領域統合

多数決フィルタを用いることによって、かなりの領域数を減らす事ができるが、まだ領域数が多く、対象を切り出すという目的を考えた時に不十分である。そこでこの処理では、領域の大きさが画像全体の画素数の $S_{th}\%$ を下回る小さい領域を隣接する最も大きな領域に統合する。

D. エッジ比較による領域統合

ここまでの処理によって、領域分割の精度は高くなっている。しかしながら、例えば空のように緩やかなグラデーションがある様な大きな領域では K 平均法で生成された不要な境界線が残ってしまう。意味のない多数の領域は、誤認識を引き起こす恐れがあるのでここで統合する。

そこで小領域統合まで終えた分割結果と、元画像のエッジとの比較を行う。エッジ画像との比較の際の画素のズレを考慮し、扱うエッジ画像にはあらかじめぼかし処理を加えておく。エッジ比較による領域統合の概念図を図 2.7 に示す。ここでは、分割結果の境界線上にある強度 E_{th} 以上のエッジが領域の境界線の $R_{th}\%$ 以上なければ、それらの領域は単一のものとして統合する。

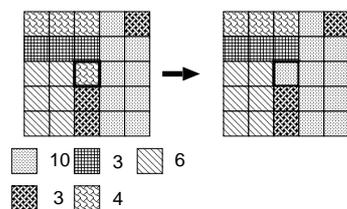


図 2.6 多数決フィルタ (5×5)

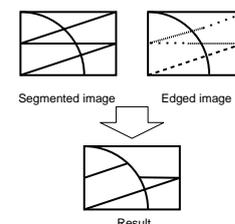


図 2.7 エッジ比較による領域統合

2.3.3 画像認識過程

領域分割された画像は、この過程でそれぞれの領域に対して認識ラベルが与えられる。この過程では、FINN[74]を画像認識用に拡張したネットワークを使用し、本章ではこのネットワークをFINN-R(Recognition)と呼ぶ。

FINN-R の入出力

FINN-Rの入力および、推論によって出力される認識ラベルを表2.3にまとめ、図2.8に構造図を示す。

FINN-Rは画像分割によって分けられた各領域ごとに、予め予備実験によって選択された領域の大きさ、位置、形状、色、テクスチャ性の計20次元を入力として受け、認識ラベルである10種類のうち一つを推論結果として出力する。本システムで利用したテクスチャ情報は、「木目、砂地、芝生の様にその要素も規則もまったく判然としないものを解析するのに有効で、その画像の統計的性質により特徴を記述する」統計的解析で用いられる特徴量である[95]。

このFINN-Rによる領域ごとの推論を、すべての領域で行うことによって画像全体の認識を終える。FINNは、それぞれの入力に対してメンバシップ関数が学習によって自動的に調整される。メンバシップ関数の幅が小さいことは、その入力値の変動に対してシステムが敏感であることを表すため、この幅の大きさによって各入力が有効なものか否かを判別できる。入力情報の選択の予備実験には、この性質を利用した。

表 2.3 FINN-R の入出力

入力	
領域の大きさ	L
領域の重心位置	x, y
領域の形状	σ_x, σ_y
領域の平均色	V (輝度), H (彩度), S (色相)
テクスチャ性	contrast (コントラスト) $0^\circ, 45^\circ, 90^\circ, 135^\circ$ uniformity (一様性) $0^\circ, 90^\circ$ correlation (相関) $0^\circ, 90^\circ$ entropy (エントロピー) $0^\circ, 45^\circ, 90^\circ, 135^\circ$
出力	
出力ラベル	sky, clouds, water, mountains, rocks woods, grass, light & shadow, snow, soil & sunset

ネットワークの並列化

本システムでは、この FINN-R を学習画像の種類により複数使用している。提案システムで用いた、複数の FINN-R を用いた認識システムの概念図を図 2.9 に示す。

図中のそれぞれのルール層が、一つの FINN-R に対応しており、入出力は一つにまとめられている。入力された特徴情報はすべての FINN-R に入力され、最もルール適応度 ρ (式 (A.3) 参照) が高いルールを持つネットワークの出力を利用する。FINN の特性と、この複数のネットワークを利用する事によって、以下の様な多数のメリットが得られる

- 領域毎に適切なネットワークの出力が得られるため正確さが向上する。
- 各ネットワーク規模を小さくでき、学習できる合計画像数が大幅に向上する。
- 各領域の認識の信頼度をルールの適応度 ρ から推定できる。
- 学習画像の追加 (ネットワークの追加) が容易であり、汎用性向上が可能。
- 追加するネットワークに予め、ファジィルールの形で知識を埋め込む事ができる。

本システムでは FINN-R を 10 個用いている。

ネットワークの数を変化させる実験により、並列化による汎用性の向上ならびに、正確さの向上が確認されている。この結果は 2.3.7 節に掲載している。

FINN は学習した画像の種類に、推論結果が大きく左右される。そこで、ネットワークを複数用意して、学習画像数を増やす事によってこの問題を解決すると同時に、各ネットワークに学習させる画像について考慮する必要がなくなる。また、大規模なネットワークを作成するには大変長い学習時間を必要とし、かつ学習容量の面でも小規模のネットワークを複数作成した方が優れている。

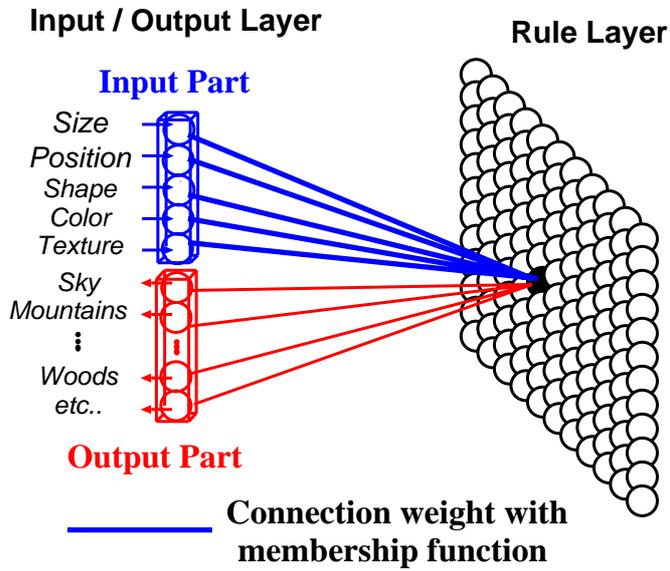


図 2.8 FINN-R の構造図

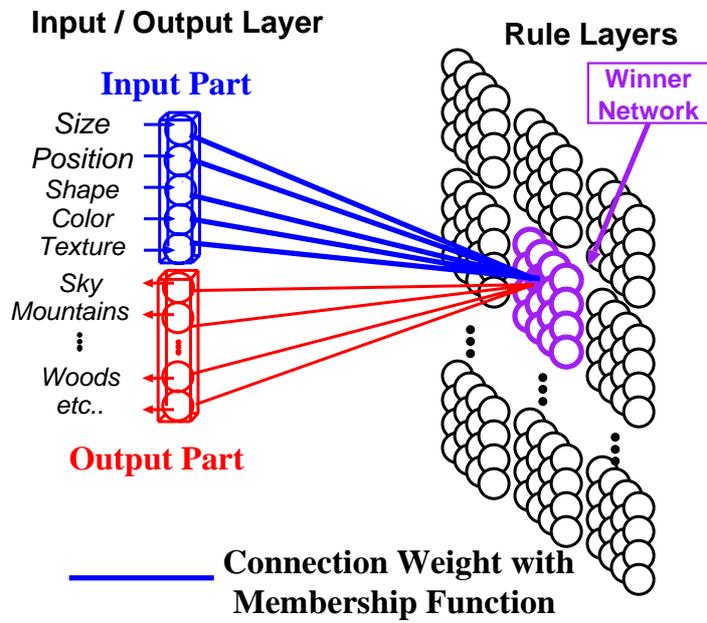


図 2.9 画像認識ネットワーク

2.3.4 画像解釈過程

画像解釈過程の概要を図 2.10 に示す。画像認識過程が終了すると各領域には、表 2.3 の出力ラベルの内どれか 1 つが割り当てられる。この画像解釈過程においては、各ラベル毎の領域情報が取り出され、画像解釈用の FINN に入力されることにより、詳細なラベル付けが行われる。この詳細なラベル付けの結果、各画像は詳細なラベルによってインデキシングされる。本論文ではこの与えられた詳細のラベルの集まりを、画像の解釈結果としている。使用した解釈ラベルを表 2.4 に示す。

この過程の推論に用いる画像解釈用 FINN を FINN-I (Interpretation) と呼ぶことにする。FINN-I は各認識ラベル毎に用意され、本章で扱うシステムのの場合認識ラベルと同数の 10 個使用する。

FINN-I の入力には、大きさ (L)、位置 (x, y)、形状 ($\sigma_x^2, \sigma_y^2, \sigma_x^3, \sigma_y^3$)、色情報として HSV 基底の色相 (H)、彩度 (S)、輝度 (V) さらに、色分散 ($\sigma_H^2, \sigma_S^2, \sigma_V^2$) の計 13 次元を用いた。出力には各ラベル毎に適当と思われる解釈語 2~4 種類に、それぞれその他を加えた計 3~5 種類を使用した。

各ラベル毎の FINN-I からもっとも高い推論結果がそのラベルの解釈結果となる。そして、全 FINN-I の出力の推論値 y_k (式 (A.4)) 参照) の高い上位 5 位までで、かつ 0.5 以上のもののみが解釈文に使われる。例えば、空の FINN-I (FINN-Isky) が「広大な空」(0.7)、山の FINN-I (FINN-Imountain) が「雄大な山」(0.8)、水の FINN-I (FINN-Iwater) が「湖」(0.4) とそれぞれ出力を示した場合、「広大な空」「雄大な山」という解釈結果が得られる。

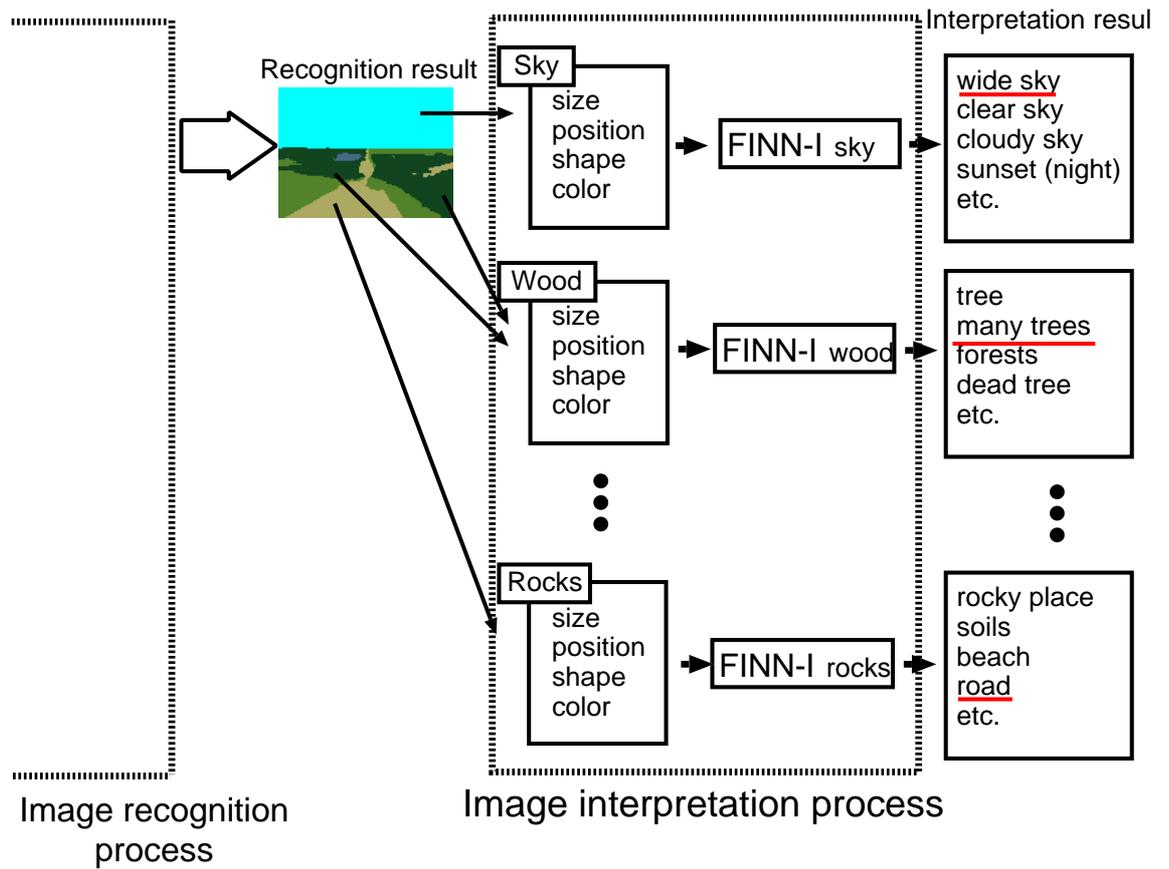


図 2.10 解釈文生成部の流れ

表 2.4 付与する解釈語 (FINN-I の出力)

学習ラベル	付与する言葉	学習ラベル	付与する言葉
空	広大な空 晴れ渡った空 曇った空 夕焼けの(夜の)	木	森 木がたっている 沢山の木がある 枯木(紅葉)がある
雲	晴れた 雲のある 曇った 夜の空	草	草原 花畑 暗い草原 畑
水	海 湖 川 その他の水	影	暗い 影のある 山の影の
山	雄大な山 険しい山 遠くに山のある 半島状の	雪	雪山の 雪がつもった
岩	岩山 砂利(土)のある 砂浜の 道状の	光	太陽 何かが輝いている 光がまぶしい

2.3.5 画像認識、解釈に関するルール抽出

領域単位認識システムでは、推論の基礎に FINN を用いているために学習後のネットワーク状態からファジィif-then ルールを抽出できる。つまり、ネットワークを解析することによって、どのような入力条件からどのような出力が得られているかが導出できる。

本システムでは、認識過程で認識された各領域毎の結果を用いて解釈を行っているので、認識過程部の入力を得られるルールの前件部、解釈過程で得られる結果をルールの後件部とした。ルール抽出の流れを図 2.11 に示す。

システム全体像である図 2.5 および、図 2.11 を用いながらルール抽出のプロセスを説明する。画像認識過程において、領域分割された画像の各領域の情報は FINN-R に入力される。すべての FINN-R の中で、最もルール適応度 ρ が高いルールをもつ FINN-R が選択され、その領域の推論は選択された FINN-R で行われる。図 2.11 の例においては、上部にある空領域に対して「とても大きな領域で青く、相関が高ければ空」というルールが最も大きなルール適応度を持ち、そのルールを持つ FINN-R が推論に用いられている。

続く解釈過程において「空 (sky)」と認識された領域は、解釈段階で空領域の認識 FINN-Isky によって「広大な空 (wide sky)」と解釈される。この認識段階で選択された FINN-R のルールの前件部を抽出するルールの前件部。解釈段階で選択された FINN-I のルールの結果を抽出するルールの後件部として統合した。従ってこの場合、抽出されるルールは「大きな領域で青く、相関が高ければ広大な空」となる。

これらの獲得されたルールは、別途他の知識としてルールベースのシステムに組み込むことも可能である。ルールベースのシステムでは、こうした多くのルールを人間が入力しなければならないが、本システムでは学習によって自動的にこれらのルールを獲得し、利用することができる。

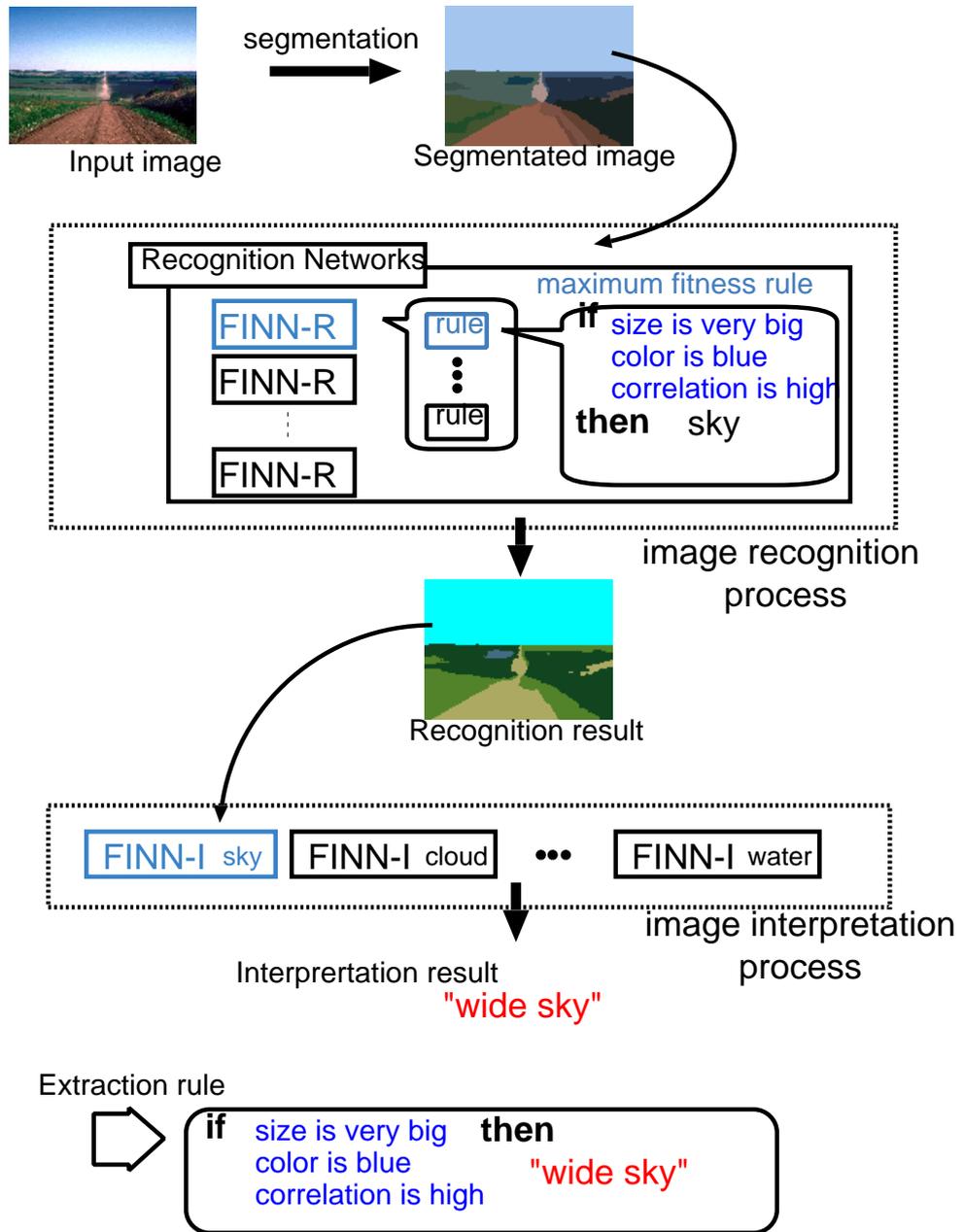


図 2.11 システムからのルール獲得の流れ

2.3.6 計算機実験

提案システムをコンピュータ上に実装し、評価を行なった。

A. 実験条件

表 2.5 に画像分割過程のパラメータを示す。

画像認識用 FINN(FINN-R) および、画像解釈用 FINN(FINN-I) の構造パラメータを以下の表 2.6 に示す。また FINN-R, FINN-I の学習パラメータを表 2.7 に示す。学習画像には無作為に選んだ 80 枚を、10 個の FINN-R に 8 枚ずつ振り分けた。

表 2.5 画像分割、統合時のパラメータ

K 平均法の初期クラスタ	3×3
多数決フィルタの格子	5×5
$k_l : k_p$ (式 2.2, 付録 C.1 参照)	100 : 1
領域毎の最小ピクセルの画像全体 S_{th}	0.4
エッジ統合時の有効最低輝度値 E_{th}	30 (max255)
エッジ統合時の統合閾値 R_{th}	20%

表 2.6 FINN-R, FINN-I の構造パラメータ

項目	FINN-R	FINN-I
ネットワークの数	10	10
入力部のニューロン数 N_1	20	13
ルール層のノード数 N_2	225 (15×15)	400 (20×20)
出力部のニューロン数 N_3	10	5
ネットワークが学習した画像の枚数	計 80 枚 (各 8 枚 \times 10)	計 40 枚

表 2.7 FINN-R, FINN-I の学習パラメータ

項目	FINN-R	FINN-I
入力次元数 N_1	20	13
出力次元数 N_3	10	10
ルール層のサイズ N_2	225 (15 × 15)	400 (20 × 20)
自己組織化学習係数 ϵ_{self}	0.5	0.5
自己組織化学習回数	4000	4000
α_{in} (式 A.18) 参照	1.0	1.0
α_{out} (式 A.18) 参照	4.0	4.0
LMS 学習係数 (重み) ϵ_{LMS}^w	0.001	0.001
LMS 学習係数 (メンバシップの幅) ϵ_{LMS}^w	0.01	0.01
LMS 学習回数	6000	3000
学習パターン (領域) 数	150 ~ 200 程度	40

B. 領域分割・統合結果

図 2.12, 2.13 に領域分割・統合の例を示す。各 (c)-(f) の注釈の後にある数値は、それぞれ領域数を表す。画像中の主な対象毎に切り出しが行われており、風景画像認識の前処理として良好な分割結果と考えられる。他の様々な画像に対しても、ほぼ同様の結果が得られている。複雑な画像の場合、 K 平均法および領域統合処理に時間がかかるが、長くて 20 秒程度であり、最終的な領域数はほぼ一定である。

なおパラメータに関しては予備実験結果から、この多数決フィルタのサイズを 5×5 とし、小領域統合の閾値 $S_{th} = 0.4\%$ 、エッジ統合時のパラメータをそれぞれ $E_{th} = 30/255$, $R_{th} = 20\%$ としている。

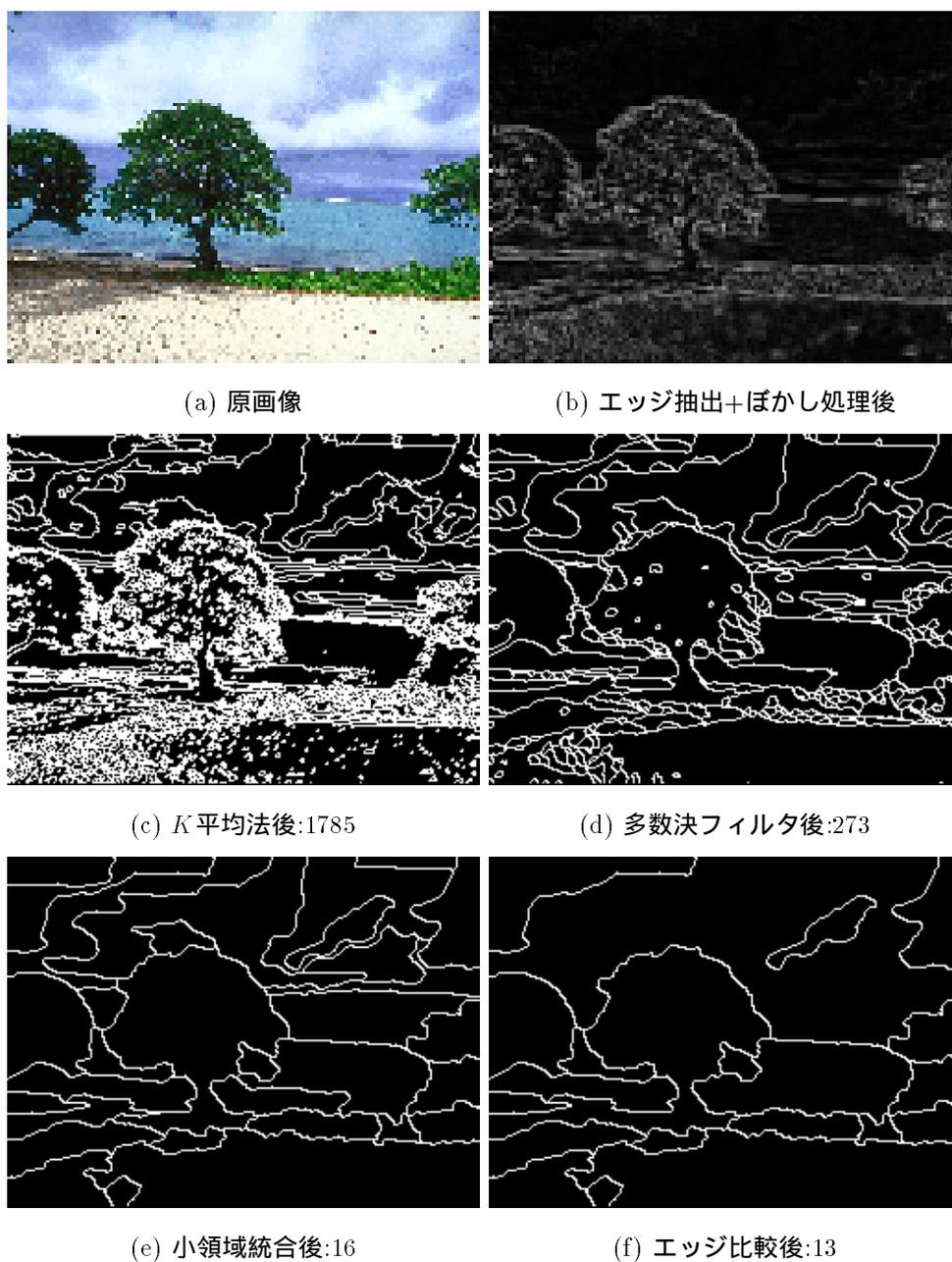


図 2.12 画像分割例 その1

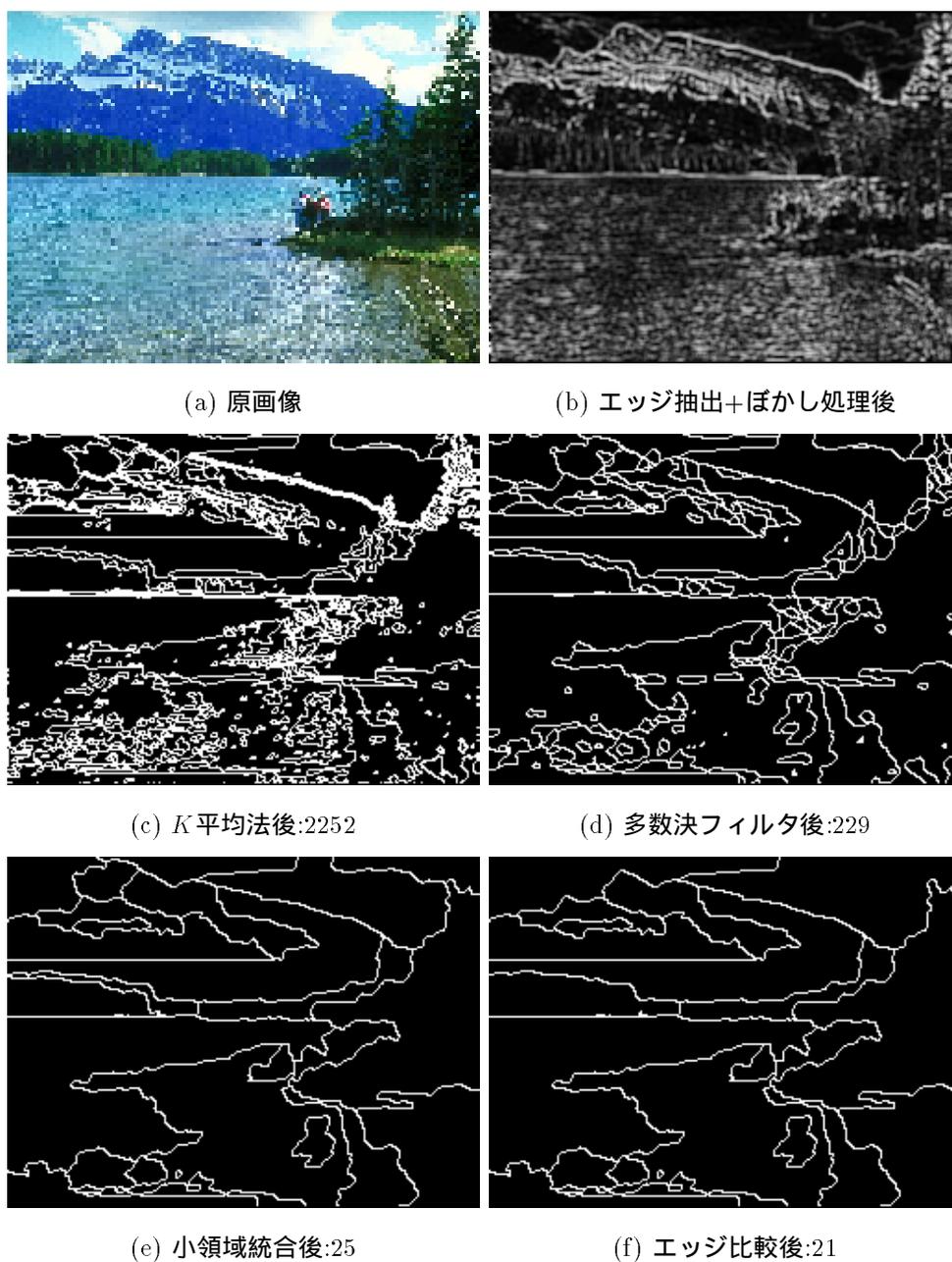


図 2.13 画像分割例 その2

C. 画像認識・解釈結果

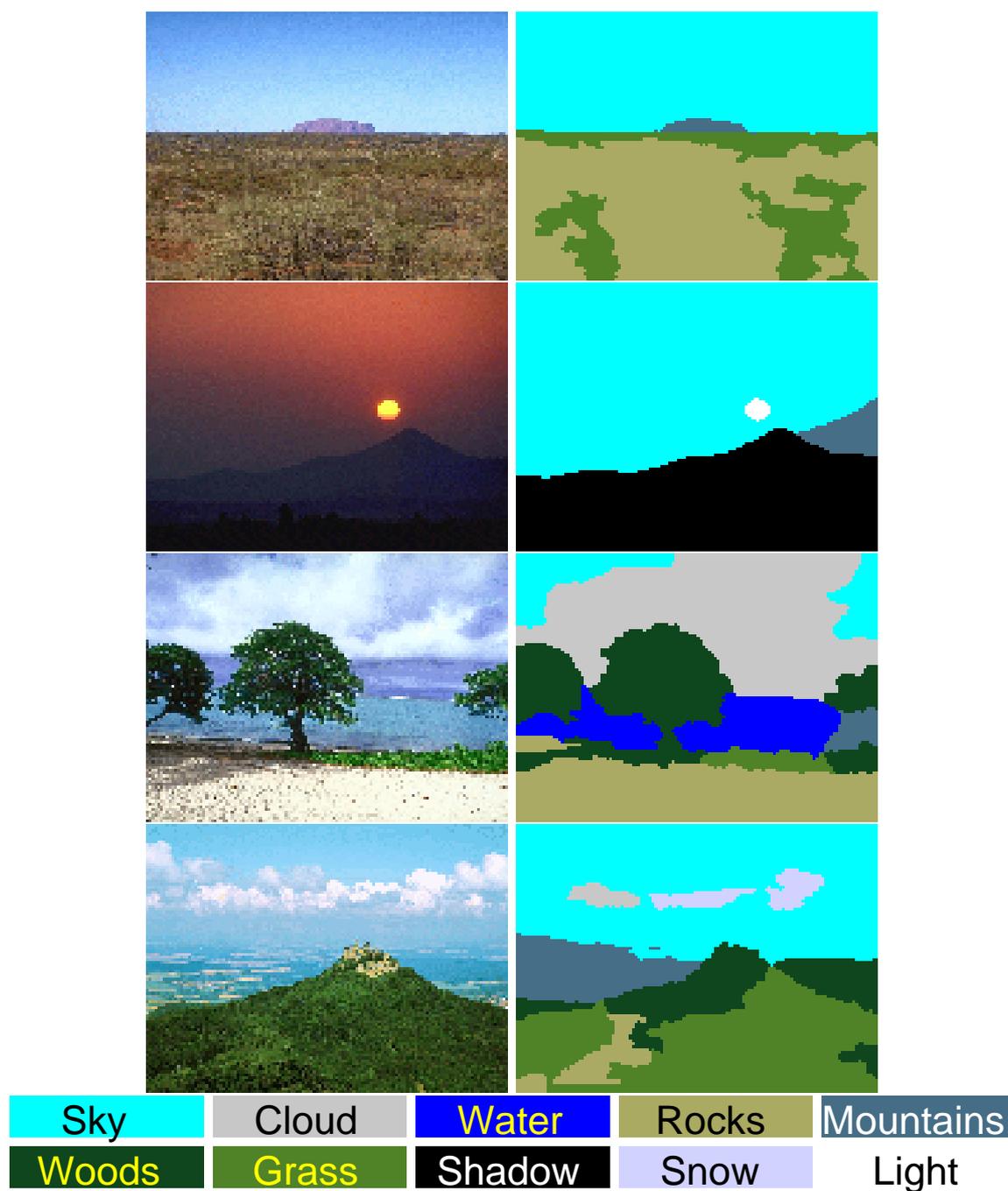
未学習学習データに対する認識・解釈結果の例を図 2.14~2.16に示す。それぞれにおいて、左側が入力画像、右側がそれに対応した認識結果である。

図の下にある解釈語群が、それぞれの画像の解釈結果であり、数字は人間が認識した結果とシステムの出力を比較した際の一致したピクセル数の割合 (ピクセル認識率) である。認識性能の絶対的指標にはならないが、参考までに併記している。

また、50 枚の未学習画像の認識結果と人間の認識結果の一致ピクセル数の割合の平均値は 71.9%であった。

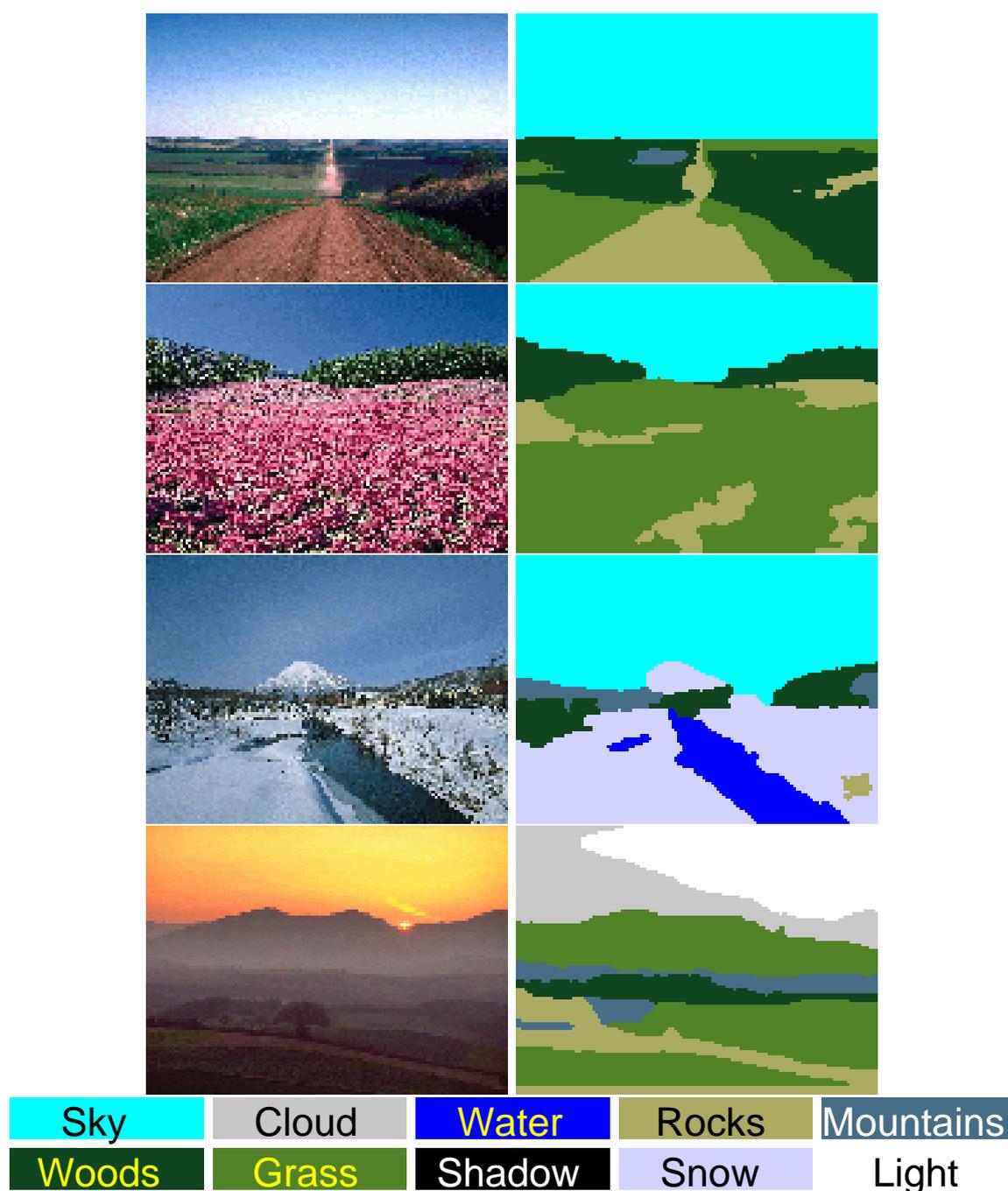
人間の認識との一致度が約 70%という値は、認識対象の多様性を考慮すると学習のみによって到達できるレベルとしては良好な結果と考えられる。また、画素ごとに処理を行った手法 [76] と比較 (59.0%) しても大きく向上している。また、提案システムは FINN を基本要素としているので、認識の誤りの部分、あるいは、ルールの適応度が低い部分に対して、ルールを新たに組み込む事ができる。

画像の認識・解釈問題は、人間の知識がなければ困難な場合が多い。本システムが認識・解釈を行い、それに付加する形で組み込み式のルールを利用できれば、さらに精度は向上できると考えられる。



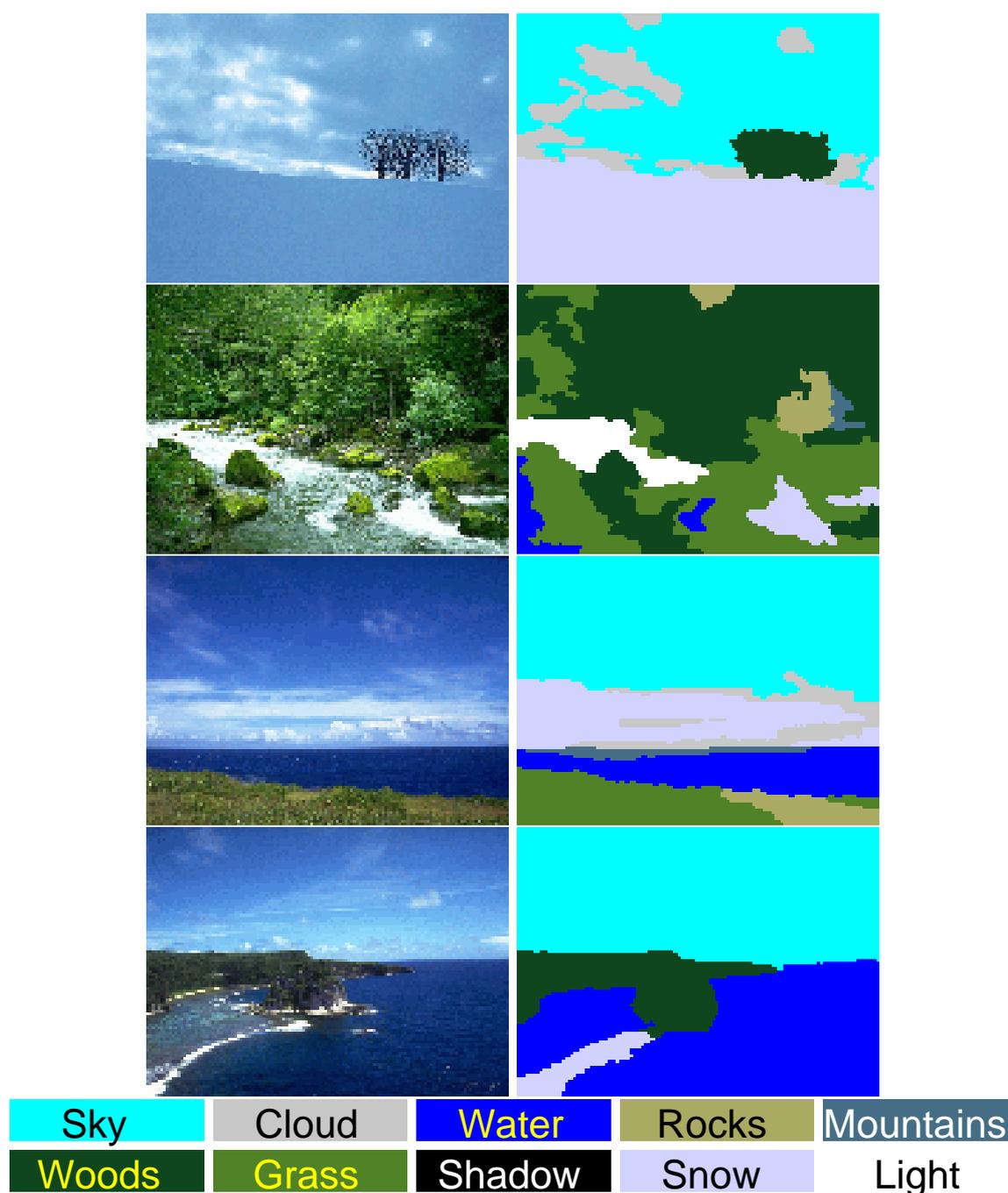
- (1 列目の画像) 広大な空・草原・遠くに山 ピクセル認識率：92.6%
- (2 列目の画像) 夕焼け空・太陽が輝く・暗い ピクセル認識率：97.3%
- (3 列目の画像) 曇った空・海・砂浜・木 ピクセル認識率：90.2%
- (4 列目の画像) 広大な空・沢山の木・遠くの山 ピクセル認識率：49.3%

図 2.14 画像認識・解釈結果例 1 (未学習)



- (1 列目の画像) 広大な空・道・沢山の木 ピクセル認識率：75.4%
- (2 列目の画像) 晴れ渡った空・沢山の木・花畑 ピクセル認識率：89.2%
- (3 列目の画像) 広大な空・海・紅葉(枯木)・雪が積もった ピクセル認識率：75.5%
- (4 列目の画像) 夕焼け空・曇った・遠くの山・砂利・沢山の木・太陽が輝く ピクセル認識率：53.2%

図 2.15 画像認識・解釈結果例 2 (未学習)



(1 列目の画像) 曇った空・木・雪が積もった ピクセル認識率：64.8%

(2 列目の画像) 森・川・雪が積もった ピクセル認識率：62.4%

(3 列目の画像) 晴れ渡った・雲・海・雪山 ピクセル認識率：79.4%

(4 列目の画像) 晴れ渡った空・海・沢山の木 ピクセル認識率：96.7%

図 2.16 画像認識・解釈結果例 3 (未学習)

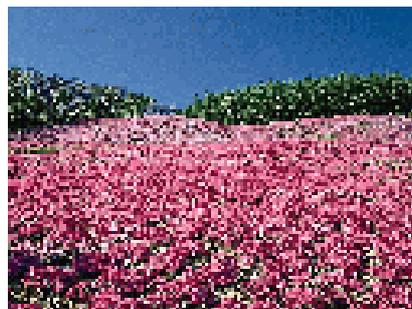
D. ルール抽出結果

本システムによって得られた画像認識、解釈に関するルールの例を図 2.17～2.20に示す。ごく自然なルールが抽出できていることが確認できる。またシステムは、これらの学習によって自動的に獲得したルールを用いて画像の認識、解釈を行っており、その結果が良好なことからも抽出されたルールが妥当なものであると言える。



- 草原：黄色・彩度が低い・一様性が低い
- 広大な空：上にある・青色・画素値のばらつき (エントロピー) が小さい

図 2.17 ルール抽出例 1



- 花畑：やや下の方・ピンク色・一様性が低い・画素値のばらつき (エントロピー) が大きい

図 2.18 ルール抽出例 2



- 海：輝度が高い・水色・水平方向以外のコントラストが高い・一様性が低い
- 木：輝度が低い・緑色・画像の中央部付近にある・画素値のばらつき (エントロピー) が大きい・一様性低い

図 2.19 ルール抽出例 3



- 積もった雪：下の方にある・輝度が高い・横方向に一様性が高い

図 2.20 ルール抽出例 4

2.3.7 考察

A. システムの並列性について

本システムは、FINN-R を並列化させる事で認識精度を高めている。

図 2.21 に 50 枚の未学習画像に対する FINN-R の数と、平均ルール適応度 $\bar{\rho}$ 、認識精度の関係のグラフを示す。

平均ルールの適応度 $\bar{\rho}$ は、画像中の各領域におけるファジィ推論において使用したルールの適応度 ρ の平均値である。この値が高いことは、システムがその入力に対する適切な処理方法を獲得していることを意味する。

図 2.21 より、用いた FINN-R の数が増えるほど平均ルール適応度 $\bar{\rho}$ および、認識精度が上昇している事が確認できる。このネットワーク追加による精度の向上は、本システムの拡張の有効性を示しており、汎用性が求められるシステムには重要な機能である。

また、平均ルール適応度 $\bar{\rho}$ と、認識精度の関係を調べたグラフを図 2.22 に示す。図 2.22 は FINN-R の数を 1 ~ 10 まで変化させ、それぞれ各 50 枚ずつ、のべ 500 枚の未学習画像に対するルールの適応度と認識精度の関係図である。これらの図より、平均ルール適応度 $\bar{\rho}$ と、認識精度には正の相関があることが確認できる。図中の青いプロットが、FINN-R の数が 10 個の提案システムの結果 ($\bar{\rho} = 0.64$, 平均認識精度 = 71.9%) を表しており、全体的に適応度、認識精度共に高い場所に多く存在している。

平均ルール適応度 $\bar{\rho}$ が 0.7 を超える領域では、平均認識精度が 80% を超える結果が得られている。したがって、ルールの適応度は認識結果の信頼度を表す指標になっていると考えられる。

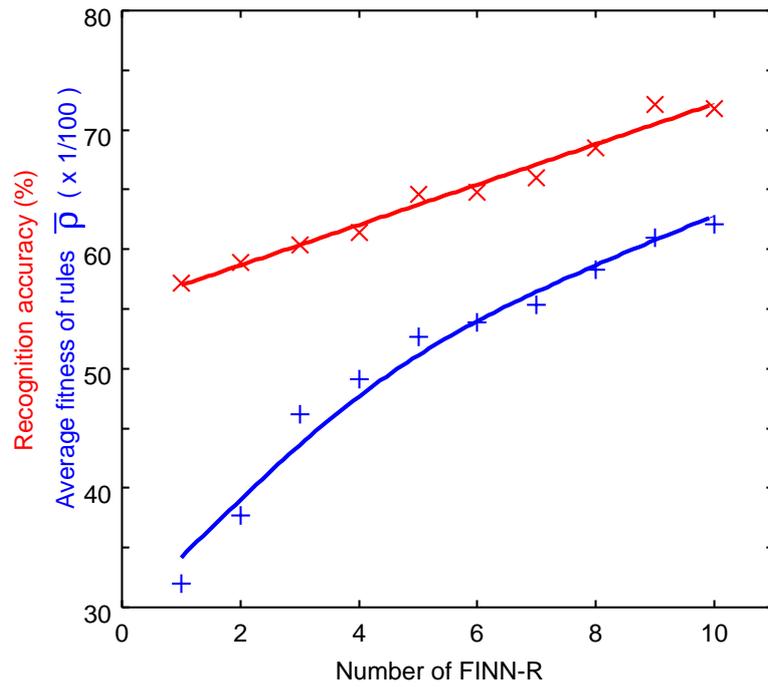


図 2.21 FINN-R の数とルール適応度・認識精度の関係

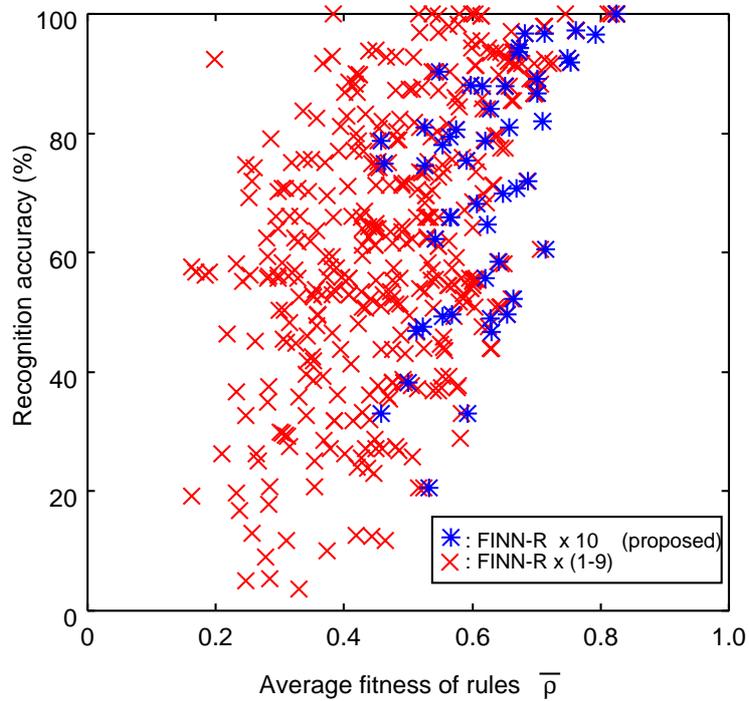


図 2.22 ルール適応度と認識精度の関係

B. 認識結果の分析

表 2.8 は、提案システムにおいて 50 枚の未学習画像に対して認識を行った際の人間の認識結果と、システムによる結果の比較結果である。表中の値は、50 枚の画像の総ピクセル数 $((256 \times 192) \times 50)$ を 100 とした 100 分率で表され、対角成分の数値が多い程認識精度が良いと言える。また、各認識ラベルの認識の詳細についての結果を得るために、その要素ごとを 100 とした結果を表 2.9 とした。

表 2.8 より、認識テスト画像には空領域や草原領域が相対的に多く存在している事が分かる。また多かった誤りは、草を岩との間違い、またその逆のケースであった。これらは色、テクスチャ共に似たものが多く、また領域分割によって形が現れにくいという点が原因として考えられる。

次に、各項目ごとに着目して結果について考察する。空領域は、認識一致度が 90% を越えており、領域分割によって適度な特徴を抽出でき、それらを認識過程で有効に利用できたと考えられる。一方で実験した画像すべてに対して、山と考えられる領域は全体の 5% 程度であるが、山の認識は非常に難しく、認識精度は 4 割弱となっている。山は、木や水、空等との誤りが多くなっている。遠くにある山は青い色をしているものが多く、その中でテクスチャ性が類似するものが誤って水と判断されたと考えられる。また一方で、山領域と多くの木が存在する領域の厳密な区別は人間にとっても難しい。これは、空と雲の認識が難しいのと同じである。雲に関しては、空、雪と誤る場合が多い。似た色を持ち、存在している場面も多く、形も捉えにくい。学習のみのシステムでこれらの問題の解決を計るならば、有効な特徴量の追加導入が必要であると考えられる。

表 2.8 システムと人間との認識の差 (全体) (%)

		システムによる認識										
		空	雲	水	山	岩	木	草	影	雪	他	合計
人間による認識	空	28.3	0.7	0.5	0.9	0.1	0.3	0.0	0.0	0.0	0.0	30.8
	雲	1.6	5.7	0.0	0.1	0.1	0.2	0.3	0.3	1.5	0.1	9.9
	水	0.9	0.0	5.5	0.2	0.1	0.2	0.8	0.0	0.5	0.0	8.2
	山	0.7	0.4	0.7	2.1	0.1	1.1	0.6	0.0	0.0	0.0	5.7
	岩	0.1	0.5	0.4	0.2	6.6	0.7	1.7	0.1	0.8	0.1	11.2
	木	0.1	0.1	0.0	0.2	0.8	6.0	2.7	0.0	0.0	0.0	9.9
	草	0.0	0.0	0.0	0.0	3.4	1.5	7.9	0.0	0.0	0.2	13.0
	影	0.0	0.0	0.0	0.0	0.1	0.0	0.0	4.9	0.0	0.0	5.0
	雪	0.0	0.2	0.2	0.0	0.1	0.0	0.0	0.0	3.5	0.0	4.0
	他	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	1.4	1.7

表 2.9 システムと人間との認識の差 (ラベル毎) (%)

		システムによる認識									
		空	雲	水	山	岩	木	草	影	雪	他
人間による認識	空	90.2	2.3	1.5	2.8	0.2	0.8	0.0	0.0	0.4	1.9
	雲	15.8	57.6	0.4	0.8	0.9	1.9	3.0	3.4	15.6	0.5
	水	10.9	0.6	67.3	2.0	1.1	3.0	9.5	0.0	5.6	0.2
	山	11.9	6.8	12.3	37.0	1.4	19.1	10.6	0.0	0.8	0.0
	岩	1.2	4.2	3.6	1.4	59.4	6.2	15.7	0.5	7.2	0.5
	木	1.5	0.7	0.4	2.2	7.9	60.0	27.4	0.0	0.0	0.0
	草	0.0	0.0	0.1	0.2	26.3	11.2	60.6	0.0	0.0	1.5
	影	0.0	0.0	0.0	0.3	1.2	0.0	0.9	97.6	0.0	0.0
	雪	0.9	4.4	5.5	1.0	1.5	0.0	0.0	0.0	86.3	0.4
	他	0.0	0.0	0.0	0.0	14.7	0.0	0.0	0.0	3.8	81.5

2.4 結論

本章では、画像認識および解釈をファジィ推論ニューラルネットワークの学習機能を利用することによって行う手法について述べた。この研究は画像理解実現のためにコンピュータの自動学習の側面からアプローチをしたものであり、例題として風景画像を扱っている。画像認識や理解問題は、知識ベースのシステムだけでは汎用性に限界があり、高い汎用性実現のためには自動学習機能が必要不可欠であると考えられる。

画素単位の認識手法では、画素の色情報と位置情報という最も基本的な情報のみから、未知の入力画像に対する認識が行える点と、認識に関する適切なルールの抽出が行えることが確認できた。これは学習機能を持つ FINN が入力データ群と、認識結果であるラベルに対する入出力関係を学習でき、かつその結果を外部で解釈可能であることを意味している。精度の限界はあるものの、認識対象に対する明示的な知識を人間が与えることなく、画素の色と位置という非常に単純な情報のみから、画像の認識が行えたことは、学習システムの有効性を示唆する結果であると言える。

続く領域毎にデータを扱う認識システムでは、形状やテクスチャ性など、より多くの特徴量を扱う点とネットワークの並列化により、さらに大幅に良好な結果を得る事ができた。また、階層的な推論を行うことにより、認識対象に対してより詳細なラベル付けが可能であることを示した。これらのシステムは構成要素に FINN を用いているため、トップダウンでの知識を埋め込むことが可能となっている。

本章では、知識組み込みの拡張性を持つ学習による認識システムについて述べた。システムは学習により、認識に関するルールを自動的に獲得するため、画像に対する明示的な知識の付与を必要とせずに、認識を行うことが可能である。

第 3 章

適応ファジィ推論ニューラルネットワーク (AFINN)

本章では、学習によって適切な処理構造や、パラメータを自動的に決定し、獲得したルールを抽出できる適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference neural network)[79] について述べる。

前章で述べた研究では、FINN[74] を基礎ネットワークとして用い、風景画像を例題に、外部からの明示的な知識の付与なしに良好な認識結果を得た。しかし、風景画像以外の一般的な画像の認識も行う場合、認識対象に関する制約が大幅に減少するため、より高度な推論機能が必要となる。従来扱って来た FINN は対象に関するモデリング機能の不足などから、獲得できるルールの一般性等に課題が残っていた。

本章で述べる AFINN は、多様な実アプリケーションの基礎コンポーネントとして使用できるよう、学習データに応じて処理に適したネットワーク構造やパラメータを自動的に決定する。また従来ファジィニューラルネットワークでは困難であった多次元データの処理も可能である。

本章では AFINN の構造と動作、モデリング機能について述べる。続いて AFINN のファジィモデル獲得能力を確認するために、例題としてまず数式の近似問題を扱った。また、UC Irvine が一般に公開している試験データ [91] を用いて 600 次元を越える音声認識問題や車の評価問題を扱い、優れた分類精度と、抽出されたルールの妥当性を確認した。

3.1 はじめに

本研究は、ファジィニューラルネットワークの学習機能と知識獲得機能に着目し、画像認識や理解を目指す上での基礎ネットワークの強化を目的とした研究である。

前章で基礎ネットワークとして用いて来た FINN[74] は、他の多くのファジィニューラルネットワークと同様、自己組織化によって初期ルールを決定した後、ルール間のユークリッド距離を基に不要なルールの削除を行う。その後、LMS(least mean square) アルゴリズムを用いてパラメータの調整ならびに、精度の向上を実現している。しかしながら FINN には、有力な入力要素を選択するようなモデリング機能が備わっておらず、大規模データを一括して扱うには課題が残されていた。

一方、適切なファジィ推論モデルを構築するためには以下の問題を解決する必要がある [67]。

1. 適切な入力特徴の選択
2. 最適なルール数の決定
3. パラメータの決定

適切な入力特徴の選択は、ファジィモデリングに限らず全てのパターン分類問題に共通する問題で、大きな次元数を持つ情報を扱う際の入力要素の選択および圧縮は必須である [13]。有名な入力選択手法には PCA(principal component analysis) や MDS(multi dimensional scaling) を用いる方法があげられる。これらの手法は入力要素の選択に、要素間の線形の結合や相関を元に選択が行う。このため、非線形の相関を持つデータセットから適切な選択が行えない。一方、線形、非線形問わずどんな方法にも広く利用できる stepwise 法は、元の入力次元数が増えるに従って階乗的に計算量が増加するため、大規模システム、特に非線形システムでの適用は現実的ではない。一方、適切なルール数の設定はニューラルネットワークのモデリングの課題として残っていた。

本章では、構成が簡易で汎用性の高い FINN[74] をベースに上記の特徴を実装した、適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference neural network) について述べる。

また、他のファジィニューラルネットワークでも当てはまるが、これまでのネットワークは動作アルゴリズムの中でファジィ推論を行っているために、およそ数十次元以上の多次元の入力に際して、学習が十分に行えない欠点があった。AFINN ではこの部分についても推論アルゴリズムの若干の変更によって解決している。

3.2 AFINN の構造と動作

3.2.1 AFINN の構造

AFINN は、入出力層とルール層の 2 層から構成され、構造的にはファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network) と全く同様である。FINN の構造については、付録 A.1 に詳述する。

3.2.2 AFINN の動作

AFINN の推論アルゴリズムも、FINN とほぼ同様であるため、変更部分のみを示す。AFINN では、従来のファジィ推論では困難であった高次元推論の実現のために、FINN でのルール適応度 ρ の計算式 (A.3) を以下の式 (3.1) に変更している。

$$\rho_j = \prod_i^{N_1} \mu_{ij}^{\frac{1}{N_{adj}}}. \quad (3.1)$$

ここで N_{adj} は、入力次元に応じた数で $N_1/4$ 程度である。ファジィ推論においてルールの適応度 ρ は、メンバシップ値の最小値を利用するものと、積を利用するものが多い。ただしこれらのファジィ推論の定義に基づいた方法は、入力次元が多くなった場合に、 ρ が大変小さくなり、場合によっては計算機の精度を越え 0 となることも多い。この場合学習が行えないばかりでなく、誤動作の原因となる。AFINN では入力次元に応じたバイアス N_{adj} を導入することによって、こうした問題を解決している。詳細は 3.3.4 節で述べる。

3.3 AFINN のモデリングと学習

AFINN は以下の 4 段階から構成され、適切な入力選択およびルール生成、パラメータの調整が行われる。

1. 初期ルール仮決定段階
2. 入力要素選択段階
3. ネットワーク構造決定段階
4. パラメータ調整段階

以下、それぞれの段階について述べる。

3.3.1 初期ルール仮決定段階

初期ルール決定段階では、Kohonen の自己組織化アルゴリズム [62] を簡易化したアルゴリズムを用いて AFINN の初期ルールを仮決定する。このアルゴリズムは、入力データによってより動的にルール数を決定でき、また Kohonen の自己組織化アルゴリズムよりも動作が高速である。以下にそのアルゴリズムを示す。

準備:

入力空間と出力空間を総合的に扱うために、 l 番目の入力ベクトル U^l とそれに対応する希望のベクトル Y^l

$$U = (u_1^l, u_2^l, \dots, y_k^l, \dots, u_{N_1}^l)^T. \quad (3.2)$$

$$Y = (y_1^l, y_2^l, \dots, y_k^l, \dots, y_{N_3}^l)^T. \quad (3.3)$$

を以下のように $(N_1 + N_3)$ 次元の学習ベクトル I^l に統合する。

$$I^l = [U^l, \mathbf{0}]^T + [\mathbf{0}, Y^l]^T. \quad (3.4)$$

ここで l は、 $1 \leq l \leq L$ であり、 L は学習するベクトルの個数である。つまり学習ベクトル I は L 個の $N_1 + N_3$ 次元のベクトルとなる。AFINN ではこれを $[0,1]$ に正規化してシステムへの入力とする。

また、表記の簡易化のために入力部、出力部それぞれからマップ層上のニューロン j への結合ベクトル w_{ij}, w_{jk} をまとめて重みベクトル W_j ($N_1 + N_3$ 次元) と書くことにする。

$$W_j = [w_{ij}, \mathbf{0}]^T + [\mathbf{0}, w_{jk}]^T. \quad (3.5)$$

現在のルール数 N_r を 0 とし、 $l = 1$ とする。このアルゴリズムの動作の中で、 j 番目のルールが更新される回数を M_j と定義して、すべての j に対して $M_j = 0$ で初期化する。

step1:

まず、最初の入力ベクトル $I^l (= I^1)$ を最初のルール W_1 として登録する

$$W_1 = I^1. \quad (3.6)$$

step2:

l 番目の入力ベクトル I^l に対して、すでに生成されているすべての j 番目のニューロンへの重みベクトル W_j とのユークリッド距離を計算し、最も距離の小さくなる勝ちニューロン s を探す。

$$\|W_s - I^l\| = \min_j \|W_j - I^l\|. \quad (3.7)$$

step3:

step2 で得られた勝ちニューロン s の重みベクトルと、入力 I^l とのユークリッド距離 $\|W_s - I^l\|$ が閾値 ξ_{SELF} 以下であれば、勝ちニューロンの重みを更新する。そうでなければ、その入力ベクトルを新しい重みベクトルとして登録し、 $N_r = N_r + 1$ とする。

$$W_s(t+1) = W_s(t) + \epsilon_{SELF}^j (I^l - W_s(t)) \quad (3.8)$$

; where $\|W_s - I^l\| < \xi_{SELF}$,

$$W_{new} = I^l \quad ; elsewhere. \quad (3.9)$$

ここで ϵ_{SELF}^j は、勝ちニューロンとして重みを更新した回数 M_j に応じて減少する学習係数で以下の式で与えられる。ただし、 $\epsilon_{SELF}^{j_{init}}$ は学習係数の初期値である。

$$\epsilon_{SELF}^j = \epsilon_{SELF}^{j_{init}} \left[\frac{L - M_j}{L} \right]^2. \quad (3.10)$$

まだ、入力ベクトルが残っていれば、step2 にもどる。

これらの簡易自己組織化処理を終えることにより、ネットワークの初期ルールが仮決定される。

3.3.2 入力要素選択段階

適切なファジィモデリングを行うにあたって、適切な入力要素の選択は必要である。不必要な入力情報は、ノイズになるばかりでなく、学習や推論時間の無駄につながり、またファジィのメリットである有用なファジィルールの生成も妨げる。AFINNの入力要素選択段階では、Linkenら [72][73] が行った単次元出力システムに対する手法を拡張した処理を行う。

この処理は2段階に分かれ、まず各入力が出力に与える影響の大きさを求めて、ある閾値以下のものを削除する。続いて、残された入力要素間の相関を調べて、ある一定以上の相関を持つもの同士を、入力の影響度の大きいもので代表し、一方を削除する。以下、それぞれの詳細について述べる。

A. 出力への影響度から入力の選択

初期ルール仮決定段階で得られたネットワークの初期重み (w_{ij}, w_{jk}) および、 L 個の入力ベクトル $U^l = \{u_1^l, u_2^l, \dots, u_i^l, \dots, u_{N_1}^l\}$, ($1 \leq l \leq L$) を用いて、各入力要素が、各出力 $Y^l = \{y_1^l, y_2^l, \dots, y_k^l, \dots, y_{N_3}^l\}$ へ与える影響度 z_{ik}^l を以下の計算式で求める。

$$z_{ik}^l = \frac{\sum_j^{N_2} \mu_{ij} w_{jk}}{\sum_j^{N_2} \mu_{ij}}. \quad (3.11)$$

ここで、 μ_{ij} は、入力 U と重みの前件部 w_{ij} によって求められるメンバシップ値であり、式 (A.2) で計算される。

次に i 番目の入力が k 番目の出力に与える影響度 F_{ik} を計算する。

$$F_{ik} = R_{zik} / R_k. \quad (3.12)$$

ただし

$$R_{zik} = \max_l(z_{ik}^l) - \min_l(z_{ik}^l). \quad (3.13)$$

$$R_k = \max_i R_{zik}. \quad (3.14)$$

最後に、以下の式で入力の影響度を表す F_i を求め、その最大値 $\max_i F_i$ と入力選択の閾値決定率 ξ_{input} の積を下回る入力要素を削除する。

$$F_i = \sum_k^{N_3} F_{ik}, \quad (3.15)$$

$$\max_i F_i \times \xi_{input} > F_i. \quad \Rightarrow \quad \text{該当 } i \text{ 番目の入力の削除}$$

B. 入力間の相関による不要入力の削除

ここでは各入力の相関 $\varphi(i, j), (1 \leq (i, j) \leq N_1)$ を調べ、ある一定 ξ_{corr} 以上であればそれらの入力の内 F_i の大きい方のみを利用する。これにより、必要な情報は残し冗長な入力を取り除くことができる。

$$\varphi(i, j) = \frac{1}{N_1} \frac{\sum_i^{N_1} (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sigma_{x_i} \sigma_{x_j}}, \quad (3.16)$$

ただし、 $\bar{x}_i = 1/N_1 \sum x_i$, $\sigma_{x_i} = \sqrt{x_i^2 - (\bar{x}_i)^2}$ である。

3.3.3 ネットワーク構造決定段階

これまでの処理により、システムにとって有効な入力特徴が求められたので、ここで再度自己組織化アルゴリズムによってシステムの初期ルールを生成する。これにより、そのデータを処理するための入力要素およびルール数が定まり、ネットワーク構造が決定する。ルール数の決定に際しては数々のクラスタリングの手法が検討されているが、適切なルール数が事前に不明であること、全体の構成を簡単にする目的から、3.3.2節で述べた自己組織化アルゴリズムを採用する。

3.3.4 システムパラメータの決定

この段階で、構造が決定されたネットワークに対して、LMS(least mean square) アルゴリズムを用いてパラメータ調整を行う。

AFINN を構成する各重み係数は、 ϵ_{LMS} を、LMS 学習における学習定数として

$$w_{jk}(t+1) = w_{jk}(t) + \epsilon_{LMS}^w (y_k - \hat{y}_k) \frac{\rho_j}{\sum_n^{N_2} \rho_n}, \quad (3.17)$$

$$\begin{aligned} w_{ij}(t+1) &= w_{ij}(t) + \epsilon_{LMS}^w \\ &\times \sum_k^{N_3} (y_k - \hat{y}_k) \left(\frac{w_{jk} \sum_n^{N_2} \rho_n - \sum_n^{N_2} (w_{nk} \rho_n)}{(\sum_n^{N_2} \rho_n^2)} \right) \\ &\times \frac{1}{N_{adj}} \rho_j \frac{2(u_i - w_{ij})}{\sigma_{ij}^2}, \end{aligned} \quad (3.18)$$

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) + \epsilon_{LMS}^\sigma$$

$$\begin{aligned}
& \times \sum_k^{N_3} (y_k - \hat{y}_k) \left(\frac{w_{jk} \sum_n^{N_2} \rho_n - \sum_n^{N_2} (w_{nk} \rho_n)}{(\sum_n^{N_2} \rho_n^2)} \right) \\
& \times \frac{1}{N_{adj}} \rho_j \frac{2(u_i - w_{ij})^2}{\sigma_{ij}^3}.
\end{aligned} \tag{3.19}$$

で更新される。

この時 ϵ_{LMS} は、時間とともに減少するように設定する。多くのファジィニューラルネットワークでは、ルールの適応度 ρ の計算に、各メンバシップ値の最小値 ($\rho_j = \min \mu_{ij}$) もしくは積 ($\rho_j = \prod \mu_{ij}$) をとるようになっている。入力次元数が大きい問題を扱う場合、特に積を取るモデルでは、ルールの適応度 ρ は非常に小さくなりすぎ、計算機の有効桁数を越え0になってしまうことも多い。またLMS学習では、誤差の減少のみに着目して学習が進行するため、メンバシップ関数の幅 σ が負になる事も発生し得る。こうした状況を防ぐためには、学習を途中で中断する方法や、 σ_{ij} の最小値 $\xi_{min\sigma}$ を設ける方法がある。

しかしメンバシップ関数の幅に制限を用いる場合、ある一定以上学習が進行しないという問題点があった。特にルールの適応度に最小値をとるアルゴリズムの場合、LMS学習のアルゴリズムによって重みの更新対象になるのは、 $\rho_j = \min[\mu_{ij}]$ が成り立つ j 番目のルールに関してのみであるために、ある一定以上学習が進行しないことが多い。

AFINNでは ρ_j の決定に式(3.1)を用いているために、学習はすべての重みに対して行われる。このため σ の範囲の制限である $\xi_{min\sigma}$ を設定しても学習を進めることが可能であり、また N_{adj} 導入の効果で ρ_j が0になるのを防ぐことができる。

3.4 計算機実験

AFINN の有効性を検証するために 3 つの実験を行った。本節では、実験結果について述べ考察する。

3.4.1 数値モデルの近似とモデル構築

1 つの冗長入力を含む 5 入力 3 出力の以下の非線形モデルを用いて、AFINN が適切にモデル構築を行えるか、また得られる結果が適切なものかどうかを検証した。式 (3.20) に対象とした数式モデルを挙げる。

$$\begin{aligned}y_1 &= u_1^3 + 3u_2 - 4 \sin u_3 \\y_2 &= \cos^2 u_1 - \tan^3 u_2 + u_4 \\y_3 &= \sqrt{u_1 + u_2 + u_4}\end{aligned}\tag{3.20}$$

冗長入力 u_5 を含む入力には $[0,1]$ のランダムな値を用い、出力も各次元ごとに $[0,1]$ に正規化して AFINN の入出力とした。また各パラメータは表 3.1 の通りである。

AFINN の構造の決定結果を表 3.2 に示す。また、このモデルに対して LMS アルゴリズムを用いてパラメータの調整を行った結果、ならびに 300 個の未学習テストデータで推論した結果を表 3.3 に示す。表中の数値は、 $[0,1]$ に正規化された出力に対する MAE (mean-absolute-error) を百分率で表したものである。各欄において、上段が Linkens の手法 [73] の結果であり、下段が提案手法である AFINN の結果である。学習結果の矢印を挟んだ左右の数値は、それぞれ LMS 学習前後の誤差を示す。また図 3.1 に、誤差の収束と、実行誤差について従来の学習アルゴリズムとの比較を行った。また図 3.2 は、ルール数を固定した場合の学習収束曲線と、実行誤差をまとめたグラフである。

表 3.2 より、ルール統合比率 ξ_{SELF} を大きくするにつれ、生成されるルール数が減少することが確認できる。またいずれの場合においても 5 入力のうち、必要な正しい 4 入力を選択されていることが分かる。

表 3.1 非線形関数近似問題のパラメータ

N_1	5
N_3	3
L (# of vectors)	1000
ξ_{SELF}	5-20%
$\epsilon_{SELFinit}$	0.5
ξ_{input}	0.05
$\xi_{corr.}$	0.8
$\epsilon_{LMSinit}$	0.001
σ_{init}	0.1
$\xi_{min\sigma}$	0.01
# of LMS learn.	5000

表 3.2 モデル構築の結果 (関数近似問題)

ξ_{SELF} (%)	5	7	10	12	15	20
# of rules (initial)	815	509	208	127	61	23
# of rules (N_2)	527	254	80	44	21	8
F_1	0.30	0.30	0.29	0.25	0.33	0.32
F_2	0.57	0.57	0.58	0.59	0.60	0.59
F_3	0.22	0.22	0.25	0.25	0.25	0.25
F_4	0.28	0.27	0.25	0.24	0.22	0.25
F_5	0.03	0.03	0.03	0.03	0.04	0.11
selected	u_1, u_2					
input	u_3, u_4					

表 3.3 誤差収束と実行誤差 (%)

上段:Linkens[73]、下段:AFINN

学習結果欄内の左右の数値はLMS 学習前後の誤差を示す

ξ_{SELF} (%)	5	7	10	12	15	20
# of rules	527	254	80	44	21	8
学習結果 (誤差 %)	1.61→0.11 1.17→0.18	2.35→0.33 2.36→0.33	4.60→0.65 4.84→0.31	6.27→0.77 6.55→0.30	8.31→1.09 8.52→0.33	12.29→1.82 11.95→0.72
実行結果 (誤差 %)	1.63 1.96	1.75 1.77	1.16 0.50	1.13 0.39	1.20 0.35	1.93 0.70

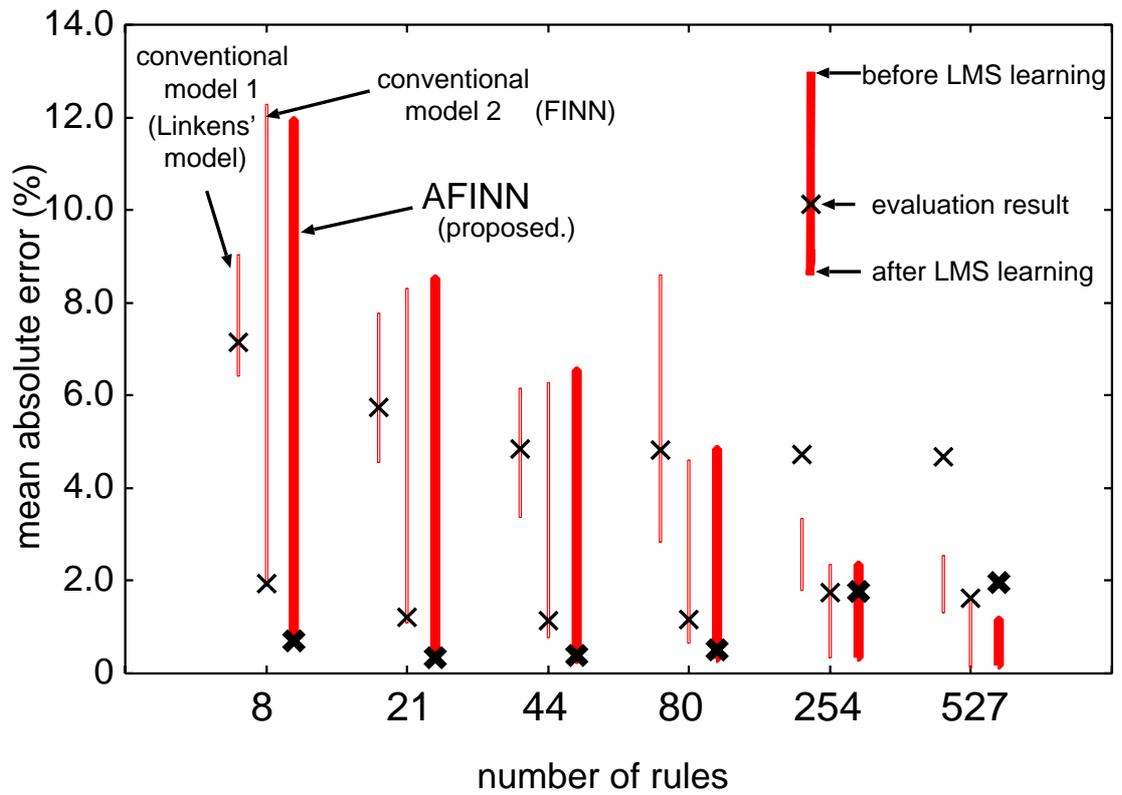


図 3.1 誤差収束の比較 (関数近似問題)

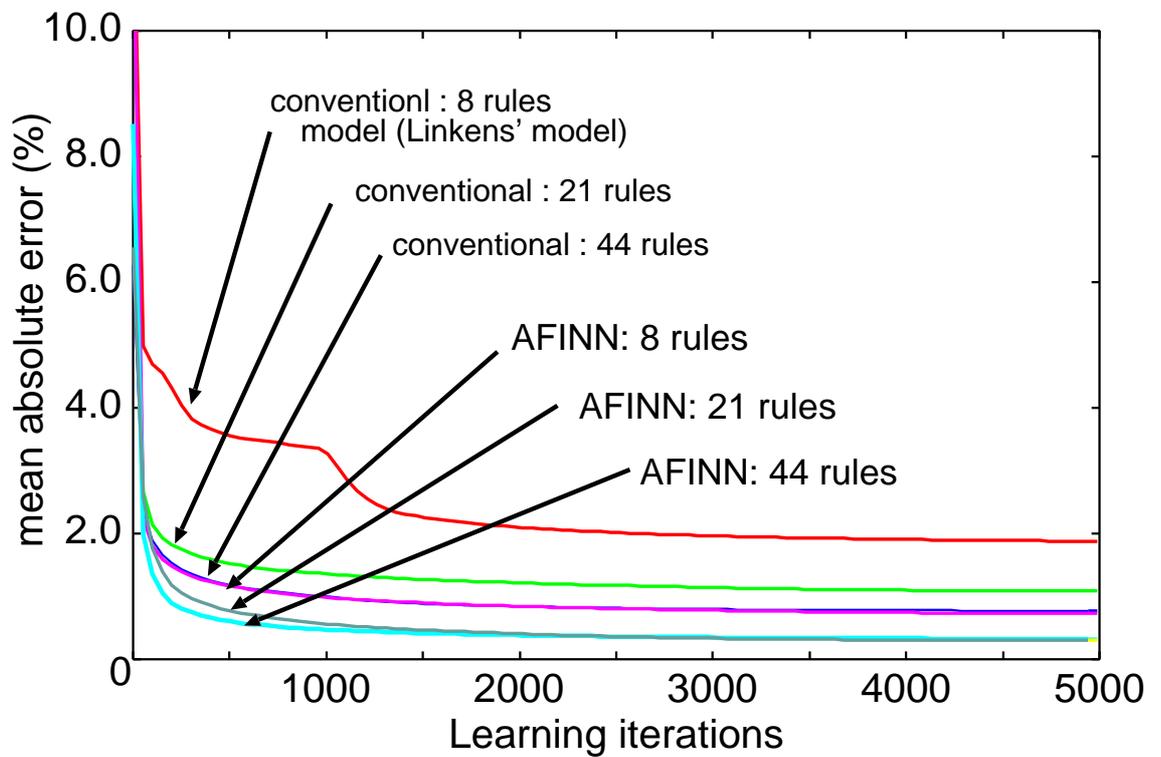


図 3.2 誤差収束の比較 2 (関数近似問題)

表 3.3 および図 3.1 より、AFINN の学習アルゴリズムの方が良好に誤差を収束させ、未学習のテストデータに対しても優れた結果を出力していることが確認できる。また、同じルール数で学習の収束を比較した図 3.2 より、AFINN は 8 ルールで、Linkens[73] の 44 ルールのシステムに匹敵する収束を見せている。

これらの結果から、この数値近似問題に対しては ξ_{SELF} は 10-20% 程度が妥当であると考えられる。必要以上のルールを持つシステムは、無意味な計算量増加をもたらし、さらに多くのルールを持たせた場合、評価データに対する誤差が逆に増加に転じる。これは BP 等でも良く知られている、ルール数の増加に伴う過学習状態に陥ったためと考えられる。

以上の非線形多入力多出力の数値データの学習実験により、AFINN は入力データから適切に入力要素の選択、ルール数の調整、パラメータの調整が行えることが確認できた。

3.4.2 音声認識問題への適応

次に AFINN が、多入力多出力の系において学習を適切に行えるかを検証する。システム評価に客観性をもたせるために、UC Irvine が公開しているデータベース [91] から音声分類問題を取り上げ、評価実験を行った。

扱ったデータは ISOLET(isolated letter speech recognition) データで、617 の特徴をもつデータを 26 のアルファベットに分類する 617 入力 26 出力の問題である。この ISOLET データセットは、F 以外の A-Z のアルファベットが各 240、F が 238 の計 6238 の学習データと、それとは別の A-Z 各 60 ずつ計 1560 の評価データがから構成される。本実験では、それらをそのままの割合で利用した。

表 3.4 に、この実験で用いた AFINN のパラメータを示す。この学習セットは、同じアルファベットに属するデータが 240 個あることから、学習回数は 100 回と少ないが、十分に誤差を収束させることができた。 $\xi_{SELF} = 12\%$ 時に、617 の入力要素の内 296 が選択され、ルール数は 450 となった。LMS 学習前、後、未学習のデータに対する各 MAE は、それぞれ 3.48%、0.49%、1.00% で、26 のアルファベットの分類精度は 93.53% であった。

450 ルールでの Linkens のモデルと AFINN の誤差収束曲線を図 3.2 にまとめた。ただし、通常ファジィ推論を行うこの Linkens のモデルはこの様な高次元の入力に対応できないため、ルール適応度 ρ を求める式を (A.3) に変更して評価している。

この結果は、これまでに報告されている結果の中で最も良好な、BP による 95.9% [92] にやや及ばないが、ファジィニューラルネットワークの有するルール生成能力やネットワークの解析利便性を考慮すれば優れた結果であると言える。

表 3.4 音声認識問題のパラメータ

N_1	617
N_3	26
L (# of vectors)	6338
ξ_{SELF}	12%
$\epsilon_{SELFinit}$	0.5
ξ_{input}	0.05
$\xi_{corr.}$	0.8
$\epsilon_{LMSinit}$	0.001
σ_{init}	0.1
$\xi_{min\sigma}$	0.01
# of LMS learn.	5000

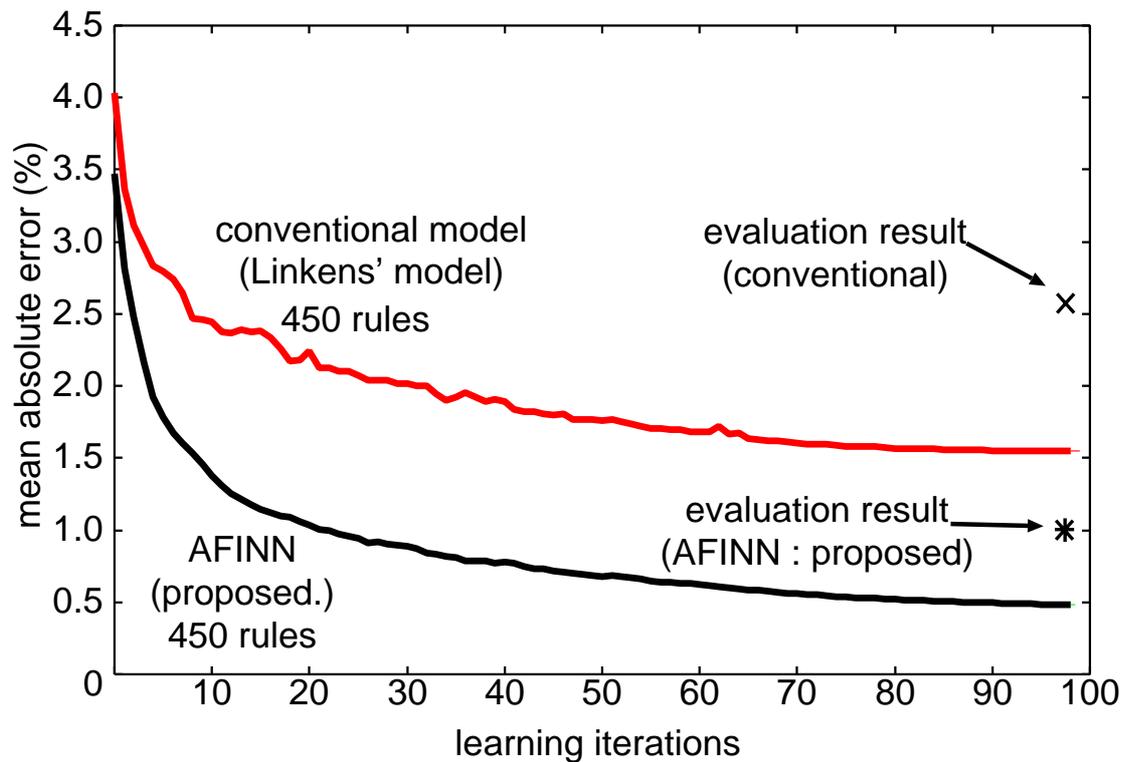


図 3.3 学習収束曲線の比較 (音声認識問題)

3.4.3 車の評価問題

最後に、AFINN が生成するルールが適切なものかどうかを確認するために、先程と同様に UC Irvine[91] の Car evaluation database を用い評価実験を行った。

このデータは、以下で述べる 6 つの基準と、4 段階の車の評価から構成され、6 入力 4 出力のクラスタリング問題として利用できる。このデータセットにおける車の特徴づける 6 つの指標は、“buying price”, “maintenance cost”, “# of doors”, “# of passengers”, “luggage size”, “safety” であり、それぞれの評価が 3~5 段階で記述されている。一方、車の評価は “very good”, “good”, “acceptable”, “un-acceptable” の 4 種類であり、それぞれ 1210,384,69,65 個からなる合計 1728 データである。本実験では、これらの 8 割を学習に用い、残りをテストに利用した。

この実験の条件を表 3.5 にまとめた。実験の結果、 $\xi_{SELF} = 20\%$ で、6 つすべての入力を選択され、計 10 個のルールが作成された。未学習データに対するテストでは、MAE が 3.19% で、分類の正答率は 96.2% と良好であった。このデータベースを用いた最も新しい報告 [93] では、ほぼ 100% の正当率を実現しているが、この手法は事前にデータを手動で複数に分類する前処理が必要とする。またこの手法は単に分類結果を提示するだけであり、そこからルールや他の情報を引き出すことはできない。

AFINN が学習によって獲得した 10 個すべてのルールを表 3.6 に示した。このデータベースで very good あるいは good と評価されている車は低コスト、低メンテナンスコスト、多くの人に乗れて、荷物がつめ、ドアが多く、安全という傾向がある。直接的に獲得されたルールが適切である点、また獲得されたルールを用いて得られた分類結果の正答率の高さから適切なルールが抽出されていると言える。

表 3.5 車評価問題のパラメータ

N_1	6
N_3	4
L (# of vectors)	1382
ξ_{SELF}	20%
$\epsilon_{SELFinit}$	0.5
ξ_{input}	0.05
$\xi_{corr.}$	0.8
$\epsilon_{LMSinit}$	0.001
σ_{init}	0.1
$\xi_{min\sigma}$	0.01
# of LMS learn.	5000

表 3.6 抽出されたルール (車評価問題)

Rule #	Price	Maintenance cost	Number of doors	Number of passengers	Luggage size	Safety	estimation
1	very high	med	4	4	med	med	unacceptable
2	very high	med	4	$5 \leq$	small	low	unacceptable
3	high	med	$5 \leq$	4	med	high	acceptable
4	low	low	3	2	med	high	unacceptable
5	low	high	$5 \leq$	4	large	high	acceptable
6	low	low	4	4	small	med	acceptable
7	low	low	4	4	large	low	unacceptable
8	low	med	4	$5 \leq$	large	high	very good
9	med	low	4	$5 \leq$	med	med	good
10	low	med	2	$5 \leq$	small	high	unacceptable

3.5 結論

本章では、学習によって処理に必要な適切な構造やパラメータを決定し、データから自動的にルールを生成できる適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference neural network) について述べた。

計算機実験の結果、非線形の多入力多出力の数式近似ならびに、音声認識、車の評価といった例題を用いて、AFINN のモデル構築が適切であること、推論が高い精度で行えること、また得られるルールの妥当性を確認した。

データの解析や推論を行う場合、与えられたデータから重要な情報のみを選択する事は、精度と汎用性確保の面において大変重要であり、ニューラルネットワークを用いた場合でも同様である。AFINN は、入力要素の選択ならびに、ファジィニューラルネットワークでは重要な要素であるルール数の決定を自動的に行うことにより、汎用性の高いルールの生成および推論を実現している。

第 4 章

適応ファジィ推論ニューラルネットワークとアクティブ探索法を用いた画像認識

本章では、従来のパターンマッチング手法と、学習システムの両方を利用した、画像認識手法 [80] について述べる。画像認識システムを構築する場合、事前に認識対象に関する全ての情報をルールや知識の形で登録しておくことは大変困難であり、これらの自動化を担う学習システムは必須であると考えられる。一方、認識精度に重点をおいた場合、特定の状況に特化した処理方法が記述されたルールベースシステムの利用や、パターンマッチングの手法が有利である。

本章では、これらの長所を融合して幅広い用途で、高い精度の画像認識実現を目指した統合型画像認識システムについて述べる。まず、パターンマッチング手法を利用して、画像中から所望の物体がある候補領域を高速に探索し、次に学習システムを利用しそれらの検証を行う。パターンマッチング手法には、目的とする物体を複数の 2 次元の参照画像を用いて高速に抽出する「アクティブ探索法」を利用し、学習システムには第 3 章で述べた知識処理が可能な適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference neural network) を用いる。AFINN は、必要な入力情報の選択等の推論モデル構築を自動で行い、学習によって推論ルールを自動的に獲得する。このため AFINN を画像認識問題に適用した場合、認識のための明示的な知識を必要とせずに認識が可能になる。また本システムは、登録する参照画像を用いて AFINN が学習を行う構成のため、それらを変えることにより認識対象を変更、追加できるという柔軟性を持つ。本章では、コンピュータやモニタ、椅子などが存在する研究室内の画像を例題に扱い、高い認識精度の実現を確認した。

4.1 はじめに

第2章で、学習機能と知識処理の両方の能力を持ったファジィニューラルネットワークを用い、風景画像を例題とした画像認識ならびに認識に関するルール抽出に関する研究について述べた。これらの研究は、コンピュータの学習機能のみを利用して高い認識精度と汎用性の両立を目指したものである。

一般にファジィニューラルネットワーク [64][73][74][79] は、学習によって任意の入出力関係を近似可能なだけでなく、ニューロンの結合加重上に獲得したルールが生成されるために、それらの抽出や解釈、加工が容易に行える。したがって、BP [58] 等の代表的な階層型ニューラルネットワークでは困難であった追加学習や動作の一部修正、ルールベースシステムとの融合が容易である。また、ファジィニューラルネットワークを利用した画像認識システムは、認識に関するルールや知識を学習によって自動的に獲得するため、システム設計時に明示的に与える必要がないという大きな特長を持つ。このように、ファジィニューラルネットワークの様々な特長は画像認識問題においても大きな効果を発揮するが、高精度での認識と汎用性の両立のためには、必要に応じてトップダウンの知識の利用やパターンマッチングなどの手法との融合も必要である。

これまで画像認識、探索の研究で主に行われてきたパターンマッチング手法には、3次元参照モデルを利用するものと、2次元参照モデルを利用するものがあったが、それぞれ計算量や精度の面で問題点があった。こうした問題に対して村瀬らは、複数の2次元参照画像を用いて非常に高速に所望の物体を抽出するアルゴリズムを提案している [29][30]。特に「アクティブ探索法」 [30] は、大きさも位置も分からない物体を、しらみ潰しの探索と同じ精度を保証しながら 10 ~ 1000 倍高速に検出できる優れたアルゴリズムである。この手法は、非常にシンプルで効果的なために、さらなる高速化の検討 [31] や音声やビデオといった時系列処理にも応用がなされている [32]。

そこで本章では、より用途の広い高精度な画像認識法を目指し、パターンマッチング手法と学習システムを統合したシステムについて述べる。本論文では、このシステムを複合型画像認識システムと呼ぶ。

本システムはまず、登録された参照画像を用いてアクティブ探索と形状比較を行い、着目している物体の候補領域を限定する。次にその候補の中から学習システムを用いて適切な部分の選択を行う。

学習システムには、第3章で述べた適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference neural network) [79] を用いる。AFINN は、学習によってルールを自動的に生成して推論に利用できるだけでなく、与えられた学習データから適切な入力要素の選択、ルール数の決定といった適切なモデリングを自動的に行うことができる。画像認識問題は、一般的なパターン分類問題に帰着でき、分類精度の向上

には有効な特徴量の選択は重要である。無駄な入力要素は系の自由度を不必要に増加させ、汎用性を低下させる。AFINN は、出力に与える影響の大きい入力要素のみを自動的に選択して利用するために、推論の汎用性を向上させ一般性の高いルールを生成することができる。本章で述べる複合型画像認識システムで用いる AFINN の学習には、パターンマッチングで使用する参照画像を用いるため、利用者は認識対象に関するいくつかの参照画像を登録するだけで別途学習データを必要としない。またその画像の種類を変えることで他の用途でも利用可能となる。

4.2 複合型画像認識システム

本章で述べる複合型画像認識システムはアクティブ探索法 [30] と第 3 章で述べた AFINN を元に構成される。

アクティブ探索法は村瀬らによって提案され、画像中にある大きさも位置も分からない 3 次元物体を、複数の 2 次元参照画像を用いて非常に高速に検索するアルゴリズムである。探索には形状変化の影響を受けにくい色ヒストグラム特徴を用い、また高速化のために、ある参照画像の照合結果を利用して他の参照画像の照合回数を適応的に削減する。

3 次元物体の照合に複数の 2 次元画像を用いる場合、拡大、縮小なども考慮すると莫大な照合回数になり、処理時間が課題となっていた。アクティブ探索法は上記の工夫により理論的に精度を保証しながら、10 ~ 1000 倍の高速化が図られている。

4.2.1 システム概要

本章では画像の認識を、「画像の中から利用者が指定した物体の位置と大きさを示すこと」と定義する。対象とする物体の種類は、計算時間が許容される範囲であれば、種類を多くすることが可能である。図 4.2.1 に 4 種類の物体を扱う例を用い提案システムの概要を示す。本システムは、以下の 2 つの過程から構成される。

1. 探索過程 (Search phase)
2. 検証過程 (Validation phase)

探索過程では、アクティブ探索法と新たに定義する形状比較処理によって、注目する物体の候補を画像中からおおまかに探索する。検証過程では、AFINN を用いて探索過程で候補になった部分の検証を行い、適切な部分の検出を行う。

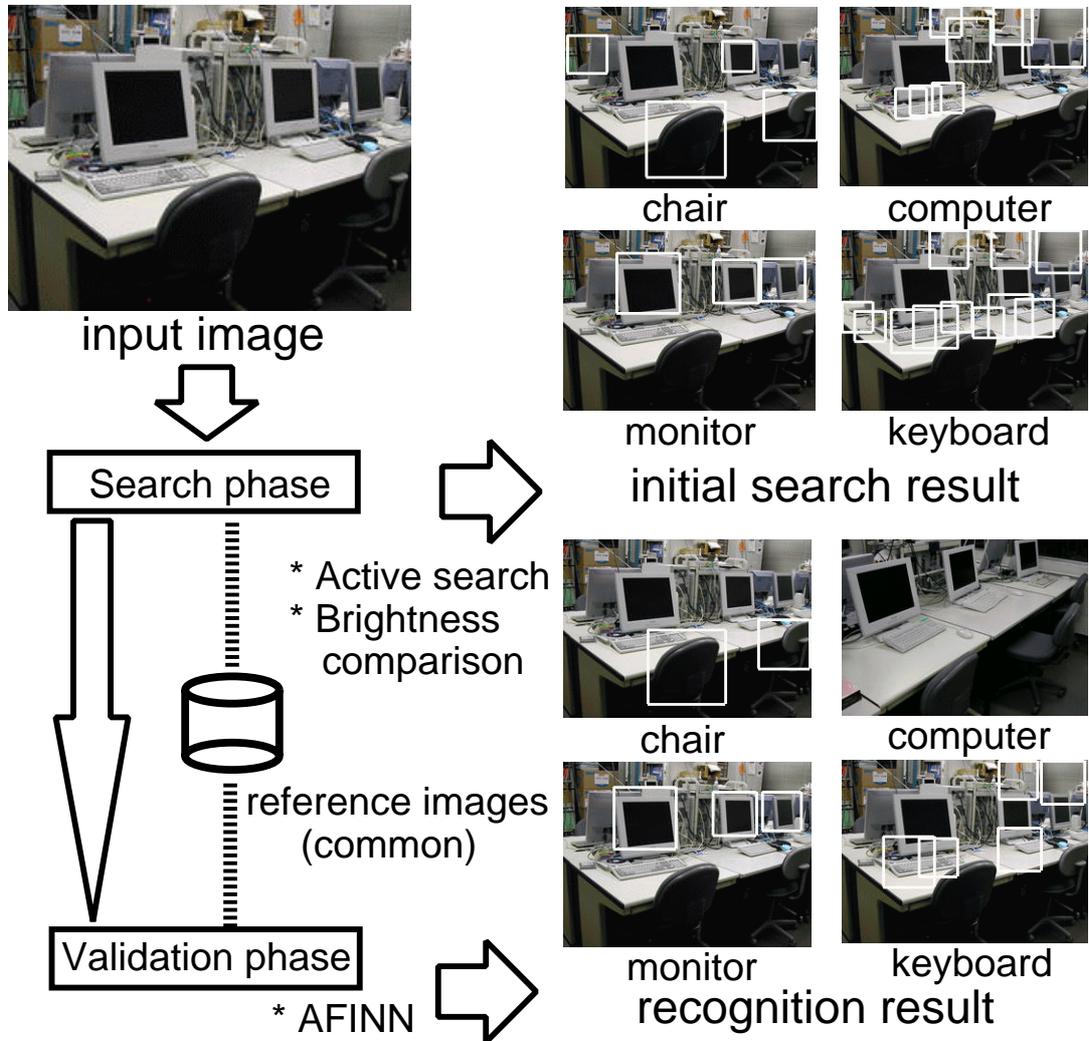


図 4.1 複合型画像認識システムの概要

4.2.2 探索過程

探索過程では、事前に登録された参照画像を元にアクティブ探索 [30] と、形状比較を行って、所望の物体を高速に抽出する。アクティブ探索法は、画像から対象とする物体を含むウィンドウ状の部分を取り出すため、画像認識の前処理として従来広く行われている領域分割等と異なり、光の状態や位置、遮蔽の影響等を受けにくい。しかしこの探索法は、色ヒストグラム特徴のみを利用したアルゴリズムであるため、参照画像と同じ特定の画像を探すには非常に効果的である反面、一般画像認識問題に単体で用いるのは十分ではない。例えば、似た形状の物体でも色が異なる場合排除されてしまう。そこで本システムでは、物体の形状を反映するために、画像の輝度分布に基づく類似の比較法を考案し、評価に加えた。

今、対象となっている領域の画素数を N とする。RGB 空間をそれぞれ均等に Q 分割し、合計 $I = Q^3$ 色の代表色を軸とした色ヒストグラムを作成する。アクティブ探索法の適応度 F_{AS} は、入力画像中から抽出された領域のヒストグラムを H 、参照画像のヒストグラムを M とすると、

$$F_{AS} = \frac{1}{N} \sum_{i=1}^I \min(H_i, M_i). \quad (4.1)$$

で求められる。ここで H_i, M_i の値は、色 $i \in I$ に属する画素数で、 $\min(A, B)$ は、 A と B の小さい方、あるいは同等ならその値を表す。

一方、形を考慮するための輝度分布の一致度 F_V を、その領域内の画素 j の輝度を V_{Hj} 、参照画像のピクセルの輝度を V_{Mj} として

$$F_V = \frac{1}{N} \sum_{j=1}^N \frac{\min(V_{Hj}, V_{Mj})}{\max(V_{Hj}, V_{Mj})}. \quad (4.2)$$

で求める。ただし、各画素の輝度 V は、 $V = 0.30R + 0.59G + 0.11B$ である。これらの積 $F = F_{AS} \times F_V$ がある閾値 ξ_{search} 以上の領域を、物体が存在する領域の候補とする。

しかしながら、これだけの情報で優れた認識を行うことは困難である。そこで探索過程では、多くの候補を高速に検出し、続く検証過程で精査する。

4.2.3 検証過程

検証過程では、探索過程によって得られた各物体の候補を、AFINN[79]を用いて精査する。AFINNには、注目している候補領域内の色やテクスチャといった特徴量が入力され、出力の各ノードは、抽出したい個々の物体に対応する。出力ノードのそれぞれの出力値は、対応する物体が注目している領域にあるかの度合を表す。AFINNの学習には、アクティブ探索法で用いた参照画像を用いる。これによりシステム利用者は、学習のための新たなデータを準備する必要がなく、参照画像を登録するだけで認識結果を得ることができる。また、より高度な認識を行うために、学習データを追加することも可能である。AFINNは学習によって自動的に生成されたルールを用いて、各対象物体の候補が適切かどうかを判断する。注目している各候補領域ごとに推論を行い、各物体に対応するAFINNの出力ノードの値が閾値 ξ_{AFINN} を超えた候補領域には、対象物体があると判断し検出する。

4.3 計算機実験

4.3.1 概要と準備

本章で述べた複合型画像認識システムの有効性を検証するために、コンピュータなどが多く設置された情報系研究室内で撮影された 256×192 画素の画像を例題に、認識実験を行った。実験には、フラッシュ等の映りこみがなく、カメラから対象とする物体までの距離が最低 1メートル以上確保された写真を利用した。

認識対象とするのは、「椅子」「コンピュータ」「モニター」「キーボード」の4種類であり、合計 30 枚の画像からオブジェクト毎に 30~40 個の参照画像を抽出して用いた。利用した参照画像の例を図 4.2 に示す。また、実験パラメータを表 4.1 に、AFINN への入力要素を表 4.2 に示す。

同じ種類の物体でも、個々の機種や色、撮影状況の違いにより様々であり、一般にこうした多様さが画像認識問題を困難にしている。探索過程でのマッチングは、参照画像を縦横共に約 3 倍に拡大した 106×106 画素から、基本サイズである 32×32 画素まで縦横比を保存しながら 6 段階に縮小しながら行った。

結果の評価には、以下の様に定義した 2 つの指標を用いた。

- 画面中の物体が抽出された割合 (SE)
- 結果として得られた領域が適切である割合 (SP)

SE(Sensitivity) は、画像中の物体が見落とされずに正しく抽出されるかの指標で、例えば画像中に 5 個のモニターがあり、その内 4 つが抽出されている場合 80%となる。SP(Specificity) は、抽出された領域の正当性を評価する指標で、例えば選択された 10 個のコンピュータ候補領域の中で、適切なものが 7 つであれば 70%となる。ここで適切な抽出とは、対象とする物体をピクセル数換算で 50%以上含んだ結果を指すものとする。SE と SP は相反する関係にあり、双方共に高い値を実現することが望ましい。

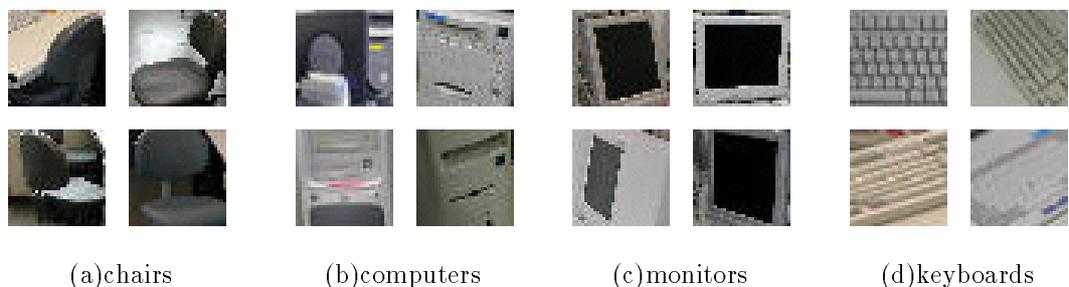


図 4.2 参照画像例

表 4.1 実験パラメータ

入力画像サイズ	256 × 192
対象画像の基本サイズ	32 × 32
ウィンドウサイズ	32×32 ~ 106×106 (6 段階)
利用した参照画像数	147 (30 枚の画像から抽出)
色ヒストグラム階調 I	125 (5^3)
ξ_{search}	0.6
N_{adj}	$N_1/4$
ξ_{AFINN}	0.3
$\epsilon_{SELFinit}$	0.5
ϵ_{LMS}	0.001
ξ_{SELF}	10
ξ_{input}	0.1
ξ_{corr}	0.95

表 4.2 AFINN の入力特徴量

色情報 (計 12 次元)	maxR, minR, E[R], σ [R] maxG, minG, E[G], σ [G] maxB, minB, E[B], σ [B] R,G,B それぞれ最大、最小、平均、標準偏差
テクスチャ性 (計 16 次元)	contrast 0 °, 45 °, 90 °, 135 ° uniformity 0 °, 45 °, 90 °, 135 ° correlation 0 °, 45 °, 90 °, 135 ° entropy 0 °, 45 °, 90 °, 135 °
エッジ強度 (1 次元)	($\sum \delta V$) エッジ強度の和
輝度ヒストグラム (計 5 次元)	V_1, V_2, V_3, V_4, V_5 輝度分布を 5 段階のヒストグラム化
合計	34 入力中 20 入力 (太字が選択された入力)

4.3.2 実験結果

50枚の未学習画像を用いた認識結果を表4.3, 4.4に、検証過程におけるAFINNのモデリング結果を表4.5に示す。AFINNの入力要素の選択結果については、表4.2において、選択された入力を太字で表記した。

また認識結果例を図4.3～図4.5に示す。それぞれ図中の上段が、探索過程の結果で、下段が検証過程結果である。なお1枚の画像の処理に要する処理時間はPentium4(2.4GHz)の計算機を用い、探索段階でおよそ5秒、検証段階でおよそ1秒であった。

表 4.3 探索過程結果

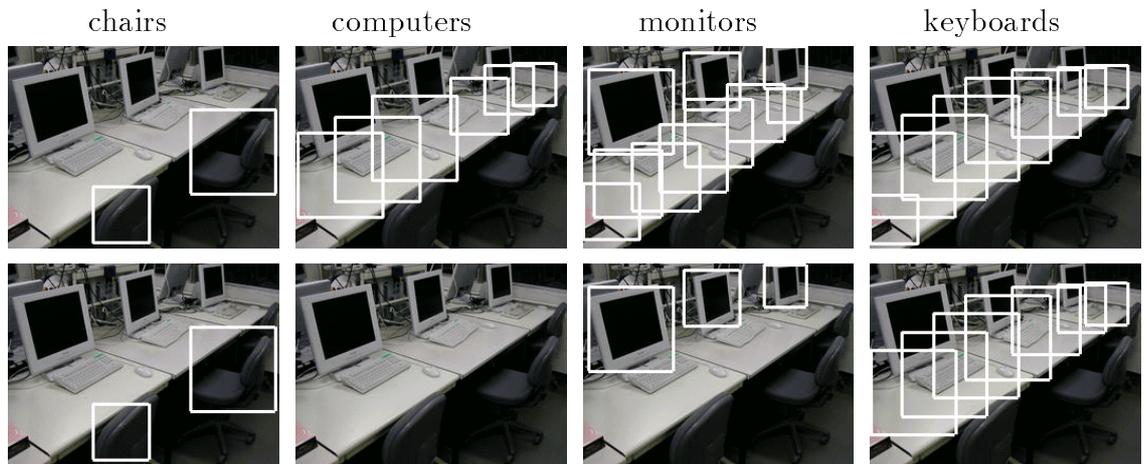
	SE	SP
椅子	80.9% (51/ 63)	73.6% (56/ 76)
コンピュータ	92.0% (23/ 25)	37.0% (162/437)
モニタ	86.2% (69/ 80)	25.4% (74/291)
キーボード	98.5% (68/ 69)	25.2% (128/507)

表 4.4 検証過程結果

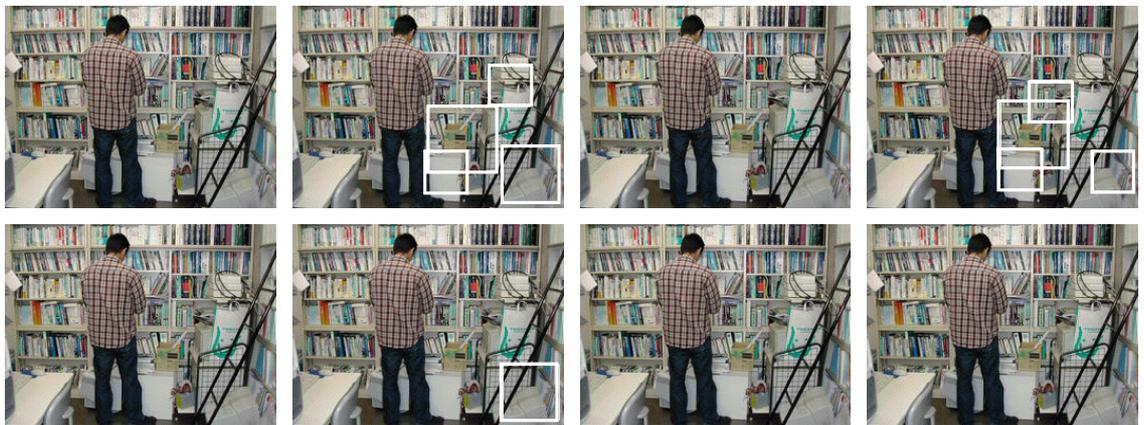
	SE	SP
椅子	80.9% (51/ 63)	74.6% (56/ 75)
コンピュータ	84.0% (21/ 25)	52.3% (33/ 63)
モニタ	71.2% (57/ 80)	89.4% (51/ 57)
キーボード	98.5% (68/ 69)	38.3% (118/308)

表 4.5 AFINN のモデリング結果

参照画像数	147
入力特徴量数	34
選択された入力要素数 N_1	20
生成されたルール数 N_2	16
出力次元 N_3	4
LMS 学習回数	10000
最終学習総誤差 (MSE:%)	1.22



(a) 結果例 1-a

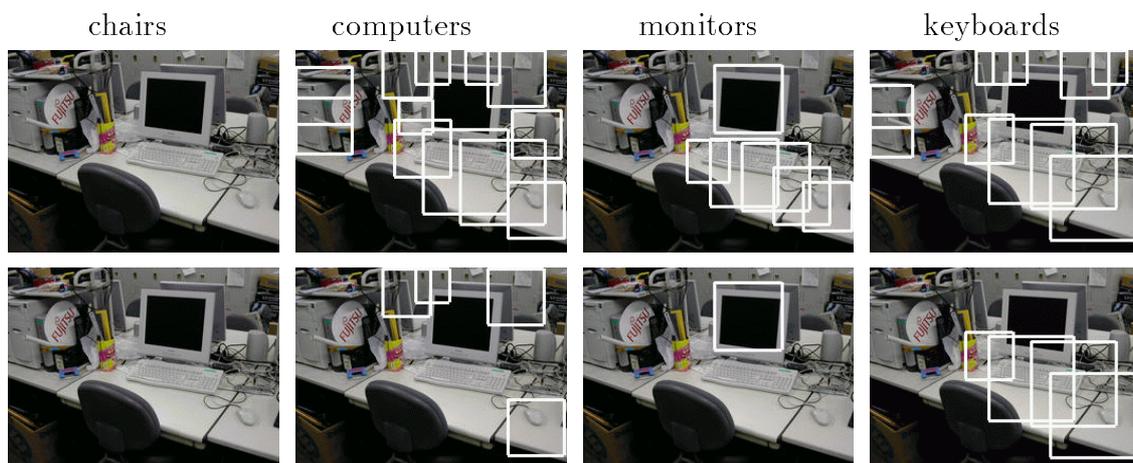


(b) 結果例 1-b



(c) 結果例 1-c

図 4.3 出力結果例 1 上段：探索過程 下段：検証過程
それぞれ左から「椅子」「コンピュータ」「モニタ」「キーボード」



(a) 結果例 2-a

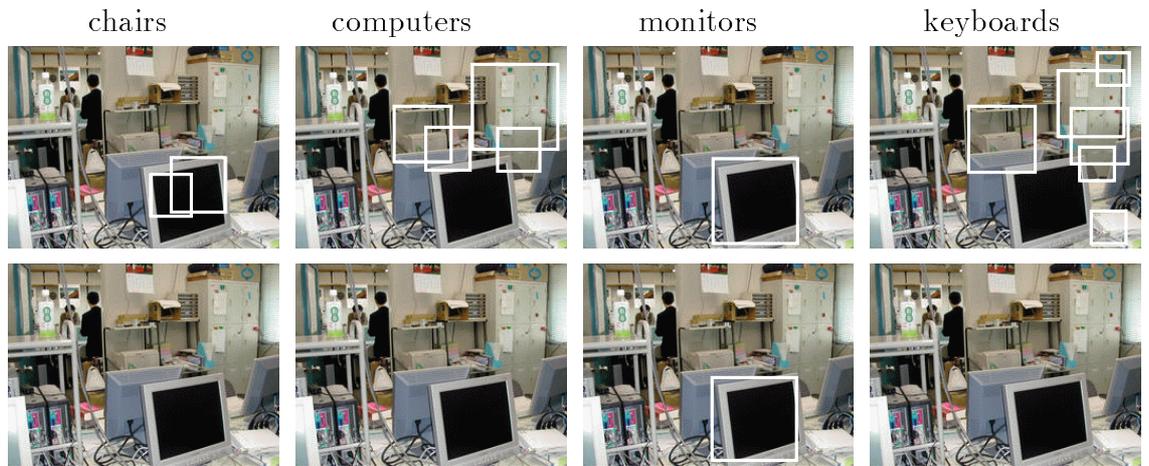


(b) 結果例 2-b

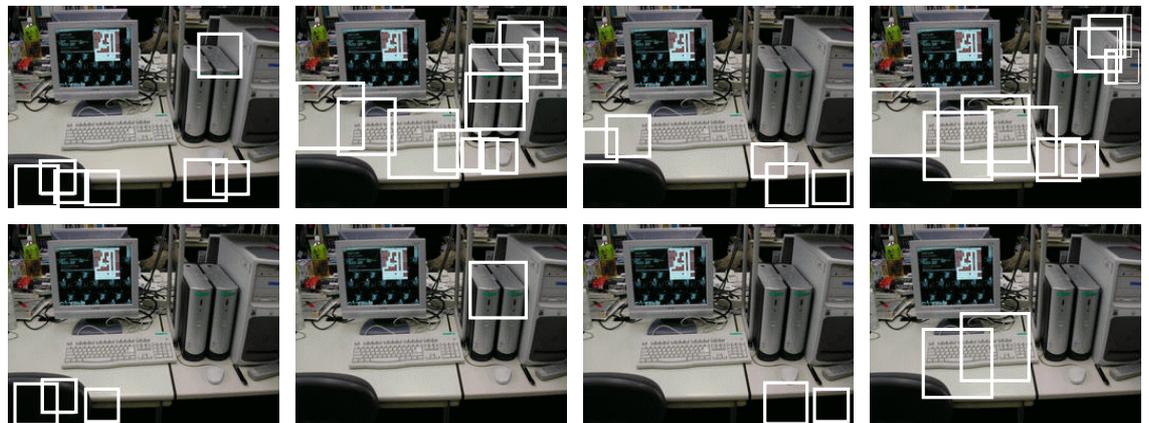


(c) 結果例 2-c

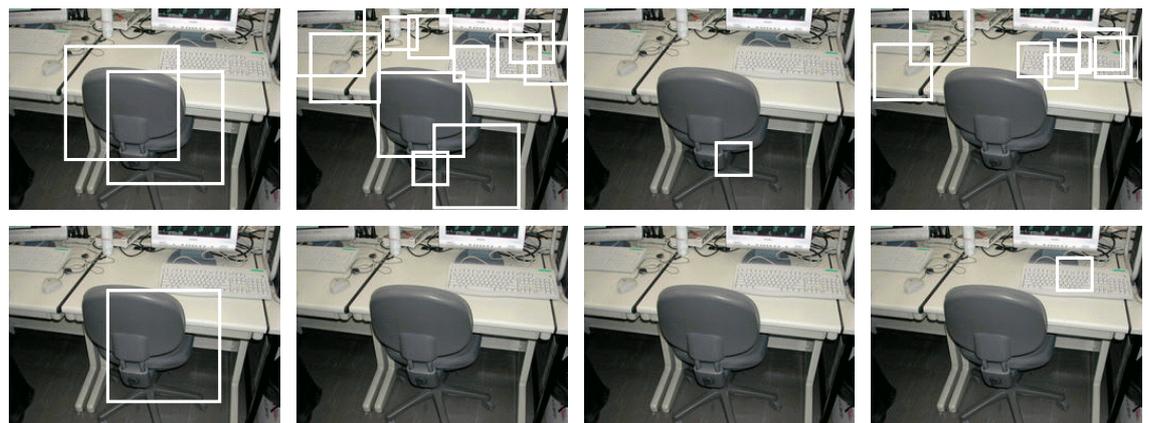
図 4.4 出力結果例 2 上段：探索過程 下段：検証過程
 それぞれ左から「椅子」「コンピュータ」「モニタ」「キーボード」



(a) 結果例 3-a



(b) 結果例 3-b



(c) 結果例 3-c

図 4.5 出力結果例 3 上段：探索過程 下段：検証過程
 それぞれ左から「椅子」「コンピュータ」「モニタ」「キーボード」

4.4 考察と課題

表 4.3 から、探索過程において必要な物体はほぼ抽出され高い SE を実現している一方で、誤抽出領域が多く SP が低くなっていることが分かる。しかし表 4.4 から確認できるように、続く検証過程において、SE の低下を少なく抑えながら大幅に SP が改善できている。これは、探索過程で必要な候補が外れてしまわないように多くの候補を抽出し、検証過程でその中から適切な候補のみが検出できているためである。

4.4.1 物体ごとの認識結果について

図 4.3 ~ 4.5 から、探索過程では不適切な部分を含む多くの領域が抽出され、続く検証過程でおおむね適切な部分のみが検出されていることが確認できる。しかし図 4.4(a) の結果では、探索段階で椅子が正しく抽出されておらず、また背景のロッカーをコンピュータとして認識してしまっている。

各物体毎の認識結果を見ると、椅子とモニタに関しては SE、SP 共に約 7 割から 9 割と高い値を実現できていることが分かる。一方で、コンピュータとキーボードは同レベルの SE を実現しようとするとも SP が低くなってしまっている。これは、ある程度もれがないように抽出しようとする、多数の他の領域も一緒に検出されてしまうことを意味している。コンピュータやキーボードは他の 2 つの物体に比べ、形状が複雑で、図 4.2 の参照画像の例にあるように機種により外見が大きく異なるためと考えられる。また本手法は正方形のウィンドウのみを利用しているため、ウィンドウ内の注目している物体以外の影響も無視できない。例えばキーボードの様に全体が収まらない場合、ウィンドウ内の背景の影響などで他の類似する部分が誤抽出され SP が低下したと考えられる。本論文では、システム設計の利便性のために正方形のウィンドウに限定しているが、他の形状の利用についても今後、検討が必要である。

4.4.2 認識誤りについて

誤抽出結果を詳しく見ると、白色領域の多いコンピュータ、キーボードに関する誤りが目立ち、またブラインドやロッカーなどの背景の誤抽出も多かった。誤りをより解析的に評価するために、各物体間の分類の困難さを線形判別分析を利用して比較した。今回用いた4種類の物体中の任意の2種類で判別分析を行い、判別得点から求めたグループ間分散 σ_B^2 と全分散 σ_T^2 の比(σ_B^2/σ_T^2)を表4.6に示した。この比率が高い程、判別分析において両者の分類が容易ということを示している。

この結果、この実験において椅子は比較的他の物体と区別しやすく、コンピュータとキーボードの区別は他の区別に比べて困難であることがわかった。

また、検出過程での抽出パラメータ ξ_{AFINN} をオブジェクト毎に調整することで、表4.4の結果からSEをほぼ維持しながらSPをさらに3~7%向上できた。こうしたパラメータの調整は、システム利用の利便性とのトレードオフとなる。一方で、参照画像を追加した場合にSE,SPの向上が期待されたが、現状では追加した参照画像に類似する誤抽出領域も増加し、参照画像を単純に追加するのみでは明確な精度向上は確認できなかった。確実な性能向上を図るためには、先に述べた正方形に限定されたウィンドウ形状の改善の他に、より正確に物体を抽出できる特徴量についてさらに検討する必要があると考えられる。

表 4.6 オブジェクト特徴量のグループ間分散、全分散比 (σ_B^2/σ_T^2)

	chairs	computers	monitors	keyboards
chairs	—	0.971	0.939	0.970
computers	0.971	—	0.910	0.857
monitors	0.939	0.910	—	0.976
keyboards	0.970	0.857	0.976	—

4.4.3 検証過程とシステム構成について

次に検証過程で用いた AFINN のモデリング、推論について考察する。AFINN が選択した入力要素を表 4.2 の中で太字で表示した。検証段階において、多くの候補領域から適切な領域が抽出されていることから、AFINN が適切に画像認識問題に対してモデリングを行えていることが確認できる。

また、本システムがアクティブ探索法と AFINN を用い、探索過程での検出パラメータ $\xi_{search}=0.6$ を選択した点について、その妥当性を検証するために以下の 3 つの比較実験を行い、これらの最終的な結果を表 4.7 に示した。

1. 探索過程のみ実行 ($\xi_{search} = 0.7$)
2. 提案法と同手法の実行 ($\xi_{search} = 0.5$)
3. 検証過程のみ実行 ($\xi_{search} = 0$)

(1) はアクティブ探索法のみで、認識を試みる実験である。ここでは必要な領域のみを選択できる様に、 $\xi_{search} = 0.7$ に設定した。その結果、椅子、コンピュータ、モニター、キーボードそれぞれについて $\xi_{search} = 0.6$ 時とほぼ同等の SP を実現する一方で、選択される候補領域が大幅に減少し、SE が大幅に低下した。このことから、探索過程のみでは十分な結果が得られず、本手法のより多くの特徴情報を利用した検証過程が効果的に作用していることが確認できた。

(2) は探索過程のパラメータを下げた場合の実験である。パラメータを下げることによって、探索過程において候補領域が非常に多く選択される。この影響で検証過程後に、若干の SE の上昇は見られるものの、多くの誤抽出を取り去ることができず SP が低下した。

(3) はアクティブ探索を行わず、すべての領域に対して AFINN による推論を行う実験である。つまり $\xi_{search}=0$ に相当する。探索過程がなくウィンドウを 1 画素毎ずらしながら検証が行われるため、大変冗長で非常に多くの領域が最終的に選択された。そのため今回の実験ではすべての物体に対して SE=100% となったが、不適切な領域も非常に多く抽出され、SP はどの物体に対してもさらに低下した。これは選択される領域数が極端に増加したためであり、検出部分の重なりを削除する等の改善の余地はある。しかし最も大きな問題は、検証の回数の爆発的増加に伴う処理時間の増加である。探索段階がある時には、画像 1 枚あたり検証の回数が、4 種類の物体、6 種類のウィンドウサイズに対して 20 回程度であったが、全検証の場合にはおよそ $20 \text{ 万} \times 4 \times 6 = 480$ 万回にもなる。この時処理に要した時間は 40 分にも達した。

(2)(3) の結果は、選択される領域の形状が正方形のみという制限もあるが、AFINN に与えられた 20 種類のデータのみでは十分な検出ができず、アクティブ探索と組み合

わせ、ある程度適切な領域が選択された上で利用することで効果が高められることを示している。

検証過程を構成する推論部分は、判別分析や、ニューラルネットワーク等に、stepwise法や主成分分析などによる入力選択、圧縮手法を組み合わせた方法でも実現可能である。しかしながら AFINN は、これらの処理モデルの構築から推論までを一貫して行うことができる。また、入出力の関係を明示的に記述できるだけでなく、本論文では取り上げていないが機能の一部追加やルールベースとの統合が容易という利点を持つ。汎用性の高い画像認識実現には、必要に応じた機能拡張は必須であり AFINN はこれを可能にしている。AFINN とパターンマッチングを組み合わせた統合型画像認識システムは、認識に関する明示的な知識を与えること無く、参照画像を与えただけで認識を行うことができる。これはシステム構築に際し、学習用画像を新たに準備する必要がなく、システム構築時の負担を軽減している。また、アクティブ探索と AFINN の組み合わせと適切なパラメータが、速度面でも、精度面でも効果的に作用していることが確認できた。

表 4.7 パラメータ別の認識結果

(SE,SP 共に欄内の数字は左から順に

椅子、コンピュータ、モニタ、キーボードに対応)

ξ_{search}	SE (%)				SP (%)			
0.7	29.3	45.0	32.2	86.8	100	50.6	88.8	33.3
0.6	80.9	84.0	71.2	98.5	74.6	52.3	88.1	38.3
0.5	88.0	88.0	81.6	100	43.6	30.0	64.8	29.4
0.0	100	100	100	100	34.0	7.8	50.7	23.0

4.5 結論

画像認識システムを構築する場合、事前に認識対象に関する全ての情報をルールや知識の形で登録しておくことは大変困難であり、これらの自動化を担う学習システムは必須であると考えられる。一方、認識精度に重点をおいた場合、特定の状況に特化した処理方法が記述されたルールベースシステムの利用や、パターンマッチングの手法が有利である。本章では、汎用的な画像認識システムの実現を目指し、効果的なパターンマッチング手法であるアクティブ探索法と、優れた学習機能、論理処理機能を持つ AFINN(adaptive fuzzy inference neural network) を利用した統合型画像認識システムについて述べた。本システムは、画像認識に関する知識やルールの埋め込みが不要で、参照画像を提供するだけで認識を行うことが可能である。また、登録する参照画像を変えることで、様々な状況での認識が可能であり、従来のような専用のシステム設計が不要である。

計算機実験では、コンピュータ等が設置された研究室内の画像を例題に扱い、およそ 150 枚の参照画像を利用し、例えば椅子やモニタに関しては、7 割以上の認識率を実現できることを確認した。本手法は誤抽出領域の削除や実利用時における速度面などで課題は残るが、登録する参照画像を追加や変更することで、他の用途でも適用が可能という柔軟性を有し、利点は大きい。また、本システムは AFINN を構成要素として用いているために、必要に応じて追加学習や知識ベースとの融合も可能であり、精度の向上を図ることができると考えられる。

第 5 章

結論

本論文は、ファジィニューラルネットワークの学習機能および知識処理に基づく画像認識、画像理解を目指した研究である。画像認識の実現には、対象に関する適切な知識の獲得と利用が必要である。従来の画像認識システムの多くは、あらかじめ登録されているルールやデータとのパターンマッチングにより処理を行うため、想定されていない状況では正しい処理は行えない。本論文ではこうした知識獲得、統合、利用に起因する画像認識の汎用性の限界を解決するために、ファジィニューラルネットワークの学習能力と論理的処理機能を利用する研究について述べた。

第 2 章では、風景画像を例題に認識と、学習によって自動生成される知識の抽出に関する研究について述べた。学習能力を持つニューラルネットワークと、論理的処理が可能なファジィの長所を融合した、ファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network) を基礎要素として構成された画像認識システムを作成し、獲得されたルールのみを用いて良好な認識結果および、単語の組み合わせによる画像の解釈結果を得た。これらの研究は、一般的に知識の付与が必要とされている画像認識問題に、ファジィニューラルネットワークを利用することで、明示的な知識を与えずに高精度の結果を目指したものであり、認識実験の結果からその有効性を確認した。

第 3 章では、画像認識問題で基本要素として用いて来た FINN の弱点を補う新しい処理モデルである、適応ファジィ推論ニューラルネットワーク (AFINN: adaptive fuzzy inference Neural Network) について述べた。AFINN は、学習データに応じて、処理に適したネットワーク構造やパラメータを自動的に決定することが可能であり、また従来ファジィニューラルネットワークでは困難であった多次元データの処理も可能である。評価実験の結果、非線形の多入力多出力の数式近似、ならびに公開されている音声認識、車の評価といった例題において、AFINN のモデル構築が適切であること、推論が高い精度で行えること、また得られるルールの妥当性を確認した。

第 4 章では、学習システムの持つ汎化能力と、特定の場面で高精度な認識が行えるパターンマッチングの手法の長所を融合した、複合型画像認識システムについて述べた。画像認識システムを構築する場合、事前に認識対象に関する全ての情報をルールや知識の形で登録しておくことは大変困難であり、これらの自動化を担う学習シス

テムは必須であると考えられる。一方、認識精度に重点をおいた場合、特定の状況に特化した処理方法が記述されたルールベースシステムの利用や、パターンマッチングの手法が有利である。このシステムは、これらの長所を融合して幅広い用途で、高い精度の画像認識実現を目指している。評価実験ではコンピュータなどが設置された研究室内の画像を例題に扱い、認識に関する明示的な知識やルールの埋め込みを必要とせず、参照画像を提供するだけで優れた認識率を実現できることを確認した。本手法は誤抽出領域の削除や実利用時における速度面などで課題は残るが、登録する参照画像を追加や変更することで他の用途でも適用が可能という柔軟性を有し、利点は大きい。また本システムはAFINNを構成要素として用いているために、必要に応じて追加学習や知識ベースとの融合も容易であり、精度の向上を図ることができると考えられる。

画像の認識は人間にとっても難しい部分が多く、知識処理などこれから解決すべき問題も多い。人間はそれまでの経験や学習に基づく多くの知識から総合的に判断して、状況の認識を行っており、人間に匹敵する認識、理解を実現するためには、莫大な知識の獲得とそれらの効果的な利用が必要である。これらを量的に実現するには、ある程度の自動化が必要であり、何らかの学習システムは必須であると考えられる。一方、個々の場面での対応には、状況に応じたトップダウンの処理が効果的である。これらのことから高い汎用性と正確さの両立には、ルールベースと自動学習の両方の統合が必要であると考えられる。本論文で扱ったFINNやAFINN等のファジィニューラルネットワークは、学習システムでありながらルールベースとの親和性が高いことから、画像認識という大きな課題解決に向けた効果的なアプローチの1つになると考えられる。

謝辞

本論文は、著者が慶應義塾大学理工学部から、大学院理工学研究科博士課程在籍中に同大学萩原将文教授のもとで行った研究をまとめたものです。博士論文執筆にあたり、学部4年生の頃より非常に親身になって御指導頂いてくださった萩原将文教授に心から深く感謝致します。先生には日頃から、公私にわたり様々な相談にのっていただき、また論文執筆時には、丁寧に何度も添削をしていただきました。修士課程修了後、一度は就職し研究から離れたものの、再び博士課程の学生として復帰する機会を与えてくださり、学位を取得できたのも先生の暖かい指導のお蔭です。本当にありがとうございました。また本論文の作成にあたっては、お忙しい中副査を引き受けてくださり、御指導頂いた本大学理工学部小沢慎治教授、櫻井彰人教授、斎藤英雄助教授に心から感謝致します。

また2年3ヵ月と短い期間ではありましたが、様々な経験と知識、心から楽しく充実した社会人生活を送るための環境を与えてくださった、勤務先の日本ヒューレット・パカード株式会社および、お世話になった皆様に深く感謝いたします。特に、仕事面でも私事の面でも大変お世話になり、また激務の中大変御迷惑がかかるにもかかわらず、私の博士課程進学のために2年間の休職期間を認めて、快く送り出してくださいました。上司の石積久尚さん、大野繁さん、MMPFプロジェクトの皆様、また同期の野村是孝さん、火山健次郎さんに心から感謝致します。本当にありがとうございました。

また、共同研究で大変お世話になり、様々な貴重なご意見をいただいた、本大学医学部皮膚科学教室の田中勝助教授、岡博史先生に深く感謝いたします。両先生は、大変親切に親身に指導してくださり、医学的な側面だけではなく工学的の面でも、私の研究の幅を大きく広げることができました。また、多大な御助言、御協力下さった萩原研究室の皆様、OB、OGの方々、高校時代からの知人であり、応用科学科博士課程の小山内靖さんにも感謝の意を示したいと思います。ありがとうございました。

本研究の成果は、友人や先輩、後輩の多大な支援のもとに成り立ったものです。私を支えてくださった全ての方々に改めて深く感謝致します。最後に、様々な面で私のわがままを聞いてここまで支えてくれた家族に心から感謝の意を示したいと思います。

彌富 仁

参考文献

- [1] 下田陽久他著, ”画像処理標準テキストブック”, 財団法人 画像情報教育振興協会, 1997.
- [2] 安西祐一郎, “認識と学習”, 岩波講座 ソフトウェア科学 16, 岩波書店, 1989.
- [3] 長尾真, “画像認識論”, コロナ社, 1983.
- [4] J.W.Cooley and J.W.Tukey, “An algorithm for the machine calculation of complex fourier series,” Mathematics of Computation, Vol.19, pp.297-301, 1965.
- [5] A.Guzman, “Decomposition of a visual scene into three-dimensional bodies,” AFIS Procs. of Fall Joint Computer Conference , Vol.33, pp.291-304, 1968.
- [6] A. Rosenfield, “Picture Processing by Computer,” Academic Press, 1969.
- [7] L.Wiskott, J-M. Fellous, N. Krueger and Christoph von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” Intelligent Biometric Techniques in Fingerprint and Face Recognition, CRC Press, Chapter 11, pp.355-396, 1999.
- [8] A. Jain, L.Hong, and S.Pankanti, “Biometric identification,” Communication of the ACM, ACM Press, Vol.43, No.2, pp.90-98, 2000.
- [9] Y.Gao and Maylor.K.H.Loung, “Face recognition using line edge map,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.24, No.6, pp.764-779, June 2002.
- [10] T. Kaneda, C.Thorpe and W.Whittaker, “Autonomous land vehicle project at CMU,” Proc. of the ACM 14th annual conference on Computer Science, pp.71-80, 1986.
- [11] T.P.Minka and R.W.Picard, “Interactive learning with a society of models,” Pattern Recognition, Vol.30, No.4, pp.565-581, 1997.

-
- [12] A.W.M.Smeulders, M.Worring, S.Santini, A.Gupta and R.Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.22, No.8, pp.888-905, Aug. 2000.
- [13] A.H.Ngu, Q.Z.Sheng, D.Q.Huynh and R.Lei, "Combining multi-visual features for efficient indexing in a large image database," *The VLDB Journal*, Vol.9, pp.279-293, 2001.
- [14] J.Z.Wang, "Integrated Region-Based Image Retrieval," Dordrecht Kluwer Academic, 2001.
- [15] K.Barnard, P. Duygulu, D. Forsyth, N.Freitas, D.M.Blei and M.I Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, Vol.3, pp.1107-1135, 2003.
- [16] J.Li and J.Z.Wang, "Automatic linguistic indexing of pictures by a semantical modeling approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vo.25, No.9, pp.1075-1088, Sep. 2003.
- [17] Y.Ohta, "Knowledge-based interpretation of outdoor natural color scenes," *Research Notes in Artificial Intelligence*, Vol.4, Pitman, 1985.
- [18] A.D.Bruce, T.C Robert B.John, R.H.Allen and M.R.Edward: "The scheme system," *International Journal of Computer Vision*, No.2, pp.209-250, 1989.
- [19] M.S.Thomas and A.F.Martin: "Context-based vision: Recognizing objects using information from both 2-D and 3-D imagery," *IEEE Trans. Pattern Anal. & Syst.*, vol.13, No.10, pp.1050-1065, Oct. 1991.
- [20] 平田真一, 白井良明, 浅田稔, "単一カラー画像から得られる3次元情報を利用したシーンの解釈", *信学論 (D-II)*, Vol.J75-D-II, No.11, pp.1839-1847, Nov. 1992.
- [21] J.Yamane and M.Sakauchi: "A Construction of a new image database system which realizes fully automated image keyword extraction," *IEICE trans. Inf.&Syst.*, Vol.E76-D, No.10, pp.1216-1223, Oct. 1993.
- [22] M.Mirnehdi, "Feedback control strategies for object recognition," *IEEE Trans. Image Processing*, Vol.8, No.8, pp.1084-1101, Aug. 1999.
- [23] P.Besl and R.Jain, "Three-dimensional object recognition," *ACM Computing Surveys*, Vol.17, pp.75-145, 1985.

- [24] E.Johnson and M.Hebert, "Using spin-images for efficient multiple model recognition in cluttered 3-D scenes," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.21, No.5, pp.433-449, May. 1999.
- [25] T.Funkhouser, P.Min, M.Kazhdan, J.Chen, A. Halderman, D.Dobkin and D.Jacobs, "A search engine for 3D models," ACM Trans. on Graphics, Vol.22, No.1, pp.83-105, Jan. 2003.
- [26] D.G.Lowe, "Three-dimensional object recognition from single two dimensional images," Artificial Intelligence, Vol.31, pp.355-395, 1987.
- [27] M.Flidcher, H.Sawhney, W.Niblack, J.Ashley, Q.Huang, B.Dom, M.Gorkani, J.Hafner, D.Lee, D.Petkoric, D.Steele and P.Yanker, "Query by image and video content: The QBIC system," IEEE Computer, Vol.28, No.9, pp.22-32, Sep. 1995.
- [28] D.Paulus, U.Ahlich, B.Heigl, J.Denzler, J. Hornegger, M.Zobel and H.Niemann, "Active knowledge - based scene analysis," MIT Press, Vol.1, No.4, pp.5-29, Winter 2000.
- [29] 村瀬洋, シュリーナイヤー, "2次元照合による3次元物体認識 - パラメトリック固有空間法 - ", 信学論 (D-II), Vol.J77-D-II, No.11, pp.2179-2187, Nov. 1994.
- [30] 村瀬洋, V.V.Vinod, "局所色情報を用いた高速物体探索 - アクティブ探索法 - ", 信学論 (D-II), Vol.J81-D-II, No.9, pp.2035-2042, Sep. 1998.
- [31] 川西隆仁, 村瀬洋, "色ヒストグラム特徴とパン・チルト・ズームカメラを用いた高速物体探索法 - 動的アクティブ探索法 - ", 信学論 (D-II), Vol.J84-D-II, No.8, pp.1722-1730, Aug. 2001.
- [32] K.Kashino, G.Smith and H.Murase, "Time-series active search for quick retrieval of audio and video," Proc. of IEEE ICASSP-99, VI, pp.2993-2996, 1999.
- [33] Ashiah Ghosh, Nikhi R.Pal, Sankar K.Pal, "Self-organization for object extraction using a multilayer neural network and fuzziness measures," IEEE Trans. on Fuzzy Systems, Vol.1, No.1, Feb. 1993.
- [34] Yoshikazu Nogami, Yoichi Jyo, Michifumi Yoshioka, and Sigeru Omatu: "Remote data analysis by Kohonen feature map and competitive learning," Proc. of IEEE System Man and Cybernetics, Vol.1, pp524-529, 1997.

- [35] 椋木雅之, 美濃導彦, 池田克夫, “対象物スケッチによる風景画像検索とインデックスの自動生成”, 信学論 (D-II), Vol.J79-D-II, No.6, pp.1025-1033, June 1996.
- [36] B.A.Drapen, “Learning control strategies for object recognition,” Oxford Univ. Press, pp.1-29, 1996.
- [37] J.Peng and B. Bhanu, “Closed-loop object recognition using reinforcement learning,” IEEE Trans. Pattern and Machine Intelligence, Vol.20, No.2, Feb. 1998.
- [38] M.Kruz, B.Volpel and H.Janfien, “Scale-invariant image recognition based on higher order autocorrelation features,” Pattern Recognition, Vol.29, No.1, pp.19-26, 1996.
- [39] L.A.Zadeh, “Fuzzy sets,” Information and control, Vol.8, pp.338-353, 1965.
- [40] E.H.Mamdani, “Applications of fuzzy algorithms for control of simple dynamic plant,” Proc. IEE, Vol.121, No.12, pp.1585-1588, 1974.
- [41] 菅野道夫, “ファジィ制御”, 日刊工業新聞社, 1998.
- [42] T.Takagi and M.Sugeno: “Structure identification of systems and its application to modelling and control,” IEEE Trans.on System Man Cybernetics, Vol.15, pp.116-132, 1985.
- [43] S.C.Lee and E.T.Lee, “Fuzzy sets and neural networks,” Journal of Cybernetics, Vol.4, No.2, pp.83-103, 1974.
- [44] A.Amano and T.Arisuka, “On the use of neural networks and fuzzy logic in speech recognition,” Proc. Int. Joint Conf. Neural Networks, pp.301-305, 1989.
- [45] C-T.Lin, and C.Lee, “Neural-network-based fuzzy logic control and decision systems,” IEEE Trans. Comput., Special Issue on Artificial Neural Networks, Vol.40, pp.1320-1336, Dec. 1991.
- [46] 林勲, 古橋武, “ファジィ・ニューラルネットワーク”, 朝倉書店, 1996.
- [47] 臼井支朗, 岩田彰, 久間和生, 浅川和雄, “基礎と実践ニューラルネットワーク”, コロナ社, 1995.
- [48] W.S.McCulloch and W.H. Pitts, “A logical calculus of the ideas immanent in nervous activity,” Bullet. Math. Biophysics, No.5, pp.115-133, 1943.

-
- [49] D.O.Hebb, "The organization of behavior," Wiley, New York, 1949.
- [50] F.Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, Vol.65, pp.386-408, 1958.
- [51] M.Minsky and S.Papert, "Perceptrons," MIT Press, Cambridge, 1969.
- [52] K.Nakano, "Associatron - A model of associative memory," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol.2, pp.380-388, 1972.
- [53] T.Kohonen, "Correlation matrix memories," *IEEE Trans. on Computers*, Vol.C-21, pp.353-359, 1972.
- [54] J.A.Anderson, "A simple neural network generating an interactive memory," *Math. Bioscience*, Vol.14, pp.197-220, 1972.
- [55] C. von der Malsburg, "Self-organization of orientation sensitive cells in the striate cortex," *Kybernetik*, Vol.14, pp.85-100, 1973.
- [56] K.Fukushima, "Neocognitron : A self-organizing multilayered neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, Vol.36, pp.193-202, 1980.
- [57] J.J.Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. National Academic Science, USA*, Vol.79, pp.2554-2558, 1982.
- [58] D.E.Rumelhart, G.E. Hinton and R.J.Williams, "Learning representations by back-propagation errors," *Nature*, Vol.323, pp.533-536, 1986.
- [59] T.J.Sejnowski, C.R.Rosenberg, "NETtalk: A parallel network that learns to pronounce English text," *Complex Systems*, Vol.1, pp.145-168, 1987.
- [60] B.Kosko, "Bidirectional associative memories," *IEEE Trans. on System Man and Cybernetics*, Vol.18, pp.49-60, 1988.
- [61] D.H.Ackley, G.E.Hinton and T.J.Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, Vol.9, pp.147-185, 1985.
- [62] Teuvo Kohonen: "The Self-Organizing Map," *Proceeding of IEEE*, Vol.78, No.9, pp.1464-1480, Sep. 1990.

- [63] C.M.Bishop and W.Maass, "Pulsed neural network", MIT Press, 1999.
- [64] Jyh-Shing and R.Jang, "ANFIS: Adaptive-network-based fuzzy inference system," IEEE Trans. on Systems, Man and Cybernetics, Vol.23, No.3, pp.665-685, May/June, 1993.
- [65] Li-Xin Wang, "Training of fuzzy logic systems using nearest neighborhood clustering," Proc. of the 2nd IEEE Int.Conf.on Fuzzy Systems, Vol.1, pp.93-100, 1993.
- [66] M.Sugeno and T.Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," IEEE Trans. Fuzzy Systems, Vol.1, pp.7-31, 1993.
- [67] Y.Lin and G.A.Cunningham, "A new approach to fuzzy-neural system modeling," IEEE Trans. Fuzzy Systems, Vol.3, pp.190-197, Apr. 1995.
- [68] N.R.Pal and J.C.Bedzek, "On cluster validity for c-means model," IEEE Trans. Fuzzy Systems, Vol.3, pp.370-379, 1995.
- [69] S.L.Chiu, "Selecting input variables for fuzzy models" Journal of Intelligent Fuzzy Systems, Vol.4, pp.243-256, 1996.
- [70] C-F.Juang and C-T.Lin, "An on-line self-constructing neural fuzzy inference network and its applications," IEEE Trans. Fuzzy Systems, Vol.6, pp.12-32, Feb. 1998.
- [71] C.Wong and C.C.Chen, "A hybrid clustering and gradient descent approach for fuzzy modeling," IEEE Trans. System, Man and Cybernetics, Vol.29, pp.686-693, Dec. 1999.
- [72] D.A.Linkens and M-Y.Chen, "Input selection and partition validation for fuzzy modeling using neural network," Fuzzy Sets and Systems, Vol.107, pp.299-308, 1999.
- [73] M-Y.Chen and D.A.Linkens, "A systematic neuro-fuzzy modeling framework with application to material property prediction," IEEE Trans. on Systems, Man and Cybernetics, Vol.31, No.5, pp.781-790, Oct. 2001.
- [74] T.Nishina and M.Hagiwara, "Fuzzy inference neural network," Neurocomputing, Vol.14, pp.223-239, 1997.
- [75] 北嶋啓至, 萩原将文, "自己組織化を利用した一般化ファジー推論ニューラルネットワーク", 電学論 C, Vol.117-C, No.7, pp.971-978, 1997.

- [76] 彌富仁, 萩原将文, “ファジー推論ニューラルネットワークを用いた風景画像からの知識抽出と認識”, 信学論 (D-II), Vol.J82-D-II, No.4, pp.685-693, Apr. 1999.
- [77] H.Iyatomi and M.Hagiwara, “Scenery image recognition and interpretation using fuzzy inference neural networks,” Pattern Recognition, Vol.35, No.8, pp.1793-1806, Aug. 2002.
- [78] 山村敦, 萩原将文, “ファジィ推論ニューラルネットワークによる位置関係を考慮した風景画像の認識”, 電学論 C, Vol.122-C, No.3, pp.506-pp.511, 2002.
- [79] H.Iyatomi and M.Hagiwara, “Adaptive fuzzy inference neural network,” Proc. of 10th International Conference on Human-Computer Interaction, Vol.2, pp.400-404, 2003.
- [80] 彌富仁, 萩原将文, “適応ファジィ推論ニューラルネットワークとアクティブ探索法を用いた汎用画像認識”, 信学論 (D-II) 採録決定.
- [81] S.Z.Selim and M.A.Ismail, “K-mean-type algorithms,” IEEE Trans. Pattern Anal. Mech. Intell., Vol.6, No.1, pp.81-87, 1984.
- [82] 境田慎一, 鹿喰善明, 田中豊, 湯川一郎, “k 平均アルゴリズムを利用した統合処理による画像の領域分割法”, 信学論 (D-II), Vol.J81-D-II, No.2, pp.311-322, Feb. 1998.
- [83] 越後 富夫, 飯作 俊一, “複数の GMRF モデルと併合仮説を用いたカラーテクスチャ画像領域分割”, 信学論 (D-II), Vol.J81-D-II, No.4, pp.660-670, Apr. 1998.
- [84] 岩田治幸, 長橋宏, “ニューラルネットワークを用いたカラー画像における動的領域分割”, 信学論 (D-II), Vol.80-D-II, No.11, pp.2995-3003, Nov. 1997.
- [85] J.M.Gauch, “Image segmentation and analysis via multiscale gradient watershed hierarchies,” IEEE Trans. Image Processing, Vol.8, No.1, pp.69-79, 1999.
- [86] M.L.Comer and E.J.Delp, “Segmentation of texture images using a multiresolution gaussian autoregressive model,” IEEE Trans. Image Processing. Vol.8, No.3, pp.408-420, 1999.
- [87] L.C.Kaplan, “Extended fractal analysis for texture classification and segmentation,” IEEE Trans. Image Processing. Vol.8, No.11, pp.1572-1585, 1999.
- [88] 高橋圭子, 阿部圭一, “ISODATA クラスタリング法を用いたカラー画像の領域分割”, 信学論 (D-II), Vol.J82-D-II, No.4, pp.751-762, Apr. 1999.

-
- [89] 木方 備輔, 松本 哲也, 岡田 稔, “並列強調型ニューラルネットワークの実装と画像処理への応用”, 信学技報 PRU95-226, 1996.
- [90] M.I.Jordan and R.A.Jacobs, “Hierarchical mixture of experts and the EM algorithm,” *Neural Computation*, Vol. 6, pp.181-214, 1994.
- [91] UCI Machine Learning Repository, <http://www.ics.uci.edu/~mllearn/>
- [92] M.Fanty and R.Cole, “Advances in neural information processing systems 3”, San Mateo, CA, 1991.
- [93] B.Zupan, M.Bohanec, I.Bratko and J. Demsar, “Machine learning by function decomposition,” *ICML-97*, Nashville, TN, 1997.
- [94] <http://www.google.com>
- [95] 谷口 慶治, “画像処理工学 基礎編”, 共立出版株式会社, 1996
- [96] 村上 伸一, “画像処理工学”, 東京電機大学出版局, 1996

付録 A

ファジィ推論ニューラルネットワーク (FINN)

本章では、コンピュータの学習機能、知識獲得に基づく画像認識、理解実現を目指す本研究の基礎となるファジィニューラルネットワーク (FINN: fuzzy inference neural network)[74] を紹介し、その構造、動作原理について概説する。

A.1 FINN の構造

図 A.1に FINN の構造を示す。FINN は、多入力に対しても自動的に入出力空間を分割しファジィ推論規則を生成、抽出する能力を持っている。FINN は、入出力層とルール層の 2 層から構成される。入出力層は、入力部と出力部からなり入力部とルール層間、ルール層から出力部間は全結合となっている。入力部からルール層にはメンバーシップ関数が配置され、ルール層から出力部には後件部定数が配置されている。ルール層上のノードは、それぞれが 1 つのファジィ推論規則を表している。つまり入力部からルール層で if-then 型のファジィ推論規則の前件部を表し、ルール層から出力部で後件部を表現している。

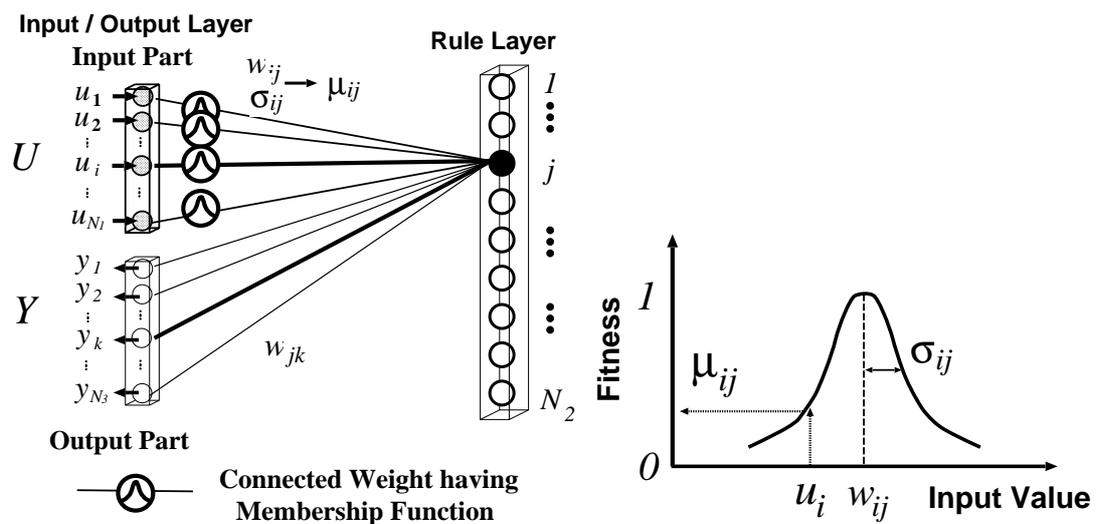


図 A.1 FINN の構造とメンバーシップ関数

A.2 FINN の動作

次に FINN の動作について述べる。入力する変数の数を N_1 、ファジィルールの数を N_2 、出力変数の数を N_3 とすると、FINN への入力は

$$\mathbf{U} = (u_1, u_2 \cdots, u_i \cdots, u_{N_1})^T. \quad (\text{A.1})$$

と表される。また入力部ノード i から、ルール層上のノード j に対しての前件部メンバーシップ関数は、次式に与えられるような Bell 型の中心値 w_{ij} 分散 σ_{ij} で与えられる。

$$\mu_{ij} = \exp\left(-\frac{(u_i - w_{ij})^2}{\sigma_{ij}^2}\right). \quad (\text{A.2})$$

$$j = (1, 2, \cdots, N_2)$$

ルール層では、各ノードに入力される『あいまいさ』 μ_{ij} の最小値をとり、『適応度』と呼ばれる確かさ ρ_j を計算する。

$$\rho_j = \min[\mu_{1j}, \mu_{2j} \cdots, \mu_{ij}, \cdots, \mu_{N_1j}]. \quad (\text{A.3})$$

そして後件部定数 w_{jk} と各ファジィ推論規則の適応度 ρ_j の荷重平均をとることによって最終的な推論出力 \hat{y}_k を得る。

$$\hat{y}_k = \frac{\sum_j^{N_2} (w_{jk} \rho_j)}{\sum_j^{N_2} \rho_j}. \quad (\text{A.4})$$

$$k = (1, 2, \cdots, N_3)$$

ファジィ推論ルール j は、次の様に与えられる。

If u_1 is \tilde{w}_{1j} , and u_2 is \tilde{w}_{2j} , and \cdots , u_i is \tilde{w}_{ij} , \cdots , u_{N_1} is \tilde{w}_{N_1j} **then** \hat{y}_k is w_{jk}

ここで、 \tilde{w}_{ij} は、 σ_{ij} の大きさによる w_{ij} の近傍である。

A.3 FINN の学習

FINN[74] の学習は、自己組織化過程、ルール抽出過程、LMS 学習過程の 3 段階からなる。

A.3.1 自己組織化過程

図 A.2 に自己組織化過程の概念図を示す。図では、概念が分かりやすいように、マップ層を 2 次元に拡張した。自己組織化過程では、Kohonen のアルゴリズム [62] を用いてメンバーシップ関数の中心値 w_{ij} と後件部定数 w_{jk} を仮決定する。

入力空間と出力空間を総合的に扱うためにまず式 (A.1) で表される入力ベクトル U とそれに対応する希望のベクトル Y

$$Y = (y_1, y_2, \dots, y_k, \dots, y_{N_3})^T. \tag{A.5}$$

を以下のように $(N_1 + N_3)$ 次元の学習ベクトル I に統合する。

$$I = [U, 0]^T + [0, Y]^T. \tag{A.6}$$

つまり学習ベクトルは $N_1 + N_3$ 次元のベクトルとなる。これを提案システムでは $[0, 1]$ に正規化してシステムへの入力とした。学習ベクトルに対して Kohonen の自己組織化特徴マップを用いることによって、類似したベクトル同士をマップ上の近くに配置することができる。

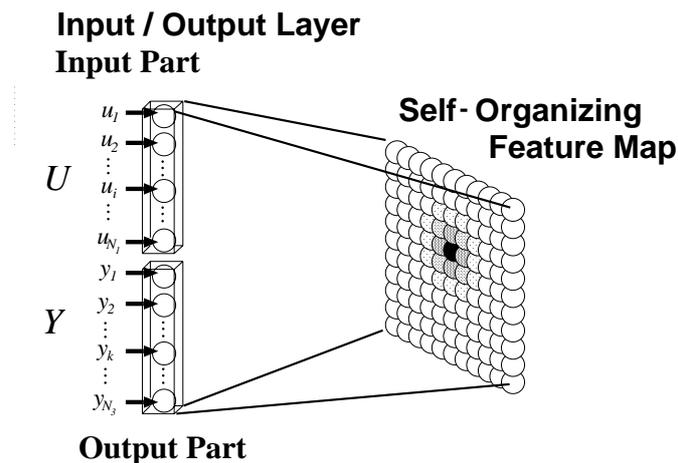


図 A.2 自己組織化段階

わかりやすくするために入力部, 出力部それぞれからマップ層上のニューロン j への結合ベクトル \mathbf{W}_j^U (N_1 次元), \mathbf{W}_j^Y (N_3 次元) を以下の式の様にまとめて重みベクトル \mathbf{W}_j ($N_1 + N_3$ 次元) と書くことにする。

$$\mathbf{W}_j = [\mathbf{w}_{ij}, \mathbf{0}]^T + [\mathbf{0}, \mathbf{w}_{jk}]^T. \quad (\text{A.7})$$

まず j への重みベクトル \mathbf{W}_j と、入力ベクトル \mathbf{I} のユークリッド距離を計算し、最も距離の小くなる勝ちニューロン s を探す。

$$\|\mathbf{W}_s - \mathbf{I}\| = \min_j \|\mathbf{W}_j - \mathbf{I}\|. \quad (\text{A.8})$$

続いて勝ちニューロン s とその近傍を以下の式で更新する。

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) + \epsilon_{self} h(j, s, t) (\mathbf{I} - \mathbf{W}_j(t)). \quad (\text{A.9})$$

ここで ϵ_{self} は、自己組織化段階における学習係数であり、 $h(j, s, t)$ は次式で表わされる近傍関数である。

$$h(j, s, t) = \exp\left(-\frac{|j-s|^2}{\sigma(t)^2}\right). \quad (\text{A.10})$$

ここで、 $\sigma(t)$ は学習の進行とともに $\sigma_{initial}$ から σ_{final} まで指数関数的に減少する値である。そのため、学習の初期では重みの更新が広範囲で行われ、次第に狭くなっていくことになる。以上の手順を繰り返すことにより、入力と出力の関係を特徴ごとに自動的に分類する。このようにして得られた、入出力層から各ニューロンの重みベクトルはファジィルールの前件部と後件部を合わせたものに相当する。

A.3.2 LMS 学習過程

自己組織化過程でおおまかにメンバーシップ関数の中心値と後件部定数を決定した。ここでは Least Mean Square(LMS) 学習則を用いて、メンバーシップ関数の分散を決定し、その中心値と後件部定数の最終的な調整を行う。過学習を防ぐためにこの過程は最小限にとどめることが望ましい。[75][76] LMS 学習では、総誤差関数 E を最小にするように各パラメータを最適化する。実際には次式で定義される誤差関数 E_p を最小化することによって総誤差 E の最小化とみなす。

$$E_p = \frac{1}{2} \sum_k^{N_3} (y_k - \hat{y}_k)^2. \quad (\text{A.11})$$

$$E = \sum_p E_p. \quad (\text{A.12})$$

ここで y_k は教師信号すなわち希望信号であり、 \hat{y}_k は推論結果であるネットワークの出力である。また p は学習に用いたパターン数であり、提案システムでは学習する画像の領域数に相当する。

最適化したい変数を x とするとき、LMS アルゴリズムによって更新される変数の変化量は $-\frac{\partial E}{\partial x}$ に比例した量になる。従って更新式は次式で与えられる。

$$x(t+1) = x(t) - \epsilon_{LMS} \frac{\partial E}{\partial x}. \quad (\text{A.13})$$

ここで、 ϵ_{LMS} は、LMS 学習における学習定数である。

(A.13) 式に基づいて後件部定数は、(A.4) 式より

$$w_{jk}(t+1) = w_{jk}(t) + \epsilon_{LMS}^w (y_k - \hat{y}_k) \frac{\rho_j}{\sum_n^{N_2} \rho_n}. \quad (\text{A.14})$$

で更新される。

メンバーシップ関数の分散の更新は同様に (A.2) ~ (A.4) 式より

$$\begin{aligned} w_{ij}(t+1) &= w_{ij}(t) + \epsilon_{LMS}^w \\ &\times \sum_k^{N_3} (y_k - \hat{y}_k) \left(\frac{w_{jk} \sum_n^{N_2} \rho_n - \sum_n^{N_2} (w_{nk} \rho_n)}{(\sum_n^{N_2} \rho_n^2)} \right) \\ &\times q_{ij} \mu_{ij} \frac{2(u_i - w_{ij})}{\sigma_{ij}^2}. \end{aligned} \quad (\text{A.15})$$

$$\begin{aligned} \sigma_{ij}(t+1) &= \sigma_{ij}(t) + \epsilon_{LMS}^\sigma \\ &\times \sum_k^{N_3} (y_k - \hat{y}_k) \left(\frac{w_{jk} \sum_n^{N_2} \rho_n - \sum_n^{N_2} (w_{nk} \rho_n)}{(\sum_n^{N_2} \rho_n^2)} \right) \\ &\times q_{ij} \mu_{ij} \frac{2(u_i - w_{ij})^2}{\sigma_{ij}^3}. \end{aligned} \quad (\text{A.16})$$

で行われる。

このとき ϵ_{LMS}^w と ϵ_{LMS}^σ はそれぞれ、後件部定数、メンバーシップ関数の中心値に対する学習係数と、分散に対する学習係数である。また q_{ij} は、

$$q_{ij} = \begin{cases} 1, & \text{for } \rho_j (= \min[\mu_{1j}, \mu_{2j} \cdots, \mu_{N_1j}]) = \mu_{ij}, \\ 0, & \text{else} \end{cases} . \quad (\text{A.17})$$

である。

A.3.3 画像認識システム適用時の FINN の変更点

提案システムで用いているネットワークは FINN を元に構成されているが、より多くのデータを扱う本研究において、動作アルゴリズムに若干の修正を加えた。本節ではこの修正について説明する。

変更点 1 : 自己組織化学習時の距離係数

提案システムで用いる FINN は入力と出力の次元数の差を考慮するために、前章と同様に、入出力にそれぞれ係数 $\alpha_{in}, \alpha_{out}$ を掛け、式 (A.8) の代わりに式 (A.18) を使用した。

$$\|W_s - I\| = \min_j (\alpha_{in} \|W_j^U - U\| + \alpha_{out} \|W_j^Y - Y\|). \quad (\text{A.18})$$

変更点 2 : LMS 学習の学習係数

FINN の LMS 学習には、式 (A.2) ~ (A.4) が用いられていたが、ここで、 ϵ_{LMS} が大きすぎると、自己組織化段階で得られたメンバーシップ関数の中心値を大きく更新してしまい、ルール層上に知識として得られるべき数値情報が正規化された $[0,1]$ から外れることがしばしば起こり得る。そうなると、適切な知識の抽出が行えなくなってしまう。逆に学習係数が小さいと学習がなかなか進行しない [76]。

そこで提案システムでは、自己組織化で得られた各ニューロンへの結合重みの値を有効利用するために、学習係数をメンバーシップの中心値と分散で別の値を用意し、前者を小さくすることでこれらの問題の解決を図った。

付録 B

画素単位画像認識システムの詳細

B.1 システム概要

論文本文、第 2 章の図 2.1 がシステム構成の概要である。図中の KEN と書かれた各要素が、FINN[74] を拡張して構成された知識抽出ネットワーク (KEN: knowledge extraction network) である。本システムでは複数の KEN が学習によりルールを獲得し、それをもとに画像の認識を行う。知識の抽出は、各ネットワークの学習が終了した時点で生成されたファジィルールが言語形式に変換されることで行われる。画像の認識は、認識したい画像を学習の終わった各 KEN に入力し、それぞれの推論結果を統合し行われる。最終的なルールの抽出は、学習段階によって各 KEN が獲得したルールを統合して得られる。

B.1.1 知識抽出ネットワークの構成

KEN は、ファジィ推論ニューラルネットワーク (FINN: fuzzy inference neural network)[74] を基本に構成される。図 B.1 に KEN の構成図を示す。知識抽出ネットワークは、知識生成の精度を高めるために、2 次元のルール層を大小 2 つ持つ。また、画像の認識率向上および、総結合数を減らすことによる学習と認識時間の低減 [89] のため、KEN を学習する画像の種類により複数用意し、最終的にそれらの出力を統合する。

各 KEN は、入出力層と大小 2 つのルール層からなる。入出力層において、入力部には注目するピクセルの x 座標、y 座標、輝度 (Intensity)、色相 (Hue)、彩度 (Saturation) が入力される。出力部にはそのピクセルの属性、すなわち空、山、樹木等といった情報が出力される。学習時においては、これらが同時に入力され FINN のアルゴリズムに基づいて学習が行われる。まず、Kohonen の自己組織化アルゴリズムによって大小 2 つのルール層にマップが形成される。その後 Least Mean Square (LMS) 学習により、抽出される知識 (ファジィルール) の調整が行われる。

大きさの異なる 2 つのルール層は、役割が異なる。大きいマップは、より細かい知識が表現されているので、細かい部分の知識抽出および画像の認識を行う。小さいマッ

プはより大局的で普遍的なルールが集まると考えられるので、大きなマップで得られた知識で矛盾のあるものや、細かいものを削除するために使われる。

知識は、LMS 学習段階が終了したときの各知識抽出部の大小 2 つのマップの持つ情報が統合され、言語に変換されることにより得られる。

未学習画像の認識時には、入出力層の入力部のみに入力され、出力部に推論結果が出力される。

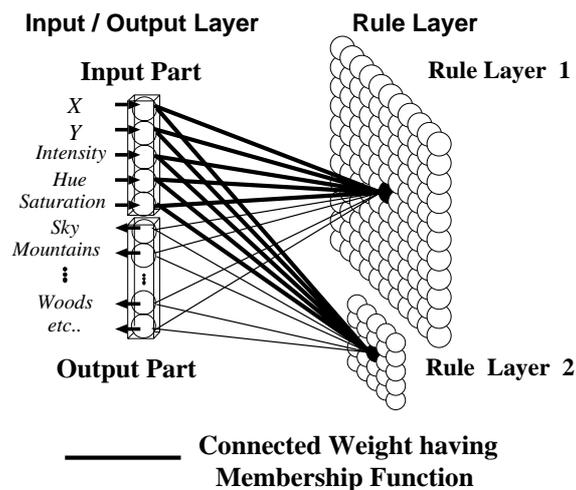


図 B.1 知識抽出ネットワークの構成

B.1.2 風景画の認識

提案システムでは、KEN が学習する画像の種類により複数個用意する。用意したネットワーク数を N_{net} とする。つまり各ネットワークは、それぞれ得意な分野を持つことになる。例えば、ネットワーク 1 は山岳部や樹木の多い画像の認識を得意とするネットワークという具合である。これは、Jordan らが提案している Mixture of Experts Network[90] の考えを利用したものである。

図 B.2 に認識過程の概念図を示す。入力画像は、水平方向に D 分割され特徴量が算出される。特徴量として用いられるのは、各領域の画素の色情報を平均して得られる RGB 色情報ベクトルである。入力画像が入力されると、 $d (= 1, 2, \dots, D)$ ブロックの 3 次元 (R, G, B) の特徴ベクトル C_d が取り出される。一方、ネットワーク $net (= 1, 2, \dots, N_{net})$ が学習画像の d ブロックより生成した特徴ベクトルを、 $L_{d,net}$ とする。

C_d と $L_{d,net}$ のユークリッド距離の比から次式により、入力画像の d ブロックにおけるネットワーク net の適応度 $F_{d,net}$ が求まる。

$$F_{d,net} \propto \|C_d - L_{d,net}\|^{-1}. \quad (B.1)$$

これらの処理により未知の画像に対して、 D 分割された部分ごとに認識が得意と考えられるネットワークの出力を、重要視することができる。

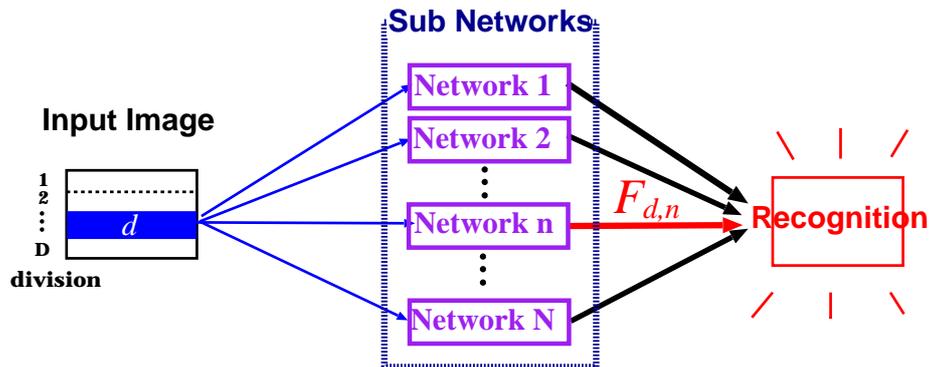


図 B.2 入力画像の認識過程

水平方向に入力画像を分割するのは、一般的な風景画では水平方向よりも垂直方向の変化が、画像中の領域に大きな影響を与えと考えられるからである。例えば、『空』の場所は画像の中で水平方向にはあまり依存しない。実際に、学習結果からも自動的に水平方向に関するメンバーシップ関数の幅が縦方向のものよりも明らかに大きくなっている。画素 (x, y) の最終的な認識結果 $R(x, y)$ は、以下のようにして与えられる。

$$I_s = \sum_{net}^{N_{net}} (\hat{y}_{s,net} \times F_{d,net}).$$

$$R(x, y) = k, \quad \text{where} \quad I_k = \max_s I_s. \quad (\text{B.2})$$

ただし、 $s, k (= 1, 2, \dots, N_3)$ は、空、山、樹木等の属性であり、 N_3 は前記の通り出力部のノード数である。また $\hat{y}_{s,net}$ は、ネットワーク net におけるそのピクセルの属性 s の推論結果である。

B.1.3 ルール抽出過程

知識抽出過程では、学習が終了したネットワークから if-then 型のルールが取り出され、言語化されルールとしてまとめられる。

本システムで抽出される認識に関する知識とは、得られたファジィルールのパラメータである入力部-ルール層間のメンバーシップ関数の中心値、および分散の値を用いて言語によるラベル付けされたものである。

いま図 B.3 のように、あるルール層上のニューロン j への結合重みを考える。ニューロン j へは、入力部からメンバシップ関数をもった重みによって結合され、ファジィルールの前件部を表している。また出力部との重みで結合され、後件部を表し、これらをまとめて一つの if-then 型のファジィルールが形成される。
 得られるファジィルールは FINN と同様に、

If u_1 is \tilde{w}_{1j} , and u_2 is \tilde{w}_{2j} , and \dots, u_i is $\tilde{w}_{ij}, \dots, u_{N_1}$ is \tilde{w}_{N_1j} then \hat{y}_k is w_{jk}

である。ここで、 \tilde{w}_{ij} は、 σ_{ij} の大きさによる w_{ij} の近傍であり、言葉で表現すると「 w_{ij} くらい」となる。

各メンバシップ関数の中心値 $w_{ij}, (i = 1, 2, \dots, N_1)$ および、の各値は $[0, 1]$ に正規化されているのでその範囲内において言語を機械的にそして適切に割り当てることが出来る。具体例を表 B.1, 表 B.2 に示す。実際には、これらの組み合わせによる処理を行っている。

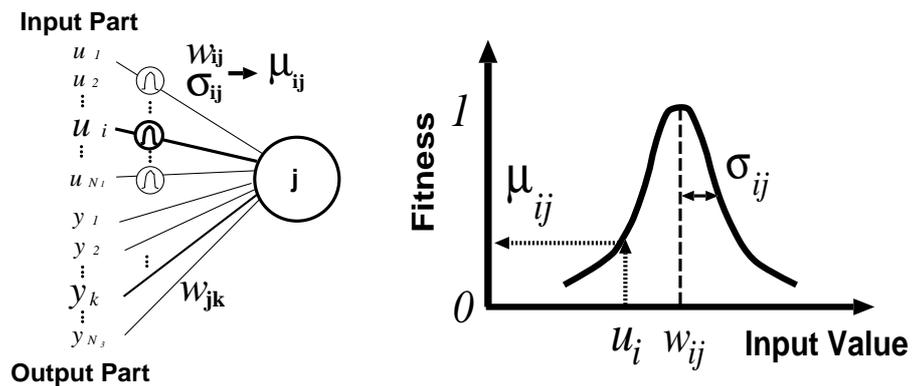


図 B.3 ニューロン j への結合とメンバーシップ関数

表 B.1 知識の数値と言語の対応の例 1

数値 \ 対象	w_{1j} (x 座標)	w_{2j} (y 座標)
0-0.15	左はじ	上
0.15-0.30	左	上の方
0.30-0.70	左右中央部	上下中央部
0.70-0.85	右	下の方
0.85-1.00	右はじ	下

数値 \ 対象	w_{4j} (色相)
0-0.01	黄色系統
0.01-0.30	緑系統
0.30-0.55	青系統
0.55-0.65	紫系統
0.65-0.80	赤系統
0.80-0.88	オレンジ系統
0.88-1.00	黄色系統

表 B.2 知識の数値と言語の対応の例 2

数値 \ 対象	σ_{1j} (左右の分散)	σ_{2j} (上下の分散)	σ_{4j} (色相の分散)
0-0.60	—	—	—
0.60-1.20	左右にはほぼ関係ない	上下にはほぼ関係ない	あまり色には依存しない
1.20-	左右には依存しない	上下には依存しない	色は関係ない

B.2 計算機実験

B.2.1 実験条件

風景画像

本論文中では棕木ら [35] の場合と同様に、風景画像をと同様に「屋外の情景を写した画像で、木・木立・湖・草原などの自然物が主な対象であるカラー画像」と定義する。また、画像は地面に対して水平に撮影されていることを仮定し、航空写真や上下の反転した画像などは考えないものとする。

対象物のラベル付け

教師あり学習と認識結果の出力のために、風景画像中に多く見られたものを大きく分ける。提案システムでは、分類不能を含む 10 種類に分類しラベル付けを行う。それぞれを表 B.3にまとめた。

利用する特徴量

画像の学習および認識に用いる特徴量として位置(座標)と色(輝度、彩度、色相)を用いた。周波数情報の利用も考えられたが、予備実験の結果、風景画像に対しては有効な特徴量が得られなかったので取り入れなかった。

入力される画像のデータは、学習に要する時間などを考慮し、 5×5 ピクセルを単位として用いた。本章の実験で扱った画像サイズは 200×140 画素なので、1 枚の画像の学習には 40×28 の計 1120 データを用いた。認識時には、必要とする解像度に応じて処理を行うことが可能である。本実験では、 2×2 画素を単位として実験を行った。

実験パラメータ

表 B.4に提案システムの構造パラメータを示す。

用意した知識抽出ネットワークの数 N_{net} と入力画像の分割数 D は、予備実験により適切と判断された 4 つにした。ルール層のニューロンの数も同様に、必要最低限と考えられる数にした。

表 B.3 画像認識で用いられるラベル

空
雲
遠くの山 (もの) 等
水 (海、湖) 等
土砂、岩山等
樹木、森等
草原等
影
雪、雪山
分類不能

表 B.4 提案システムの各パラメータ

知識抽出ネットワークの数 N_{net}	4 (同形態)
入力部のニューロン数 N_1	5
各ネットワークのマップ層 1 のノード数 $N_{2_{map1}}$	1024 (32 × 32)
各ネットワークのマップ層 2 のノード数 $N_{2_{map2}}$	25 (5 × 5)
出力部のニューロン数 N_3	10
各ネットワークが学習した画像の枚数	30 枚 (9 + 10 + 5 + 6)
画像の分割数 D	4

Sky	Woods
Cloud	Grass
Mountains	Shadow
Water	Snow
Rocks	Unknown

図 B.4 使用ラベル

B.2.2 学習結果

全部で 30 枚の画像を学習させたシステムに対して、認識能力の評価を行った。その結果を掲載するとともに、自己組織化過程、LMS 学習過程のそれぞれの途中の状況にもここで触れる。

自己組織化段階

自己組織化過程で用いたパラメータを表 B.5 にまとめる。

学習回数および学習係数を決めるにあたって、予備実験を行い、学習後のネットワークの結合荷重を調べた。学習回数は 10 回と一般的に Kohonen のアルゴリズムが用いられる場合に比べて著しく少ない。しかし学習の 1epoch に、画素数である 40×28 に枚数 10 を掛け合わせた数である約 10,000 個の 15 次元のベクトルが入力されることになる。一般的な風景画像の場合、これらの入力ベクトル同士の相関が高く、実際に画像の認識上分類されるべき特徴の数は、1epoch のベクトル数に比べて大幅に少ない。従って、学習回数は少なくとも 1epoch 中に同じようなベクトルが何度も入力されていることから、等価的に多くの学習がなされていると考えられる。

学習係数 ϵ_{LMS} および、学習の近傍関数 σ の初期値、最終値も予備実験の結果、適当と思われる値に設定した。

なお、10 枚の画像の自己組織化に要した時間は DEC CELEBRIS GL-2 (PentiumII 266MHz) を用いて約 180 分であった。

表 B.5 自己組織化段階の各パラメータ

自己組織化過程の学習回数 C_{self}	10
学習係数の初期値 (自己組織化段階) $\epsilon_{self}(t=0)$	0.1
学習係数の最終値 (自己組織化段階) $\epsilon_{self}(t=C_{self})$	0.001
学習の近傍関数 $h(j,s,t=0)$	0.5× マップ層の辺の長さ
自己組織化での位置重要度係数 α	1.3

LMS 学習段階

LMS 学習で用いたパラメータを表 B.6 にまとめる。LMS 学習段階はパラメータの最終調整およびファジールールの抽出の為に用いられるものである。誤差を減らすことで画像の認識結果も良くなるが、学習を進めすぎると学習画像固有のネットワークが出来上がってしまい汎用性に欠けることになる。そこで学習の回数を画像の理解結果と比較しながら 5 回が最適であると判断し、この回数にとどめた。

なお、10 枚の画像の LMS 学習段階に要した時間は、DEC CELEBRIS GL-2 (PentiumII 266MHz) を用いて約 90 分であった。

表 B.6 LMS 段階の各パラメータ

LMS 学習過程の学習回数 C_{LMS}	10
メンバーシップ関数の分散の初期値 $\sigma^{initial}$	0.5
メンバーシップ関数の中心値の学習係数の初期値 $\epsilon_{LMS}^w(t=0)$	0.1
メンバーシップ関数の分散の学習係数の初期値 $\epsilon_{LMS}^\sigma(t=0)$	0.01
メンバーシップ関数の学習係数の最終値 $\epsilon_{LMS}^{w,\sigma}(t=C_{LMS})$	0.001

B.3 知識抽出と画像認識結果に対する考察

B.3.1 知識抽出に関して

学習によって得られた知識は、KEN の結合荷重の形であらわされる fuzzy if-then 形式で表現される。このため、ルールを構成するメンバーシップ関数の形状から、色や位置情報といった各入力要素が、結果にどの程度影響されるかも含んだものになっている。またルール層上その知識が発火した回数によってその知識自体の重要性、普遍性が判断できる。

本手法は、大小2つのルール層を用いた階層的な知識の処理を行うことにより、言語による知識表記の矛盾などの問題解決を目指した。その結果、小さいルール層で得られた普遍的なルールと考えられる情報から、矛盾を含む細かい知識の削除が行えるようになった。人間が見て自然な形での知識が得られ、また、これらの知識を利用して行われた画像認識結果が良好であることから、適切な知識が得られたと考えられる。本手法は FINN を基本コンポーネントとして用いているために学習および、知識抽出が容易であり、またその構成が簡単であるため適用できる範囲が広い。

しかしながら、人間が持っているような知識は当然その様な情報だけではなく、一般的な知識と呼ぶにはまだまだ未熟である。本研究における知識抽出は、学習画像というパターン情報のみから、知識という形でテキスト形式の情報を抽出する、パターンからシンボルへの自動変換の初歩であると言える。

B.3.2 画像認識に関して

図 2.2 ~ 2.4 (第 2 章) から画像は、未学習画像であるにもかかわらず、良好に各領域が認識されていることが言える。

認識結果について細かく見ると、図 2.3(a) は認識結果は良好そうであるが、人間による認識結果との一致度が良くない。これは空とも雲とも取れる上半分の広い領域を人間が雲と判断したのに対し、システムは空と認識したからである。図 2.3(b)(c) では夕焼け空の画像であるが良好に認識されている。これは、夕焼け画像を扱う専門のネットワークの存在によると考えられる。

表 2.1 (第 2 章) からは、空と雲、木と草などについては認識誤差が多いことが分かる。しかしながらこれらは、人間にとっても厳密な違いの認識が難しいと考えられる部分である。

B.3.3 画素単位認識システムの課題

システム構成として、本手法の様に得意な認識領域を考慮するのではなく、万能な大きなネットワークを用意し、数多くの画像を学習させて処理させることも試みた。ところが学習する画像数が大きくなると、ネットワークの規模もそれに応じて大きくなるため、拡張性と、学習および処理時間短縮のために複数のネットワークの並列処理形式を採用した。

本手法は、ピクセルごとの処理を行うため、従来の前処理での領域分割や、その領域に対する知識を全く必要としない。そのため、各画像の認識は精度の面で従来の手法に劣る部分もあるが、基本的な手法として様々な場面で利用できると考えられる。

欠点としては、識別結果が学習データに大きく依存してしまう点がまず挙げられる。また、ピクセルごとの認識で特徴として用いているのが位置情報と色情報のみであるので、経験による状況判断ができる人間に比べて、どうしても認識能力が劣ってしまう事が挙げられる。図 2.2(c) を例にとると、山にある雪の部分が空と認識されている。人間が見れば周囲の状況から雪であると判断できるが『上のほうにあって白いものは空、雲』という知識を持つネットワークはこれを空と判断した誤った例である。

また、物体の形状に考慮していないため、人工物などの対応に大きな課題が残る。今後、より高い精度での認識を行うためには、周囲の情報の考慮や、効果的な特徴量の獲得などを行いより洗練されたルールの獲得が必要である。

付録 C

領域分割とテクスチャ

C.1 K 平均アルゴリズム

K 平均アルゴリズム [81] は、画像内の各画素の有する特徴量でクラスタリングを行う代表的な手法である。K 平均アルゴリズムには、色情報や位置情報など多様な特徴量を一元的に取り扱うことができ、簡単な計算で処理が実行できるという利点がある。以下に、K 平均アルゴリズムによる領域分割法を示す。

1. 初期状態として画面を矩形に分割したクラスタ (領域) を与える。
2. 各クラスタごとに特徴量の平均値を計算する。クラスタリングの特徴量として画素の色情報と位置情報を用いる。
3. 全画面の画素ごとにその画素の特徴量がどのクラスタの特徴量の平均値と最も近いかを式 (C.1) の距離 L によって調べ、そのクラスタに統合することで、クラスタを再構成する。

$$L = \sqrt{k_l \{(I_1 - \bar{I}_1)^2 + (I_2 - \bar{I}_2)^2 + (I_3 - \bar{I}_3)^2\} + k_p \{(x - \bar{x})^2 + (y - \bar{y})^2\}} \quad (\text{C.1})$$

4. 再構成されたクラスタごとに平均値を調べ、同様な処理を繰り返しても平均値が変化しなくなったときクラスタリングを終了する。

領域の分割は、初期クラスタの形状、数、またパラメータ k_l, k_p の比により大きく変化する。 K_p の値が大きくなると領域分割するにあたって色よりも位置の情報が重要視され、結果が初期のクラスタ形状に大きく依存する事になる。本研究では、遠く離れた同じ領域を別のものとして扱い、後に統合を行うために、意図的に K_l の値を大きくし色情報を重要視する。

C.2 テクスチャ

テクスチャ(texture) は、要素が一定の配列規則に従って配列された繰り返しパターンである。画像処理においてテクスチャ情報は衛星写真による地形や森林の解析、生体組織の顕微鏡写真の解析などで重要視されている [95]。

テクスチャ解析法を適用する対象によって大きく分けると、1) 統計的手法 2) 構造的手法 3) スペクトル的手法 の3つに分類される。

本研究では、要素も、規則もはっきりしない風景画像の様な画像の解析に用いられる、統計的テクスチャ解析手法を用いた。

その統計的手法の中に、濃度変化の様子を画像上のあらゆる点について統計的に調べて、これを総括的に見ることの出来る手段の1つに共起行列 [95][96] を利用するものがある。

本システムでは、画素の輝度と位置をパラメータとした同時濃度生起行列 (gray level co-occurrence matrix) に基づくテクスチャ情報を利用した。

これはまず、輝度 i の画素から距離 r 角度 θ にある輝度 j である頻度 $P_\delta(i, j)$ を要素とする同時生起魚行列を求め、 $[0,1]$ に正規化してテクスチャ特徴量を計算する。

本システムでは、 δ を $r = 1, \theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ の場合に限って行列を求め、各行列から各角度において以下の特徴量を求めた。

$$q_1 = \sum_i \sum_j (i - j)^2 P_\delta(i, j) \quad (C.2)$$

$$q_2 = \sum_i \sum_j P_\delta(i, j)^2 \quad (C.3)$$

$$q_3 = \frac{ij P_\delta(i, j) - \mu_i \mu_j}{\sigma_i \sigma_j} \quad (C.4)$$

$$q_4 = - \sum_i \sum_j P_\delta(i, j) \log P_\delta(i, j) \quad (C.5)$$

$$(C.6)$$

ただし、 μ_k, σ_k は k についての平均と標準偏差である。

q_1, q_2, q_3, q_4 は、それぞれ コントラスト (contrast), 一様性 (uniformity), 相関 (correlation), エントロピー (entropy) であり、一様性と相関については 0° と 90° についてのみ計算し、計 12 種類の特徴量として抽出した。

関連文献

〈 主論文 〉

- ・ 彌富 仁, 萩原将文, “適応ファジィ推論ニューラルネットワークとアクティブ探索法を用いた汎用画像認識”, 信学論 (D-II) 採録決定.
- ・ H.Iyatomi and M.Hagiwara, “Adaptive fuzzy inference neural network,” Journal of Pattern Recognition (Accepted).
- ・ H.Iyatomi and M.Hagiwara, “Scenery image recognition and interpretation using fuzzy inference neural networks,” Pattern Recognition, Vol.35, No.8, pp.1793-1806, Aug.2002.
- ・ 彌富 仁, 萩原将文, “ファジィ推論ニューラルネットワークを用いた風景画像からの知識抽出と認識”, 信学論, Vol.J82-D-II, No.4, pp.685-693, 1999-4.

〈 主論文に関連する国際会議 〉

- ・ H.Iyatomi and M.Hagiwara, “Adaptive fuzzy inference neural network,” Proc. of 10th International Conference on Human-Computer Interaction, Crete, Greece, Vol.2, pp.400-404, June 2003.
- ・ H.Iyatomi and M.Hagiwara: “Knowledge Extraction from Scenery Images and the Recognition using Fuzzy Inference Neural Network,” , Proc. of IEEE Systems Man and Cybernetics, San Diego, USA, pp. 4486-4491, Oct.1998.

〈 主論文に関連する研究会 〉

- ・ 彌富仁, 萩原将文, “適応型ファジィ推論ニューラルネットワークとアクティブ探索法を用いた汎用画像認識”, 電子情報通信学会, 画像認識, メディア理解研究会 (PRMU), 信学技報, Vol.103, No.206, pp.17-22, 2003 年 7 月.
- ・ 彌富仁, 萩原将文, “風景画像の認識とその解釈” , 電気学会 産業システム情報化研究会 , IIS-99-30, pp.37-46, 1999 年 11 月.

- ・ 彌富仁, 萩原将文, “ファジィ推論ニューラルネットワークを用いた風景画像からの知識抽出と認識”, 第 14 回 ファジィシステムシンポジウム, pp.393-396, 1998 年 6 月.
- ・ 彌富仁, 萩原将文, “ファジィ推論ニューラルネットワークを用いた風景画像からの知識抽出と認識”, 電気学会全国大会, 3-31,32, 1998 年 3 月.

〈 その他の論文 〉

- ・ H.Oka, M.Hashimoto, H.Iyatomi, G.Argenziano, H.P.Soyer and M.Tanaka, “Internet-based program for automatic discrimination of dermoscopic images between melanomas and Clark nevi,” *British Journal of Dermatology*. (accepted).

〈 その他の国際会議 〉

- ・ H.Oka, M. Tanaka, S.Kobayashi, H.Iyatomi, M.Hagiwara, G.Argenziano, H.P.Soyer and T.Nishikawa, “Digital analysis of dermoscopy with multivariate stepwise discriminant analysis between early Melanoma and Nev,i” *Journal of European Academy of Dermatology and Venereology*, Barcelona, Spain, Vol.17, No.3, p.109, Oct. 2003.
- ・ Michiaki Itani, Hitoshi Iyatomi and Masafumi Hagiwara, “Additional Learning Machine Framework for Multipurpose Image Recognition,” *Proceedings of 4th International Symposium on Advanced Intelligent Systems*, Jeju Island, Korea, pp.480-483, Sep.2003.

〈 特許 〉

- ・ 岡博史、田中勝、田中敏幸、彌富仁、萩原将文、”色素沈着部位の遠隔医療システム”, 特願 2004-004747. (出願中).

〈 著書 〉

- ・ 小澤慎治編著 “ニューアルゴリズムによる画像処理システム事例解説”, トリケッ
プス, 第 3 章第 2 節分担執筆 (pp.41-72), 2003 年 6 月,