

論文 / 著書情報  
Article / Book Information

題目(和文)	適応的な写像獲得アルゴリズムとその人間支援システムへの応用に関する研究
Title(English)	
著者(和文)	平野敬
Author(English)	Takashi Hirano
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第7288号, 授与年月日:2008年3月26日, 学位の種別:課程博士, 審査員:小杉 幸夫
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:甲第7288号, Conferred date:2008/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

適応的な写像獲得アルゴリズムと  
その人間支援システムへの応用に関する研究

東京工業大学

大学院総合理工学研究科

メカノマイクロ工学専攻

学籍番号 06D34069

平野 敬

# 目次

<b>第1章 序論</b> .....	<b>4</b>
1.1 本研究の背景 .....	5
1.1.1 集合論における写像 .....	5
1.1.2 意味論における写像 .....	7
1.1.3 脳における写像 .....	10
1.1.4 ニューラルネットにおける写像.....	15
1.1.5 写像のモデル化における制約.....	15
1.2 本研究の目的 .....	17
1.3 本論文の構成 .....	18
<b>第2章 身体モデルへの写像獲得とアーム制御への適用</b> .....	<b>19</b>
2.1 アーム制御の背景・目的.....	20
2.2 双方向 FEATURE MAP によるアーム作業空間の学習的構築.....	22
2.2.1 Counter Propagation モデル .....	22
2.2.2 双方向 Feature Map モデル .....	23
2.2.3 双方向 Feature Map モデルの写像特性 .....	24
2.2.4 アーム制御への適用 .....	27
2.3 適応的な自己増殖による離散的写像への対策.....	32
2.4 第2章のまとめ .....	38
<b>第3章 形状モデルへの写像獲得と FAX-OCR システムへの適用</b> .....	<b>39</b>
3.1 FAX-OCR システムの背景・目的 .....	40
3.2 帳票モデルと帳票画像間の写像推定方式.....	42
3.2.1 帳票モデル作成処理 .....	42
3.2.2 帳票読み取り処理 .....	44

3.2.3	ロバストな変形量の推定 .....	45
3.3	実験結果 .....	49
3.4	FAX-OCR システムへの実適用 .....	58
3.5	第 3 章のまとめ .....	59
<b>第 4 章</b>	<b>知識概念への写像獲得と文書処理システムへの適用.....</b>	<b>60</b>
4.1	文書とカテゴリ間の写像.....	61
4.1.1	カテゴリへの写像の背景と目的.....	62
4.1.2	ページ記述言語 .....	64
4.1.3	提案方式の概要 .....	66
4.1.4	PDL 解析処理.....	67
4.1.5	文書論理構造抽出処理 .....	71
4.1.6	評価実験 .....	74
4.1.7	メタ情報抽出処理とカテゴリへの写像.....	81
4.1.8	文書とカテゴリ間写像を用いた人間支援システム.....	83
4.2	文書と関連知識間の写像.....	84
4.2.1	文書と関連知識間の写像に関する背景と目的.....	85
4.2.2	プッシュ型設計支援システム ( PSDSs ) の構成.....	87
4.2.3	設計知識とオントロジ .....	88
4.2.4	写像獲得処理 .....	90
4.2.5	ガイダンスメッセージの確認・レポート出力機能.....	93
4.2.6	実験結果 .....	94
4.2.7	文書と知識間の写像を用いた人間支援システム.....	97
4.3	第 4 章のまとめ .....	98
<b>第 5 章</b>	<b>事象概念への写像獲得と交通映像監視システムへの適用.....</b>	<b>99</b>
5.1	交通映像監視システムの背景・目的.....	100
5.2	意味論を利用した映像と交通事象との写像獲得.....	101

5.2.1 交通映像監視システム .....	101
5.2.2 画像解析部の処理 .....	102
5.2.3 意味解析部の処理 .....	104
5.3 実験結果 .....	109
5.3.1 仮想的なシナリオを用いたシミュレーション .....	109
5.3.2 交差点の映像に対する事象抽出結果 .....	110
5.3.3 知識データベースのサイズ .....	113
5.4 事象概念への写像を用いた人間支援システム .....	113
5.5 第5章のまとめ .....	114
<b>第6章 結論</b> .....	<b>115</b>
6.1 本論文のまとめ .....	115
6.2 今後の展望と課題 .....	116
<b>参考文献</b> .....	<b>117</b>
<b>本研究に関する発表論文等</b> .....	<b>123</b>
<b>付録1 メンタル・スペース理論</b> .....	<b>125</b>
<b>謝辞</b> .....	<b>127</b>

# 第1章

## 序論

生物の脳は、網膜や皮膚などの末梢神経から取り込んだ感覚情報を処理して外界の認識や目的の決定を行い、行動を行うための指示命令を筋肉などの効果器に対して出力する。これらの処理過程には、多様な形式の写像関係が存在していることが知られている。例えば、網膜から得た映像情報は視床の外側膝状体を介して大脳皮質の視覚野に送られるが、この過程では幾何学的な位相構造を保ったトポグラフィックマッピングと呼ばれる写像が構成されている（図 1.1）。

本研究では、この“写像”という切り口で生物の情報処理を考え、代表的な写像関係をモデル化することにより、実社会に役立つ人間支援システムを開発した成果をまとめる。代表的な写像関係として、脳には「集合論的写像」と「意味論的写像」が存在することを示し、各々に対して2例ずつをモデル化する。この研究は、生物の生体機能や構造を参考に、これを工学的に利用するバイオミメティックな情報処理研究の一種と言える。

本章では、まず 1.1 節において本研究の背景を述べ、次に 1.2 節で本研究の目的および研究対象とした4つの写像関係の定義、特性について説明する。最後に 1.3 節で本論文の構成を示す。

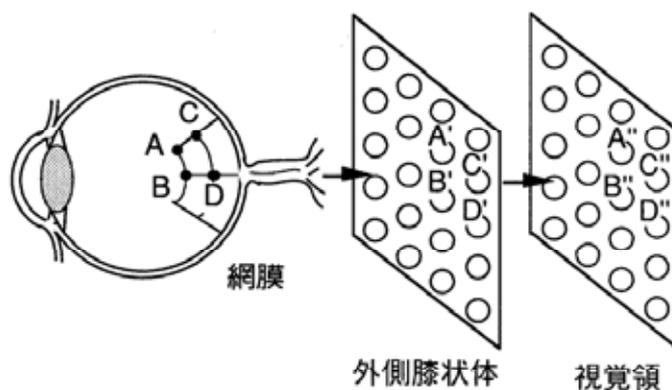


図 1.1 網膜から視覚野への写像

網膜上の関係 A-B-C-D が保持されたまま情報が視覚野に伝達される。視覚野における高次の視覚処理でも、この位相関係を保ったまま情報が処理される。

## 1.1 本研究の背景

本研究では，集合論（数学）において二つの集合間の関係を示すために定義された「集合論的写像」と，論理哲学や認知言語学の分野で意味を扱うために提唱された「意味論的写像」を扱う．本節では，これらの写像について定義すると共に，同様な写像関係が人間の脳内にも存在し，知的情報処理を行う上で重要な役割を果たしていることを示す．

### 1.1.1 集合論における写像

集合論は，集合を扱う数学理論である．“写像”という概念は本分野で最初に定義された．集合論における写像とは，二つの集合が与えられた時に，一方の集合の各元に対し一つずつ，他方の集合の元を指定して結びつける対応関係を示す．例えば，図 1.2 のように集合 A の各元に対して，集合 B の元をただ一つずつ指定するような規則  $f$  が与えられているとき， $f$  を始域（source）A から終域（target）B への写像であると言う．ここで集合 A の元  $a$  が  $f$  によって B の元  $b$  に移されるとき， $b$  を  $f$  による  $a$  の像（image）と呼び， $b = f(a)$  と表す．ここで写像の形態は，関数  $f$  の形により線形写像と非線形写像に分類される．また，関数  $f$  の連続性により連続写像と離散写像に分類される．

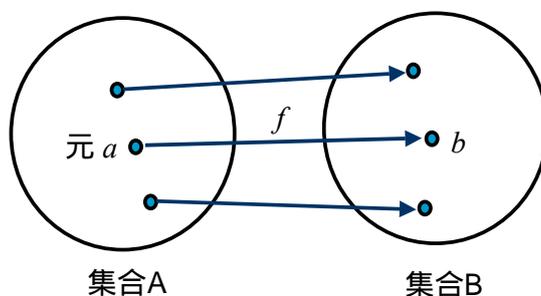


図 1.2 集合論における写像

#### (1) 線形写像と非線形写像

線形写像とは写像関数  $f$  が一次式であるか，一次式の和で表される写像であり，線型変換とも呼ばれる．線形写像以外の写像は非線形写像と呼ばれる（図 1.3）．

例えば  $V$  と  $W$  が同じ体  $K$  の上のベクトル空間であり， $V$  から  $W$  への写像  $f$  が，任意の  $x, y \in V$  と， $K$  の部分体  $F$  における任意の元  $c$  に対して式 1.1 と 1.2 を満たす場合， $f$  を  $F$  上の線形写像 という．この線形写像の工学的な応用例として，画像認識の分野で拡張や回転，平行移動を伴う物体の検出に利用されるアフィン変換がある．

$$f(x+y) = f(x) + f(y) \tag{1.1}$$

$$f(cx) = cf(x) \tag{1.2}$$

(2) 連続写像と離散写像

連続写像は位相空間上において元の値が連続的に変化する写像である。逆に、元の値が非連続的に変化する写像は離散写像と呼ぶ(図1.4)。

例えば  $f: X \rightarrow Y$  ( $X, Y$  は位相空間) において  $f$  が  $x \in X$  で連続写像であるとは、 $f(x) \in Y$  の近傍を  $V$  とした場合、 $x$  の近傍  $U_x$  において、近傍の像  $f(U_x)$  が  $V$  に含まれることをいう。これに対し、多変量解析や K-平均法などのクラスタリング手法を用いて、入力されたデータを特定のクラスタに分類するような写像は離散写像となる。

神経回路を模擬したニューラルネットは写像関係を獲得するモデルの一種と考えられる。ここで教師信号に従い写像を獲得する階層型ニューラルネットは、一般に連続写像を獲得する。教師なし学習を行う学習ベクトル量子化法 (Learning Vector Quantization :LVQ) などの自己組織化ニューラルネットは、一般に離散写像を獲得するモデルである。

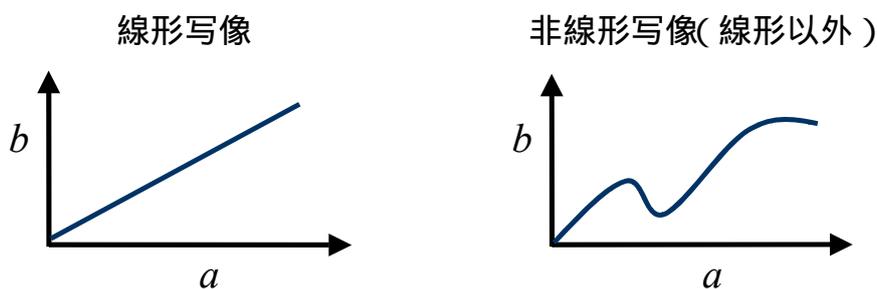


図 1.3 線形写像と非線形写像

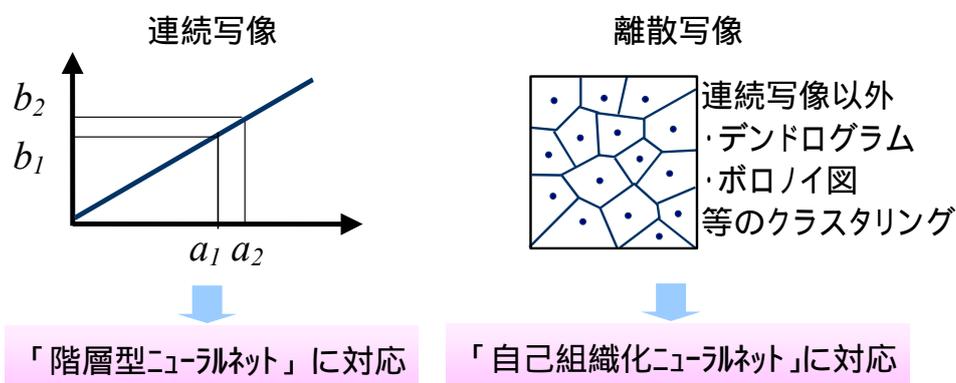


図 1.4 連続写像と離散写像

### 1.1.2 意味論における写像

意味論は、ものごとや言語の“意味”を扱う学問分野である。意味論における写像とは、現実世界と、それを認識して脳内で構築された世界モデルとの対応関係を表す。意味論的写像の概念は論理哲学の分野で提唱され、数理論理学や認知言語学を経て進展している。

#### (1) 論理哲学における写像

オーストラリアの哲学者 Wittgenstein は、論理哲学の研究が盛んになりつつあった 1921 年に「論理哲学論考」を著した[1-1]。ここで Wittgenstein は、人間が感覚器から得た情報を元に外界の状況を理解するというプロセスにおいて、言語と世界とを切り離すことはできず、世界そのものが言語であるという考え“写像理論”を提唱した。

写像理論では、言語世界を一種の幾何学的写像関数として、現実世界を構成する個々の事実を、世界モデルの構成要素である“像”に写像する(図 1.5)。ここで“像”は個々の事実を言語的に表した諸命題である。そして個々の像は真偽の値を持ち、その像に対応した事実と同じ論理構造を持つと述べた。

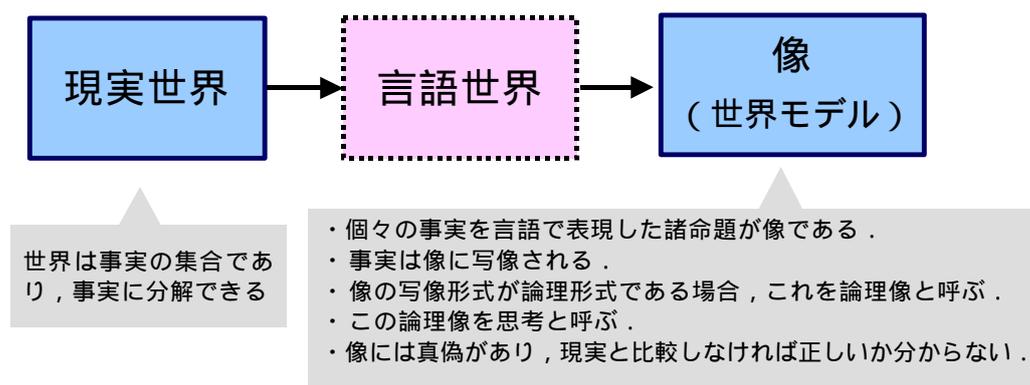


図 1.5 写像理論における現実世界と像の関係

#### (2) 数理論理学における写像

数理論理学は、言葉を文字や記号の列で表し、文法が数学的に厳密に定義された数式(形式言語)を用いて現実世界と人間の思考との関係を解析する学問分野である。数理論理学では、例えば「風が吹いた」という概念を記号 A で、「桶屋が儲かる」という概念を記号 B で、「風が吹けば桶屋が儲かる」という概念を数式  $A \rightarrow B$  で表すことで、記号と数式を用いて三段論法などの推論処理を行うことができる。

前記の写像理論は、この数理論理学の分野で具現化が検討された。例えば、この述語言語と推論処理を用いて現実世界から実在物 (Entity) の集合  $E$  への写像  $f$  を表現し、かつ  $E$  と真偽の集合  $\{0, 1\}$  との直和を像とすることで写像理論の具現化が試みられた。

実際にこの処理をプログラミング言語に実装した例が Prolog (PROgramming in LOGic : 論理プログラミング) である。Prolog は 1972 年に A. Colmerauer らにより開発された[1-2]。Prolog は述語論理を応用した非手続型の言語であり、推論機構を簡潔に記述できるため、人工知能の研究に広く利用された。日本では第五世代コンピュータの主要言語にも採用された。

ただし Prolog は、「知識（事実の集合）が明確に 0 と 1 に切り分けられており、全ての知識が記述されており、かつ記述されていない知識はすべて嘘（偽）である」という仮定のもとでのみ、論理演算の答えを出すことができる。しかし、未知の知識や明確な分類が難しい知識が存在する現実世界に対しては上手く適用することが困難であった。この問題に対して、人工知能向けプログラミング言語 Planner の設計者である C. Hewitt は“論理的推論はオープンシステム（現実世界）の並行計算を実行することが不可能だ”と述べた[1-3]。このように Prolog は現実世界に対する意味論的写像を扱うことが難しく、そのため狭い閉世界に限ったエキスパートシステムとして利用されるに留まった（図 1.6）。

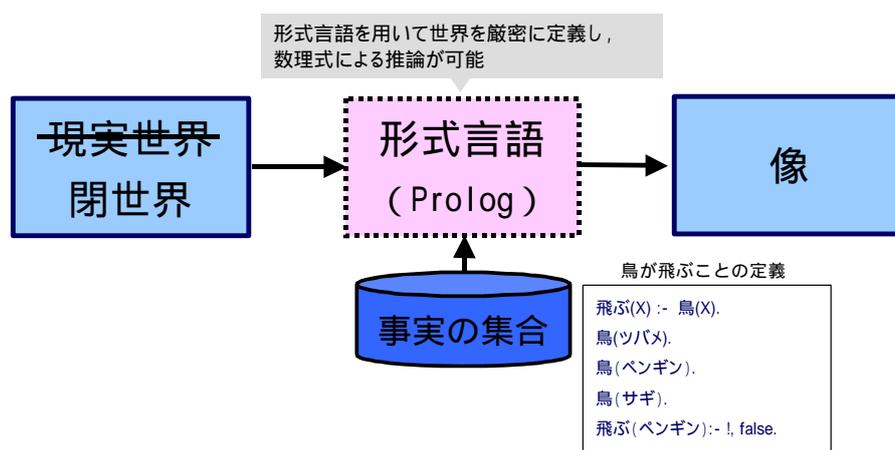


図 1.6 数理論理学における写像

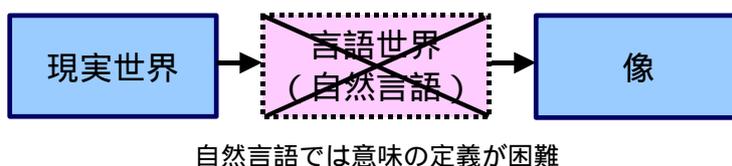
### (3) 言語学・認知言語学における写像

数式ではなく、より人間に密接な自然言語を用いた意味論の検討も行われた。しかし、自然言語では“意味”を明確に定義することが難しく、かつ機械翻訳分野では意味を考えずにある程度の翻訳結果を得ることができたため、意味論の必要性自体が問われた。結果として自然言語処理の分野では、意味論の検討はそのままに、文字列に動詞や名詞といったラベル付けを行う統語論の検討が進んだ。

しかし 1980 年代以降、認知言語学(Cognitive Linguistics)が立ち上がり、自然言語処理を用いた意味論が再検討されている。認知言語学は「言語のあり方は、認知の主体である人間が外的世界をどのように認識しているによる」とする言語研究である。認知言語学の方

法論として、1990年に Fauconnier はメンタル・スペース理論を提唱した[1-4]。この理論では、自然言語は単独で意味を定義できないと考え、これを補うためにメンタル・スペースという新しい場を設けた(図1.7)。この理論において、メンタル・スペースは一種の心的表現とみなされる。そして言語世界はメンタル・スペースの構築を行うための不完全な指令とみなされる。自然言語の文は写像構築の出発点であり、最終的な像は、文が表す不完全な意味に対して、一般的知識や推論、文脈からの補足、スキーマの抽出など様々な認知操作を加えることで得られる。メンタル・スペース理論は、メンタル・スペースと言う中間レベルを仮定することで、記述の曖昧性や意味論的スコープ、前提の投射など、意味論における幾つかの問題に対する解法を示し、自然言語を用いた意味の理解方法を説明した(メンタル・スペースの詳細は付録1に示す)。

### 言語学における意味論



### 認知言語学における意味論 (メンタル・スペース理論)

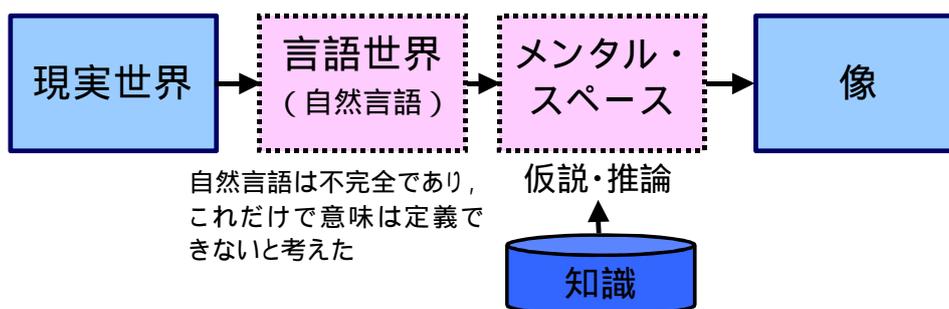


図 1.7 認知言語学における写像

### 1.1.3 脳における写像

人間の脳において、末梢神経から得られた感覚情報は視床を介して大脳皮質へに伝達される。視床を介した情報伝達は特殊投射核によるものと、非特殊投射核によるものがあり、各々が集合論的写像と意味論的写像に対応すると考える。

#### (1) 大脳皮質の機能局在と視床

大脳皮質は随意運動や感覚、思考等の高度な機能を担当する。ここで大脳皮質上の特定領域は、特定の機能を担当している。例えば、随意運動を担当する運動野は前頭葉の中心前回に、皮膚感覚を担当する体性感覚野は頭頂葉の中心後回に、視覚を担当する視覚野は後頭葉に、聴覚を担当する聴覚野は側頭葉にある（図 1.8）。このように脳内の特定領域が特定の機能を実現していることを、脳の機能局在と呼ぶ。表 1.1 は代表的な機能局在を示しており、言語に関する機能局在も確認されている。

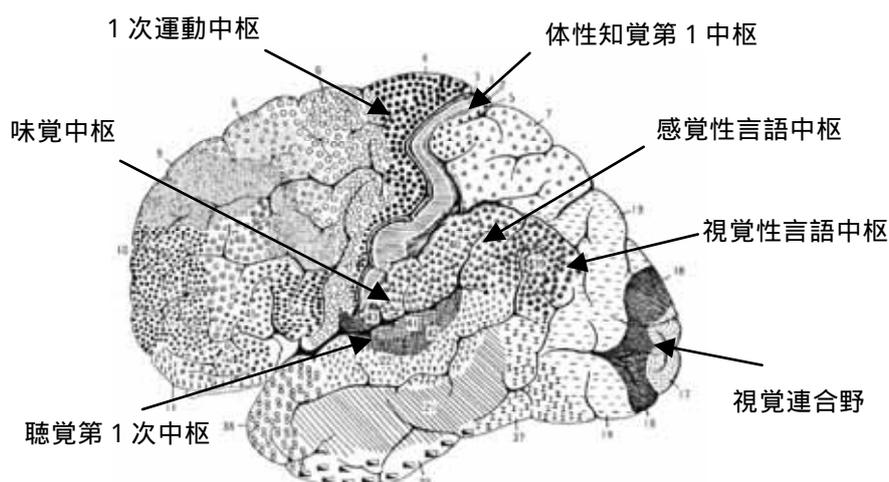


図 1.8 大脳皮質の機能局在

出典：“*Human Physiology*”, Springer-Verlag に一部追記

一方、視床は末梢神経から得た感覚情報（嗅覚以外）を大脳皮質の関連する機能部位に伝達したり、機能部位間の情報伝達に介在する。視床の内部は、その線維結合の様式により特殊投射核(specific nuclei)と非特殊投射核(nonspecific nuclei)に分類できる（図 1.9）。これらは後述の理由により、集合論的写像および意味論的写像と対応関係を持つと考える。

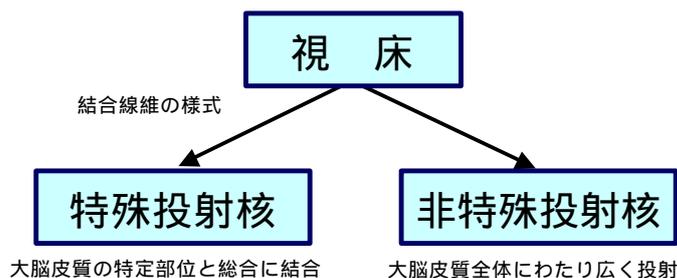


図 1.9 視床の特殊投射核と非特殊投射核

表 1.1 大脳皮質の機能局在例

領域	機能
運動中枢	随意運動を司り、前頭葉の中心前回中心旁小葉(Brodmann 4野)に1次中枢がある。体部位局在性(顔面野が下部、下肢野が中心傍小葉)を持ち脳幹・脊髄に投射する。
体性知覚第1次中枢	頭頂葉の中心後回、中心旁小葉(3-1-2野 その他)にあって、視床からの投射線維を体部位局在性(顔面野が下部、下肢野が上部)を持って受ける。
視覚第1次中枢	後頭葉の有線領(17野)にあって、外側膝状体からの投射線維を受取る。その周囲の、18野、19野は視覚連合野である。
聴覚第1次中枢	側頭葉の横側頭回(41野)にあり、内側膝状体からの投射線維を受取る。
嗅覚中枢	側頭葉の鉤を中心とした梨状葉、前梨状葉、扁桃体にある。
言語中枢	運動性言語中枢と聴覚性言語中枢、および視覚性言語中枢の三つがあり、通常は左脳に言語野(領)がある。 ・ <u>運動性言語中枢</u> : 左側の下前頭回(44, 45野)にありプロ - カ中枢(Broca's center)と呼ばれる。ここが損なわれると言葉を話すことが出来なくなる。 ・ <u>感覚性(聴覚性)言語中枢</u> : 左側の縁上回(22と40野)にあり、ウエルニッケ中枢(Wernicke's center)と呼ばれる。ここが損なわれると聴いている言葉の意味が理解できなくなる。音が聞こえるのは聴覚野の機能だが、聞こえた言葉の意味が分かるには聴覚性言語中枢が関与する。 ・ <u>視覚性言語中枢</u> : 左側の角回(39野)にあると言われている。ここが損なわれると書物に書かれた言葉が理解できなくなる。図形が見えるのは視覚野の機能だが、見えた文字の意味が分かるには視覚性言語中枢が関与する。
連合野	上記のような特定機能に関与していない部位。連合皮質は色々な情報を元に思考・推論を行い、判断を下す統合処理を行う。

(2) 特殊投射核と集合論的写像

視覚情報処理では、網膜が光を神経活動に変換し、この神経活動が大脳皮質の視覚野に届くことで周囲が見える。特殊投射核は、このような末梢神経と大脳皮質の特定部位との間の感覚伝道を中継する。視覚と同様、皮膚と体性感覚野や、耳（蝸牛）と聴覚野などの感覚伝道も特殊投射核を介して行われる（図 1.10）。図 1.11 は特殊投射核に関する情報伝達の流れを示しており、水色が視床への情報入力、赤色が大脳皮質の特定領域への情報出力を示す。図 1.11 に示した体性感覚野部分では、身体の感覚情報が位相構造を保つ形で写像されている（図 1.12）。同様に、視覚野部分では、網膜上の位相構造を保つ形で視覚情報が写像されている（図 1.13）。

このように特殊投射核に関係した脳内の写像は幾何学的な位相構造を保持した形態が多く存在している。そのため集合論的な写像との関係性が高いと言える。

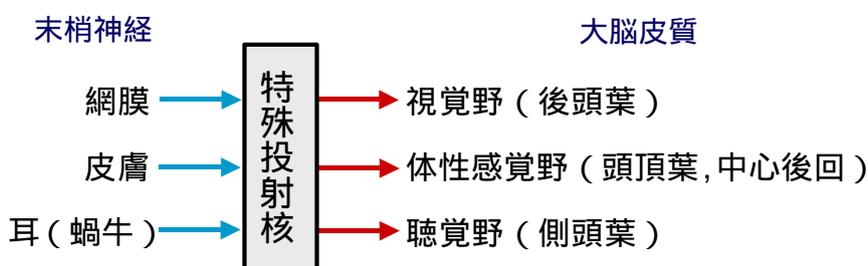


図 1.10 特殊投射核による末梢神経と大脳皮質特定部位間との情報伝達

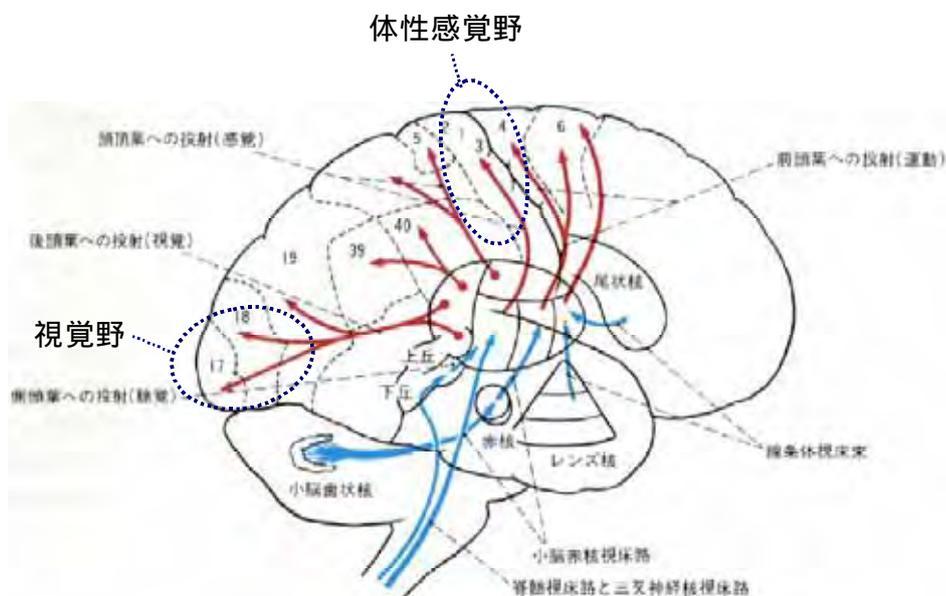


図 1.11 特殊投射核を介した情報伝達の流れ

出典：“脳解剖学”，南江堂 一部追記

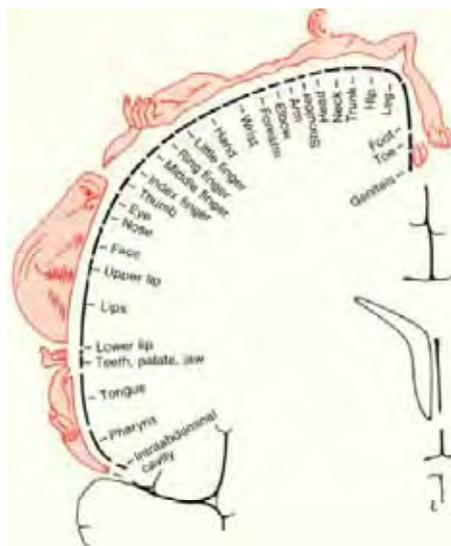


図 1.12 Penfield の調査による身体から体性感覚野への写像  
出典：“*Human Physiology*”，Springer-Verlag

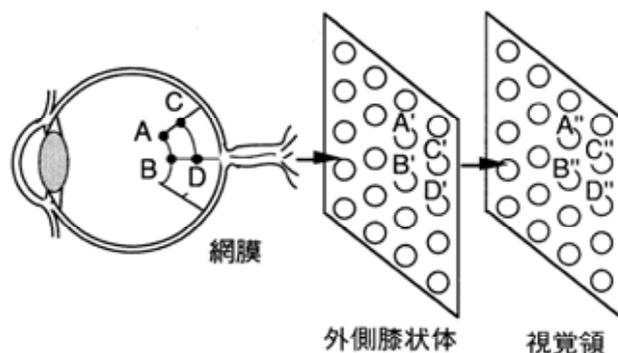


図 1.13 網膜から視覚野への視覚情報の写像

### (3) 非特殊投射核と意味論的写像

非特殊投射核は末梢神経からの感覚情報を大脳皮質全体に投射することで大脳皮質全体の機能を高めて意識レベルを上昇させる。また、大脳皮質の機能部位から他の機能部位への情報伝達にも介在することで抽象概念を扱う(図 1.14)。例えば、網膜からの神経活動が視覚野に届いただけでは周囲が「見える」だけだが、非特殊投射核を介して海馬や大脳皮質に蓄積させた記憶と照合するなどの過程を経ることにより、周囲の様子が「分かる」ようになる。周囲の様子が「分かる」という高次の認知機能は意味論的写像において像を求めることと同等と考えられる。

高次の認知機能が脳内でどのように実現されているかについては PET(Positron emission tomography ポジトロン断層法)を用いて調査が行われている。例えば、Bookheimer (1995)らは物体の呼称と単語の読みの過程を[1-5]、Kosslyn, Alpert, and Thompson (1995)は階層的

カテゴリーの検索過程を[1-6] ,Fletcher(1995)らは心の理論に関する処理過程を調べた[1-7] .  
 また, 脳内のどこに文法知識が貯蔵され利用されているかについて, Damasio(1992)がブローカ失語症の典型的な症状が文法規則の欠如に関係することを指摘し[1-8] , Stromswold (1996)らは左の44野(ブローカ野の一部)が文法処理に関わっていることを明らかにした[1-9] . また乾(1997)らは, 日本語独特の特性に着目して構文処理に関わる脳活動を調べた[1-10] .

これらの研究から, 脳内で言語処理・文法構造を担当する機能部位が存在することが明らかになっている. メンタル・スペース部の認知操作が脳内でどう実現されているかは未知な部分が多いが, 非特殊投射系の中ではメンタル・スペース理論に対応したような意味論的写像が実現されている可能性があると考えられる.

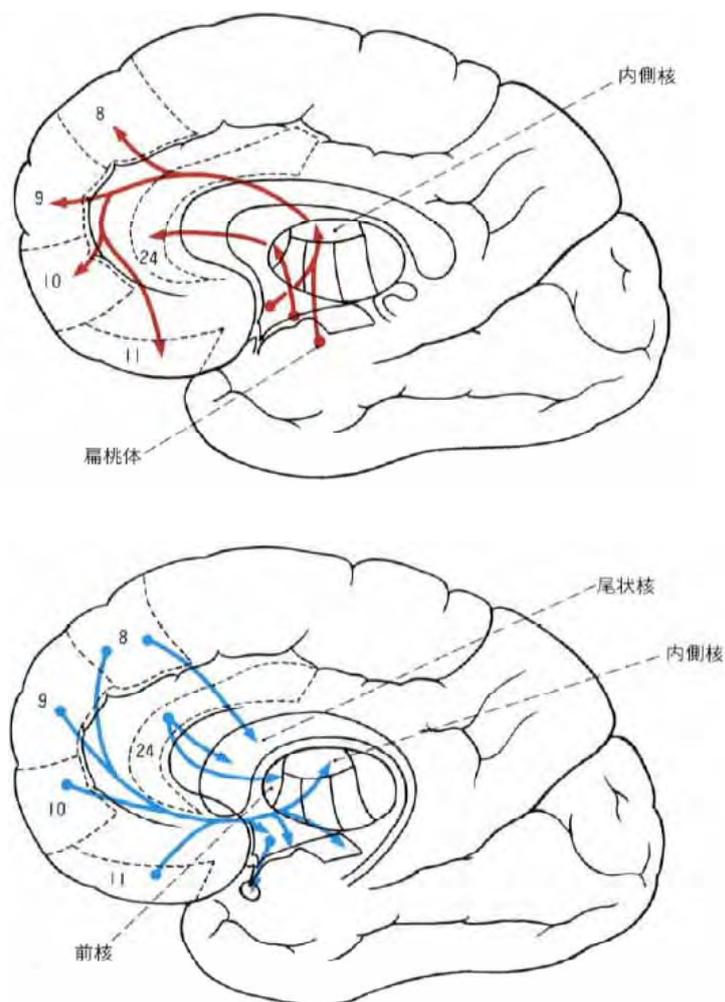


図 1.14 非特殊投射核を介した情報伝達の流れ  
 出典：“脳解剖学”，南江堂

### 1.1.4 ニューラルネットにおける写像

ニューラルネットは脳神経系を工学的にモデルにした情報処理システムであり、パターン認識やクラスタリングなど 様々な分野で利用されている。ニューラルネットの種類は、教師あり学習モデルと教師なし学習モデルに大別できる。

教師あり学習モデルの代表例にパーセプトロンがある。単層パーセプトロンは最初に考案された単純な構造のニューラルネットモデルである。M. Minsky と S. Papert らは[1-11]、単層パーセプトロンは線形非分離な問題を解けないことを指摘したが、Rumelhart らは誤差逆伝播法という学習方式を提唱し、この問題を克服した[1-12]。D. Marr は小脳はパーセプトロンであるという説を提唱し、脳機能の調査により、その説が正しいことが確認されている[1-13]。

一方、教師なし学習モデルの代表的例には、Kohonen の提唱した Feature Map [1-14] がある。これは大脳皮質の視覚野が持つコラム構造をモデル化したニューラルネットである。

これらのニューラルネットは何れも集合論的写像を構築する。意味論的な写像については、文法や概念を扱うニューラルネットの研究が齊藤、片岡、萩原らにより実施されている[1-15][1-16][1-17]。しかし、比較的単純な文法が概念の表現に留まっており、実用システムへの適用は困難である。

### 1.1.5 写像のモデル化における制約

本研究において集合論的写像（特殊投射核に対応）と、意味論的写像（非特殊投射核に対応）をモデル化する。ここではモデル化を容易にするため、脳機能や認知言語学の知見から写像の形に制約を設ける。

#### (1) 集合論的写像の制約

集合論的写像のモデルは、必ず幾何学的な位相構造を保持する。

この制約は、特殊投射核を介した視覚や触覚の情報処理において、幾何学的位相構造を保持したトポグラフィックマッピングが構成されている知見に基づく。

#### (2) 意味論的写像の制約

意味論的写像のモデルは、自然言語処理部と推論処理部との連携により写像を構築する。この推論処理では知識として、物理法則や一般常識等のアприオリな知識<sup>1</sup>を用いる。

ここで、自然言語を扱う能力は母国語で異なる可能性があり、また物理法則や一般常識

<sup>1</sup> アプリオリな知識：狭義には、哲学や生物学的な立場から人間が生得的に持っている先天的知識のことを示す。ここでは“誰もが、あたり前と思う知識”もアプリオリな知識に含める。

は学習的に獲得される後天的な知識であるため、この制約だけでは普遍性がないモデル化が行われる恐れがある。

しかし、Chomsky(1986)は言語習得の能力は生物的、遺伝的なものだとする生成文法(Generative grammar)の概念を提唱した[1-18]。この仮説は「人間の子供が、幼児期に触れる母国語の種類に依存せず、誤りを含む不完全な例文を聞くうちに、複雑な文法規則に従って言葉を使いこなせるようになるのはなぜか?」という問題に回答する。この事実は、言語を獲得しようとする子供の脳中に、それを可能にさせる生得的な器官の存在を示唆する。この生得的な言語機能を普遍文法(Universal Grammar)と呼ぶ。Chomsky の仮説には賛否があるが、本仮説が正しいとすると、母国語や人で大きく異なる普遍性のある写像が形成可能であると考えられる(図 1.15)。

また、物理法則や一般常識などの知識は、全てが後天的に獲得されたものではない。例えば乳幼児が表 1.2 に示すような単純な物理法則を知っていることを Baillargeon (1994) [1-19]や Spelke (1990) [1-20]が実験により示した。そのため物理法則や一般常識などのアプリアリな知識を推論に用いることは、写像の普遍性を損なはないと考えられる。

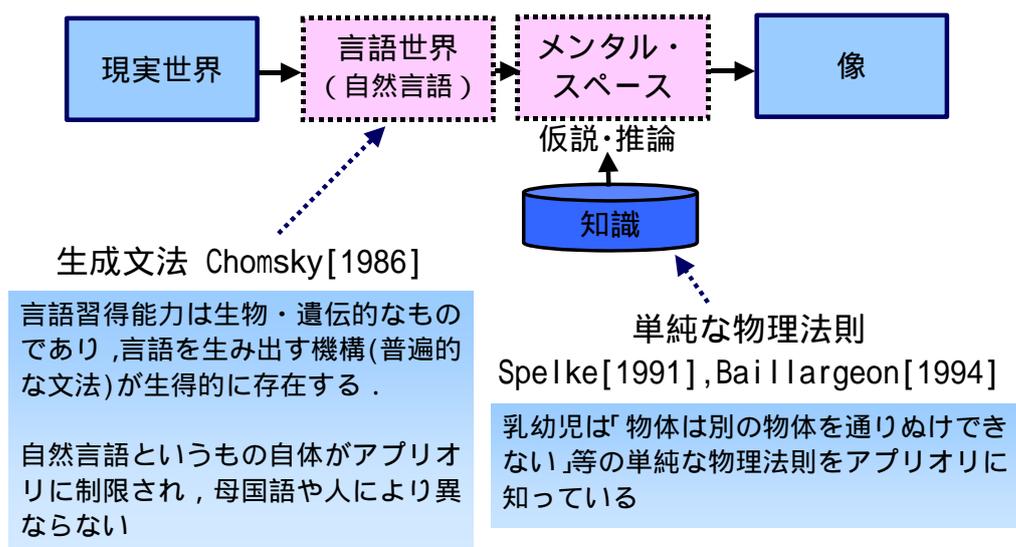


図 1.15 意味論的写像の制約 (アプリアリな知識)

表 1.2 乳幼児が知っている単純な物理法則の例

(1)物体は、外力によって動かされない限り、同じ場所に存在する
(2)物体は別の物体を通り抜けできない。物体が壁に突き当たった場合、壁に穴をあけて突き抜けない限りは壁の前で止まる
(3)物体は他の力を加えられない限り、自分自身でその形を変化することはない
(4)一つの物体は全体がまとまって動く(一部をつかめば全体がついてくる)

## 1.2 本研究の目的

本研究では，“写像”という切り口で生物の情報処理を考え，写像をモデル化することにより，実社会に役立つ人間支援システムを実現することを目的とする．モデル化は，特殊投射核に対応した集合論的写像から2例，非特殊投射核に対応した意味論的写像から2例を対象とする（図 1.16）．



図 1.16 本研究の目的

ここでは集合論的写像として「身体モデルへの写像獲得とアーム制御への適用」および「形状モデルへの写像獲得と帳票解析システムへの適用」について検討を行う．そして，この開発成果をアーム制御ソフトウェアや FAX-OCR 装置といった人間支援システムに適用する．

また意味論的写像として「知識概念への写像獲得と文書処理システムへの適用」および「事象概念への写像獲得と交通監視システムへの適用」について検討する．開発した写像モデルを文書・図面管理システムや FA 用表示装置，設計品質向上ソリューション等の実現に利用することにより，実例を持ってその有用性を示す．

### 1.3 本論文の構成

本論文の構成を図 1.17 に示す。まず第1章では本研究の概要や背景と目的を説明する。次に第2章と第3章で集合論的な写像について述べる。この第2章では「身体モデルへの写像獲得とアーム制御への適用」について示し、第3章では「形状モデルへの写像獲得と帳票解析システムへの適用」について説明する。その後、第4章と第5章では意味論的な写像について述べる。ここで第4章では「知識概念への写像獲得と文書処理システムへの適用」について記し、第5章では「事象概念への写像獲得と交通監視システムへの適用」について説明する。最後に第6章の結論において、本研究のまとめと今後の課題を述べる。

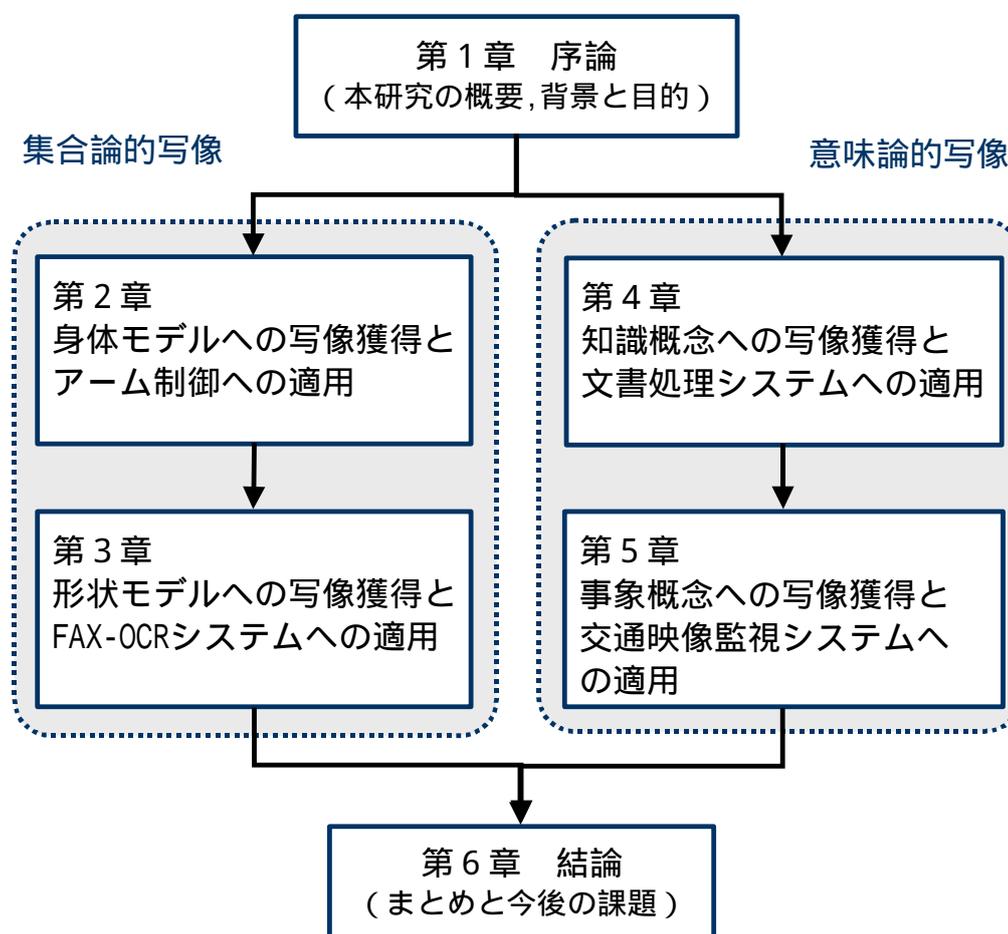


図 1.17 本論文の構成

## 第 2 章 身体モデルへの写像獲得とアーム制御への適用

本章では，人間が所望の位置に手を動かす動作に必要な写像関係をモデル化する．具体的には，ニューラルネット「双方向 Feature Map」を提案し，ロボットのアーム先端の空間座標とアームの身体座標間の写像を学習的に獲得する(図 2.1)．この写像モデルは表 2.1 に示す特長を持つ．実際に，この写像モデルをロボットアームの制御ソフトウェアに適用し，シミュレーションにより有効性を確認した．以降，2.1 節でアーム制御の背景と目的を示す．次に 2.2 節で双方向 Feature Map モデルとアーム制御への適用結果を示す，また 2.3 節では，外界に適応的に素子数を増加させるモデル拡張について述べ，最後に 2.4 節でまとめとする．

表 2.1 身体モデルへの写像

写像関係	アーム先端の空間座標と，アームの身体座標（関節角）
適応性	<ul style="list-style-type: none"> <li>・学習的に写像関係を獲得する．</li> <li>・外界に適応的に素子を自己増殖して写像精度を高める．</li> </ul>
写像の種類	幾何学的な位相情報を保つ集合論的写像であり，離散性と双方向性を有する．
脳内の関係部位	運動野/体性感覚野/頭頂連合野
人間支援システム	数理モデルの設定が困難な，自由度が大きく剛性が低いロボットアーム (パソコンでシミュレーションを実施，製品は未実現)

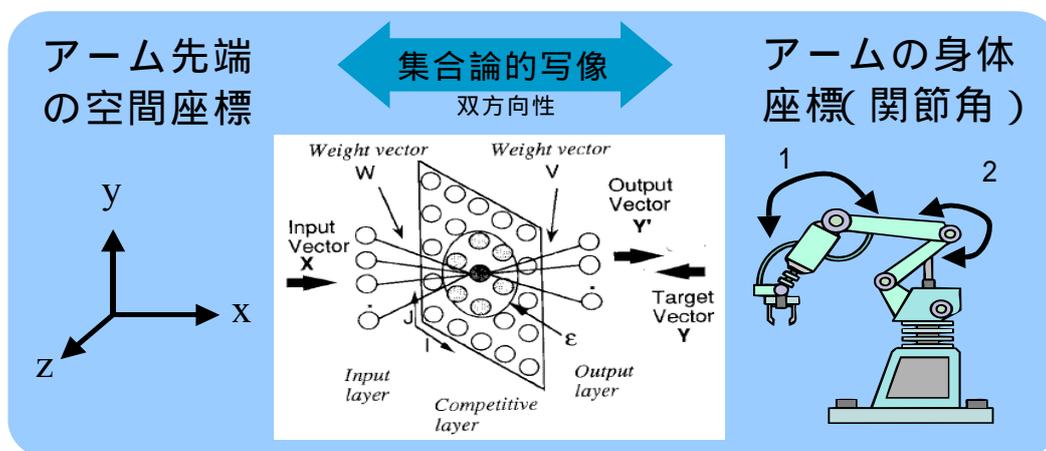


図 2.1 身体モデルへの写像獲得方法

## 2.1 アーム制御の背景・目的

自由度の大きいアームや、剛性が弱くて非線形領域を含む制御を行う必要があるアームでは、厳密な数理モデルの設定が困難な課題がある。そのため、生物の随意運動制御を模倣して、学習的に制御モデルを獲得する研究が成されている[2-1][2-2][2-4][2-6][2-7][2-11]。

生物における腕の随意運動制御は、「運動の目標設定」、「目標空間座標での軌道から身体座標上の軌道への写像」、「軌道を実現するアーム制御信号の発生」という階層構造を持つことが、川人らにより示されている[2-2]。また、目標軌道と実軌道との補正や制御モデルの学習を目的として、腕の深部感覚<sup>2</sup>を元に、身体座標から現在の空間座標を逆推定する経路も存在すると考えられる(図2.2)。このため、生物を模倣したアーム制御に必要な写像モデルは、以下の二条件を満足する必要があると考える。

### 写像の条件

- (1) 空間座標と身体座標間の双方向な写像の獲得が可能である
- (2) 制御しながら、同時に学習を行うことができる

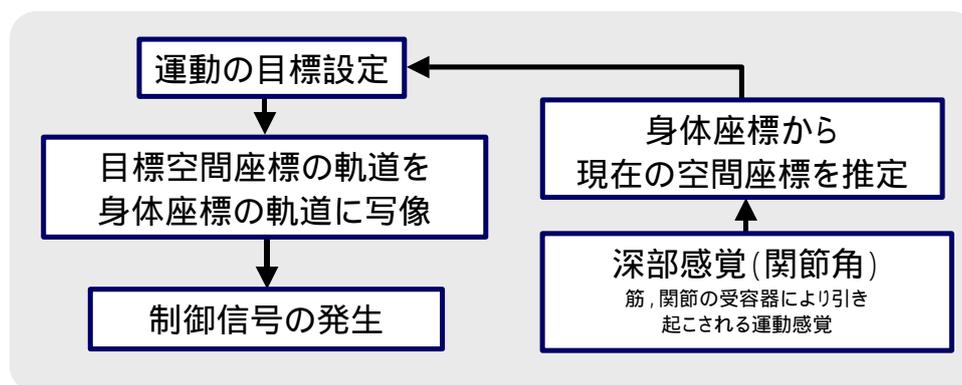


図2.2 生物における腕の随意運動制御フロー

実際に、ニューラルネットを用いて空間座標から身体座標への写像を獲得する研究がいくつか成されている。それらは誤差逆伝播法 (Back Propagation, BP) を用いた教師学習有りのニューラルネットと、教師信号が不要な自己組織化ニューラルネットを用いたものに分類できる。

BP は連続性のある写像を形成する有効な特長を持つ。しかしネットワークのサイズに対して十分な数の学習パターンが得られない場合、未学習点における幾何学的位相情報の保存は保障されない。この未学習点における BP の性能については Baum と Haussler [2-3] が検討している。本対策として、Jordan は学習データが少ない場合の未学習点における位

<sup>2</sup> 深部感覚: 体性感覚の一部であり、筋や腱、関節にある受容器によって引き起こされる感覚。運動感覚が主であり、振動、体の姿勢、位置、動きなどが感覚される

相情報の保持を保障するため、状態素子群に含まれるパラメータを調整し、内分操作を行う処理を提案した[2-4].

一方、自己組織化ニューラルネットである Feature Map[2-5]は未学習点においても幾何学的な位相関係を保存できる有効な特長を持つ。ただし教師無し学習であるために、対象とする写像関係が空間座標と身体座標との関係のように明確な場合、Martinetz[2-6]や Kuperstein[2-7]のモデルのような付加的構造を必要とする。また、写像が離散的であり空間分解能が素子数に依存する課題がある。

本研究では、Hect-Nielsen が提案した Counter Propagation モデル[2-8]に、Feature Map モデルを融合した構造を持つ、双方向 Feature Map モデルを提案する。本モデルは、空間座標と身体座標間の双方向写像を学習的に構築できる。また、幾何学的位相構造を保持する特性を持つため学習データが少ない場合でも安定的に制御を行いながら学習を進めることができる。

## 2.2 双方向 Feature Map によるアーム作業空間の学習的構築

### 2.2.1 Counter Propagation モデル

Counter Propagation モデルはパターン変換/分類に有効であり，双方向連想記憶装置として利用できる層状のニューラルネットである．図 2.3 はその構造を示す．

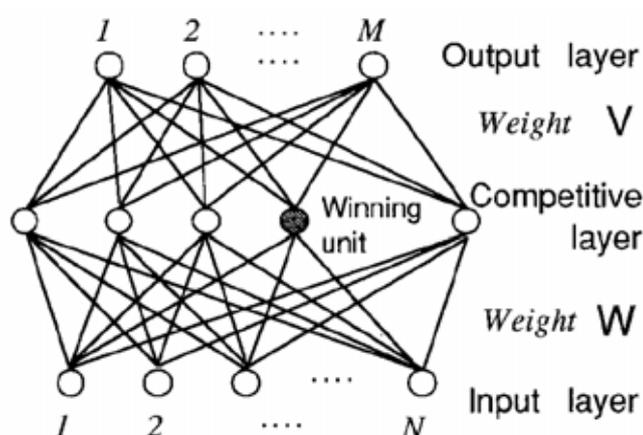


図 2.3 Counter Propagation モデル

本モデルは入力層，競合層，出力層の三層構造を持ち，各層間で全ての素子は各々シナプス荷重を持って全結合している．入力信号次元数を  $N$ ，出力信号次元数を  $M$  とすると，入力層から競合層上の素子  $j$  へのシナプス荷重  $W_j$  と，出力層から競合層上の素子  $j$  へのシナプス荷重  $V_j$  は次式で表現される．

$$W_j = (w_j^1, w_j^2, \dots, w_j^N) \quad (2.1)$$

$$V_j = (v_j^1, v_j^2, \dots, v_j^M) \quad (2.2)$$

学習過程において，時刻  $t$  に入力ベクトル  $X$  と出力ベクトル  $Y$  の学習セット  $(X, Y)$  を与えたとき，まず競合層上で  $X$  に対応する素子  $j^*$  が一個発火する．発火素子は 1 を出力し，結果として出力ベクトル  $Y' = V_{j^*}$  が出力層より得られる．ここで発火素子の探索は， $X$  と競合層上の  $j$  番目の素子が持つ荷重ベクトル  $W_j$  とのユークリッド距離  $D_j$  を調べ， $D_j$  が最小値を持つ素子  $j^*$  を発火素子と判定する．

次に，発火素子  $j^*$  が持つシナプス荷重が Hebb 則に従って次式のように更新する．

$$W_{j^*}(t+1) = W_{j^*}(t) + \eta(t) \cdot |X(t) - W_{j^*}(t)| \quad (2.3)$$

$$V_{j^*}(t+1) = V_{j^*}(t) + \eta(t) \cdot |Y(t) - V_{j^*}(t)| \quad (2.4)$$

ここで  $\eta(t)$  は時刻  $t$  での学習係数を示す。

学習終了後、このニューラルネットは入力ベクトル  $X$  を与えた時に、教師ベクトル  $Y$  の近似である  $Y'$  を出力する。ただし本モデルでは、入出力情報間の位相情報は保持されない。そこで幾何学的な位相構造を保つように本モデルを拡張した、双方向 Feature Map モデルを提案する。

### 2.2.2 双方向 Feature Map モデル

提案するモデルは Counter Propagation の競合層部分に Feature Map を適用したものである (図 2.4)。学習過程では、時刻  $t$  に入力ベクトル  $X$  と出力ベクトル  $Y$  の学習セット  $(X, Y)$  を与えたとき、Counter Propagation と同様に、競合層上で  $X$  に対応する素子  $ij^*$  を発火させる。発火素子は 1 を出力し、ニューラルネットからの出力ベクトルは  $Y' = V_{ij^*}$  となる。発火素子  $ij^*$  は、次式に示すように  $X$  と競合層上の  $ij$  番目の素子が持つ荷重ベクトル  $W_{ij}$  とのユークリッド距離  $D_{ij}$  が最小値を持つ素子とする。

$$D_{ij} = |X(t) - W_{ij}(t)| \quad (2.5)$$

$$|X(t) - W_{ij^*}(t)| = D_{ij^*} = \min \{ D_{ij} \} \quad (2.6)$$

本モデルは Counter Propagation と異なり、一回の学習過程において発火素子  $ij^*$  と、その近傍領域内の全素子が持つシナプス荷重を更新する。さらに Feature Map と異なり、一回の学習ごとに発火素子から入力層、および出力層に伸びた両方のシナプス荷重を同時に更新する。

入力層-競合層間において、近傍内素子が持つシナプス荷重  $W_{ij}$  は入力ベクトル  $X$  に近づくように式 2.7 に従い更新する。同様に競合層-出力層間では出力ベクトル  $Y'$  が教師ベクトルに近づくよう、近傍内素子の持つシナプス荷重  $V_{ij}$  を式 2.8 に従い更新する。

$$W_{ij}(t+1) = W_{ij}(t) + \eta(t) \cdot f(B(ij)) \cdot |X(t) - W_{ij}(t)| \quad (2.7)$$

$$V_{ij}(t+1) = V_{ij}(t) + \eta(t) \cdot f(B(ij)) \cdot |Y(t) - V_{ij}(t)| \quad (2.8)$$

for  $i, j$  Neighborhood area

$$f(x) = 1 - \frac{1}{1 + e^{((x-R)/2)/T}} \quad (2.9)$$

ここで  $B(ij)$  は近傍内素子  $ij$  と発火素子  $ij^*$  とのユークリッド距離を示す．また  $R$  は発火素子近傍領域の半径， $T$  は正の定数である．

本モデルは競合層で近傍関係を考慮することにより，未学習点についても位相構造を保持できる．また教師信号を用いた一般的な写像関係を学習でき，かつ出力層側に値を入力することで，出力層から入力層への逆方向写像を得る双方向連想記憶装置としても動作する．

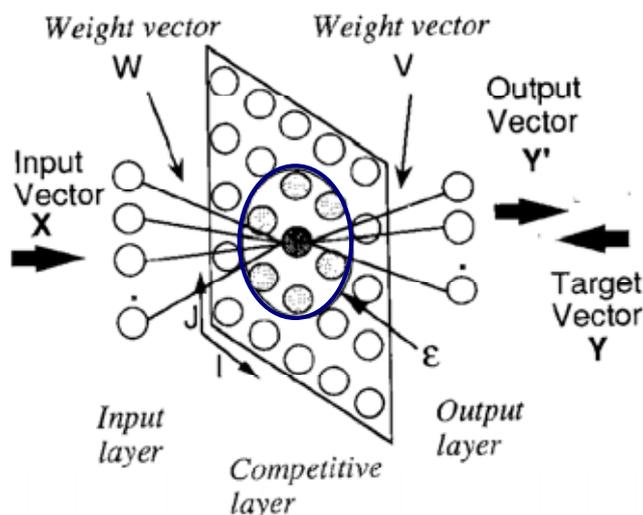


図 2.4 双方向 Feature Map モデル

### 2.2.3 双方向 Feature Map モデルの写像特性

以降，双方向 Feature Map の写像特性を確認するためのシミュレーション結果を記す．

#### (1) 位相情報の保存特性検証

15x15x15 個の三次元競合層を持つ 3 入力-2 出力のモデルを設定し，式 2.10 に示す写像関係を学習した．このとき，学習セット  $(X, Y)$  は式 2.11 に示す 216 個より選択し，20,000 回の学習を行った．式 2.11 において， $0.2i (i=0, 1, \dots, 5)$  は，0 から 1 までの値から 0.2 刻みの数値をランダムに選択することを意味する．なお，汎化能力を高めるために，入力最大値の 5% の一様ノイズを入力値に加えた．その結果を図 2.5 に示す．図中の黒四角形は学習された点の出力を，白四角形は未学習点の出力である．図から明らかなように未学習点にお

いてもほぼ正しい出力値が得られており、本モデルが幾何学的位相情報を保持するトポグラフィックマッピングを形成していることが分かる。

$$y_1 = \frac{1}{3} \sum_{i=1}^3 x_i, \quad y_2 = \prod_{i=1}^3 x_i \quad (2.10)$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.2i \ (i=0,1,\dots,5) \\ 0.2i \ (i=0,1,\dots,5) \\ 0.2i \ (i=0,1,\dots,5) \end{bmatrix} \quad (2.11)$$

### (2) 双方向性の検証

同じく 15x15x15 個の三次元競合層を持つ 3 入力-3 出力のモデルを設定し、 $Y = X^2$  の関係を学習させた。学習は 10,000 回実施した。その結果を図 2.6 に示す。図に示すように、順方向から逆方向、および逆方向から順方向の双方向写像を獲得できている。そのため本モデルが双方向連想記憶特性を持つことが分かる。なお、誤差量 *Error* は次式に従い計算した。

$$Error = \begin{cases} \sum_{i=1}^Q |Y_i - Y_i'| / (Q \cdot O) & \text{: 順方向} \\ \sum_{i=1}^Q |X_i - X_i'| / (Q \cdot O) & \text{: 逆方向} \end{cases} \quad (2.12)$$

ここで  $Q$  はデータの入力回数、 $O$  はネットワークの最大出力値を示す。

### (3) 誤差逆伝播法と、双方向 Feature Map との精度比較

学習点が少ない条件で、誤差逆伝播法を用いた階層型ニューラルネット（以下、BP と呼ぶ）と、双方向 Feature Map の写像精度を比較した。比較は中間層が各々 5 個、15 個、20 個の素子数を持つ三種類の BP と、20x20x20 素子の 3 次元競合層を持つ双方向 Feature Map で行った。学習させる写像関数は式 2.10 と同じものを用い、次式に示す計 27 個の入力ベクトルを学習に用いた。

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.5i \ (i=0,1,2) \\ 0.5i \ (i=0,1,2) \\ 0.5i \ (i=0,1,2) \end{bmatrix} \quad (2.13)$$

学習は、学習点に関する誤差量 *Error* の平均値が 0.01 以下になるまで実施した。この場合、BP は 10,000 回、双方向 Feature Map は 8,500 回の学習が行われた。学習後の BP (中間層の素子数 5 個) と双方向 Feature Map に、ベクトルの各要素が同じ値を持つ入力ベクトルを連続的に入力した結果、図 2.7 に示す写像関係を得た。ここで図中の黒点が学習点を、

白点が双方向 Feature Map の出力値を表す．図からわかるように，BP は双方向 Feature Map に比べて実際の写像とのずれが大きくなっている．これは表 2.2 に示すように，中間層の素子数を多くした場合も同様である．このことから双方向 Feature Map は未学習点に対する性能が BP より高く，かつ大きく誤った出力値を出さない．そのため制御を行いながらの学習に適していることが分かる．

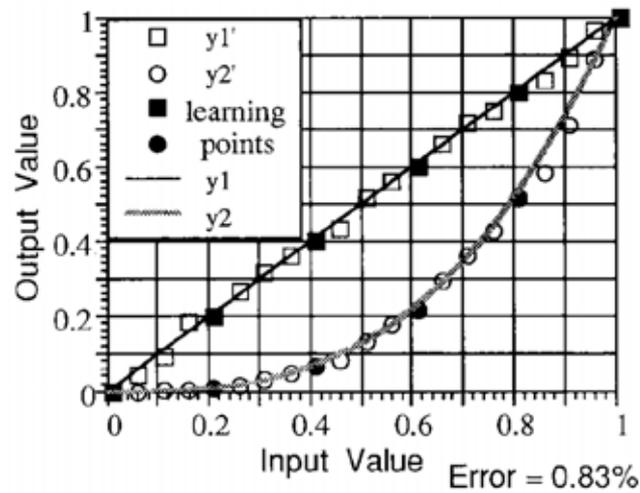


図 2.5 位相情報の保存特性

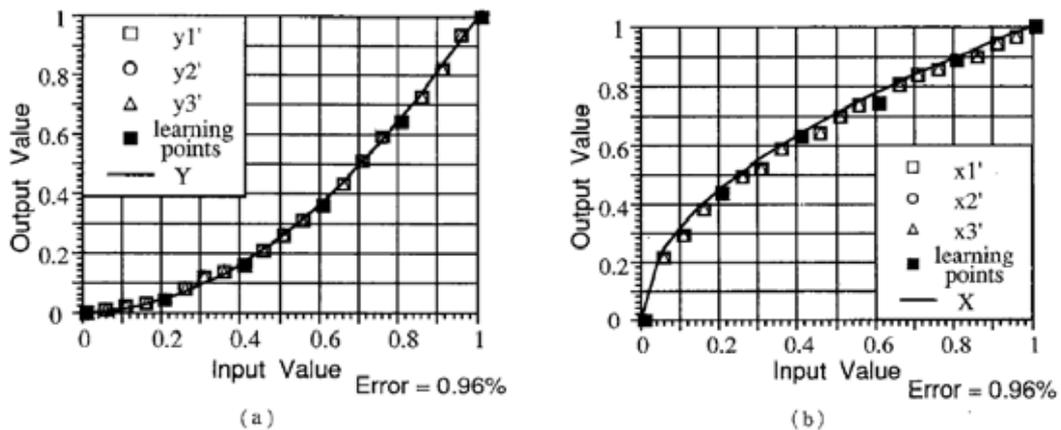


図 2.6 双方向性の検証 (a)順方向写像特性，(b)逆方向写像特性

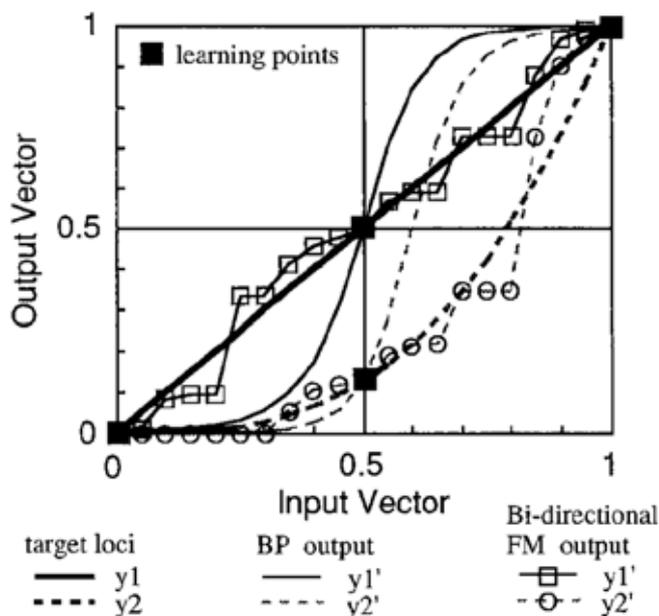


図 2.7 少ない学習点により得られた写像関係 (BP と双方向 Feature Map との比較)

表 2.2 BP と双方向 Feature Map との比較

ニューラルネット	誤差の平均値			
	中間層素子数	学習点	未学習点	全体(学習点 + 未学習点)
BP	5	0.009	0.253	0.218
	15	0.005	0.235	0.202
	25	0.002	0.244	0.209
双方向 Feature Map		0	0.084	0.071

## 2.2.4 アーム制御への適用

アーム制御機構は図 2.8 に示す階層構造を成す 3 個のネットワークで構成する。それは運動の目標から空間座標中の軌道を決定する第 1 のネットワーク，空間座標での軌道を身体座標の軌道に変換する第 2 のネットワーク，および軌道を実現するためのトルクを得る第 3 のネットワークである。

第 1 のネットワークには Flash と Hogan による加速度変化率最小モデル[2-10]を用いて軌道を決定する。この軌道から時刻  $t$  でのアーム先端の空間座標( $X, Y, Z$ )を得ることができる。第 2 のネットワークは双方向 Feature Map で構成し，アーム先端の空間座標から，アーム身体座標（関節角）への写像変換を行う。これにより第 2 のネットワークは目的とするアーム先端の空間座標を実現するための最適な身体座標（関節角）を出力する。最後に第 3 のネットワークは，その身体座標を実現する制御信号を出力する。ここで，第 2 にネットワークでは，アーム先端が特定の空間座標を指し示す場合，それを実現する関節角の

組み合わせが複数存在する問題（逆運動学での不適切問題）が起こりえる。

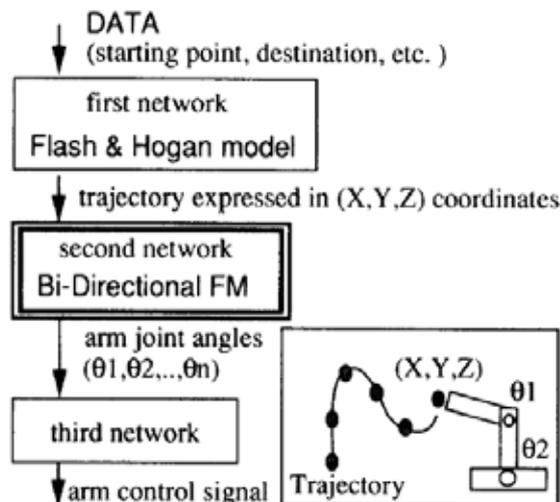


図 2.8 双方向 Feature Map を用いたアーム制御機構

(1)逆運動学での不適切問題への対応

逆運動学での不適切問題が発生すると，双方向 Feature Map において競合層内の複数個の素子が発火する．例えば，図 2.9 は 3 つの関節を持つアームを制御する場合の競合層の状態を示す．図に示すように，ある空間座標を指定した場合，No.1 ~ 11 までの複数個の関節角の候補が存在する．本問題への対策として，ここでは素子の競合活動に制約条件を導入し，複数の発火素子中から最適な素子を選択することで，本問題を回避する．

ここでは，制約条件として式 2.14 に示す擬似エネルギー<sup>3</sup>  $E$  を定義し，この値が最小となる素子を選択する．ここで  $E_d$  はアームを移動するために必要な動的擬似エネルギー量を， $E_s$  は現在の姿勢を保つために必要な静的擬似エネルギー量を表す．静的擬似エネルギー量は，生物が，その姿勢を取る際の苦痛度合いを表す．

$$\Delta E = E_d - \kappa(E_d) \cdot (1 - E_s) \tag{2.14}$$

$$E_d = \frac{1}{N} \sum_{i=1}^N f_i(\theta(t)) \cdot (\theta_i(t+1) - \theta_i(t))^2 \tag{2.15}$$

$$E_s = \frac{1}{N} \sum_{i=1}^N (\theta_i(t+1) - \theta_i^{opt})^2 \tag{2.16}$$

$$\kappa(x) = \cdot(1 - x) \quad 0 \leq x \leq 1 \tag{2.17}$$

<sup>3</sup>運動に必要なエネルギー量だけでなく，生物における姿勢の苦痛度合いもあらかずものであり，物理的なエネルギーの次元とは異なる．

ここで  $N$  はアームの関節数，  $(t)$  は時刻  $t$  におけるアームの関節角である．関数  $f$  は関節駆動時の負荷量を示すものであり 腕の長さや質量，慣性モーメントなどの影響を受け，関節毎に異なる．また，  $i^{opt}$  はアームが安定姿勢を取る場合の関節角を表す．

動的擬似エネルギー量を用いたアーム制御の制約は，宇野[2-11]や Jordan[2-4]が定義した「滑らかさ」を表す制約条件に近い．式 2.14 では  $(x)$  の影響により，動的擬似エネルギー量が小さい場合（アーム先端が目的位置に到達した場合など）に静的擬似エネルギー量を強く考慮し，動的擬似エネルギー量が大きい場合は静的擬似エネルギー量をあまり考慮しないで発火素子を選択する．これにより，単位時間あたりのアーム移動距離が大きい場合は動的擬似エネルギーが強く考慮され 移動に要する動的擬似エネルギーが少ない関節角が選択される．また，アームの先端位置が固定された場合においても，その姿勢を保つために必要な静的擬似エネルギー量が少なくなるように関節角を変化させる．

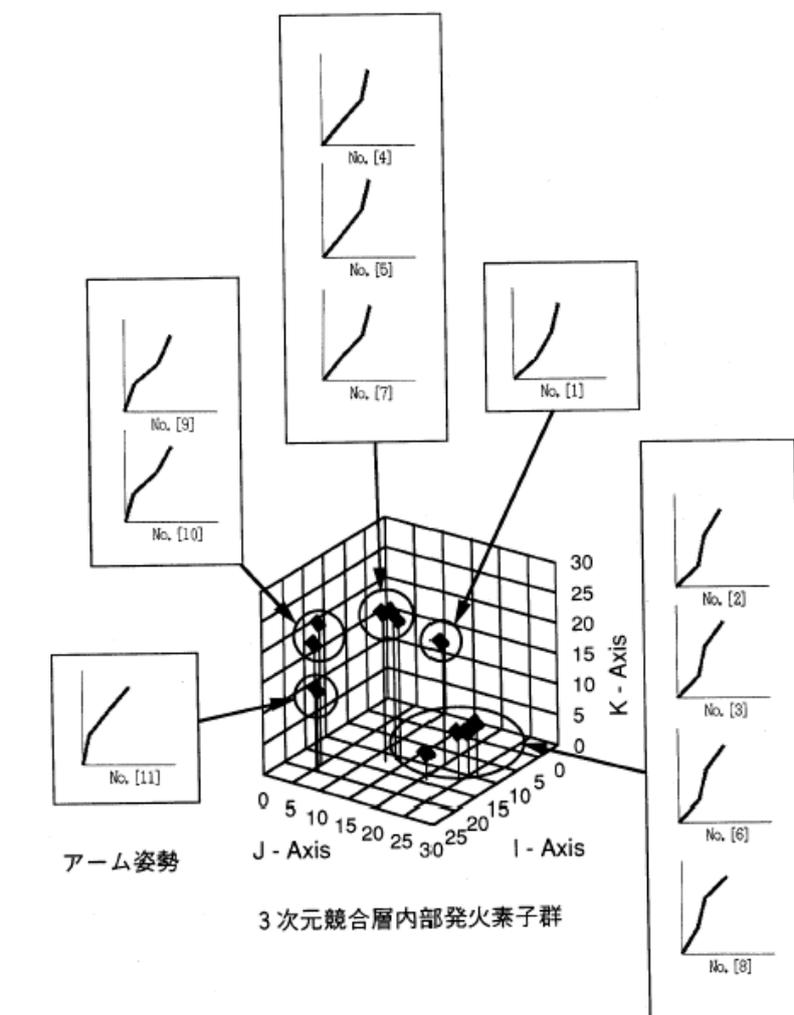


図 2.9 ある空間座標を実現するアーム関節角度の候補

(2)アーム制御のシミュレーション結果

ここでは図 2.10 に示す 3 つの関節を持つアームモデルを定義してシミュレーションを行う．使用した双方向 Feature Map は，競合層の素子数 27,000 個 (30x30x30) からなる 3 入力-2 出力のモデルとした．

学習過程：本モデルへの入力信号はアームの関節角( $\theta_1, \theta_2, \theta_3$ )とし，教師信号として関節角に対応したアーム先端座標( $x, y$ )を与えた．学習は，これらの教師信号をランダムに入力し，80,000 回の学習を行った．

制御過程：アームの先端を初期位置 A 点[550,150]から，目標位置 B 点[350,350]まで直線的に等速度で移動する軌道を設定し，各時間におけるアーム関節角を求めた．ここでは簡単のため関節角の負荷関数  $f$  はアームの姿勢によらず値 1 と仮定した．また図 2.11 に示す状態をアームの安定姿勢  $i^{opt}$  と定義した．その結果を図 2.12(a)に示す．図からわかるようにアームは滑らかにその姿勢を変化させた．一方，アームの先端位置を C 点[400, 400]に固定した場合の結果を図 2.12(b)に示す．ここでは静的擬似エネルギー  $E_s$  の影響を受けるため，アームの先端位置を固定したまま，安定状態に近づくようにその姿勢が変化した．そして  $t=4$  において，最も安定した姿勢で収束した．

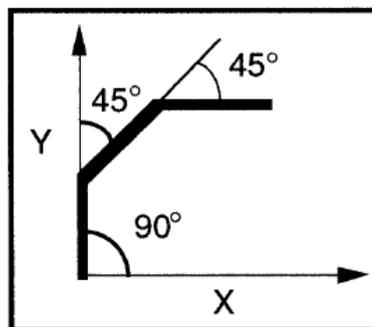
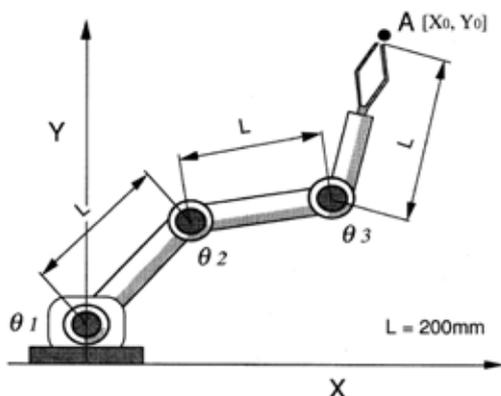


図 2.10 シミュレーションに用いたアームモデル

図 2.11 アームの安定姿勢

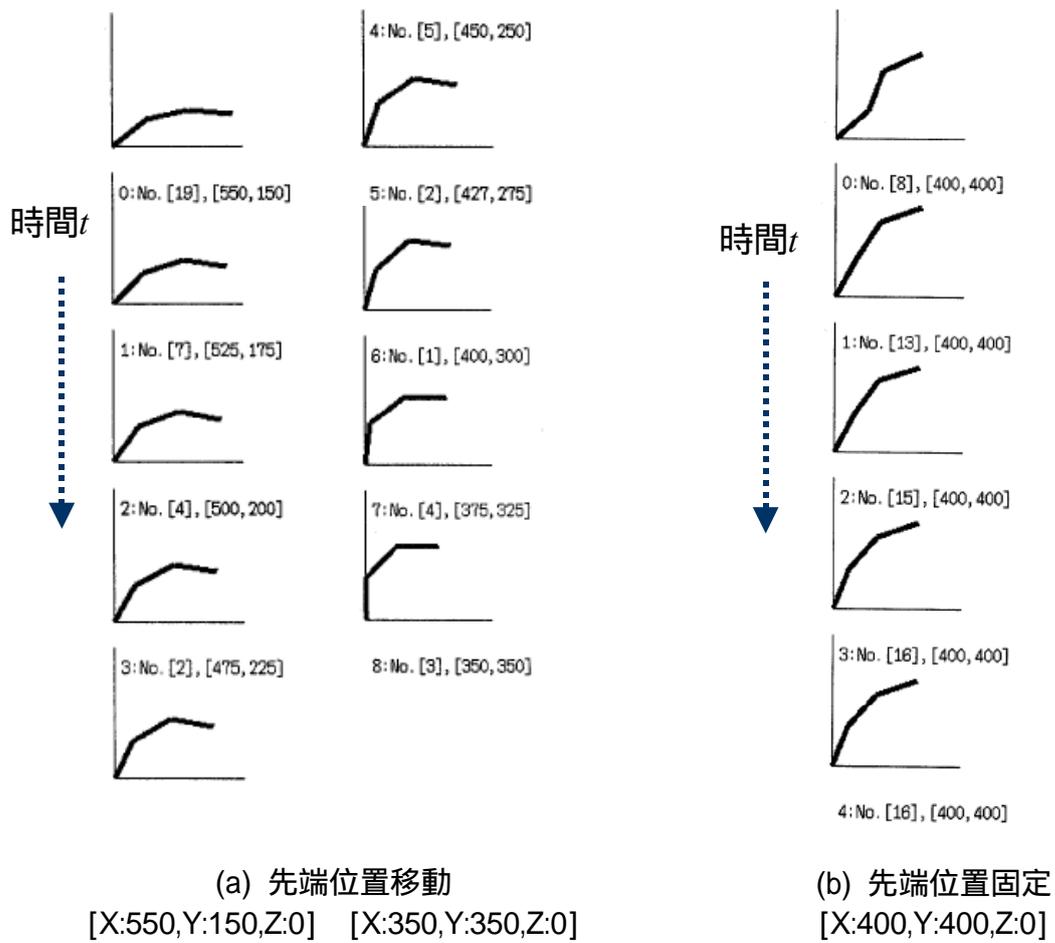


図 2.12 アーム制御のシミュレーション結果

## 2.3 適応的な自己増殖による離散的写像への対策

前記の双方向 Feature Map は写像が離散的であり、その空間分解能が競合層の素子数により決まる課題がある。Feature Map は脳の大脳皮質の構造を模擬しているが、大脳皮質の体性感覚野では、口や指先といった人間の生活環境のなかで使用頻度が高く、重要な部位の皮質が大きな面積を示す(図 1.12)。大脳皮質の面積は素子数の多さに関係することから、この皮質領域には、高い情報分解能を得るために他領域よりも多数の素子が割り当てられていることが分かる。このような素子数の差は、具現化された脳の進化の一例であり、外界からの情報入力確率に従い、情報処理精度を向上させるような外界への適応が生じたと考えられる。ここでは双方向 Feature Map のモデルを拡張して、外界からの情報入力頻度に応じて素子数を適応的に増殖させる機能を追加し、必要な部分の空間分解能を高める拡張を実施した。

### (1) 適応的な増殖方式

競合層の各素子は、その内部メモリに、増殖のための指標である Split Level を持つ。学習過程において、Split Level が臨界値を越えた場合に素子が増殖する(図 2.13)。増殖した素子は、親の責任部位を受け継いだ複数の素子から成る下位層を形成する。下位層は、親の素子をゲートとして入力信号を受ける小さな双方向 Feature Map と考えられる。増殖により形成される下位層内の素子数は一定であり、その特性と内部メモリは親と同じである。

ここで Split Level には、ある一定学習回数中に、その素子が発火した回数を用いる。これにより高い確率で入力信号を受ける素子が増殖する。そのため、入力頻度が高い空間に対する分解能を効率的に向上させて、写像が離散的である問題を抑制できる。

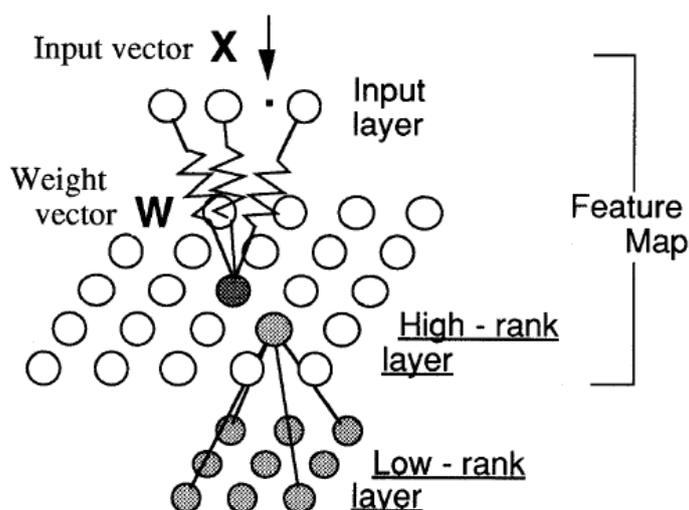


図 2.13 素子の増殖例

双方向 Feature Map の競合層内素子は内部メモリを持つ(図 2.14). ここにはシナプス荷重ベクトル  $W$  と[Split Level]の他に, その素子の誕生日[Birthday], 年齢[Age]を保持する. 学習初期段階では  $W$  にランダムな値を与え, その他の値は 0 に初期化する.

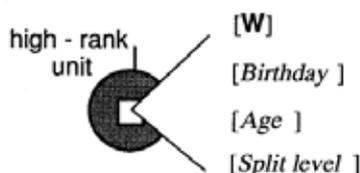


図 2.14 競合層素子の内部メモリ

親の素子が増殖した場合, 増殖直後の下位層素子は図 2.15 に示す内部メモリを持つ.

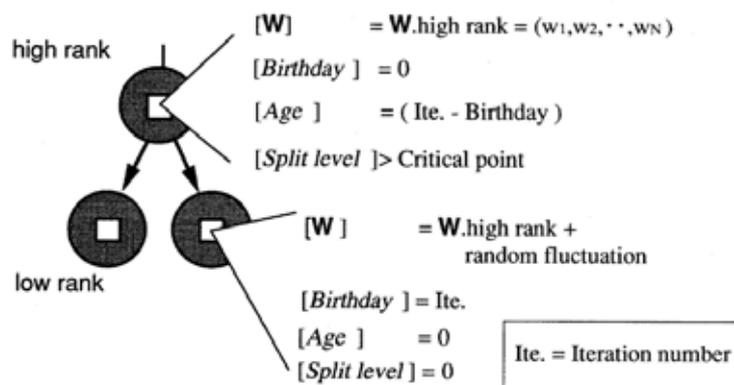


図 2.15 増殖直後の競合層素子の内部メモリ

## (2) 学習アルゴリズムの拡張

発火素子の探索では, まず上位層内で発火素子を探索する. 発火素子が下位層を持たない増殖していない素子である場合は, これを発火素子とする. これは従来の双方向 Feature Map と同じ処理である. 上位層の発火素子が親素子であり下位層を持つ場合, 親素子は下位層へと入力信号を伝播する. この親素子に結合された下位層内において双方向 Feature Map と同じ発火素子の探索処理を行い, 下位層内での発火素子を最終的な発火素子とする. ここで, 発火素子が見つかった後のシナプス荷重の変更は以下の手順で行う.

### 発火素子が上位層の素子である場合

学習は式 2.18 に示す Kohonen 則に従い行う. 図 2.16 の網掛け部分が, シナプス荷重を変更する近傍領域を示す. もし近傍領域内に下位層を持つ親素子が存在するなら, 下位層内素子が持つシナプス荷重も式 2.19 に従い更新する. これは下位層内素子群が親素子のシナプス荷重更新に引きずられて矢印方向に移動するシフト効果を生む.

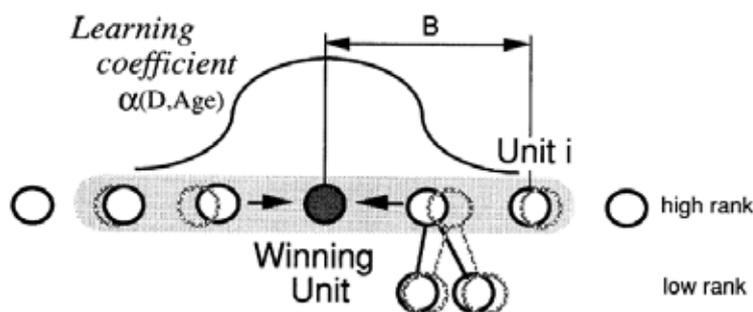


図 2.16 発火素子が下位層を持たない場合の近傍関係

$$W_{highrank}^{t+1} = W_{highrank}^t + \alpha(Age) \cdot |X - W_{highrank}^t| \quad (2.18)$$

$$W_{lowrank}^{t+1} = W_{lowrank}^t + \beta \cdot \alpha(Age) \cdot |X - W_{highrank}^t| \quad (2.19)$$

ここで  $W_{highrank}$  と  $W_{lowrank}$  は上位素子と下位素子の持つシナプス荷重を、 $X$  が入力ベクトルを、 $\beta$  は  $0 < \beta < 1$  の定数を示す。双方向 Feature Map では近傍領域の半径  $B$  を固定とし、学習係数  $\alpha$  は学習回数関数の関数としたが、増殖を行うモデルでは発火素子の年齢  $Age$  を考慮して次式に従い決める。

$$\alpha = 1 - \frac{1}{1 + \exp\{(0.5M - Age)/M\}} \quad (2.20)$$

$$B = \begin{cases} 3: Age < 0.15M \\ 2: 0.15M \leq Age < 0.3M \\ 1: 0.3M \leq Age \end{cases} \quad (2.21)$$

ここで  $M$  は素子の成熟度(Maturity)を意味する固定値である。学習係数  $\alpha$  は年齢の増加に従い低下して、年齢が成熟度を超えると非常に小さい値となる。なお、上位素子の年齢  $Age$  は学習回数に比例して増加するが、下位層素子の年齢  $Age$  は上位素子の発火回数に比例して増加させる。

#### 発火素子が下位層の素子である場合

発火した上位素子は式 2.18 に従い自分のシナプス荷重を更新し、かつ下位素子の発火素子と、その近傍領域内の素子が持つシナプス荷重を式 2.22 に従い更新する(図 2.17)。

$$W_{lowrank}^{t+1} = W_{lowrank}^t + \beta \cdot \alpha(Age) \cdot |X - W_{lowrank}^t| \quad (2.22)$$

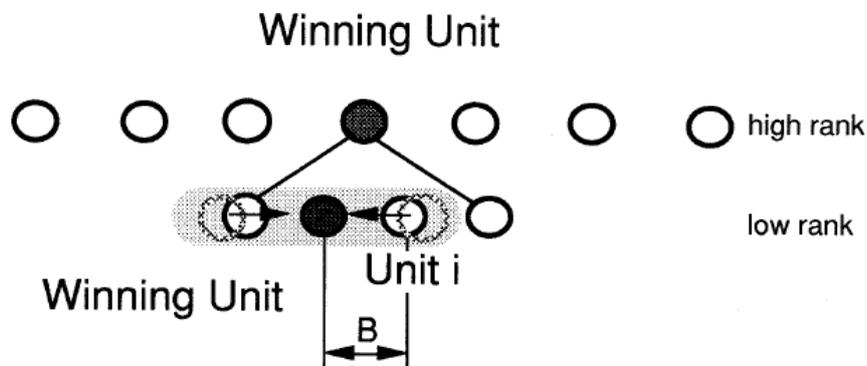


図 2.17 発火素子が下位層である場合の近傍関係

### (3) 適応的な増殖のシミュレーション

情報空間上で確率分布が偏在している入力ベクトルを、その確率分布に従って自己増殖機能を持つよう拡張した双方向 Feature Map に入力した。学習すべき写像関係は、式 2.23 とし、入力ベクトル空間 $(x_0, x_1)$ に対する情報入力確率（頻度）の分布は、S 字上の領域に対する情報入力確率を高くした（図 2.18）。

$$y_0 = \prod_{i=0}^1 x_i \quad , \quad y_1 = x_0 \quad (2.23)$$

ここで、入力ベクトルの要素  $x_i = 0 \sim 1$  ( $i = 0, 1$ ) とする。

実験には、表 2.3 に示すような自己増殖機能を持つ双方向 Feature Map モデルを用いた

図 2.19 は、この自己増殖機能を持つ双方向 Feature Map における、競合層の素子が持つシナプス荷重の状態を示す。図からわかるように、学習が進むことで情報入力頻度の高い領域において自己増殖が発生した。学習後の平均二乗誤差は 0.031698 となった。これは同一条件で学習を行った自己増殖機能を持たない双方向 Feature Map の平均二乗誤差 0.047297 よりも小さい。このことから、自己増殖機能を持つことで、外界に対して適応的に必要な領域に対する空間分解能が向上し、写像精度が向上することが確認できた。

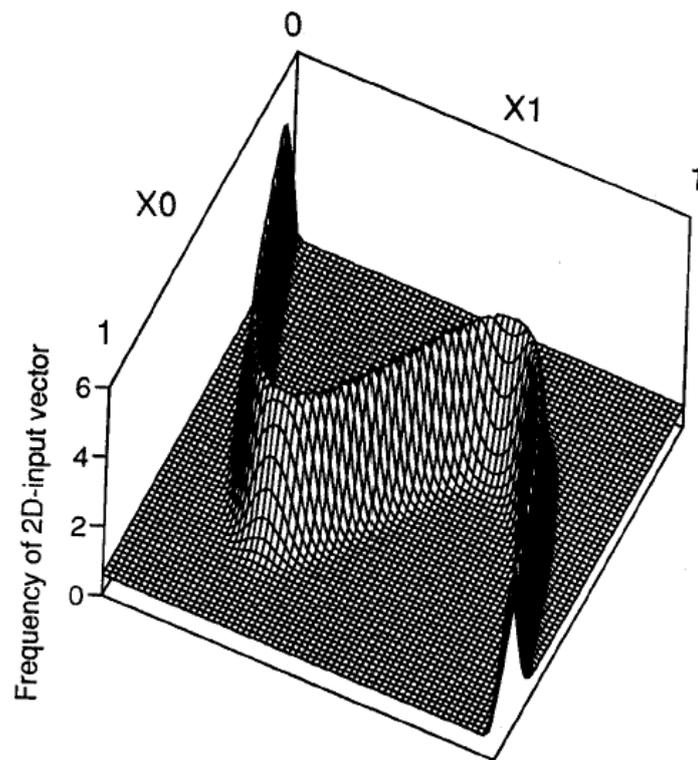


図 2.18 入力ベクトル空間( $x_0, x_1$ )に対する空間情報入力確率

表 2.3 実験に用いた自己増殖機能を持つ双方向 Feature Map モデル

ニューラルネットワークモデル	2 入力-2 出力の双方向 Feature Map 自己増殖機能有り
上位層素子数	64 個(8x8)
下位層素子数	25 個(5x5) 一つの親素子に結合する素子数
成熟年齢	25000
学習回数	最大 70000

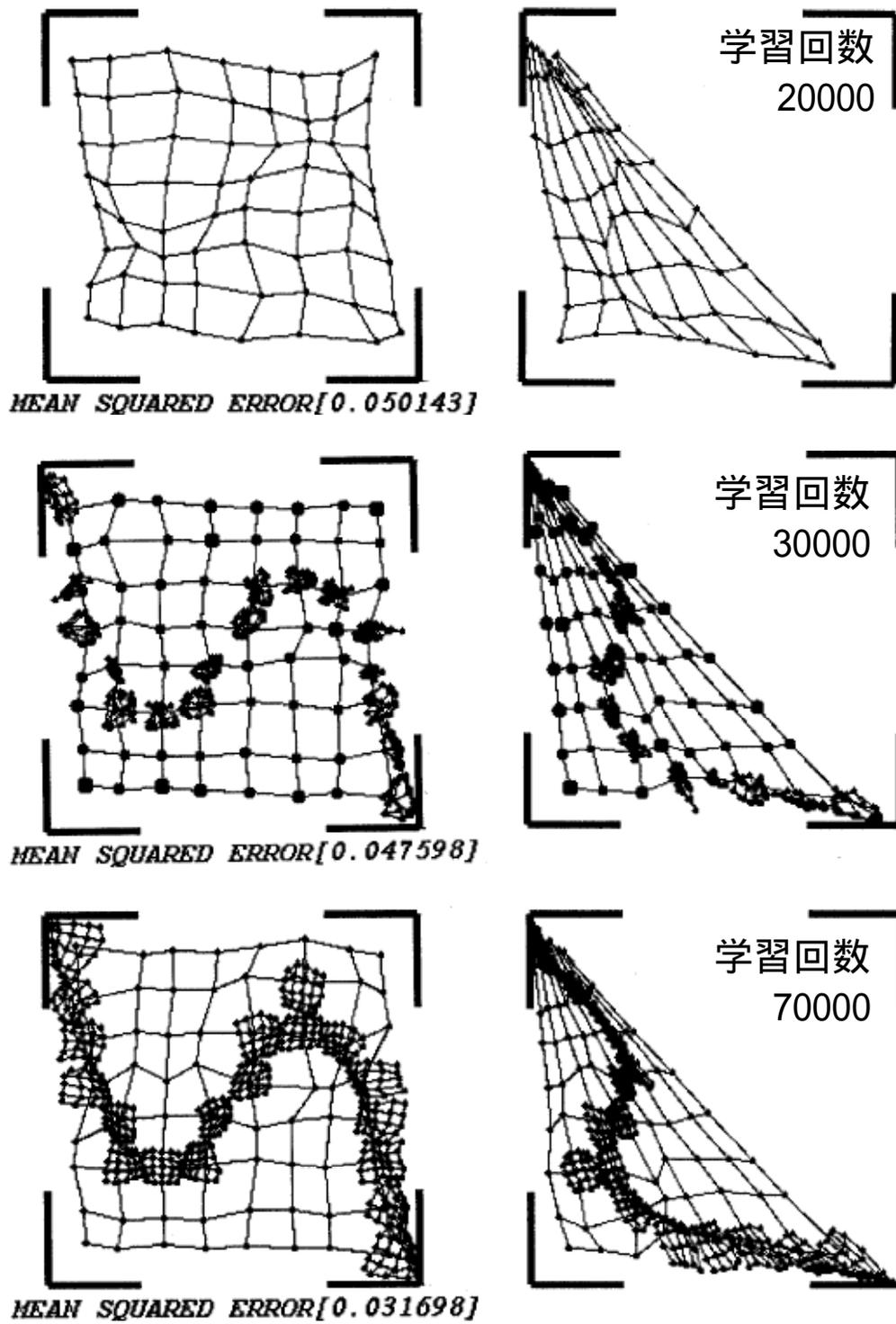


図 2.19 自己増殖機能を持つ双方向 Feature Map の学習過程におけるシナプス荷重状態

## 2.4 第 2 章のまとめ

本章では、集合論的写像の一つとして、ニューラルネット「双方向 Feature Map」を提案し、ロボットのアーム先端の空間座標とアームの身体座標間の写像獲得に適用した。シミュレーションにより、双方向 Feature Map モデルが異なる次元間の双方向写像を学習的に獲得できることを示した。また、幾何学的写像関係を保持する特長により、従来の誤差逆伝播法を用いたニューラルネットに比べて、未学習点における誤差が小さく、不正な出力信号を出力しないことを示した。これにより、生物の随意運動を模したアーム制御に必要とされる「双方向性を持つ写像の学習的構築」と「制御しながら学習が可能である」の条件を満たすことを確認できた。

ただし、本モデルは、写像が離散的であり空間分解能が競合層の素子数に依存する課題が存在した。これに対して、生物が進化の過程で外界に適応的するのと同様に、情報入力頻度の高い素子を増加させる自己増殖機能を検討した。この自己増殖機能を用いることにより、情報入力頻度が高い領域に対して空間分解能が増加し、写像の離散性の問題を抑制できることを実験により示した。

# 第3章 形状モデルへの写像獲得と FAX-OCR システムへの適用

人間は画像中から特定の形状を探し出す際、画像中に存在する形状と、脳内に存在する形状モデルとの形状照合を行っていると考えられる。本章では、人間の形状照合時に実施しているであろう、画像と形状モデルとの位置合わせを行う幾何学的な写像変換の獲得処理をモデル化する。具体的には、FAX 送信により変形した帳票画像と帳票モデルとを照合して、帳票画像に加えられた位置ずれや回転、拡大縮小などのアフィン変形量をロバストに推定するアルゴリズムを検討する(表 3.1)。以降、3.1 節で FAX-OCR システムの背景と目的を示し、3.2 節で帳票画像と帳票モデル間のロバストな写像推定方式を述べる。次に 3.3 節では実験結果を示し、3.4 節で製品適用事例を述べる。最後に 3.5 節でまとめとする。

表 3.1 形状モデルへの写像

写像関係	変形した帳票画像と、帳票の形状モデル間の写像。
適応性	-
写像の種類	帳票に加わる変形がアフィン変換であると仮定した、幾何学的な位相関係を保つ連続写像
脳内の関係部位	感覚野/視覚野/頭頂連合野
人間支援システム	FAX 送信された帳票中の文字を自動的に読取る FAX-OCR システム(製品)に適用

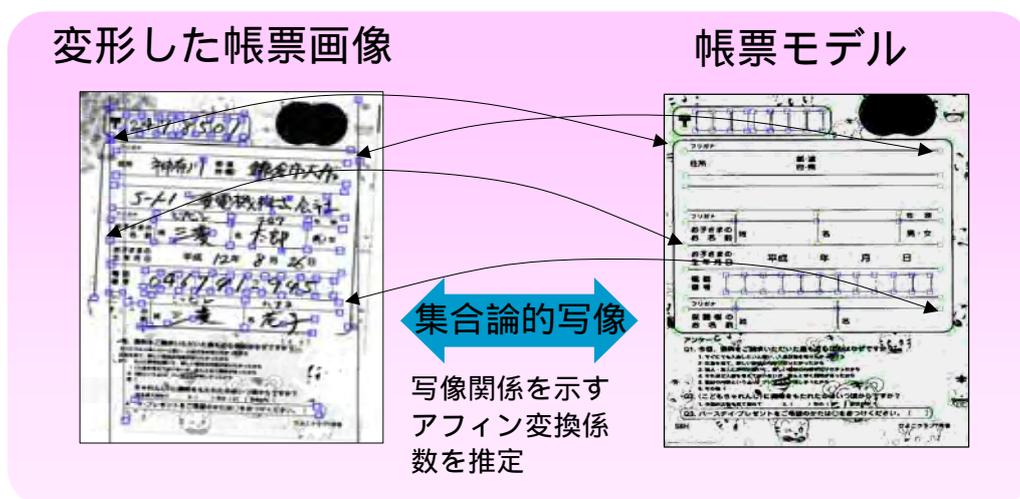


図 3.1 形状モデルへの写像獲得方法

### 3.1 FAX-OCR システムの背景・目的

FAX送信された帳票上の手書き文字を認識するFAX-OCR装置は、帳票モデルとの照合により帳票画像上から文字認識する領域（以降、フィールドと呼ぶ）を抽出し、そのフィールド内に記入された文字を認識する。しかしFAX送信された帳票画像には平行移動、回転、拡大、縮小等の変形が加わる。またFAXに特有なノイズ（ゴマ塩状のノイズ、カスレ、縦方向の黒線等）が重畳される（図3.2）。

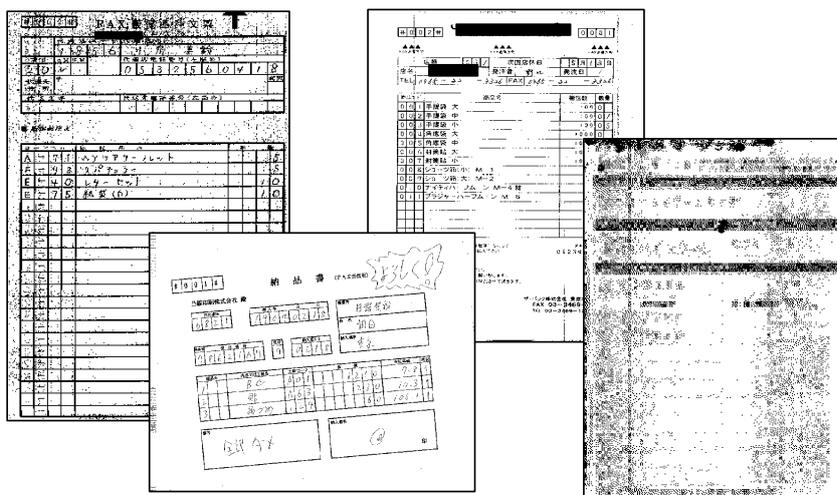


図3.2 変形やノイズが重畳された文書画像

このような帳票画像の劣化により、FAX送信された帳票画像からフィールドを正確に抽出することは困難である。そのため、従来はフィールドの抽出が容易になるように設計した専用帳票が用いられた。この専用帳票は2種類に大別できる。一つは機械が判別可能なマークを帳票の隅に印字した専用帳票（図3.3a）である。この帳票では、帳票画像から検出したマークの位置から帳票画像の変形量を推定し、FAXで検知できない薄い色（ドロップアウトカラー）で印刷されたフィールドの位置を特定する[3-1]。もう一つは表形式の専用帳票（図3.3b）である。この専用帳票を読み取る方式の多くは、帳票画像から抽出した表罫線と帳票モデルが持つ表罫線との対応付けに基づいて帳票画像の変形量を推定する[3-2],[3-3],[3-4]。著者らの研究でも、以前に同様な方式を検討して良好な結果を得た[3-5]。この形式の専用帳票は見た目が一般帳票に近いが、表罫線の対応付けを容易にするための制約を持つ。例えば文献[3-3],[3-5]では、外枠は矩形の表であり、かつ表罫線の両端は直交する表罫線に接触する必要がある。また文献[3-4],[3-6]では、線幅が太い表罫線を一定数以上必要とする。

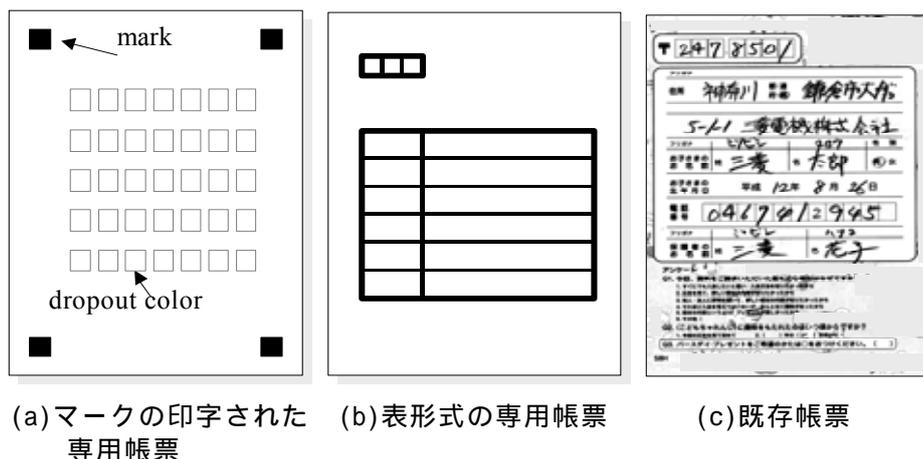


図3.3 帳票画像の例

これに対して、スキャナで取得した帳票画像を読み取る帳票OCRでは、帳票画像の劣化が少ないために帳票形式の制限が緩やかである。例えば、文献[3-7]や[3-8]は四方が表罫線で囲まれた枠を探して、その大きさと位置からフィールドを得る。そのためフィールドが表罫線で囲まれていれば抽出可能である。また文献[3-9],[3-10],[3-11]は表罫線の対応付けに基づいてフィールドを抽出するが、表の外形や表罫線の接続関係に対する制限はない。更に、文献[3-6]はプレプリント文字の対応付けを行うため表罫線の無い帳票も許容する。

しかし、FAX送信された帳票画像は劣化が激しいため、これらの方式をそのまま適用することは困難である。例えば、表罫線の対応付けによる手法では、ノイズで表罫線が切断されると対応付けを誤る場合がある。

本研究では、これらの方式を拡張することで、FAX送信された多様な形式を持つ一般帳票（例えば図3.3c）にも対応したフィールド抽出方式を提案する。本方式は帳票画像から表罫線の端点や交点、プロブ（黒く塗りつぶされた矩形領域）を特徴点として抽出し、これを帳票モデルの特徴点と対応付けることで帳票画像の変形量を推定する。このような特徴点は人間の視覚野で抽出されているものと同様であり、表や文字列、イラスト、下線等から抽出できる。そのため多様な形式を持つ一般帳票の多くに適用できる。また、特徴点は表罫線に比べてFAX送信によるノイズの影響を受け難い。例えば、表罫線の途中が大きく切れても端点は影響を受けない。加えて、本方式は帳票画像の劣化による特徴点の消失や、余計な特徴点の湧き出しに対してロバストに変形量を推定する。そのためノイズ等による帳票画像の劣化に強い特性を持つ。

以下では、本方式の処理内容を詳しく述べ、実験によりその有効性を検証する。

## 3.2 帳票モデルと帳票画像間の写像推定方式

提案方式の処理フローを図 3.4 に示す．本方式は帳票モデル作成処理と，帳票読取処理から成る．帳票モデル作成処理は，未記入の帳票画像から帳票モデルを作成する．また帳票読取処理は，帳票モデルを参照して入力された帳票画像を読み取る．以降，その処理内容を順に説明する．

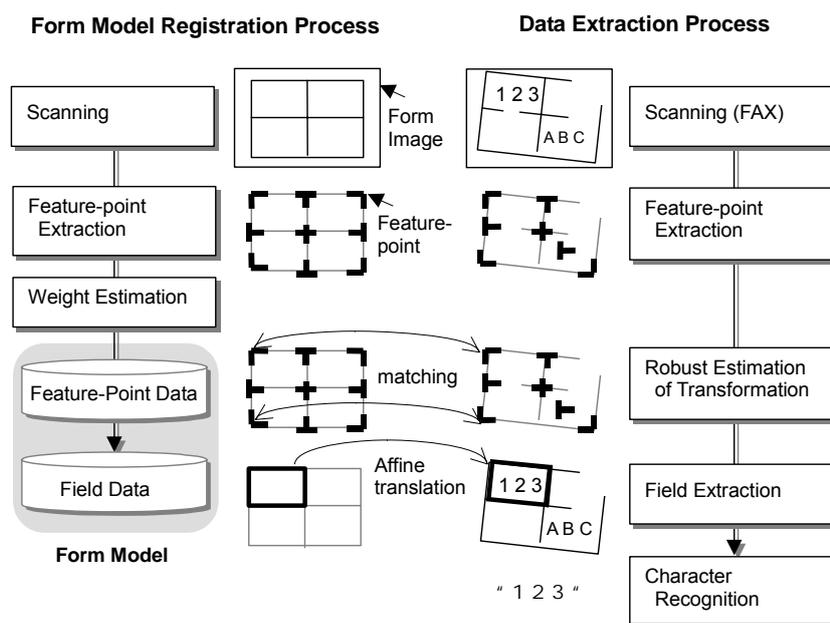


図 3.4 提案方式の処理

### 3.2.1 帳票モデル作成処理

帳票モデルは，フィールドと特徴点のデータで構成する．このフィールドのデータは，未記入の帳票画像上におけるフィールドの位置( 矩形領域の左上点座標と幅高さ )である．特徴点のデータは，帳票画像から抽出した特徴点の種類と位置，および重みの値である．この重みの値は各特徴点の重要性を示す．

#### (1) 特徴点の種類と位置

抽出する特徴点の種類は図 3.5 に示す 14 種類である．この特徴点は人間の視覚野で抽出されている特徴と同等である．図 3.6 は実際の一般帳票から抽出した特徴点の例であり，図中の小さな矩形が特徴点を示す．ここでは表罫線や下線，文字列から多数の特徴点が抽出されていることが分かる．文字列上で抽出された特徴点の多くは，文字列の一部を破線として抽出した際の端点である．これらの特徴点は，以下の手順で抽出する．

- (1) 帳票画像を  $n \times n$  画素の小領域に分割し、各小領域が持つ方向性を調べる。具体的には、小領域内で黒画素が横に  $n$  画素連続する場合は横方向性、縦に連続する場合は縦方向性を持つとする。小領域の大きさは経験的に紙上の  $0.5 \times 0.5\text{mm}$  程度とする。FAX 送信された画像では紙上の  $1\text{mm}$  が約 8 画素に相当するため、ここでは  $n=4$  とした。
- (2) 横方向性を持つ小領域が、横に  $N$  個以上連続した領域をラスタとする。この際、ノイズで分断された罫線や破線も抽出するために短い不連続は埋め込む。
- (3) ラスタに対して 4 近傍ラベリングを行い、得られたラスタの集合を横罫線とする（縦罫線も同様に抽出）。
- (4) 抽出した縦・横罫線の端点と交点を求め、これを特徴点とする。ここでは罫線の交差状態等から、特徴点を端点 4 種類、交点 9 種類に分類する（図 3.5）。
- (5) 最後に、方向性を持つ小領域に対して 4 近傍ラベリングを行い、黒画素の濃度が高く、幅高さが閾値より小さなラベル領域をプロブと判定する。そしてプロブの左上点を特徴点とする。

ここで、罫線の端点や交点は、帳票に含まれる表や下線を変形量推定の基準として利用するために特徴点として採用した。プロブから得た特徴点は、帳票に含まれる文字やイラストからも特徴点を抽出して、罫線から得た特徴点を補足すると共に、罫線を含まない帳票もできるだけ処理可能にすることを目的に採用した。例えば、図 3.3a に示す罫線の無い専用帳票でも、マークをプロブとして検出することで読み取りできる。

## (2) 特徴点の重み

前記手段で抽出した特徴点には、変形した帳票画像からも安定的に抽出されるものや、逆に抽出され難いものがある。そこで、各特徴点の抽出安定性を求め、これから各特徴点の重要性を示す重みの値を得る。具体的な重みの算出手段を以下に示す。

- (1) 未記入の帳票画像にアフィン変換を加え、4 枚の変形画像を作成する（2%拡大，2%縮小，1 度右回転，1 度左回転）。
- (2) 変形画像から特徴点を抽出する。
- (3) 未記入の帳票画像から抽出した特徴点  $f$  の座標値  $(X_f, Y_f)$  を、変形画像の作成に用いたアフィン変換により座標値  $(X'_f, Y'_f)$  へ写像する。
- (4) 変形画像上の座標値  $(X'_f, Y'_f)$  の近傍に特徴点が存在する場合、特徴点  $f$  が変形画像からも抽出できたと判断する。
- (5) 特徴点  $f$  の重み  $W_f$  = 特徴点  $f$  が存在した変形画像数/4 とする。これにより、 $W_f$  の値は特徴点  $f$  の抽出安定性が高い場合に 1、低い場合に 0 となる。

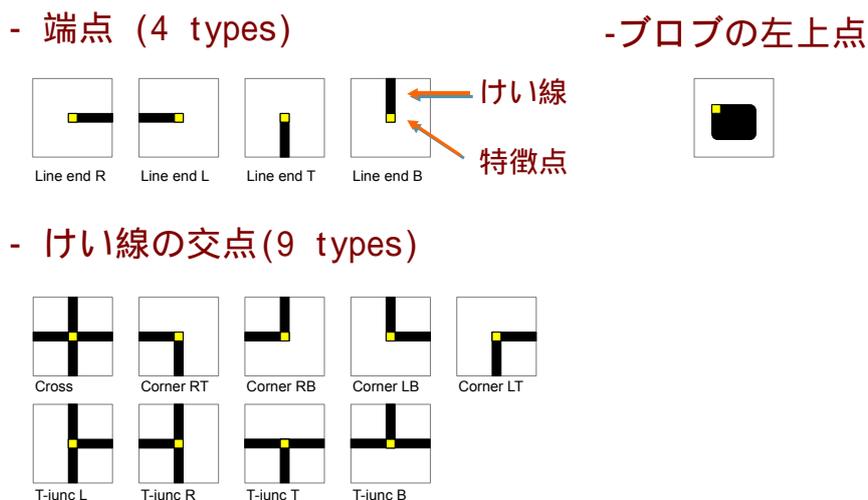


図 3.5 特徴点の種類 (端点 4, 交点 9, プロブ 1 種類)

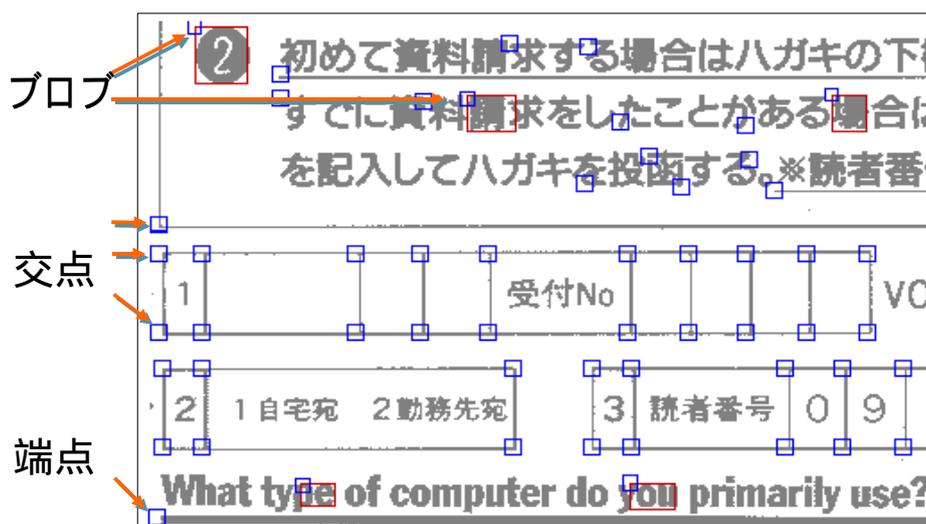


図 3.6 一般帳票からの特徴点の抽出例

### 3.2.2 帳票読み取り処理

FAX 送信された帳票画像を読み取る際は、まず帳票画像から特徴点を抽出する。特徴点の抽出処理は帳票モデル作成時と同じである。次に、抽出した特徴点と、帳票モデルの特徴点との対応付けに基づいて帳票画像の変形量を推定する(詳細は次節で述べる)。この変形量は式 3.1 のアフィン変換で表現する。

$$\begin{pmatrix} X'_f \\ Y'_f \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_f \\ Y_f \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (3.1)$$

ここで、 $X_f, Y_f$ は帳票モデルの座標点、 $X'_f, Y'_f$ はこの座標点に対応した帳票画像の座標点である。アフィン変換を用いることで、FAX に特有な、平行移動、回転、拡大、縮小等の複合した変形を表現できる。

最後に、推定したアフィン変換係数を用いて、帳票モデルが持つフィールドの位置を帳票画像上に写像し、そのフィールド内に記入された文字を認識する。すべてのフィールドに対して認識を行った後、処理を終える。

### 3.2.3 ロバストな変形量の推定

本方式の特長の一つは、劣化した帳票画像から高精度に帳票画像の変形量（アフィン変換係数）を推定する点である。これは、下記のコンセプトにより実現した。

- (a) 特徴点の多対対応を防ぐ対応付け評価指標の適用と、特徴点の重みを考慮した特徴点の選択により、特徴点の対応付け誤りを抑制する。
- (b) 特徴点の対応付け誤りがあっても、ロバストにアフィン変換係数を推定する。
- (c) 推定したアフィン変換係数に誤差が含まれていても、広義のニュートン法に従い推定処理を繰り返すことで、段階的に誤差を減少させる。

以降、図 3.7 の処理フローに従い、変形量推定処理の内容を詳しく説明する。

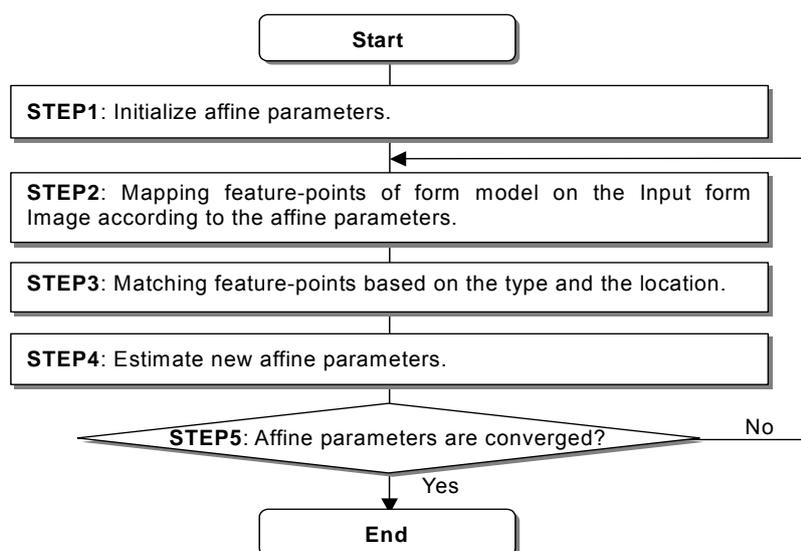


図 3.7 変形量推定処理のフロー

(1) 初期推定

まずアフィン変換係数を初期化する (図 3.7 STEP1). この際, 変形量推定処理の収束性を高めるため, 平行移動量を荒く推定しておく. 具体的には, 帳票画像の最も上にある黒画素の Y 座標と, 最も左にある黒画素の X 座標をオフセットとする (図 3.8). そして, 帳票モデルと帳票画像とのオフセットの差を荒い平行移動量と考え, アフィン変換係数を以下のように設定する.

$$a_{11}=1.0, \quad a_{12}=0.0, \quad a_{21}=0.0, \quad a_{22}=1.0,$$

$$b_1=O'x - O_x, \quad b_2=O'y - O_y$$

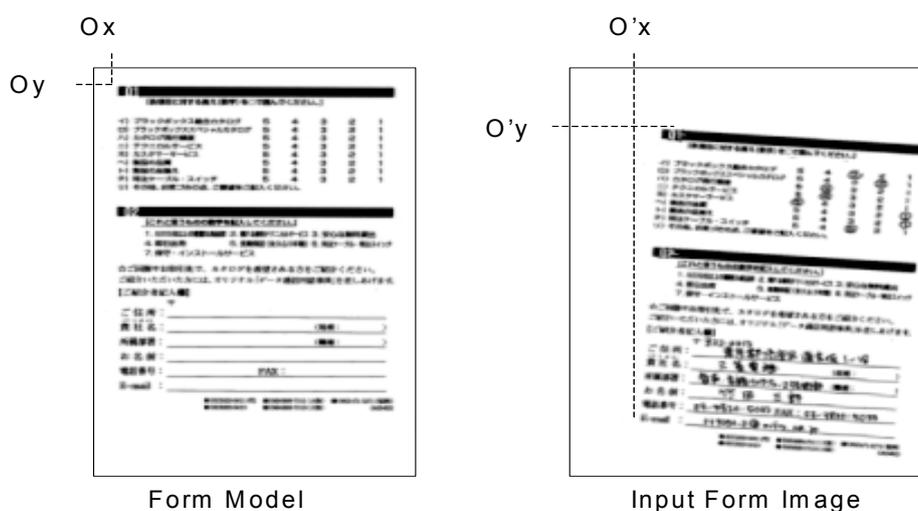


図 3.8 アフィン変換係数の初期推定

(2) 特徴点の写像

推定したアフィン変換係数に従い, 帳票モデルが持つ特徴点  $f$  の座標値  $(X_f, Y_f)$  を, 帳票画像上の座標値  $(X'_f, Y'_f)$  に写像する (図 3.7 STEP2).

(3) 特徴点間の対応付け

次に, 帳票モデルが持つ特徴点  $f$  と帳票画像から抽出した特徴点  $i$  とを対応付ける (図 3.7 STEP3). ただし, 対応付け誤りを抑制するため, 特徴点  $f$  が持つ重み  $W_f$  の値が閾値  $T_w$  より低い場合は (変形した帳票画像からの抽出安定性が低い場合は), 特徴点  $f$  を対応付けの対象から除く. まず特徴点  $f$  と特徴点  $i$  との類似度  $P_{fi}$  を,

$$P_{fi} = G(r_{fi}, T_f, T_i) \tag{3.2}$$

$$r_{fi} = \sqrt{(X'_f - X_i)^2 + (Y'_f - Y_i)^2} \quad (3.3)$$

と定義する．ここで  $T_f$  と  $T_i$  は各々特徴点  $f$  と  $i$  の種類， $r_{fi}$  は特徴点  $f$  を帳票画像上に写像した座標点と特徴点  $i$  の座標点との距離を示す．関数  $G$  は距離  $r_{fi}$  が小さく，特徴点の種類が同じ場合に大きな値を返す（図 3.9）．

次に，類似度  $P_{fi}$  の値から，特徴点を対応付ける際の評価基準  $q_{fi}$  を式 3.4 に従い求める．そして，特徴点  $f$  に対して  $q_{fi}$  の値が最大となる特徴点  $i$  を対応付ける．

$$q_{fi} = \frac{P_{fi}}{\sum_i P_{fi}} \cdot \frac{P_{fi}}{\sum_f P_{fi}} \cdot P_{fi} \quad (3.4)$$

式 3.4 では， $P_{fi}$  の値を  $\sum_i P_{fi}$  で正規化している．これは図 3.10 に示すように，特徴点  $f$  の近傍に複数の特徴点があり，どの特徴点と対応付くかを一意に判別できない場合， $q_{fi}$  の値を下げる効果を持つ．また，式 3.3 は特徴点  $f$  から特徴点  $i$  への対応度合い ( $P_{fi}/\sum_i P_{fi}$ ) に加えて，逆に特徴点  $i$  から特徴点  $f$  への対応度合い ( $P_{fi}/\sum_f P_{fi}$ ) を考慮している．これは，特徴点の多対一対応や一對多対応を抑制する効果を持つ．

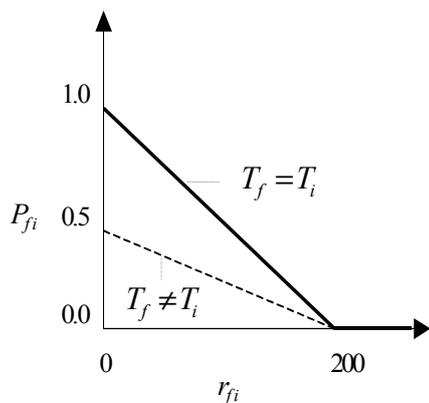


図 3.9 特徴点の類似度

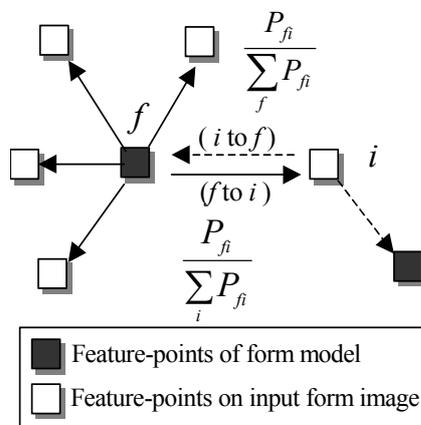


図 3.10 対応付けの例

#### (4) アフィン変換係数の推定

求めた特徴点の対応関係に従い，帳票画像の変形量を示すアフィン変換係数を推定する（図 3.7 STEP4）．アフィン変換係数は最小 2 乗法を用いて推定可能だが，最小 2 乗法の結果は例外値（誤った対応付け）の影響を受けやすい．特に，劣化した帳票画像では例外値が生じやすい．そこで式 3.5 の基準値  $M$  を最小化する，重み付き最小 2 乗法でアフィン変換係数を推定する．

$$M = \min \sum_f \rho(\varepsilon_f) \quad (3.5)$$

$$\varepsilon_f = 1 - q_{\beta} \quad (3.6)$$

ここで  $\varepsilon_f$  は対応付いた特徴点間の相違度[0-1]、 $\rho$  は重み関数である。  $\rho(x) = x^2$  の場合、式 3.5 は最小 2 乗法と等価である。ここでは、単純に  $\rho$  の値を、

$$\rho(x) = \begin{cases} x^2 & : \text{when } x < TH \\ 0 & : \text{otherwise} \end{cases} \quad (3.7)$$

とした。これは  $\varepsilon_f$  が閾値  $TH$  以上の対応付け結果を例外値として除き、残りの対応付け結果を用いて最小 2 乗法で変形量を計算することを意味する。

ただし、 $TH$  の値が大きいと例外値が除去できずに残り、逆に小さいと正しい対応付けの多くが除去されて、推定誤差が増加する恐れがある。ここでは、この対策として、ロバスト推定法の一つである LMedS 推定法のアプローチを利用して  $TH$  の値を決定する。LMedS 推定法では、推定したモデルとデータとの誤差  $\varepsilon_f$  の値から、誤差の標準偏差  $\sigma$  を、

$$\sigma = C \cdot \text{med} \sqrt{\varepsilon_f^2} \quad (3.8)$$

と推定する。ここで  $\text{med}$  は中央値、 $C$  は係数である（簡単のため  $C=1$  とする）。ここでは  $TH = \sigma$  とすることで、誤差の分布状態に応じて  $TH$  の値を決める。これにより例外値を効果的に除去し、アフィン変換係数の推定精度を高める。

#### (5) 推定処理の繰り返しによる誤差の低減

上記処理で求めたアフィン変換係数には、帳票画像の劣化により、推定誤差が含まれている可能性がある。そこで、推定したアフィン変換係数を用いて再び図 3.7 の STEP2 ~ STEP5 を繰り返し、推定したアフィン変換係数が収束した時点で処理を終える。

このような段階的な推定処理は、ニュートン法の近似となる[3-12]。そのため初期の推定誤差があまり大きくなければ、推定精度を段階的に高めることができる。

### 3.3 実験結果

実際の一般帳票に対して本方式を適用し、その有効性を検証した。また、ノイズを重畳した専用帳票を用いて、従来方式と比較した。

#### (1) 帳票画像

評価データとして、葉書～A4 サイズの一般帳票 50 種類をスキャナから読み込み、これに FAX 特有な 13 種類の変形を加えた帳票画像 650 枚を用いた。帳票画像は G3FAX と同じ 200DPI の二値画像とした。使用した帳票例を図 3.11 に示す。表 3.2 は変形の種類を示す。変形には文字記入や平行移動、拡大、縮小、回転等の複合変形を用いた。

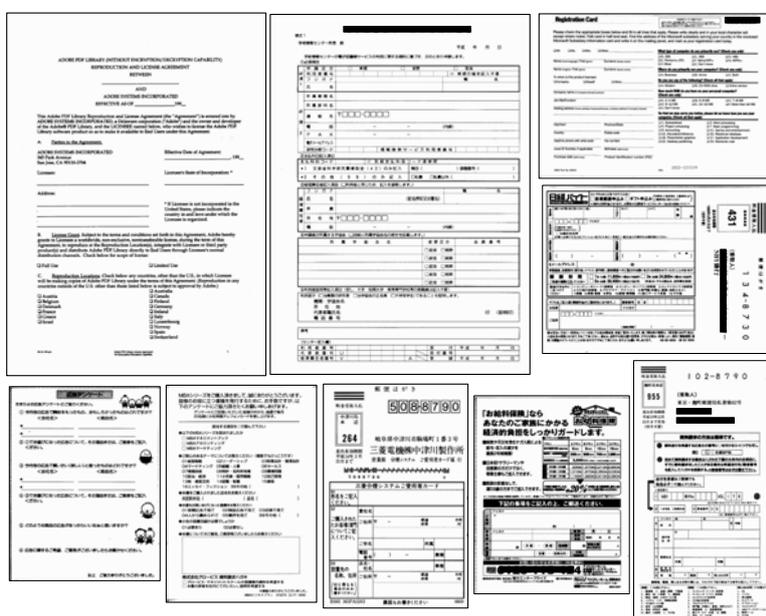


図 3.11 一般帳票の例

表 3.2 帳票画像に加えた変形の種類

<i>Transformation Type</i>	<i>Note</i>
T0. Blank (Master) form image	
T1. Filled form image	
T2. Parallel translation	T1 + 100dots translation to right
T3. Rotation	T1 + 2 degree clockwise rotation
T4. Reduction of size	T1 + 5% reduction of size
T5. Enlargement of size	T1 + 5% enlargement of size
T6. Parallel translation and rotation	T2 + T3
T7. Parallel translation and reduction of size	T2 + T4
T8. Parallel translation and enlargement of size	T2 + T5
T9. Rotation and reduction of size	T3 + T4
T10. Rotation and enlargement of size	T3 + T5
T11. Parallel translation and rotation and reduction of size	T2 + T3 + T4
T12. Parallel translation and rotation and enlargement of size	T2 + T3 + T5

## (2) 評価指標

二つの評価指標を用いる。一つは変形量の推定精度を示す推定位置ずれ量  $\Delta r$  , もう一つは入力された全帳票の内、処理に失敗した帳票の割合を示す帳票棄却率  $Er$  である(システム的な性能指標に用いられる)。

- ・推定位置ずれ量  $\Delta r$  : 推定したアフィン変換係数を用いて、帳票モデルの座標点を帳票画像上に写像した際、本来写像されるべき正しい座標点との位置ずれ量(帳票モデルの全座標点での平均値)。

- ・帳票棄却率  $Er$  :  $Er = \text{読み取りに失敗した帳票数} / \text{入力した帳票数}$  . 本評価では  $\Delta r \geq 10$  の帳票画像を、読み取りに失敗した帳票と仮定した。これはフィールドの位置が 10 画素以上ずれた場合、文字認識処理に悪影響を及ぼす場合があるためである。

## (3) 段階的な変形量推定の評価

本方式は、変形量の推定精度を向上させるため、広義のニュートン法に従いアフィン変換係数の推定を繰り返す。ここでは、その効果を検証する。評価は  $T_w=0$  の条件で行った。

図 3.12 は、帳票画像に加えた変形別に、推定処理の繰り返し回数  $Iteration$  と  $\Delta r$  との関係調べた結果である。ここで  $Iteration=0$  は帳票モデルの座標点をそのまま帳票画像上に恒等変換した状態を、 $Iteration=1$  は初期推定処理(図 3.7 STEP1)により帳票画像のオフセットを荒く補正した状態を示し、以降の  $Iteration$  の増加に伴いアフィン変換係数の推定が繰り返される。図 3.12 の結果から、 $Iteration$  の増加に伴い  $\Delta r$  の値が徐々に低下していることが分かる。また、図 3.13 は  $Iteration$  と  $Er$  の関係であり、ここでも  $Iteration$  の増加に伴い  $Er$  の値が小さくなっている。これらの結果から、アフィン変換係数の推定を繰り返すことで、初期の推定誤差が徐々に低下し、一般帳票の変形量を精度良く推定できていることが分かる。

## (4) 例外値の除去の評価

提案方式は、例外値を効果的に除去するため、例外値の除去に用いる閾値  $TH=\sigma$  とし、 $Iteration$  毎に  $TH$  の値を動的に変更する。この有効性を検証するため、 $TH$  を固定値とした場合と比較した ( $TH=1.0, 0.95, 0.9, 0.85, 0.8, 0.7, 0.6$ )。ここで  $TH$  の値が小さいほど多くの対応付け結果が例外値として除去される。

図 3.14 は、 $TH$  をパラメータとした  $Iteration$  と  $\Delta r$  の関係を示す。ここで  $\Delta r$  は全評価データの平均値であり、測定は  $T_w=0$  の条件で行った。図 3.14 の結果から  $Iteration$  の増加に伴い、すべての  $TH$  の条件で  $\Delta r$  の値が減少していることが分かる。また、各  $Iteration$  における  $\Delta r$  の値は、 $TH$  が固定値の場合よりも、 $TH=\sigma$  の方が小さい。このことから  $TH=\sigma$  とすることで、変形量の推定精度が向上していることが分かる。

変形量の推定が終了した  $Iteration=19$  の時点で  $Er$  の値を比較すると、 $TH$  が固定値の場合、 $Er$  の値はスリバチ型の曲線を描く(図 3.15)。これは  $TH$  の値が大きいと例外値が除去できず、逆に  $TH$  の値が小さいと正しい対応付け結果まで多数除去されることの影響と

考えられる。これに対して  $TH = \sigma$  の場合,  $Er$  の値はスリパチ型の曲線下に位置する。このことから  $TH = \sigma$  とすることで,  $TH$  を固定値とした場合よりもフィールド抽出に失敗する帳票数が減少していることが分かる。

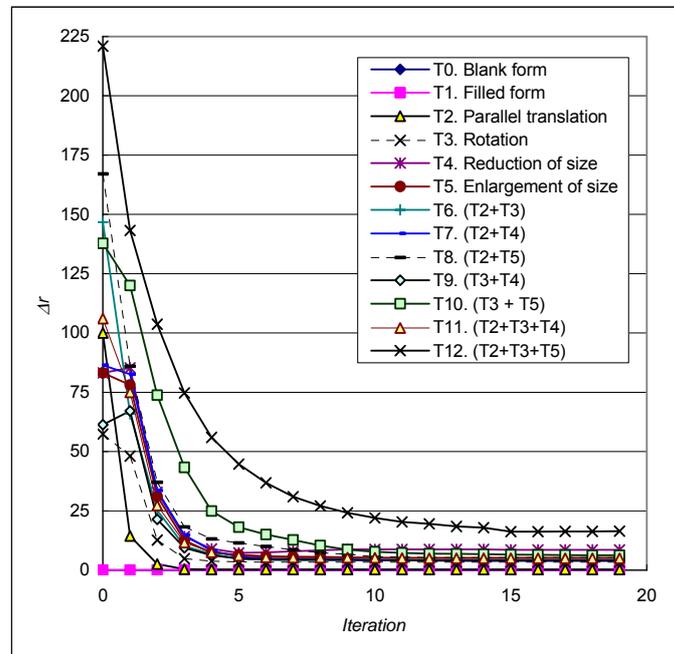


図 3.12 Iteration と  $\Delta r$  の関係

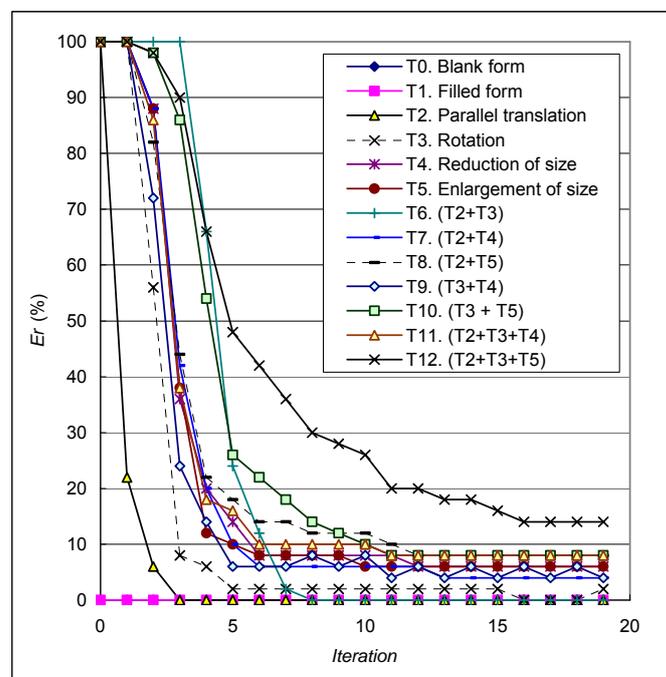


図 3.13 Iteration と  $Er$  の関係

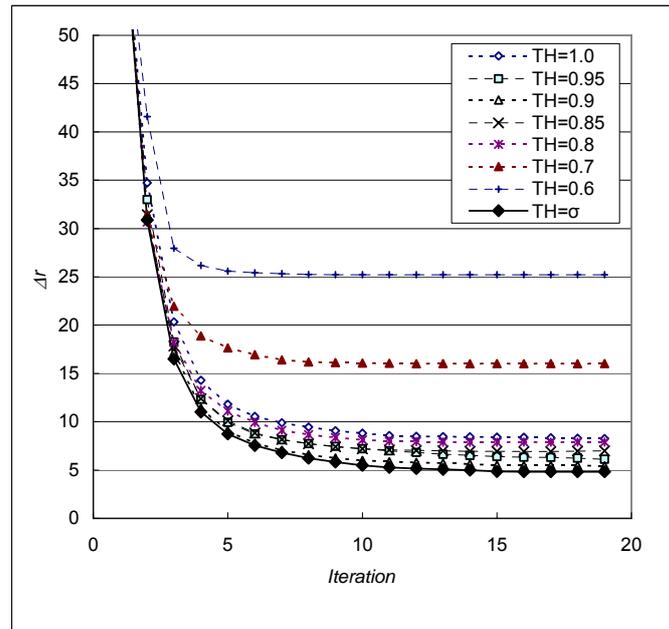


図 3.14 例外値の除去による  $\Delta r$  への効果

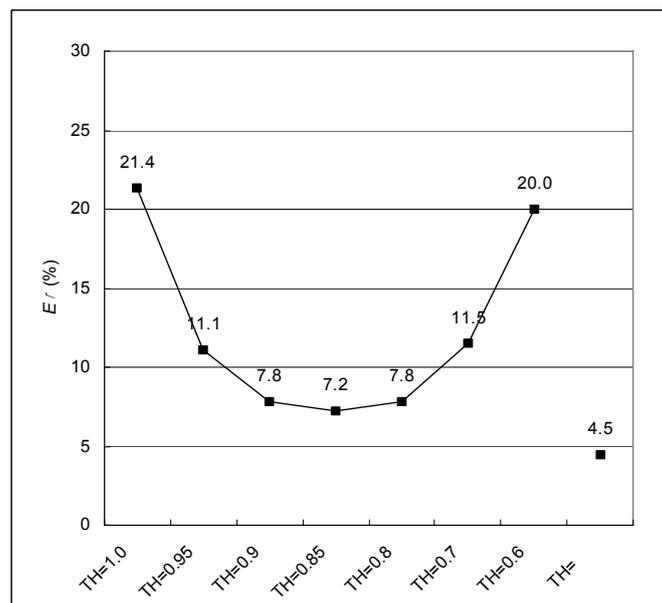


図 3.15 例外値の除去による  $E_r$  への効果

(5) 評価基準による対応付け誤りの抑制

本方式では、特徴点の多対一対応や一対多対応等の対応付け誤りを抑制するため、対応付けの評価基準  $q_{fi}$  を式 3.4 で定義した。この効果を検証するため、式 3.9 に示す単純な評価基準と比較した。前述のとおり式 3.4 は特徴点  $f$  から特徴点  $i$  に対する対応度合いと、特

徴点  $i$  から特徴点  $f$  に対する対応度合いを考慮している．これに対して，式 3.9 は特徴点  $f$  から特徴点  $i$  への対応度合いのみを考慮している．

$$q_{fi} = \frac{P_{fi}}{\sum_i P_{fi}} \cdot P_{fi} \quad (3.9)$$

図 3.16 は， $Iteration=19$  における  $Er$  の値（全評価データの平均値）を示す．評価は， $TH$  の値をパラメータ（ $TH=1.0, 0.95, 0.9, 0.85, 0.8, 0.7, 0.6, \sigma$ ）とし， $T_w=0$  の条件で行った．図から分かるように，式 3.9 よりも式 3.4 における  $Er$  の値が小さく，その有効性が明らかである．割愛するが  $\Delta r$  の比較でも同様な傾向が見られた．

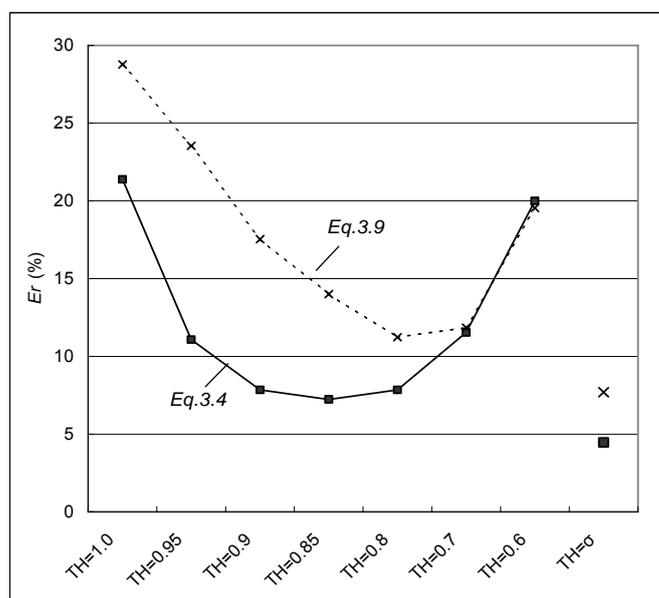


図 3.16 評価基準の効果

#### (6) 特徴点の重みによる対応付け誤りの抑制

本方式では，対応付け誤りを抑制する別の手段として，重みの値が閾値  $T_w$  より小さな特徴点を対応付けの対象から除いた．その効果を検証する．

図 3.17 は， $T_w$  を 0～1.0 まで 0.25 間隔で変更した場合の  $\Delta r$  と  $Er$  の値を示す．ここで  $\Delta r$  と  $Er$  の値は全評価データの平均値であり，測定は  $TH=\sigma$  の条件で行った．図 3.17 の結果から，重みを考慮せずに全特徴点を対応付ける  $T_w=0$  の場合よりも，重みを考慮した  $T_w=0.5$  の場合で  $\Delta r$ ， $Er$  の値が小さい．そのため重みの考慮により変形量の推定精度が向上していることが分かる．ただし， $T_w>0.5$  では対応付ける特徴点の数が極端に減少して，推定精度が低下する悪影響も見られた．

ここで、重みの考慮により読み取り可能となった帳票の多くは、図 3.18 のような文字列が多い帳票であった。図 3.18 の例では、帳票モデルの文字列上で抽出された特徴点の多くが変形した帳票画像からは安定的に抽出されないため、近くにある別の特徴点との誤った対応付けが生じて変形量が正しく推定できなかった。しかし  $T_w=0.5$  では、抽出安定性の低い特徴点を除くことで誤った対応付けが減少し、正確にフィールド抽出できた。

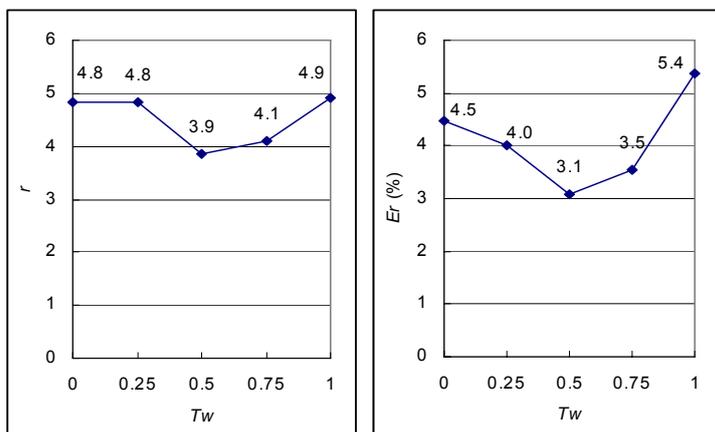


図 3.17 特徴点の重みの効果

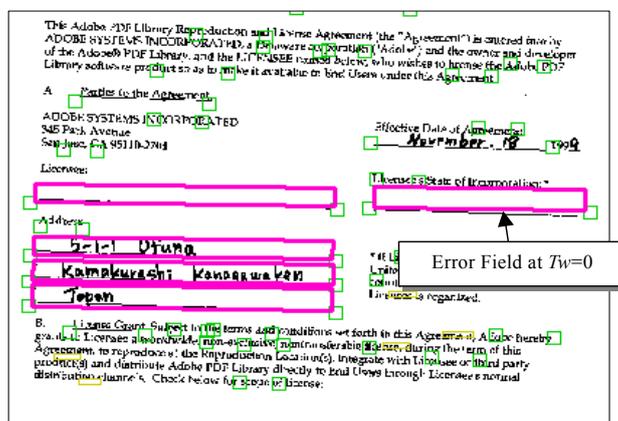


図 3.18 特徴点の重み考慮による成功例

(7) フィールド抽出結果例

フィールド抽出の成功例を図 3.19 に示す。図から分かるように、表形式の帳票（図左）に加え、文字列とイラストから成る表を含まない帳票（図右）でも正しくフィールド抽出することができた。図 3.20 は失敗した帳票画像の一部である。この例では、横線の右端点から抽出された特徴点が、帳票モデルの特徴点と上下に 1 個ずれて対応付いたことが原因で変形量の推定を誤った。



図 3.19 成功例

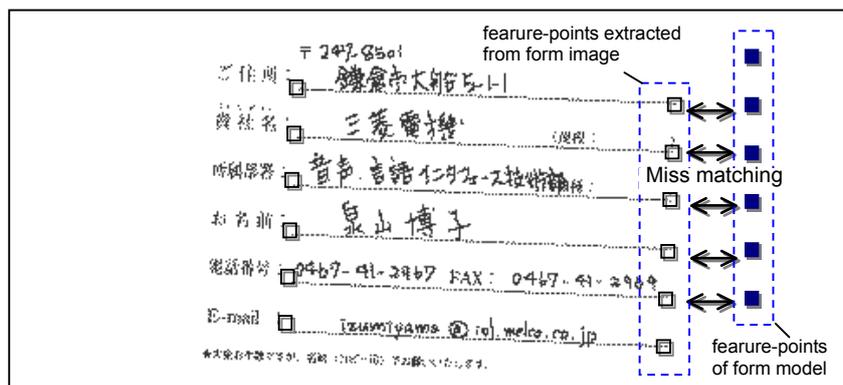


図 3.20 失敗例

(8) 従来方式との比較

従来方式との比較結果を示す．従来方式には，著者らが以前に開発した表形式の専用帳票を読み取る方式[3-5]を用いた．この従来方式は，まず表罫線の位置や長さ，および表罫線間の接続関係を元に，帳票画像から抽出した表罫線と帳票モデルの表罫線とを動的計画法に従い対応付ける．そして表罫線の対応関係から，表の内部に設定されたフィールド位置を求める．

評価には，従来方式で読み取り可能な専用帳票を用いた（図 3.21）．具体的には，文字を記入した専用帳票 200 枚を FAX から入力し，そこに人工的なノイズを加えた帳票画像を評価データとした．人工的なノイズは，FAX で頻繁に生じる「カスレ」と「離散ノイズ」を模擬した．図 3.22 はノイズを加えた帳票画像の例である．

- (1) カスレ：帳票画像上の各小領域(4×4 画素)をノイズ発生確率  $P_g$  で白画素に置換．
- (2) 離散ノイズ：帳票画像上の各小領域を，ノイズ発生確率  $P_g$  で白か黒画素に置換．

図 3.23 はノイズ発生確率  $P_g$  を変えて、提案方式と従来方式の  $E_r$  を比較した結果である。提案方式は  $T_w=0.5$ ,  $TH=\sigma$  の条件で測定した。ここで提案方式は  $P_g=20\%$  でも  $E_r$  の値が 0% に近く、ノイズへの耐性が従来方式よりも強いことが分かる。これは従来方式が表罫線に着目した手法であるため表罫線の切れに影響を受け易いことに対し、提案方式が着目した特徴点は表罫線の切れの影響を受け難く、かつ特徴点の対応付け誤りに対してロバストに変形量を推定する効果と考える。

この表罫線の切れの問題は、画像処理的な表罫線の切れの補正により、ある程度は対応できる。実験で用いた従来方式も、表罫線を抽出する際に数画素の表罫線切れを補正している。ただし画像処理的な補正には限界があり、ノイズが大きいと対応が困難である。別の対策として、石谷[3-2]はグラフマッチングにより、切れを考慮した表罫線の対応付けを提案している。この方式は画像処理的な補正よりも表罫線の切れに強いと考えられる。ただし  $P_g>10\%$  の場合のように表罫線が何箇所も大きく分断されると、やはり十分な性能を得ることが困難と考えられる。実際に FAX 送信された帳票画像では  $P_g=10\sim 30\%$  程度の劣化は、まれではない。

3/01 01 16:38 0467 46 5083 ファイル名: NO.2154 001 P01

#023# ○Xスゴークラフ人會申込書 個人会日 年 月 日

〒

ご住所

お名前

生年月日 性別 年齢 全 身 健 別

年	月	日	男	女	全	平	子	子	子	子
---	---	---	---	---	---	---	---	---	---	---

身体のサイズ			筋 力			特記事項
身長	cm		握力	kg		
体重	kg		腕力	kg		
胸囲	cm		反復握力	回		
脚囲	cm		立位体前屈	cm		

図 3.21 専用帳票

(9) フィールドデータによる評価

最後に、複数台の FAX から一般の公衆回線を通して送信されたフィールドデータを用いて、本方式に対する帳票棄却率  $E_r$  を調べた。測定は  $T_w=0.5$ ,  $TH=\sigma$  の条件で行い、帳票は A4 サイズ、フィールド数 24 個の専用帳票を用いた。その結果、入力された帳票画像 11465 枚中の 404 枚を除いて正しくフィールド抽出でき、 $E_r=3.5\%$  となった(表 3.3)。これに対し、著者らが既に実用化した専用帳票対応の FAX-OCR システムでは、各種のフィールドデータに対して  $E_r=7\sim 15\%$  程度の結果を得ている ( $E_r$  の値は帳票の形式や、FAX の機種・整備状態等で変動する)。本方式が示した  $E_r$  の値は実際の FAX-OCR システムで得られた値よりも小さく、処理速度も 1 枚あたり 294msec と高速なことから、本方式は実用的な性能を持つといえる。

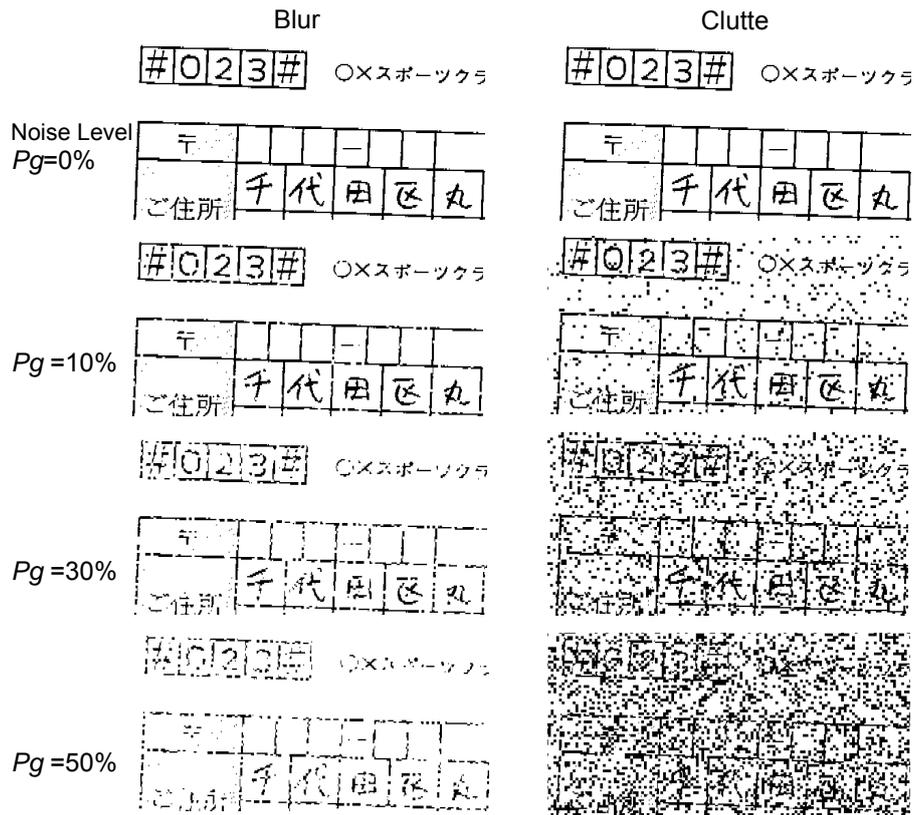


図 3.22 ノイズを重畳した帳票画像

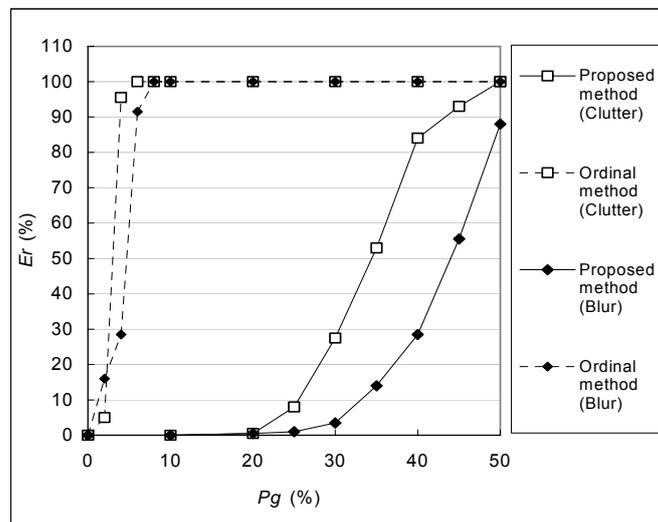


図 3.23 ノイズ重畳画像に対する従来方式との比較

表 3.3 フィールドデータでの評価結果

Input Images	Error Images	Form Reject $E_r$	Processing time for Field Extraction
11465	404	3.5%	294 msec/page

OS:WindowsNT4.0, CPU:PentiumIII 1GHz, Memory:512MBで測定

### 3.4 FAX-OCR システムへの実適用

本フィールド抽出する方法は、人間が文字を読み取る処理を代替する FAX-OCR システムに実装した(1997～)。本提案方式を採用することにより、従来は困難であった自由な帳票を読み取ることができ、製品の適用範囲を広げた(図 3.24)。



図 3.24 FAX-OCR システム

## 3.5 第3章のまとめ

本章では、人間が画像中から大きさや回転角度が異なる特定の形状を見つけるために、画像と形状モデルとの写像関係を獲得する処理をモデル化した。具体的には、FAX送信により変形した帳票画像と帳票モデル間の写像を、それら両方から抽出した特徴点同士の対応付け結果に従い算出する方式を提案した。

本方式では、視覚野で抽出されているのと同様な特徴点を用いた。この特徴点は表や下線、文字列、イラスト等から抽出できるため、多様な形式の既存帳票に適用できる。また、特徴点自身が従来方式で利用されている表罫線よりもノイズに対して安定である（表罫線が切れても影響を受け難い）。さらに特徴点の消失・湧き出しに対してロバストに変形量を推定するため、FAX送信による帳票画像の劣化に強い特性を持つ。

評価では、50種類の一般帳票にFAX特有の多様な変形を加えた650枚の帳票画像に対して本方式を適用し、変形した一般帳票からもフィールド抽出が可能なことを確認した。さらにノイズを重畳した専用帳票を用いて従来方式との比較を行い、本方式が従来方式よりも高いノイズ耐性を持つことを確認した。また本方式を人間が帳票上の文字を読み取る処理を代替するFAX-OCRシステムに適用し、実用化した。

## 第 4 章

# 知識概念への写像獲得と文書処理システムへの適用

本章では、文書と知識概念間の写像関係を獲得する意味論的な処理をモデル化する。これは、人間が文書を読んで既存のカテゴリに分類したり、文書から関連知識を想起する機能を実現するものである。ここでは、以下の二例について検討を行う。

- (1) 文書とカテゴリ間の写像
- (2) 文書と関連知識間の写像

以降、4.1 節では文書とカテゴリ間の写像獲得について考える。また 4.2 節において文書と関連知識間の写像獲得を述べ、最後に 4.3 節でまとめとする。

## 4.1 文書とカテゴリ間の写像

文書とカテゴリ間の写像では、文書中から氏名や地名といったメタ情報を抽出し、このメタ情報を元に文書を特定のカテゴリへと分類（写像）する。本写像獲得のため、ここでは多様なファイル形式の文書から漏れなく論理情報を抽出可能な文書解析方式を提案する。抽出した文書情報をアプリオリな知識（例えば、“氏名”という文字列の近くにある名詞句は人の名前である等）に基づいて解析することで文書中からメタ情報を抽出し、文書をカテゴリに分類する。

表 4.1 文書とカテゴリ間の写像モデル

写像関係	文書と、カテゴリ（地名や人名）間の写像。
適応性	-
写像の種類	文書の論理構造と言語処理、およびアプリオリな知識を元に文書内容からメタ情報を抽出し、これを元に写像を行う。
脳内の関係部位	言語野/連合野（前頭葉）
人間支援システム	多量の検索結果を自動的に分類・表示する文書管理システムに適用（複数製品に実装）



図 4.1 文書とカテゴリ間の写像獲得方式

### 4.1.1 カテゴリへの写像の背景と目的

インターネットやイントラネットに蓄積された情報の活用を目的として文書のキーワード検索機能が広く利用され、これに伴い文書からテキストを抽出するための規格・技術が開発されている。しかし近年、単純なキーワード検索よりも有効な知識活用方法が求められている。この実現方式として、文書中に人間が予めメタ情報を付与しおくことで、機械が容易に文書の内容を解釈できる Semantic Web の検討が進められている。例えばメタ情報を用いることで、検索された多量の文書をカテゴリ毎に分類して整理できる。ただしメタ情報付与に知識や労力を要する等の問題があり、Semantic Web の普及は進んでいない。そのために既存の文書から、自動的に内容情報を取得して、メタ情報まで自動取得できる文書解析方式が必要とされている。

従来の文書解析方式は、スキャナ等で取得した文書画像を対象とした場合と、電子文書や図面を対象にした場合で異なる。文書画像を対象とした文書解析方式は、文字認識処理を用いて画像中の文字をコード化することで内容情報を取得する。文字認識は計算量が多く、しかも結果に誤りを含む課題がある。しかし、画像解析時に得たレイアウト情報を用いることで文書の論理構造情報が抽出できる効果もある。例えば、文献[4-1][4-2]は文字列の位置情報を元に関連した文字列の集合をブロック化して、文書タイトルや段組等の論理構造を抽出する。また文献[4-3][4-4][4-5]は画像から検出した表罫線を用いて表の位置や構造を解析することで、文書中の表位置や、表内の論理構造情報を抽出する。

一方、電子文書や図面を対象とした文書解析では、情報抽出用フィルタが一般的に利用される。情報抽出用フィルタは特定のファイル形式の文書からテキストを取得するソフトであり、入出力仕様が規格化されている。この規格として Microsoft 社の iFilter [4-6]が著名である。情報抽出用フィルタは検索エンジンで利用されており、ワープロソフトメーカーは規格に合致した自社の情報抽出用フィルタを公開することで、自社のファイル形式を各種の検索エンジンで検索可能としている。情報抽出フィルタの利点は高速かつ安定した情報取得が可能である。ただし取得できる情報は文書内にあるコード化されたテキスト部分であり、文書に貼り付けた画像中の文字や、図面に線分で文字を描いたベクトルフォント文字、業務用印刷のためアウトライン化した文字などの非コード化文字は抽出できず、情報が欠損する。また、レイアウト情報や文書論理構造情報も抽出が困難である(図4.2)。

ただし、情報抽出フィルタでは抽出困難な文書論理構造も、ファイル形式を限定した専用の解析処理を作成すれば抽出可能な場合がある。例えば、リッチテキストフォーマット(RTF)のような仕様が公開されたファイル形式や、文書データを操作する DOM(Document Object Model)のインタフェースが公開されているファイル形式では、専用の解析処理を作成することで章節構造などの文書論理構造情報を抽出できる。ただしファイル形式毎に専用処理の開発を必要とするため、多様なファイル形式を扱うことは難しい。

このような多様なファイル形式を統一的に扱う課題に対しては、PostScript 解析方式 [4-7]が提案されている。この方式は、文書を擬似的に印刷処理して多様なファイル形式をプリンタが解釈できる PostScript に変換し、内部のテキストオブジェクトからテキストを取得する。これによりファイル形式に依存せずテキストが抽出できる。ただし、非コード化文字は抽出対象とならず、レイアウト情報や文書論理構造情報も抽出しない。

ここでは多様なファイル形式の文書からコード化文字と非コード化文字を抽出し、同時にレイアウト情報と文書論理構造情報の抽出が可能な新しい文書解析方式を提案する。本方式は、まず文書を擬似的に印刷処理してプリンタが印刷を行うための情報を格納したページ記述言語データを作成する。そして内部のテキストオブジェクトとイメージオブジェクト、パスオブジェクトなどの基本的なグラフィック要素を解析することでコード化文字と非コード化文字の情報を漏れなく抽出する。また、印刷時のレンダリング情報を用いてレイアウト解析を行うことで文書論理構造も含めた詳細な内容情報を得ることができる。

本技術は、製造業における文書検索や知識活用システムのための情報抽出に適用できる。例えば、製造業のオフィスには電子文書や CAD 図面、文書画像などの多種多様な文書ファイルが存在する。従来は電子図面や電子文書からコード化されたテキストを抽出した結果と、文書画像中の文字を文字認識した結果を用いて検索が行われた。そのため、電子文書や CAD 図面中に含まれる非コード化文字部分が検索対象とならず、必要な情報にアクセスできない問題が発生していた。本提案技術では、文書中のコード化文字と非コード化文字を漏れなく抽出するため検索漏れが少なく、かつ印刷可能な文書であれば文書種類を問わずに検索できるため、製造業に適した文書検索が実現できる。さらに文書画像以外から文書論理構造を抽出できるため、メタ情報を利用した文書の分類や関連文書への自動リンク付けなどが可能となり、製造業の知識活用をさらに促進できると考える。

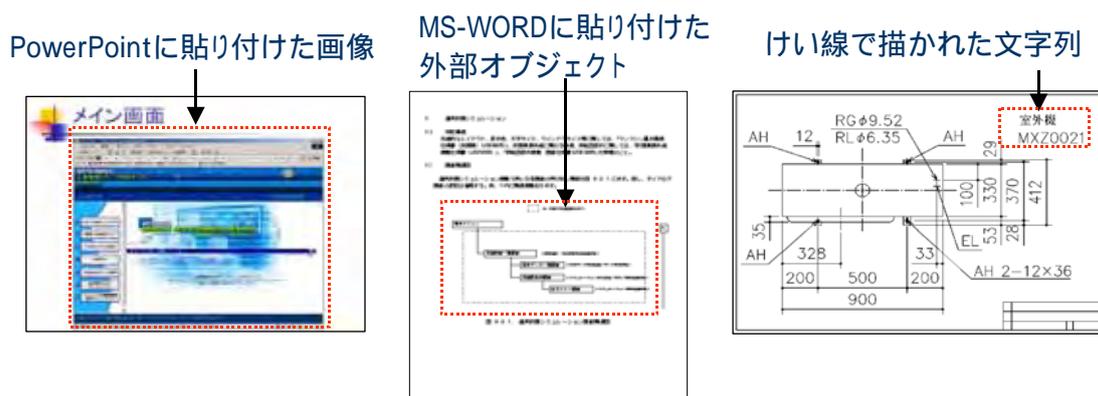


図4.2 従来方式で情報抽出対象とならない領域

### 4.1.2 ページ記述言語

#### (1) データ構造

ページ記述言語 (PDL: Page Description Language) はプリンタに文書を印刷するためのプログラミング言語であり, 1984年にAdobe Systems社が発表したPostScriptが標準である。ワープロソフトで文書の印刷処理を行うと, まずOS上のプリンタドライバが印刷装置に依存しないPostScriptデータを作成する。次に作成したPostScriptデータはプリンタに転送され, プリンタはこの記述を解釈することで紙面上へ文書を印刷する。

PostScriptのイメージングモデルでは, 文字形状 (グリフ), 幾何学形状, 線, サンプルイメージを文書の特定頁の指定した領域上に配置する。グラフィック要素同士が重なる場合は, 後に描くグラフィック要素で先に描かれたグラフィック要素を上書きすることで頁の印刷状態を表現する。ここでPostScriptのデータを構成する主要なグラフィック要素には下記の3つがある (図4.3)。

- ・ **テキストオブジェクト**: コード化されたテキストを構成する各文字のコードと文字形状 (グリフ), および印字位置を表す
- ・ **パスオブジェクト**: 座標点の集合により, 直線や矩形, 3次元ベジェ曲線で構成した形状を表す
- ・ **イメージオブジェクト**: 写真や画像等のサンプリングイメージのデータと, その描画領域を示す

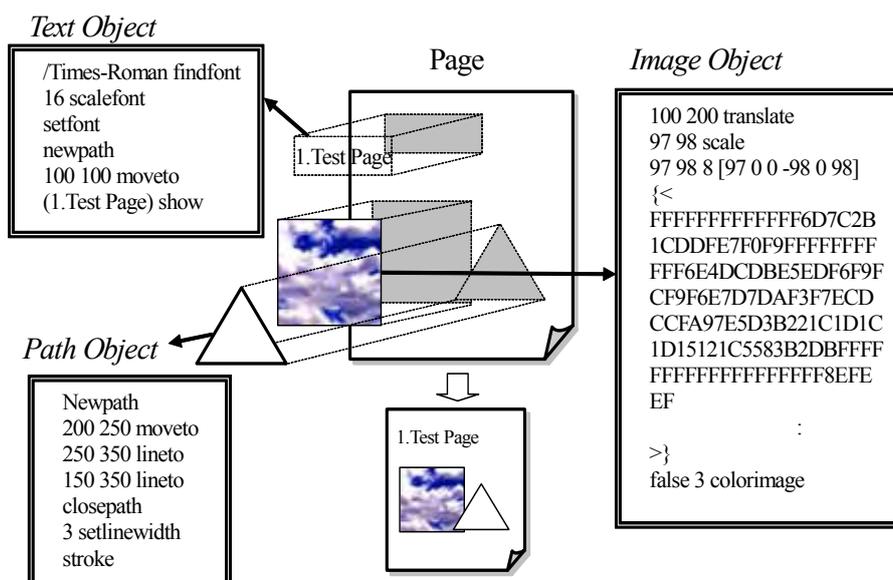
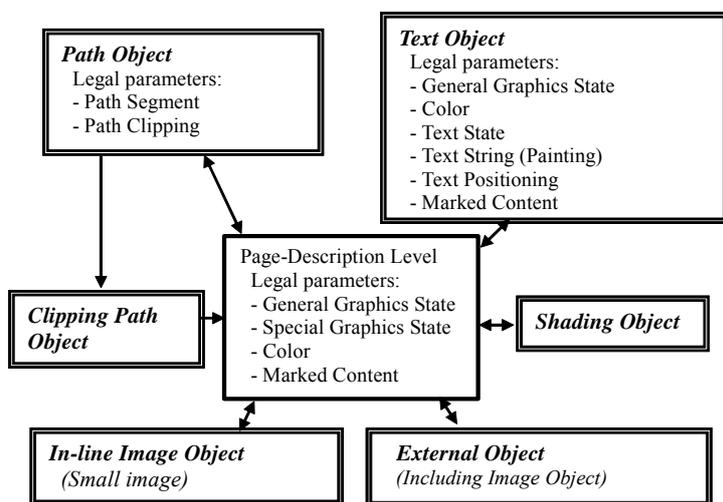


図 4.3 PostScript のグラフィックス要素

PostScriptと同様にPDF[4-8]やDjVu[4-9]、および近年仕様が公開されたSVG[4-10]、XPS[4-11]も、データ構造は異なるがPDLの一種である。図4.4はPDFのデータ構造であり、PostScriptにない拡張用のオブジェクトを持つ。しかし、テキストオブジェクト、パスオブジェクト、イメージオブジェクトは、これら全てのPDLで共通に存在する。我々は、汎用的な文書解析方式を実現するため、これら3つのグラフィック要素を解析対象とする。



(出典: "Portable Format Reference Manual Version 1.3" Adobe systems Incorporated)

図 4.4 PDF のグラフィックス要素

## (2) PDLからの内容情報抽出の課題

PDLは印刷を目的としており情報抽出用のデータではない。また、印刷結果が同じでも印刷するアプリケーションによりPDLデータの記述内容が異なる場合がある。このためPDLからの内容情報抽出には以下の課題がある。

- テキストオブジェクトから抽出したテキストは不正な単語を含む。例えば太字や影付き文字を印刷するため、同じ文字を数ドットずらした位置に複数回印字している場合がある。また、文字列の読み順の情報を持たないため、抽出結果が読み順とおりにならない場合がある。
- 一枚の画像が複数のイメージオブジェクトに分断され、イメージオブジェクトを個々に文字認識しても正確な情報が得られない場合がある。例えばあるアプリケーションは、一つの画像を細長い短冊状のイメージオブジェクト群に分断する。また別のアプリケーションはファイルの圧縮率を高めるために画像中の文字記述領域と写真領域を分割して圧縮保持し、これらが別々のイメージオブジェクトとなる。
- グラフィック要素が重畳するため、重なったグラフィック要素の種類や、重なり順を考慮して情報抽出しないと、情報の欠損や不必要な情報抽出が発生する。

## 4.1.3 提案方式の概要

提案方式の処理フローを図 4.5 に示す．本方式は以下に示す 3 つの処理から構成する．

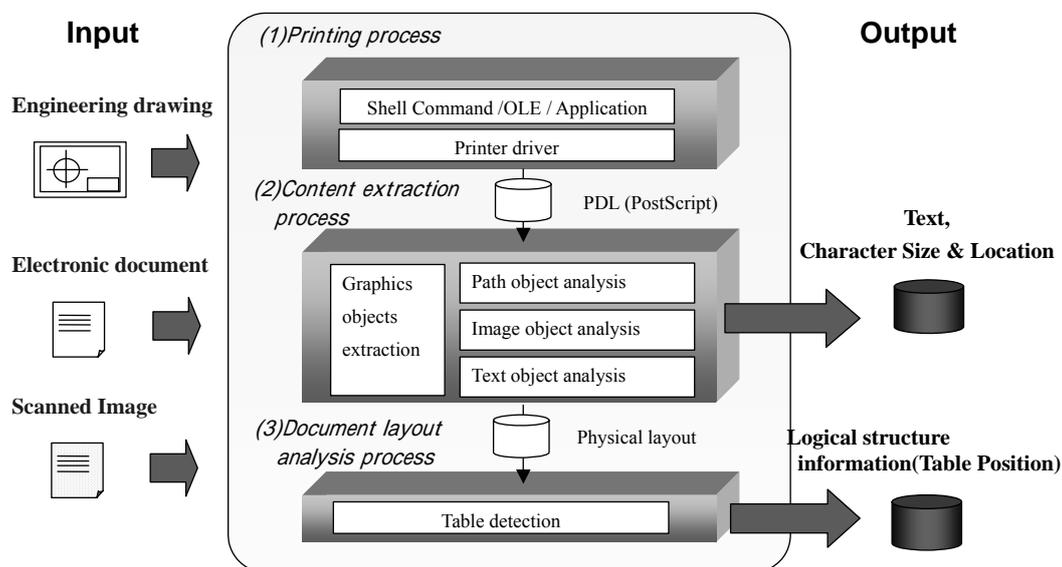


図 4.5 提案方式の処理

- (1) 擬似印刷処理：解析する文書を擬似的に印刷処理する．ファイル形式に対応したアプリケーションが外部から制御可能な印刷命令を提供している場合は、これを用いる．その他の場合は Windows OS の印刷命令 ( ShellExecuteEx ) を用いて印刷する．印刷した文書はプリンタドライバを介して PostScript 形式に変換される．
- (2) PDL 解析処理：作成した PostScript ファイル内にあるデータを Adobe のライブラリを利用して取得し、グラフィックス要素の内容を解析することでテキストとレイアウト情報を抽出する（詳細は 4.1.4 節に後述）．これにより文書中の文字のコードと位置座標、大きさ、および罫線の位置と長さを得る．
- (3) 文書論理構造抽出処理：PDL 解析処理で求めたテキスト情報とレイアウト情報から、レイアウト解析的な手法により文書論理構造情報を抽出する（詳細は第 4.1.5 節を参照）．文書論理構造情報として、文書の章節・パラグラフ構造と表構造を抽出する．この抽出結果は XML 形式でファイル出力する．

### 4.1.4 PDL 解析処理

#### (1) テキストオブジェクトの解析

まず PDL 内のテキストオブジェクトからコード化されたテキストを抽出する．これは PostScript のテキスト抽出方式[4-7]と同じ処理である．ただし PDL 特有の課題で，太字や影付き文字部では文字が複数回印字されており，誤ったテキストが取得される場合がある（図 4.6）．例えば太字の文字列“ABC”は，位置をずらして“ABC”“ABC”“ABC”と文字列を繰り返して印字したり，各文字を“AAA”“BBB”“CCC”と繰り返して印字することで描かれる．この文字列をそのまま抽出すると文書の検索精度が低下する．そのため本処理では，式 4.1 に合致するような，印字位置が近くて同じコードを持つ文字を抽出結果から除去する．

$$D(i, j) + S(i, j) \leq \theta \quad \text{and} \quad \text{Code}(i) = \text{Code}(j) \quad (4.1)$$

ここで  $i, j$  はテキスト先頭からの文字のオフセット番号である． $D$  は文字  $i$  と  $j$  の左上点の印字位置座標の差， $S$  は文字サイズの違い， $\theta$  は固定の閾値を示す．

さらに PDL では透明なテキストが描かれる場合や，頁の表示領域外にテキストが存在する場合がある．これらのテキスト部分が検索でヒットした場合，ユーザはそのヒット理由が文書を見ても分からない．そのため文字色に透明色を持つテキストや頁の表示領域外に存在するテキストは抽出対象外とする．

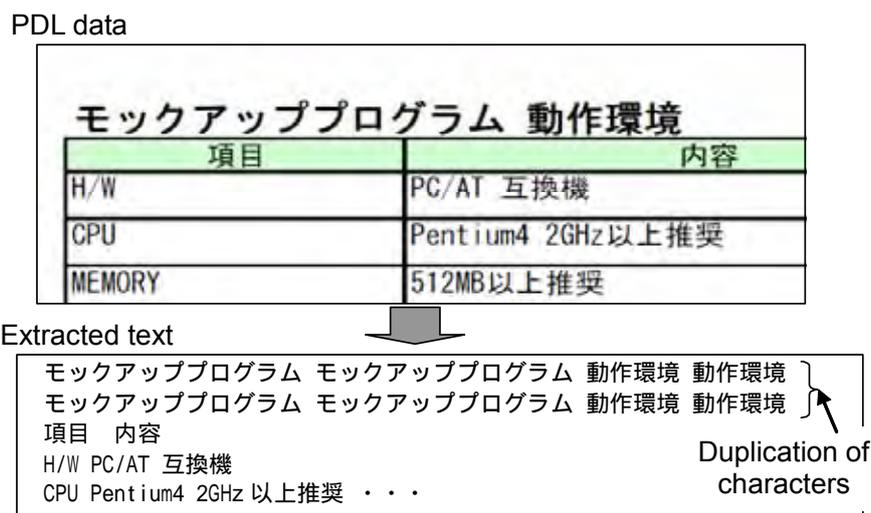


図 4.6 PDL からのテキスト抽出結果例

## (2) イメージオブジェクトの解析

イメージオブジェクトの解析では、まず PDL の各頁からイメージオブジェクトの外接矩形の左上座標と、幅・高さ、解像度、圧縮方式の情報を得る。次に PDL の各頁をレンダリング処理して文書を印刷したのと同じ状態の BMP 画像を作成する。画像は文字認識処理が可能ないように、一般的なスキャナと同程度の解像度 400dpi, 256 階調色とする。次に BMP 画像中のイメージオブジェクトの領域内に対して文字認識処理を適用することで、イメージオブジェクト内の文字をコード化する。

ただし通常の文書画像と異なり、PDL 特有の課題としてオブジェクト同士の重なりを考慮する必要がある。例えばテキストオブジェクトが上に重畳されているイメージオブジェクトは文字の背景画像と判定し、文字認識処理の対象から除く。これは図 4.7 に示すように、テキストオブジェクトが重畳されたイメージオブジェクトの領域を文字認識した場合、テキストオブジェクトの文字列まで文字認識して情報が二重に得られてしまうためである。ただし透明なテキストが重畳されたイメージオブジェクトは文字認識対象とする。

また PDL では、1つの大きな画像を小さな短冊領域に分割して複数のイメージオブジェクトで表現する場合がある。この現象は特定のアプリケーションで発生する。このようなイメージオブジェクトを個々に文字認識しても文字列が分割されているため正しい結果が得られない。そのため、式 4.2 に従って隣接したイメージオブジェクト群や、内包関係にあるイメージオブジェクト群を 1個に統合し、その領域を文字認識処理する(図 4.8)。

$$\begin{aligned} & \text{If } \left\{ \begin{array}{l} \text{(the area of } IO_i \text{ is involved in the area of } IO_j \text{) or} \\ \text{(} IO_i \text{ comes in contact with } IO_j \text{)} \end{array} \right\} \\ & \text{then } IO_i \text{ and } IO_j \text{ are integrated} \\ & \text{where } IO_i \notin \mathbf{X} \cap IO_j \notin \mathbf{X} \end{aligned} \quad (4.2)$$

ここで  $IO$  はイメージオブジェクト、 $\mathbf{X}$  は背景画像と判断されたイメージオブジェクトの集合である。なお、文字認識処理では図 4.9 に示すように、まず抽出されたイメージエリアの画像を二値化して白黒の BMP 画像を作成し、次に白黒の BMP 画像中から文字列を抽出する。その後、文字列内から個々の文字領域を切り出して認識することで、文字コードと文字の位置、大きさ情報を得る。ここで文字列抽出と文字領域の切り出しには、文献[4-17]に示すような X, Y 方向への射影ヒストグラムを利用した方式を用いる。また切り出した領域の文字認識には文献[4-12]の手法を実装した日本語文字認識エンジンを用いる。

## (3) パスオブジェクトの解析

パスオブジェクトは、一般の文書では、イラストや罫線や表を描くために利用される。ただし、CAD では図面中の文字を描くために用いる場合がある(図 4.10)。同様に雑誌やカタログなどの業務用に作成された PDF ファイルでは、フォントの有無で見た目が変

わからないように文字をアウトライン化してパスオブジェクトに変換している場合がある。このようなパスオブジェクトの文字はコード化されていないためテキスト抽出できない。本対策として、パスオブジェクトのみを含む頁は文書画像と同じく頁全体を文字認識処理する。具体的には、頁全体をレンダリング処理して BMP 画像に変換し、これに前述の文字認識処理を適用する。その結果、パスオブジェクトで描かれた非コード化文字もコードとして抽出し、同時に文字の位置座標と大きさ情報を得ることができる。

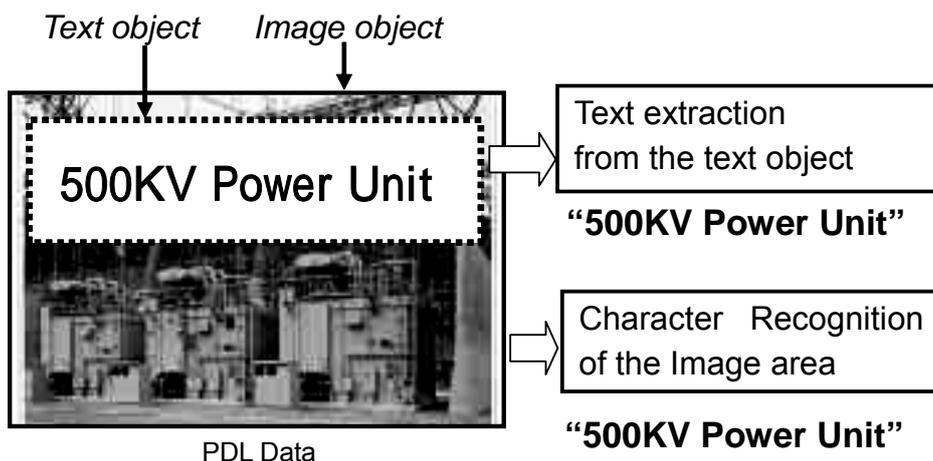


図 4.7 背景画像の文字認識

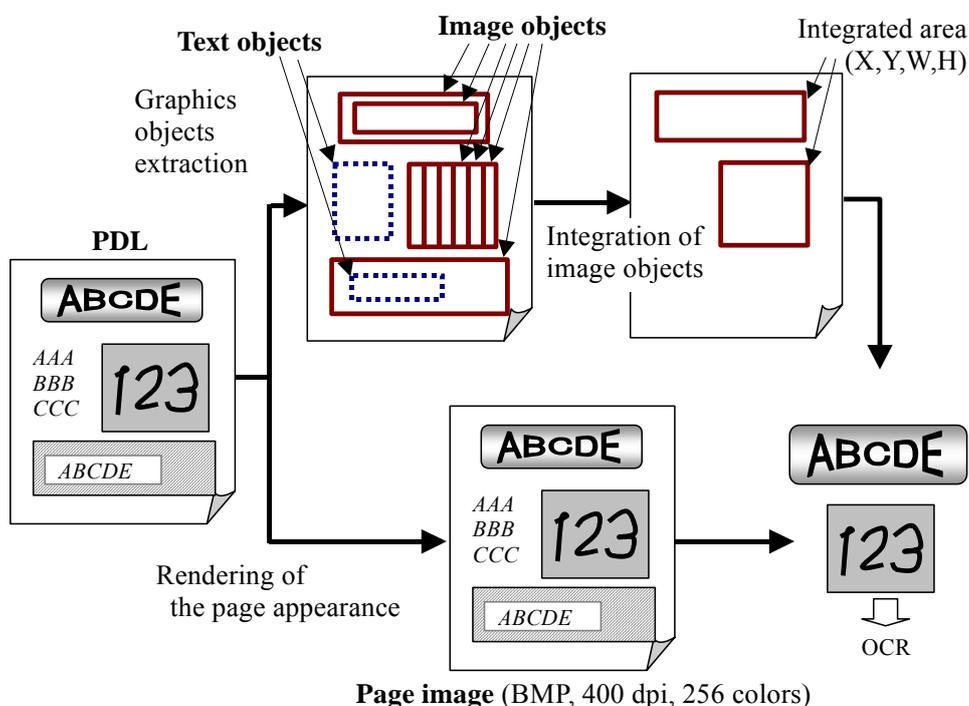
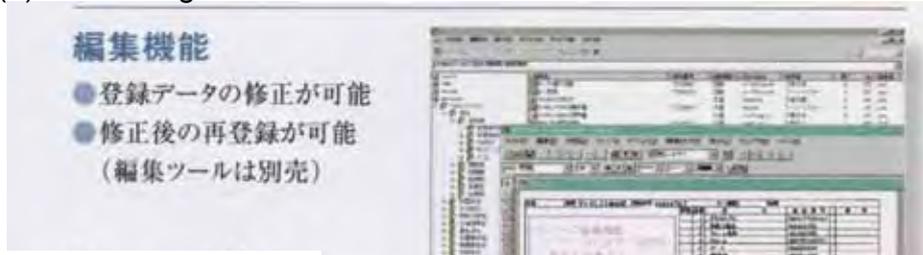


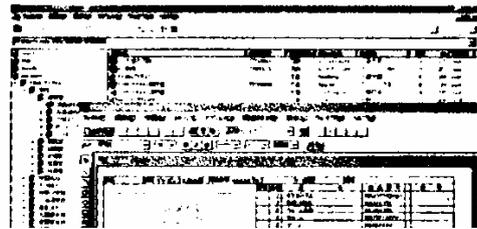
図 4.8 イメージオブジェクトの解析フロー

(a) Color Image



(b) Binary Image

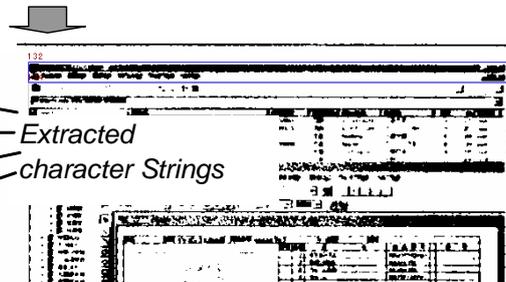
**編集機能**  
 ●登録データの修正が可能  
 ●修正後の再登録が可能  
 (編集ツールは別売)



(c) Extracted Strings

**編集機能**  
 ●登録データの修正が可能  
 ●修正後の再登録が可能  
 (編集ツールは別売)

← *Extracted character Strings*



Printed and Handwritten Character Recognition

Recognition Result

“編集機能 登録データの修正が可能 修正後の再登録が可能(編集ツールは別売り)”

図 4.9 イメージエリアの文字認識処理例

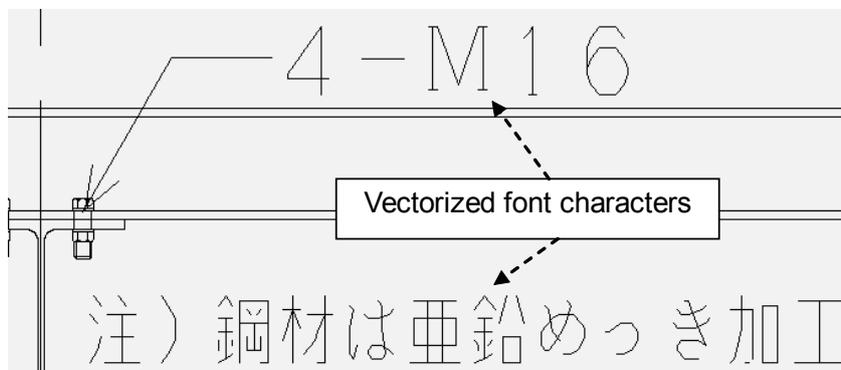


図 4.10 C A D 図面中の文字例

### 4.1.5 文書論理構造抽出処理

#### (1) 処理概要

PDL から抽出した文字コードの位置座標と大きさ, およびパスオブジェクトの情報を元に文書論理構造を抽出する. ここでは文書論理構造として章節・パラグラフ構造と表構造を抽出する. 表は四角い外形の単純な表のみを抽出対象とした. 図 4.11 は文書論理構造の抽出例である.

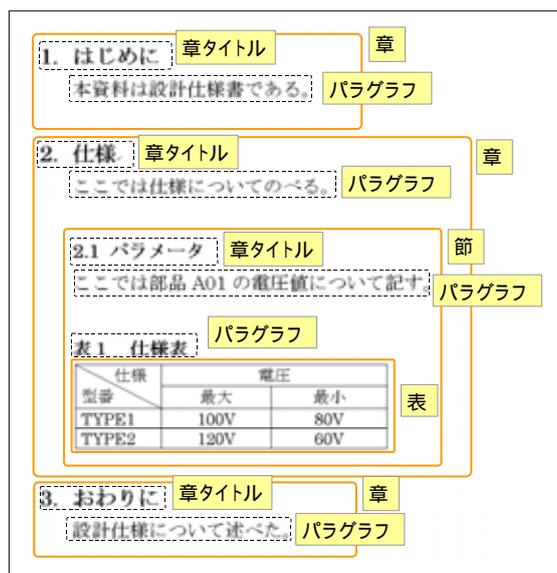


図 4.11 文書論理構造

図中の実線で囲んだ領域はブロックと呼び, ブロック内の破線で囲んだ領域はリージョンと呼ぶ. 各ブロックは「章節」か「表」の属性を, リージョンは「章節タイトル」か「パラグラフ」の属性を持つ. 抽出した結果は XML 形式のファイルに変換して出力する. この XML ファイルは, 文書論理構造と, PDL 解析処理で抽出したテキスト情報, およびリージョンや文字の位置座標と大きさを示すレイアウト情報を含む (図 4.12).

文書論理構造を抽出する既存方式として, 石谷[4-1]らは文書画像に文字認識処理や画像処理を適用して得たレイアウト情報を用いて章節構造, 箇条書き構造, 表構造, 図構造を抽出し, XML 形式で出力する方式を提案している. また渡邊[4-13]らは複雑な形式の表を解析する表構造解析法を提案している. ただし, これらの方式は文書画像を対象としている. また, 電子文書の種類である PDF ファイルから数式の構造を抽出する方式[4-14]が提案されているが, これは PDF を文書画像に変換して数式の構造を解析し, その際に文字認識誤りを PDF から取得したテキストで補正するものであり, 文書画像の解析を主としている. 同じく PDF ファイルに対しては表解析を行う方式[4-15]が提案されているが, 多様な電子文書を対象として表以外の文書論理構造も同一に処理しているものではない. また文

献[4-16]では文書画像だけでなく PDL データを対象とすることを想定して、グラフ構造を元にした表解析を提案している。

本研究では、これら既存の文書論理構造解析方式を応用することで、多様な文書から文書論理構造を抽出する。ただし文書画像に対する従来技術と異なり、PDL には読み順とは異なる順番でテキストが格納される。そのため抽出した論理構造に誤りが生じないように、文字列の位置を考慮して、近くの文字列は同一の論理構造内に含むように拡張した。

## (2) 章節・パラグラフ構造の抽出

以下の手順で処理を行い、段組のない文書に対して章節・パラグラフを抽出する。

- (a) 文字列の先頭位置が閾値よりも左側にあり、かつ先頭の文字が数字とピリオドのパターンを持つ文字列を探す。これを「章節タイトル」の属性を持つリージョンと判定する。
- (b) 残りの文字列からパラグラフを求める。具体的には、図 4.13 に示すように対象とする文字行  $i$  とその上にある文字列  $i-1$  との距離  $d2$  が 1 行以上あくか、文字列同士の重なり長さ  $d1$  が 0 の場合、これをパラグラフの先頭行と考える。先頭行から次の先頭行までの文字列群を「パラグラフ」の属性を持つ 1 個のリージョンと判定する。なお、文字列  $i$  と  $i+1$  は、テキスト抽出された文字列の順位ではなく、その印字位置が上下方向に隣接した文字列の並びを示す。PDL データでは、読み順にテキスト情報が格納されない場合があるが、文字列の印字位置を考慮して論理情報を抽出することで、読み順の誤りに起因した論理情報の抽出誤りを抑制する。

## (3) 表構造の解析

ここでは以下の処理に従い外形が四角形の表を抽出し、その内部構造を解析する。

- (a) 水平方向と垂直方向に細長いパスオブジェクトを選択して罫線候補とする（図 4.14）。
- (b) 罫線は、印刷した際は 1 本に見える。しかし実際は図 4.14 に示すように罫線同士が交差する位置で複数個のパスオブジェクトに分断されている場合が多い。そのため罫線候補を長さ方向に数画素分伸ばし、接触した同一方向を持つ複数の罫線候補群を 1 本の罫線と判定する。
- (c) 罫線同士の交点の形状を求める。ある 4 本の罫線が四角形を構成し、交点形状が「┌┐└┘」である場合、これらの罫線で囲まれた領域を「表」属性を持つブロックとする。そしてブロック内を縦・横方向に分割する罫線がある場合、この罫線で表内部を行、列に分けてセルを抽出する。最後に、セルの領域座標を元にして、そこに含まれる文字列を抽出する。

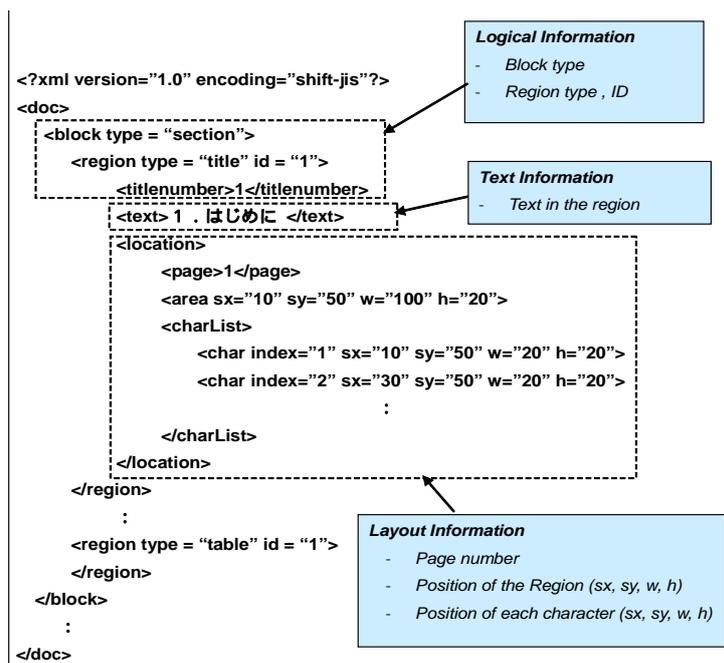


図 4.12 文書論理構造の抽出結果例

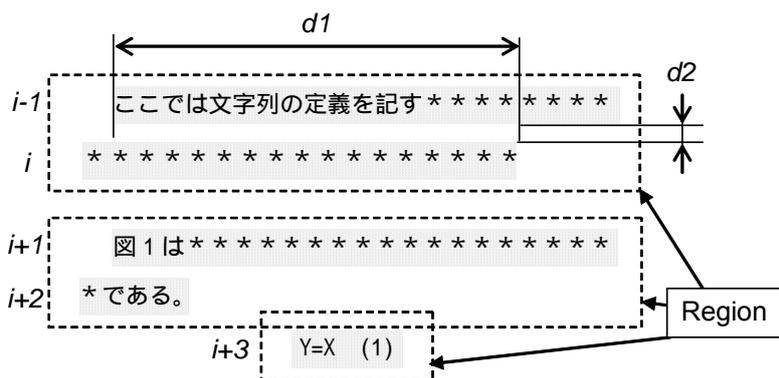


図 4.13 パラグラフの抽出例

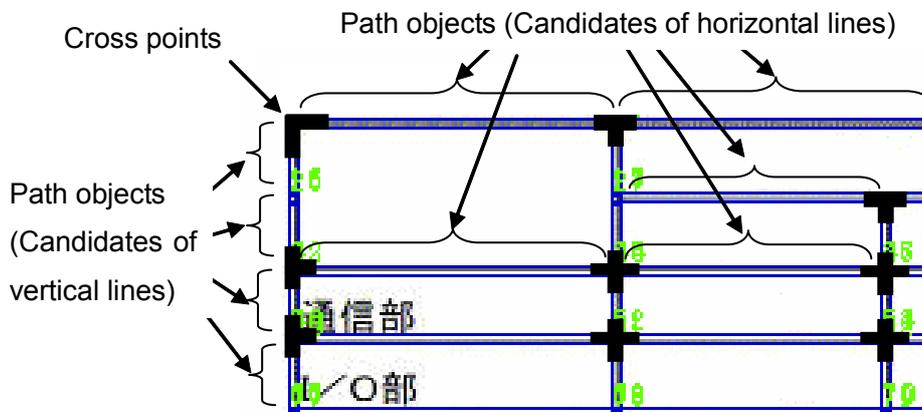


図 4.14 表構造解析例

## 4.1.6 評価実験

## (1) 内容情報の抽出率

複数のファイル形式に対して文字情報の抽出率を評価した。比較のため、iFilter [4-6]を用いた情報抽出方式と PostScript からの情報抽出方式[4-7]を従来方式として用いた。表 4.2 は評価に用いたファイル形式を示す。ここでは電子文書 9 種類、CAD 図面 5 種類、文書画像 2 種類の計 16 種類のファイル（1 ファイル/種類）を評価対象とした。実験は、表 4.3 に示すパソコンで実施し、このパソコンには印刷処理に必要なアプリケーションをインストールした。

表 4.2 評価用のファイル形式

Type	Document File Format(Extension of File Name)
Electronic document	Microsoft MS-WORD (DOC), EXCEL(XLS), PowerPoint(PPT), Visio(VSD), Adobe PDF(PDF), Fuji Xerox DocuWorks(XDW),Justsystem ICHITARO(JHT),Justsystem HANAKO(JHD), Deneva Canvas8(CNV)
Engineering drawing	Drawing Interchange Format(DXF), Autodesk AutoCAD(DWG), Autodesk AutoCAD(DWF), AutodeskInventor(IDW), Scadec exchange format(P21)
Scanned image	Tag Image File Format(TIFF), BMP/DIB(BMP)

表 4.3 実験環境

OS	Windows 2000 Professional
Memory	2GB
CPU	Intel Xeon 3GHz
Applications	MS-Office2003 (WORD, EXCEL, PowerPoint), Visio2003, Acrobat7.0, DocuWorks5.0, Ichitaro13, Hanako13, Canvas8, AutoDesk Voloview3, SXF Browser

提案方式と従来方式を用いて、電子文書、CAD 図面、および文書画像に対する内容情報の抽出率を評価した。評価は文書中から正しく抽出できた文字列数で行った。全文字列数は 3290 行で、これは人間が対象文書を目視して数えた。図 4.15 は評価結果を示す。ここで図の縦軸は文書中から正しく抽出できた文字列数であり大きいほど抽出率が高いことを示す。ここで提案方式は従来方式に比べて高い文字列抽出率を示した。これは、iFilter を用いた情報抽出方式と PostScript からの情報抽出方式が、文書に含まれる画像中の文字や図面中のパスオブジェクトで描かれた文字などの非コード化文字を抽出できないが、提案方式ではこれらを抽出可能なことから生じた差である。この結果から提案方式は従来方

式に対して文字情報を欠損なく抽出し、内容情報の抽出率が高いと言える。

ここでテキストオブジェクトから文字列は、理論的に全ての方式で同じである。しかし、提案方式と PostScript からの情報抽出方式で抽出した文字列数が、iFilter を用いた情報抽出方式で得た文字列数よりも多くなった。この差異は、今回の実験に対して一部のファイル形式に対応した iFilter プログラムが利用できなかった（存在していない、もしくは有償であった）ことに起因する。iFilter を用いた情報抽出方式はファイル形式に対応した個別の iFilter プログラムが必要だが、提案方式と PostScript からの情報抽出方式は多様なファイル形式に適用できるため、このような問題は発生しない。

なお、文書画像に対する PostScript からの情報抽出方式[4-7]と提案方式は同じ文字列抽出精度となった。これは文献[4-7]に記載の通り、TIFF 形式などの文書画像については PostScript 化を行わずに通常の文字認識処理を適用したためである。

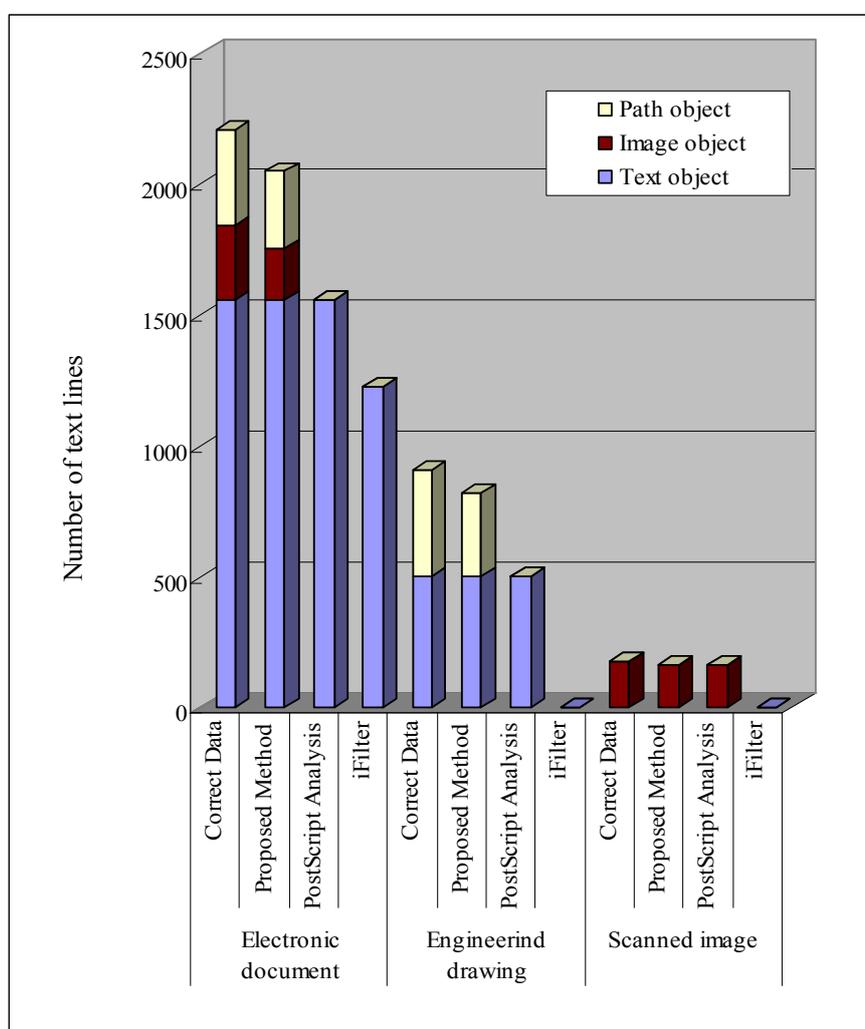


図 4.15 文字列抽出結果

(2) 内容情報の抽出時間

内容情報抽出に要する処理時間を従来方式と比較した。測定には既存のファイル 1831 個、計 8113 頁を用いた(表 4.4)。その結果を図 4.16 に示す。図の横軸は 1 ファイルからの情報抽出に要した平均処理時間である(文書構造解析処理時間は含まない)。ここで CAD 図面と文書画像に対する iFilter の情報抽出速度を”not measured”としているが、これは CAD 図面と文書画像に対する iFilter が利用できなかったことによる。結果から iFilter が最も高速であり、特定のファイル形式を多量に処理するには有効であることが分かる。提案方式は最も処理時間を要した。PostScript を用いた情報抽出法と比較すると、提案方式において PDL 解析処理時間が長い。これは提案方式がテキストオブジェクト以外のグラフィック要素も解析しているためである。なお、文書種類毎の比較では CAD 図面で最も長い処理時間を要した。これは図から分かるように、CAD のアプリケーションにおける擬似的な印刷処理が、電子文書のアプリケーションに比べて重いことが主な原因である。

表 4.4 処理時間測定用の実験ファイル

Electronic document	MS-WORD (DOC) 892 files, EXCEL (XLS) 351 files, PowerPoint (PPT) 3 files, Adobe PDF (PDF) 297 files
Engineering drawing	Drawing Interchange Format (DXF) 102 files, AutoCAD (DWG) 96 files, Scadec exchange format (P21) 6 files
Scanned image	Tag Image File Format(TIFF) 82files

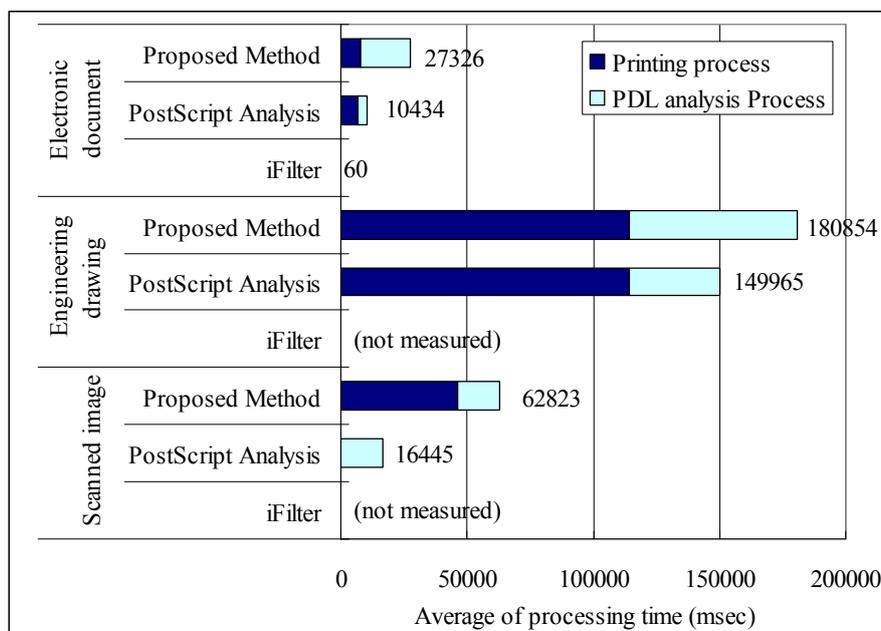


図 4.16 情報抽出処理時間

(3) 内容情報の抽出結果例

PDL 解析による内部情報の抽出結果例を図 4.17 と図 4.18 に示す . 図 4.17 はパスオブジェクトから抽出した罫線であり , 電子文書中の表や数式中の罫線が位置情報と共に抽出されていることが分かる . 図 4.18 はテキストオブジェクト , イメージオブジェクト , パスオブジェクトを含む頁から内部情報を抽出した例である . 図中の小さい矩形は抽出された 1 文字の位置と大きさを示す . ここでテキストオブジェクトとイメージオブジェクト ( グラフの画像 ) の両方から文字が抽出されている . この結果から , コード化文字と非コード化文字の両方に対して , 文字の印字位置と大きさが正しく特定されていることが分かる .

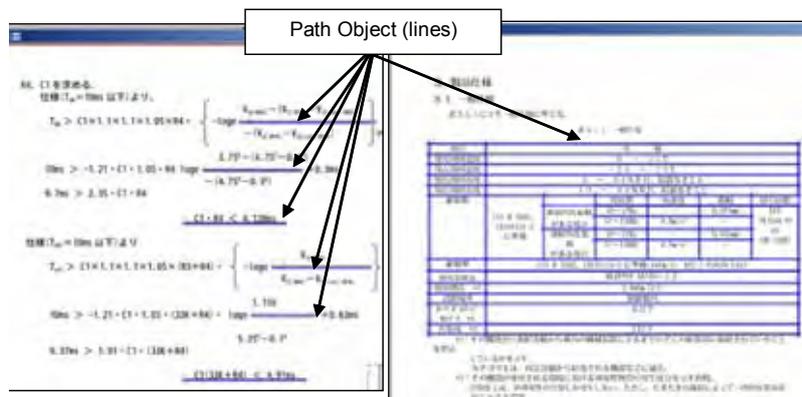


図 4.17 パスオブジェクトからの罫線抽出例

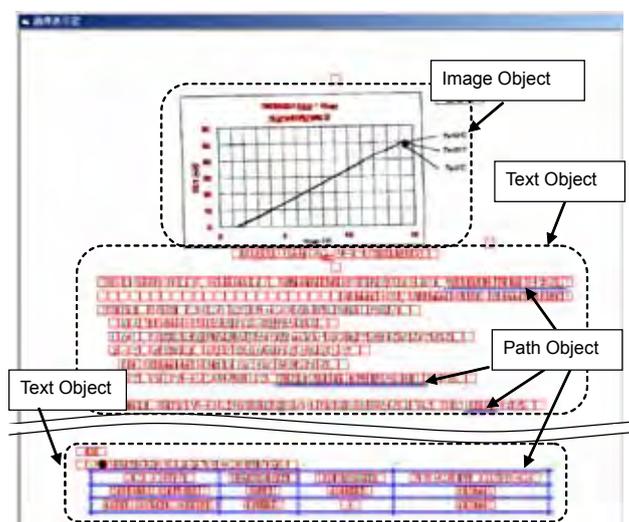


図 4.18 頁からの文字抽出例

(4) 文書論理構造抽出精度

表 4.5 に示す 3 個の MS-WORD ファイルを対象として電子文書に対する文書論理構造情報の抽出精度を検証した . これらの文書は合計 114 頁 , 章 35 個 , 節 59 個 , 表 82 個を含む .

パラグラフは人により数がばらつくため計測対象から除いた。これらの文書は電気回路系の仕様書であり、章節の論理構造が明確でレイアウトが比較的整理されている(図4.19)。

これらの文書から文書論理構造を抽出した結果を表4.6に示す。ここで章節構造は100%抽出でき、1709個のパラグラフが抽出された。表については抽出精度が94%であり5個抽出に失敗した。抽出に失敗した表の3個は、表の外形が複雑で矩形でないことが原因であった。残りの2個は表の途中で罫線の太さが変化しており、そこで分断されたパスオブジェクトの統合が罫線太さの差による座標値のズレのために失敗したことが原因であった。

表 4.5 文書論理構造抽出の評価用文書

	File size (KB)	Pages	Sections	Sub-sections	Paragrahs	Tables
Document A	1,656	47	14	15	-	30
Document B	17,908	31	8	34	-	29
Document C	11,395	36	13	10	-	23
Total	30,959	114	35	59	-	82



図 4.19 文書論理構造抽出に用いた文書の頁例

表 4.6 文書論理構造の抽出結果

	XML file size (KB)	Section	Sub-section	Paragraphs	Tables	Characters
Document A	2344	14 (100%)	15 (100%)	665	27 (90%)	40245
Document B	1549	8 (100%)	34 (100%)	752	27 (93%)	25223
Document C	2252	13 (100%)	10 (100%)	292	23 (100%)	39715
Total	6145	114 (100%)	35 (100%)	1709	77 (94%)	105183

最後に、実際に抽出した文書論理構造を XML 形式で出力した例を図 4.20 に示す。章節や表の構造が抽出され、かつテキストと位置情報が出力されている。この結果から、比較的単純な文書論理構造であるが、従来はレイアウト解析が困難であった電子文書からも文書論理構造情報が抽出できていることが分かる。

```

- <block type="section">
- <region type="title" id="31">
  <titlenumber>3</titlenumber>
  <text>3. 製品仕様 </text>
  <location>
    <page>5</page>
    <area sx="311" sy="316" w="527" h="77" />
  - <charList>
    <char index="40" sx="311" sy="316" w="77" h="77" />
    <char index="41" sx="388" sy="316" w="77" h="77" />
    <char index="42" sx="465" sy="316" w="77" h="77" />
    <char index="43" sx="444" sy="316" w="77" h="77" />
    <char index="44" sx="522" sy="316" w="77" h="77" />
    <char index="45" sx="600" sy="316" w="77" h="77" />
    <char index="46" sx="683" sy="316" w="77" h="77" />
    <char index="47" sx="761" sy="316" w="77" h="77" />
  </charList>
  </location>
</region>
- <block type="section">
- <region type="title" id="32">
  <titlenumber>3.1</titlenumber>
  <text>3.1. 一般仕様 </text>
  <location>
    <page>5</page>
    <area sx="311" sy="450" w="532" h="66" />
  - <charList>
    <char index="48" sx="311" sy="450" w="66" h="66" />
    <char index="49" sx="377" sy="450" w="66" h="66" />
    <char index="50" sx="394" sy="450" w="66" h="66" />
    <char index="51" sx="461" sy="450" w="66" h="66" />
    <char index="52" sx="472" sy="450" w="66" h="66" />
    <char index="53" sx="505" sy="450" w="66" h="66" />
    <char index="54" sx="572" sy="450" w="66" h="66" />
    <char index="55" sx="638" sy="450" w="66" h="66" />
    <char index="56" sx="711" sy="450" w="66" h="66" />
    <char index="57" sx="777" sy="450" w="66" h="66" />
  </charList>
  </location>
</region>
- <region id="33">
  <text>表 3.1.1 に示す一般仕様に従って、表 3.1.1 一般仕様 </text>

```

(a) 章節の記述例

```

- <block type="table">
- <region type="body" id="79">
- <row>
- <col colitem="" rowitem="" colspan="1" rowspan="1">
  <region id="80">
    <text>No. </text>
    <location>
      <page>10</page>
      <area sx="388" sy="2283" w="189" h="83" />
    - <charList>
      <char index="245" sx="444" sy="2300" w="55" h="55" />
      <char index="246" sx="472" sy="2300" w="55" h="55" />
      <char index="247" sx="500" sy="2300" w="55" h="55" />
      <char index="248" sx="533" sy="2300" w="55" h="55" />
    </charList>
    </location>
  </region>
</col>
- <col colitem="" rowitem="" colspan="1" rowspan="1">
  <region id="81">
    <text>名 称 </text>
    <location>
      <page>10</page>
      <area sx="577" sy="2283" w="706" h="83" />
    - <charList>
      <char index="249" sx="816" sy="2300" w="55" h="55" />
      <char index="250" sx="872" sy="2300" w="55" h="55" />
      <char index="251" sx="933" sy="2300" w="55" h="55" />
      <char index="252" sx="988" sy="2300" w="55" h="55" />
      <char index="253" sx="1050" sy="2300" w="55" h="55" />
    </charList>
    </location>
  </region>
</col>
- <col colitem="" rowitem="" colspan="1" rowspan="1">
  <region id="82">

```

(b) 表の記述例

図 4.20 生成された XML ファイルの例

## (5) 内容情報抽出の課題と考察

本方式で抽出した内容情報抽出結果を調べたところ、以下の課題が確認された。これらは印刷を主目的としたPDLから内容情報を抽出したことにより発生した。

(1) PDLではテキストが読み順通りに格納されない場合がある。また図や画像からも文字列が抽出できるため、抽出された文字列が混在して読み順が損なわれる場合がある。このような読み順を誤った文字列は、文書検索時の漏れやゴミの要因となる。図4.21は、Document Bの回路設計に関する記述部分（10頁、図や画像を含む）を対象に、抽出した文字列の連続性を調べた結果を示す。ここでは文字列の連続性を文字領域と図・画像領域に分けて調べた。文字領域から得た260文字列では、93%が正しい連続性を得た。残りの7%は、図領域の文字列と文字領域の文字列が混在することにより連続性を誤った。例えば、図4.22(c)に示す文字領域と図・画像領域の混在箇所、文字列の順序を と誤った。一方、図・画像の領域では正しい連続性を得た文字列は67%となった。残りの22%は図4.22(a)の に示すように位置が離れた複数個の文字列を“COM+XOYIF”という1個の文字列に誤結合したもの、1%は図4.22(b)における“ドライバIC”と連続すべき文字列を“ドライバRDIC”と抽出するような図内での抽出順番の誤り、10%は図領域と文字領域から得た文字列が混在することによる誤りであった。このような文字列連続性の誤りについては、今後、提案技術の後段にレイアウト情報を用いた文字列連続性の判断処理を追加するなど、対策を検討する必要がある。

(2) PDF文書から作成したPDLは、テキストオブジェクト内に文字コードを持たず、グリフ（文字の形状情報）のみを保持する場合があります。このテキストオブジェクトからはテキストを抽出できない課題がある。これについては、文字コードを持たないテキストオブジェクト領域をイメージ化して文字認識処理する対策が考えられる。

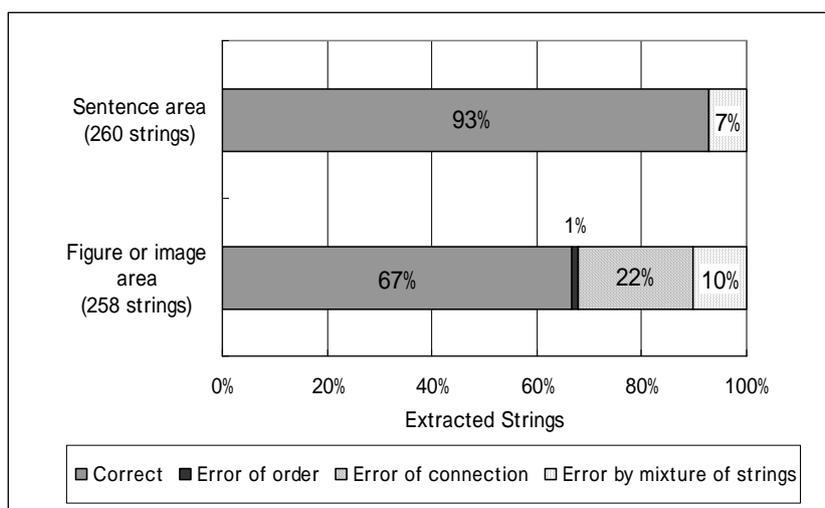


図 4.21 文字列の連続性

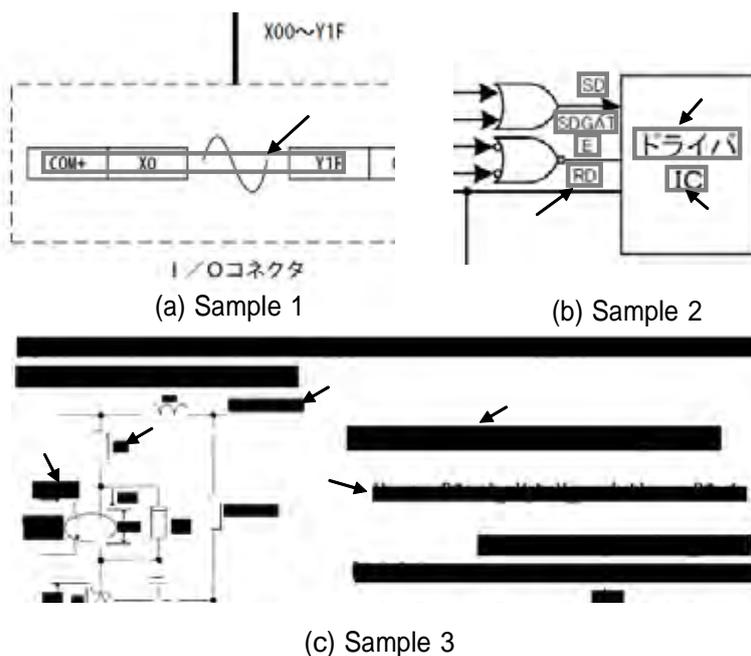


図 4.22 文字列の連続性誤り例

#### 4.1.7 メタ情報抽出処理とカテゴリへの写像

最後に、文書から抽出した論理情報と、テキストに対する形態素解析処理結果、およびアプリアリな知識に従いメタ情報を抽出する。アプリアリな知識は以下の例に示すように人間が文書を読んで内容を理解する上で、あたり前と思われる一般常識的な知識(ルール)を適用した。

##### メタ情報抽出のためのアプリアリな知識例

- ・“担当”という文字列の近くにある名詞句は<担当者名>である。
- ・数字+“年”+数字+“月”というパターンに合致した文字列は<日付>である。
- ・国名、市町村名辞書と一致する文字列は<地名>である。
- ・“費用”という文字列の右側にある数字は<金額>である。

このメタ情報抽出結果を利用してカテゴリ毎に分類する。図 4.23 は、その利用例を示す。ここではエアコンの点検記録書(手書き、電子データが混在)に対して上記のアプリアリな知識を適用してメタ情報を抽出した。このメタ情報を用いて文書の検索結果を地域別な視点や、エアコンのユニット別の視点、機種別の視点、室外機部品別の視点といった、異なるカテゴリで分類できる。製品や社内利用などで、Web の文書を地図に写像したり、設計関連文書を組み立て図やグラフに写像するシステムを開発し、有用性を確認した(図 4.24)。

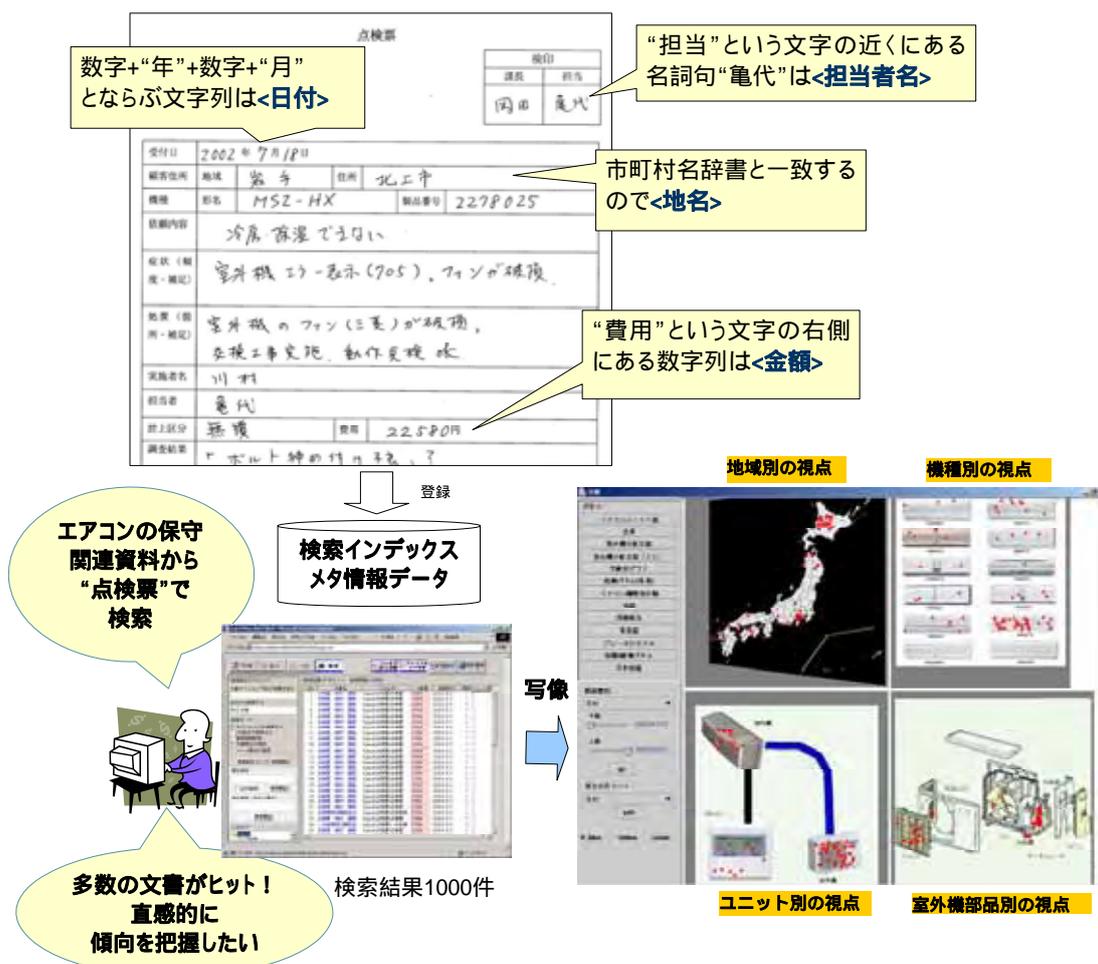


図 4.23 エアコン点検記録表に対するメタ情報抽出・写像例

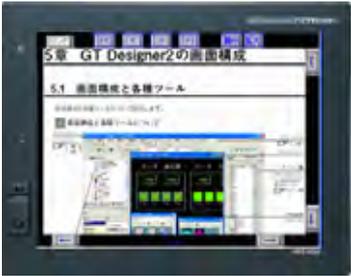
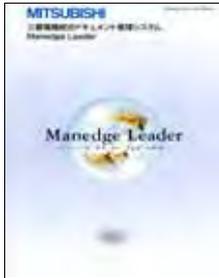


図 4.24 メタ情報抽出・写像を行う その他のシステム例

### 4.1.8 文書とカテゴリ間写像を用いた人間支援システム

ここで提案した文書から論理情報抽出技術とメタ情報抽出は、文書の検索や検索結果の自動分類機能を備えた4種類の文書検索・知識管理システムに実適用した(表4.7)。

表4.7 文書からの論理情報抽出技術，メタ情報抽出技術を用いた人間支援システム

人間支援システム	画面例
自治体,中央官庁 向など特殊用途向け 文書・図面管理シス テム(00年～)	
F A表示装置用 組込み文書閲覧・検 索機能 (05年～)	
企業向けドキュメ ント管理パッケージ (03年～)	
社内 設計文書共 有システム 4製作所 (03年～)	

## 4.2 文書と関連知識間の写像

文書と関連知識間の写像では、人間が文書の内容から関連する知識（情報）を想起する処理を模擬する。具体的には、文書と知識間において内容の構造的な類似性を判定し、これらの写像を獲得する技術を提案する。この写像の特長を表4.8に示す。本技術を用いることで、設計書を作成している最中に、注意すべき過去の不具合事例や設計知識をシステム側からプッシュ型で自動提示する設計不具合防止支援システムを実現する（図4.25）。

表4.8 文書と関連知識間の写像モデルの特長

写像関係	文書と関連知識との写像。
適応性	設計知識の作成が容易である。また、オントロジは既存の部品データベースや資材データベースを利用して作成可能。そのため製造場所への適応が容易。
写像の種類	用語同士の構造的な類似性を用いて、文書と知識との関連性を計算する。部品情報をオントロジ（知識）として簡易推論を行うことで、記述の抽象度が異なっても関連性を計算できる。
脳内の関係部位	言語野/連合野（前頭葉）
人間支援システム	作成中の設計資料に関連した、注意事項などをプッシュ型で自動提示する設計不具合防止支援システム

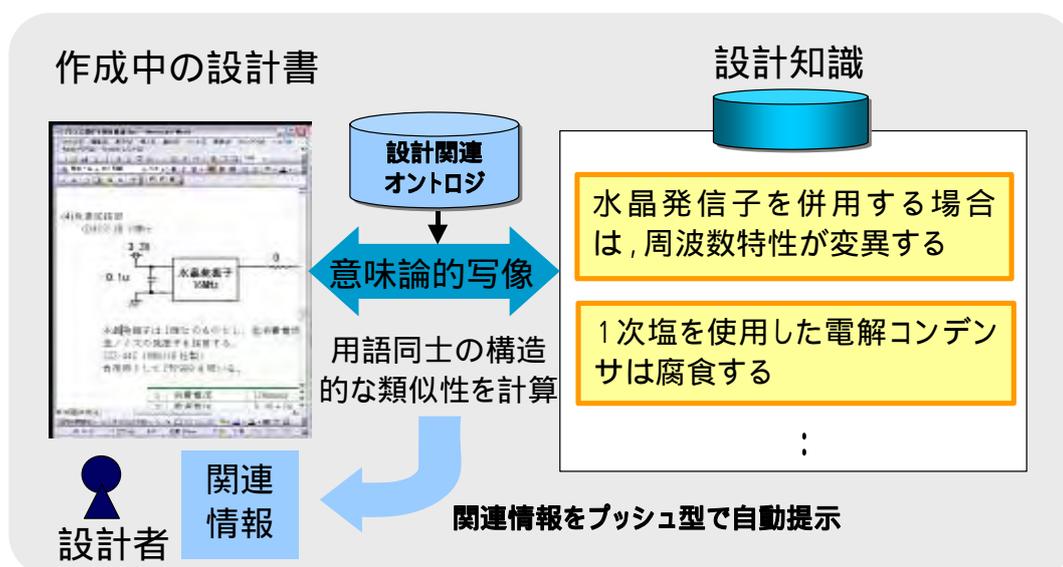


図4.25 文書と関連知識間の写像概念図

#### 4.2.1 文書と関連知識間の写像に関する背景と目的

近年、製造物に対するリコール問題や基準不適合など、設計段階でのミスや見逃しによるトラブルが社会問題となっている。一方、製造業の開発・設計現場においては、製品のライフサイクル短縮、多品種少量生産、製品の高機能化により開発期間の短縮が余儀なくされ、設計者への負担がますます大きくなっている。また、ベテラン設計者の退職による技術伝承の問題（2007年問題）の顕在化も懸念されている。また、Cusumano は不具合と製品コストの関係では、市場に流失した不具合の対応に要するコストは、設計段階で対応したコストの約100倍近いという調査結果を示した[4-18]。さらに著者らの調査によれば、製品の不具合に含まれる約半分が、過去に発生した不具合事例の何らかの再発であった。これらの背景から、設計者の負担をかけずに、開発の初期段階で設計書の品質を向上させる要求が高まっている。

このような設計初期段階での品質向上を目的として、従来は検索エンジンを搭載した不具合情報蓄積システムが広く利用されている。このシステムでは不具合事例をデータベースに蓄積しておき、新規設計時において設計者がキーワードにより関連する不具合事例を探すものである。このシステムでは不具合事例を登録するだけであり、設計知識の作成とメンテナンスは比較的容易である。ただし、設計者の熟練度により適切なキーワードを入力できずに重要な情報が漏れるなど、取得した設計知識の精度が低い課題がある(図4.26)。

これに対して、ルールベースの推論を行うエキスパートシステムを用いることで取得した設計知識の精度を高めることができる[4-19][4-20][4-21]。このような設計支援のためのエキスパートシステムとして Tamura ら[4-20]は、設計不具合の発生要因を因果連鎖として記述する Stress-Strength Model (SSM)を提案した。このSSMでは不具合の発生要因を分解して因果連鎖のための要素を作成し、これら要素を自動的に結合することにより、未知の不具合の発生を推論できる。また Suto らは因果連鎖の結合を行う Physical Causal Network(PCN)を提案し、設計支援システムを構築した[4-21]。これらのシステムでは、十分な設計知識が提供された場合、設計に関する問題を効果的に抑制できる。しかし、このような設計知識を作成するには、熟練設計者が不具合の発生要因を分析し、そこから因果連鎖の要素を適切に作成していく必要がある。そのため、実用においては、知識作成とメンテナンスのコストが多大となる課題がある。

これらの課題に対応するため、我々はプッシュ型設計支援システム (Push-Style Design Support System: PSDSs) を提案する。PSDSs は設計者が設計書を作成中に、関連した設計知識を自動提示する。PSDSs は、ユーザのタスクに関連して情報提供を行う Schilit[4-22] や Yunwen[4-23]らが提案した Criticism systems の一種である。このシステムではキーワード入力が必要ないため設計者の必要な設計知識の参照漏れを抑制でき、かつ PSDSs は設計知識の作成がルールベースのエキスパートシステムよりも容易である。

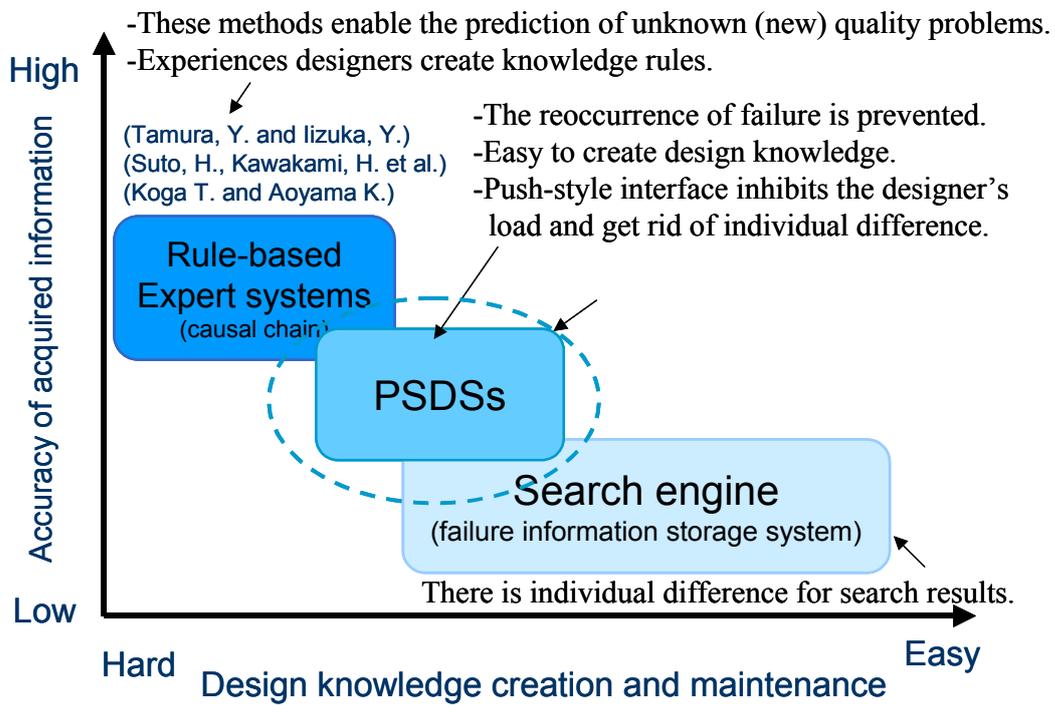


図 4.26 PSDSs のコンセプトと狙い

### 4.2.2 プッシュ型設計支援システム ( PSDSs ) の構成

PSDSs のシステム概要を図 4.27 に示す．本システムは詳細設計段階に利用する．設計者がワープロ ( MS-WORD ) で設計書を作成中，カーソル付近の文書内容を知識サーバに自動送信する．知識サーバは受信した文書内容と設計知識とを比較して類似性を計算し，類似性の高い設計知識をガイダンスメッセージとしてクライアント PC に送付する．ここで PSDSs は以下の特長を持つ．

- ・ プッシュ型の情報提供を行うため，通常の検索エンジンと異なり検索キーワードの入力が不要である．そのため設計者の負荷が減少し，かつ設計者のキーワード選択誤りによる設計知識の確認漏れが抑制できる．
- ・ 設計知識は不具合事例から作成する．設計知識は，不具合が生じた原因を複数個の用語 ( 回路名や部品名，材質など ) で表現する．因果連鎖を作成できず未知の不具合を推測することはできないが，作成が容易であり実運用に適する．
- ・ 設計書は具体的に記述されるが，設計知識は一般化され抽象的に記述される．PSDSs はオントロジを用いて抽象度の差を吸収するため，これらの高精度な照合が可能．
- ・ PSDSs は設計者がガイダンスメッセージを確認したかを記録し，その結果をレポート出力する機能を持つ．このレポートを DR 会議 ( Design rationale meeting ) に提出して関係者で確認することにより，人間系による最終確認を充実化できる．

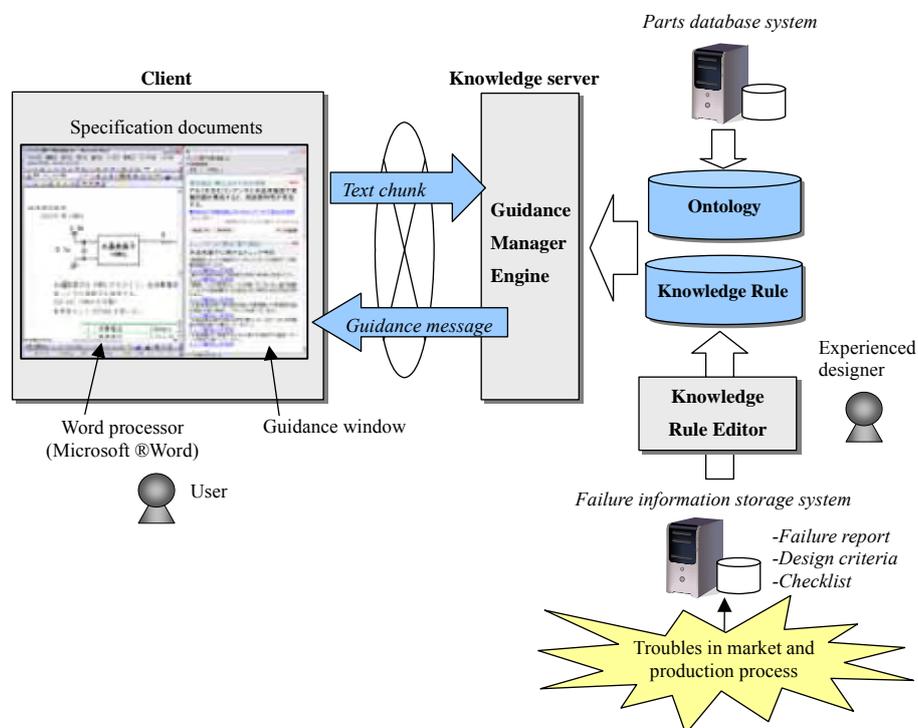


図 4.27 PSDSs のシステム概要

### 4.2.3 設計知識とオントロジ

#### (1) 設計知識

熟練設計者が不具合事例やチェックリストの内容を元に設計知識を作成する。基本的に 1 個の不具合事例から 1 個の設計知識を作成する。設計知識は「条件部」と「メッセージ部」から構成する。条件部は有向グラフを用いて不具合が発生した条件を記述する。この条件部の各ノードはキーワードを表す。キーワードは対象物名や属性名であり、例えば“Air compressor”, “Power supply circuit”, “Oscillation circuit”, “Crystal oscillator”, “Capacitor”, “Aluminum”, “Voltage”などである。ノード間のエッジは、これら対象物名や属性名間の関係性を記す。例えば、図 4.28 は不具合事例と、そこから作成した設計知識の例を示す。この条件部と類似した構造が設計書中から検出された時、メッセージ部の情報をクライアントに対してプッシュ型で送信する。

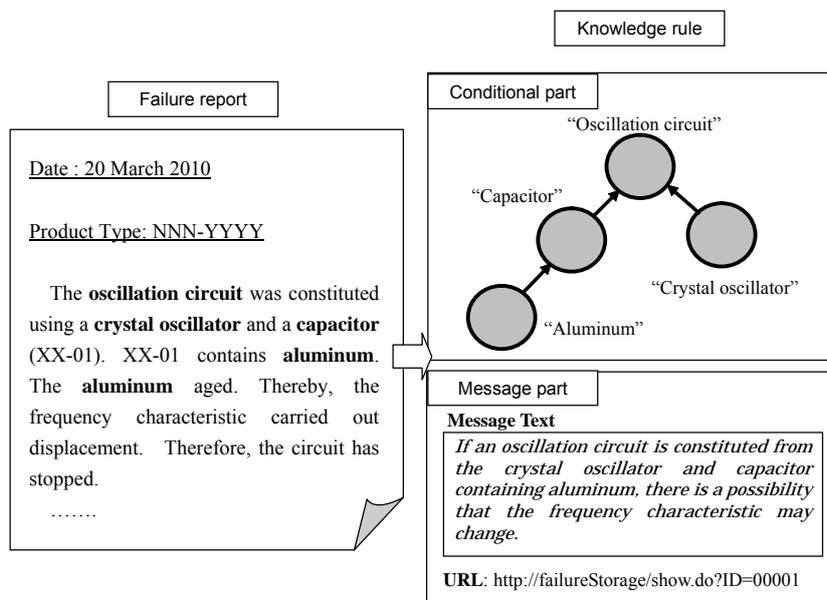


図 4.28 設計知識のサンプル ( 実例ではない )

#### (2) オントロジ

本システムはシソーラス辞書と部品データをオントロジとして利用する ( 図 4.29 ) 。部品データは設計・製造現場が持つ製品製造用の部品データベースから、必要部分を抜き出したものである。この部品データは部品名、部品型番、メーカー名、品質情報、材質などが格納される。一方、シソーラス辞書は、条件部に用いた各用語の同義語や上位語、下位語を記す。このシソーラス辞書は工場の資材部門が持つ資材調達用データベースの内容を元に、熟練設計者が情報を追加して作成する。

Level 1	Level 2	Level 3	Level 4	Level 5
ROOT				
└	Electronic parts			
	└	Capacita	Condenser	
		└	Electrolysis condenser	
			└	Aluminum electrolytic capacitor
			└	Tantalum electrolytic capacitor
			└	Nonpolarized electrolytic capacitor
			└	Solid electrolytic capacitor
			└	Paper condenser
			└	Film capacitor
			└	Ceramic capacitor
			└	Laminated ceramic capacitor
	└	Crystal oscillator	Crystal unit	Quartz crystal
	└	IC		
	└	Fan		

(a) thesaurus dictionary

Parts name	maker model name	Maker	quality	materials
Tantalum electrolytic capacitor	ZUU62000	ABC	1 A	-
Aluminum electrolytic capacitor	ZUV5800	ABC	1 A	Aluminum
Crystal unit	XAQ1 04	ABC	1 A	-
Crystal unit	PPP888	ABC	1 A	-
Crystal unit	PPP999	ABC	1 A	-

(b) Parts data

図 4.29 オントロジデータの例

### 4.2.4 写像獲得処理

知識サーバでは、設計書の内容と設計知識とを下記プロセスに従い照合し、それらの類似性を求める。そして類似性を元に、設計書の章節と設計知識間の写像関係を形成する。

(1)設計者がワープロ(MS-WORD)上でカーソルを移動させたか、文字を入力して改行した際、その章の内容と、章節タイトルを XML データに変換して知識サーバに送信する(図 4.30)。章の内容や章節タイトルの取得は MS-WORD のマクロ機能を用いて実現した。

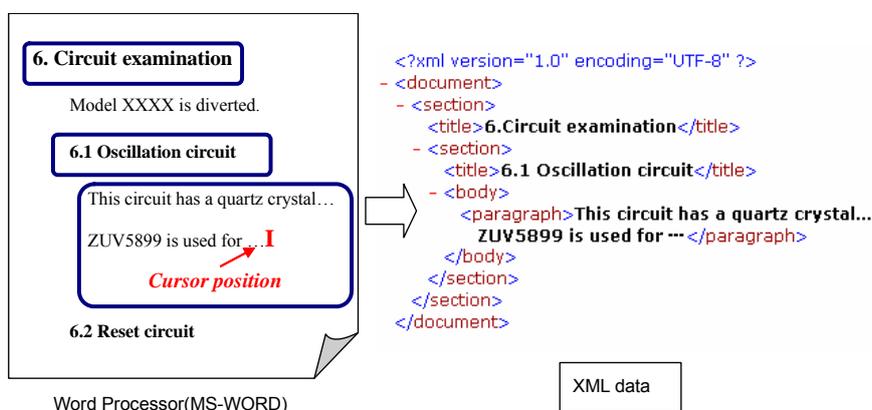


図 4.30 設計書からの XML データの抽出

(2)知識サーバは、XML データ中から部品型番を探す。部品型番が見つかった場合、オントロジの部品データを参照して、XML データ中に部品名、メーカー名、材質名などの拡張情報を追加する(図 4.31)。

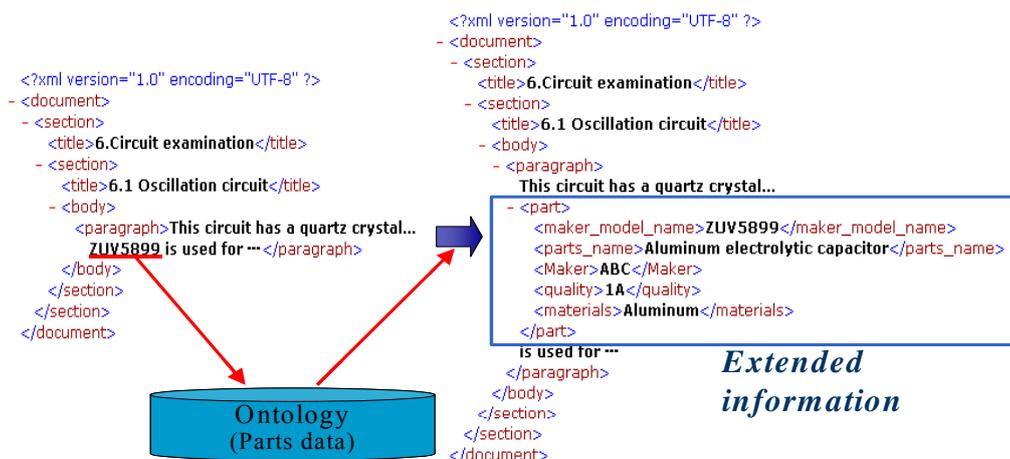


図 4.31 部品データを用いた XML データの拡張

(3)次に、設計知識を構成する「条件部」のノードを構成する用語“Oscillation circuit”，“Capacitor”などを XML データ中から探索する。この構造探索では、オントロジのシソーラス辞書を利用して用語を拡張しながら行う。ことにより、設計知識と設計書間の抽象度の差を吸収する。例えば、設計知識の“Capacitor”は XML データ中の“Aluminum electrolytic capacitor”の上位語であるため、これらの用語が対応付く。

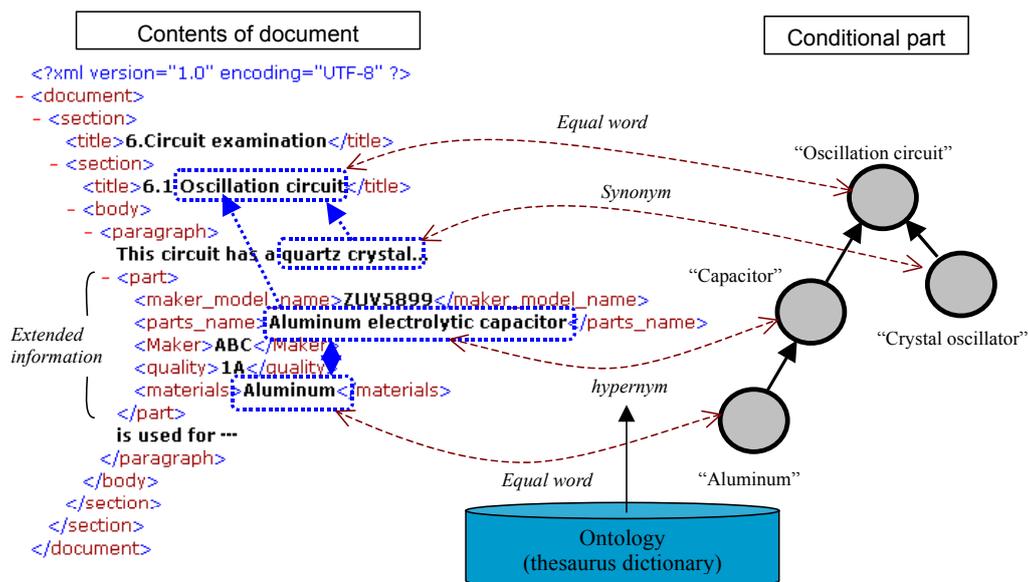


図 4.32 照合処理

(4)次に構造探索結果から類似性のスコアを計算する。このスコア計算では条件部の有向グラフの向きを考慮し、次式に従い行う。グラフの向きを考慮することにより、単純なキーワードの有無による検索処理よりも高精度に類似性を判断できる。

$$\begin{aligned}
 & \text{If}((A \leftarrow B) \cap (A' \text{ is in the upper level of } B')) \text{ Then the score is high} \\
 & \text{If}((A \leftrightarrow B) \cap (A' \text{ and } B' \text{ are in the same level})) \text{ Then the score is high}
 \end{aligned}
 \tag{4.3}$$

ここで、A, B は条件部のノードが持つ用語であり、A' と B' は XML データ中から検出された、A, B に対応付く用語である。シンボル“←”は、条件部の有向グラフにおける A, B 間のエッジの向きを示す。

(5)最後に、構造の類似性がある閾値よりも大きい設計知識が見つかった場合、この XML データを取得した設計書の章節と、設計知識間に写像を作成する。そしてメッセージ部の情報をガイダンスメッセージとしてクライアント PC に送付する。

図 4.33 は本処理によって実際に提示されたガイダンスの例である。図の左側が対象とした設計書、右側が提示されたガイダンスメッセージである。右上に「アルミを含むコンデンサと水晶発振器で発振回路を構成すると周波数特性が変位する」というガイダンスメッセージが提示されているが、設計書中には“アルミ”や“コンデンサ”という記述はない。図 4.34 は、このガイダンスメッセージの照合過程を示す。図中の赤色が条件部を式で記述したものである。また白色の部分が設計書中に存在した文字列である。水色は設計知識と設計書間を繋ぐ簡易推論の過程を示す。以下にその詳細を示す。

- ・ 領域 (A) の部分：まず文書中から文字列“ PPP999 ” が抽出され、それが部品型番だと判断される。そして部品データの内容から、その部品が“ 電解コンデンサ ”であり、材質が“ アルミ ”であると分かる。そのため赤字で記述した条件 (材質=“アルミ”, 部品名= “電解コンデンサ”) と合致する。
- ・ 領域 (B) の部分：文書中の文字列“ 水晶発振器 ”が“ 水晶発信器 ”の同義語と判断され、赤字で記述した条件 (“ 水晶発信器 ”) と合致する。
- ・ 領域 (C) の部分：設計文書の章節タイトル内に“ 発振回路 ”という単語が含まれていたため、赤字で記述した条件 (“ 発振回路 ”) と合致する。

上記のように設計書と設計知識間の抽象度の差を簡易的な推論を行い繋ぎ合わせることにより、漏れなく精度の高い写像が獲得できる。

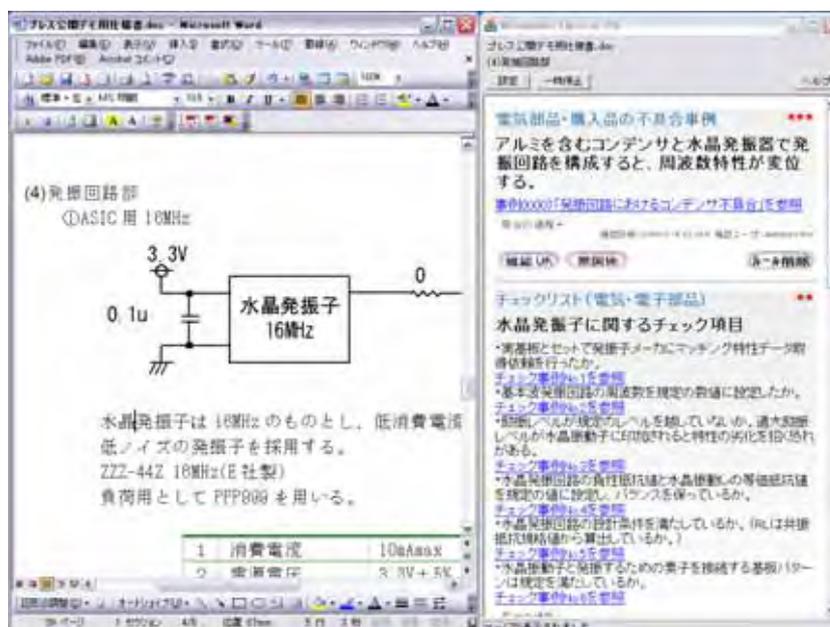


図 4.33 ガイダンスメッセージの出力例

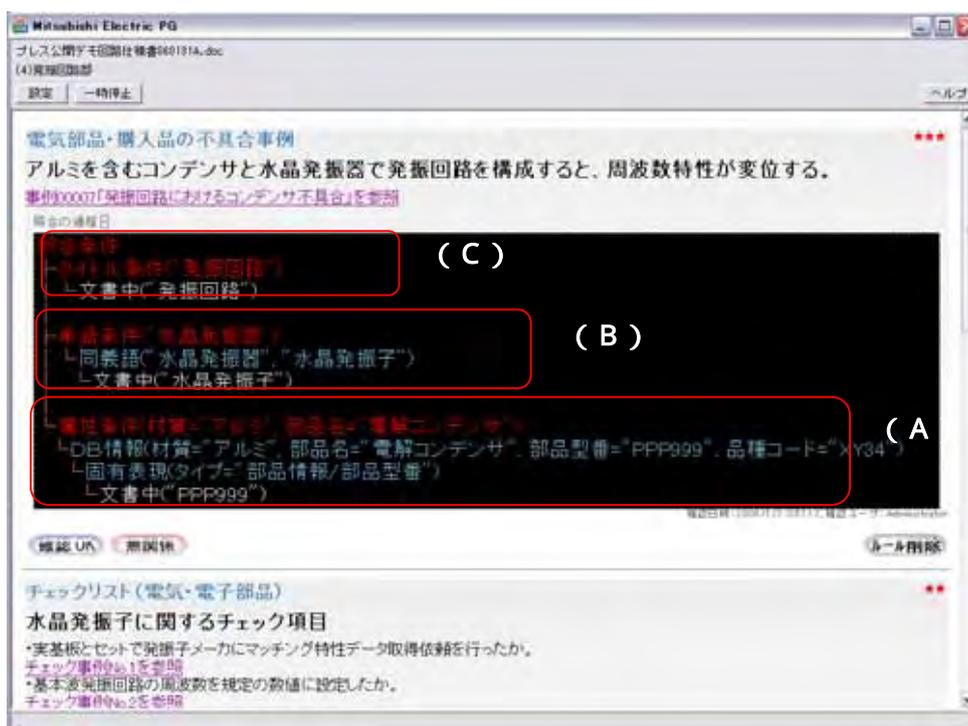


図 4.34 ガイダンスメッセージの出力例

#### 4.2.5 ガイダンスメッセージの確認・レポート出力機能

ガイダンスメッセージは、クライアント上にあるガイダンスウインドウ上に表示される（図 4.35）。設計者はガイダンスメッセージを読み、設計書にその内容が既に反映されている場合は、OK ボタンを押す。この結果は知識サーバ上に記録される。全てのチェックが完了した後、設計者は確認状況をレポート形式で出力し、DR 会議に提出する。DR 会議では本内容を元に関係者が設計品質を確認する。

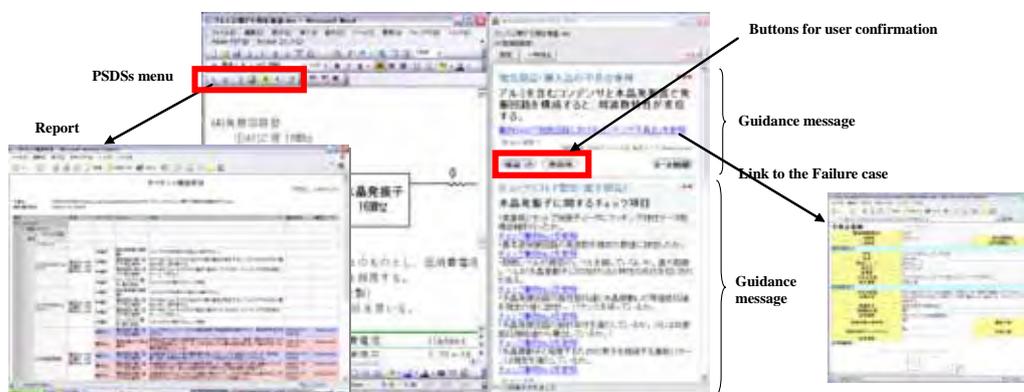


図 4.35 ガイダンスメッセージの確認・レポート出力機能

## 4.2.6 実験結果

### (1) 従来システムとの比較

ここで従来システムである不具合情報蓄積システム（検索システム）を用いた場合と、PSDSsを用いた場合とで、設計書のチェックに要する時間を比較した。図4.36は3人の設計スキルが低いエンジニアによる測定結果を示す。ここでは133個の不具合事例と132個のチェックリストから作成した合計265個の設計知識を用いた。設計書は36ページ、44個の章節から成る既存の回路設計書を用いた。ここで、不具合情報蓄積システムはWebブラウザから検索キーワードを用いて、上記の設計知識を検索することができる。図に示すように、設計者の負荷はPSDSsを利用することで49%に削減された(128/262)。

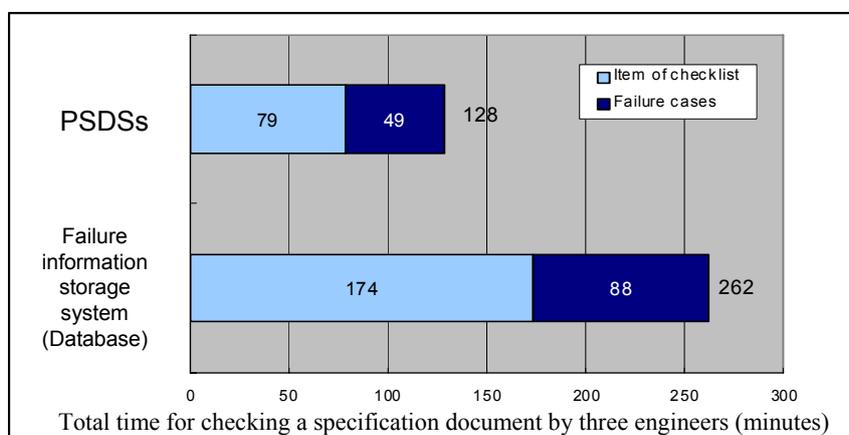


図 4.36 設計者の作業負荷比較

### (2) ガイダンスの精度

図4.37はPSDSsにより提示されたガイダンスの精度を示す。精度の測定は3人の熟練設計者が設計書の章節の内容と、その章節に対して提示されたガイダンスの内容を比較して求めた。実験対象は、昨年度に作成された設計書の一部（回路設計部分、27ページ）である。設計知識は前記の「(1)従来システムとの比較」と同じものを用いた。

ここでは合計231個のガイダンスメッセージが提示された。図に示すように、これらガイダンスメッセージの72%が適切であり、残りの28%が不適切なガイダンスメッセージであった。不適切なガイダンスメッセージにおいて、「条件部が不適切である」が13%と最も大きな理由を占めた。これは、アナログ回路の知識がデジタル回路の設計中に表示されるといった誤りであり、条件部に情報を追加することで抑制できる。第2の理由は「ガイダンスが冗長」であり6%を占める。これは同じガイダンスメッセージが複数の章節で表示されるものである。これに対しては、ガイダンスメッセージを提示する章節を限定するフィルタ条件を追加するなどの対策が考えられる。

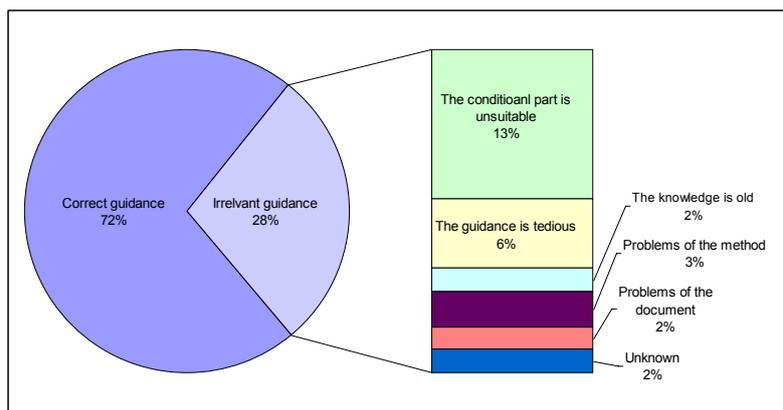


図 4.37 ガイダンスメッセージの精度

### (3) 設計者のスキルによる比較

次に、初級設計者（設計経験が5年以内）と熟練設計者の2名で、PSDSsの有効性の違いを調査した。評価には作成したばかりの実際の設計書2冊を利用した。この評価において、ガイダンスメッセージは以下の3つのクラスに分類した。

- [A] 適切であり、かつ設計者が気が付いていなかった知識
- [B] 適切だが、設計者が既に気が付いていた知識
- [C] 記述内容にあわない不適切な知識

図 4.38 に示すように、初級設計者では[A]の割合が36%を占めた。この結果から、初級設計者の作成した設計書に対して品質改善効果があることが分かる。また同時に実施したアンケートからも、このような関連情報を自動提示することで、初級設計者に対する教育効果があることが認められた。ただし熟練設計者では[A]の割合が低く、12%となった。これらの全ては不具合事例から作成した設計知識であった。この結果から初級設計者よりは劣るが、熟練設計者に対しても設計品質改善効果があると言える。ただし[C]の割合が26%と高くなっており、深い知識を持つ設計者にとっては、ゴミが多いと感じさせる結果となった。

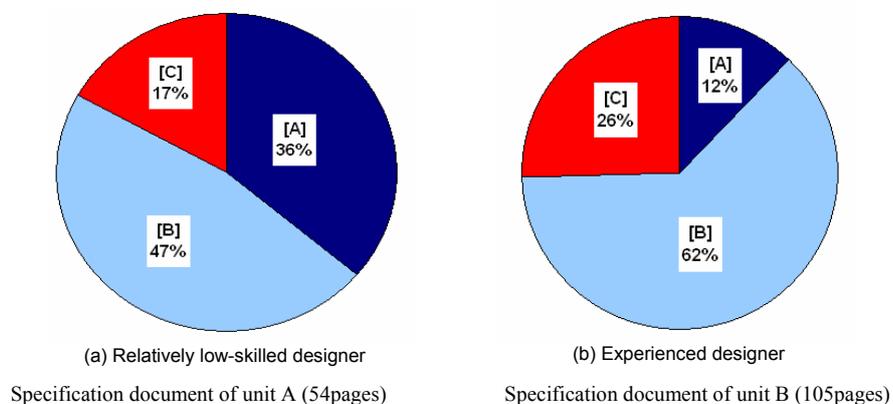


図 4.38 初級設計者と熟練設計者による有効性の比較

(4) 処理速度の評価

次に、PSDSsによるプッシュ型情報提供のリアルタイム性を検証した。検証では実際の設計書と、330個の設計知識を用いて実際の設計フロアで実験した。表4.9は実験に用いた環境・条件を示す。図4.39は処理速度の測定結果である。図の横軸は1回のチェックにおいてクライアントと知識サーバ間で送受信されたデータ量であり、縦軸は処理時間を示す。本実験では1回のチェックにおいて平均165msecでガイダンスメッセージが得られた。表示も含めた全体の応答性については、実設計者から実用上問題ないとの評価を得た。

表 4.9 実験に用いた環境・条件

クライアント	OS:Windows XP, CPU:Pentium M 1.3GHz Word processor:Microsoft MS-Word 2003
知識サーバ	OS:Windows 2003, CPU:Xeon 3.0GHz x 2, Memory: 2GB
ネットワーク	100 Base T
設計知識	330 個 (不具合事例193個, チェックリスト 137個)

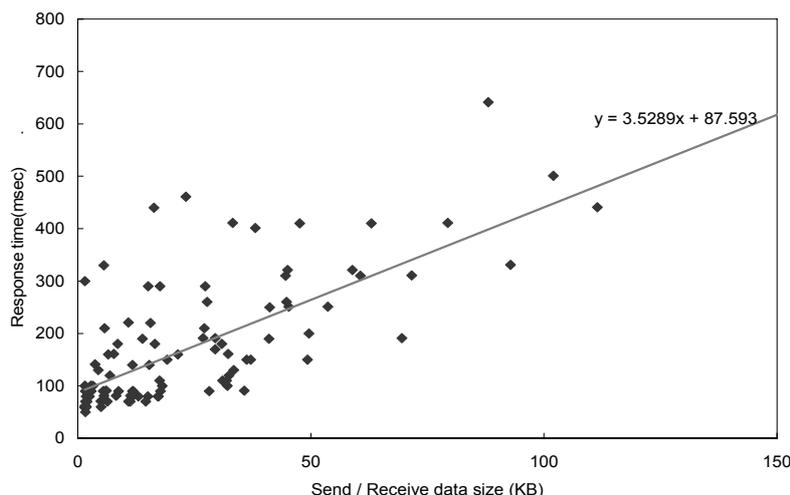


図 4.39 1回のチェックに要する処理時間

(5) 設計知識の作成コスト

最後に、設計知識を作成するための作業時間を測定した結果を表4.10に示す。熟練設計者とPSDSsの開発者が共同で設計知識を作成し、330個の設計知識を作成するために139時間を要した(2.4個の設計知識/時間)。従来の不具合情報知識システムに比べて登録に時間がかかるが、ルールベースのエキスパートシステムに比べて作成コストが小さい。

表 4.10 設計知識の作成に要した時間（合計 139 時間）

処理名	要した時間
(1)不具合事例の理解と、基本的な設計知識の作成	9人, 6時間/人 (合計 54時間)
(2)設計知識を EXCEL 表に整理	3人, 3時間/人 (合計 9時間)
(3)熟練設計者による設計知識のレビュー・修正	2人, 16.75時間/人 (合計 33.5時間)
(4)オントロジーデータの作成	2人, 10.5時間/人 (合計 20.5時間)
(5)熟練設計者によるオントロジーのレビュー・修正	1人, 4時間/人 (合計 4時間)
(6)過去の設計書を用いた動作確認・評価	6人, 3時間/人 (合計 18時間)

#### 4.2.7 文書と知識間の写像を用いた人間支援システム

ここで提案した文書と知識間の写像技術は、設計不具合を未然に防止するための設計支援システムとして製品化を行った。また社内における設計業務でも利用を開始した。

##### (1) 製造行向け、設計品質向上ソリューション (2006～)



図 4.40 設計品質向上ソリューション

##### (2) 社内適用 2ヶ所 (FA 機器設計, 空調機器設計) (2006～)

### 4.3 第4章のまとめ

本章では、文書と知識概念間の写像を獲得について検討を行い、文書をカテゴリ毎に分類（写像）する方式と、文書から関連知識への写像を獲得する方式をモデル化した。

文書とカテゴリ間の写像獲得では、文書中から氏名や地名といったメタ情報を抽出し、これを元に文書を地図やグラフ上に写像した。本写像を実現するため、文書を擬似印刷処理してPDLデータ作成し、PDLデータを構成するグラフィック要素を解析する文書解析技術を提案した。これにより多様なファイル形式の文書から漏れなくテキスト情報を抽出でき、かつ論理情報が取得できることを実験により確認した。また抽出した論理情報に対して形態素解析を適用し、その結果をアプリアリな知識と照合することで、多様な文書からのメタ情報を実現した。本方式を「自治体設備管理,中央官庁向など特殊用途向け文書・図面管理システム」や「FA用表示装置」,「企業向けドキュメント管理パッケージ」などの人間支援システムとして製品化を行い、その有用性を示した。

文書と関連知識間の写像獲得は、設計書と設計知識間の類似性を判定し、類似性の高い設計書の章節と、設計知識とを対応付けることで実現した。設計書は具体的に記載され、設計知識は一般化されて記載されるために抽象度が異なる問題があるが、オントロジを利用した簡易推定を行うことで高精度な写像獲得を実現した。本技術を用いた設計不具合防止支援システムPSDSsを開発し、社内評価によってその有効性を確認した。また、「設計品質向上ソリューション」として製品化を行った。

# 第5章 事象概念への写像獲得と交通映像監視 システムへの適用

人間は映像を見ながら、現在起きている状況を言葉にあらわすことができる。本章ではこの機能を実現するために必要な写像獲得について検討する。具体的には、自然言語処理の方法論である格文法と交通事象に関するアプリオリな知識を元に、交差点の映像情報から、現在起きている交通事象（状況）への写像を求め、これを言語化する（図5.1）。

以降、5.1節で交通映像監視システムの背景と目的を示し、5.2節で交通映像から事象への写像を獲得する方式について述べる。5.3節では実験結果を示し、5.4節で製品適用事例を述べる。最後に5.5節でまとめとする。

表5.1 事象概念への写像モデル

写像関係	交差点の映像から、交通事象（概念）への写像。
適応性	従来の画像処理を用いた監視システムは、抽出したい事象に対して抽出ルールを画像処理の専門家が作りこむ必要があった。本方式では自然言語で知識を記述することで新規事象の抽出が可能であり、適応性が高い。
写像の種類	自然言語処理に格文法を利用し、アプリオリな知識には車や人の定義、物理量と動詞の関係等を記載
脳内の関係部位	視覚野/言語野/側頭連合野
人間支援システム	交通状況監視システムに適用することで、交通事故の抑制が可能（検討中、未実用化）

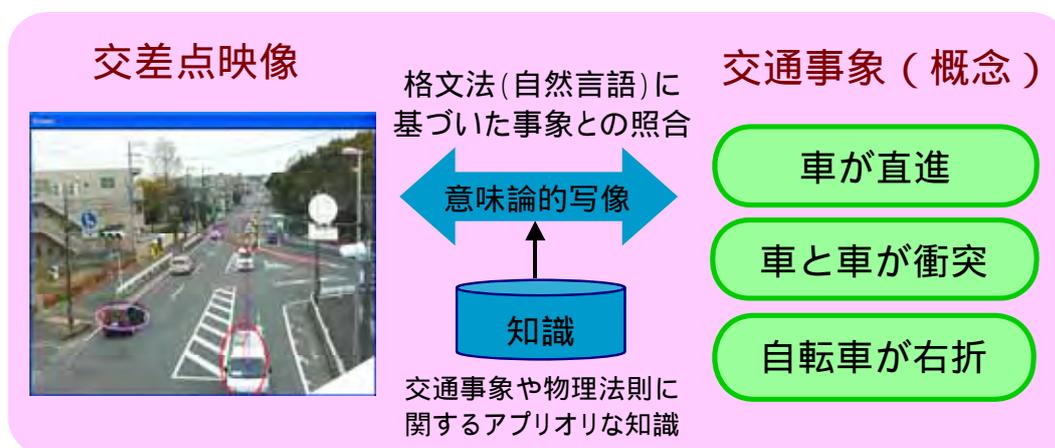


図5.1 事象概念への写像獲得方式の概要

## 5.1 交通映像監視システムの背景・目的

交通映像を自動的に監視して、交通事故の検出や事故の未然防止に利用する要求がある。特に事故の未然防止については、日本(国内)において安全運転支援システム(Driving safety support systems, DSSS)の計画があり、社会インフラの整備を含めた開発が進められている。このシステムでは、運転者から認識が困難な位置に存在する車やバイク、歩行者などを様々な感知器で検出し、その情報をカーナビなどの車載装置を通じて運転者に伝えることで注意を促し、安全運転を支援する。このため交差点の映像から状況(交通事象)を検出し、人間が理解できる形で表現する技術が必要である。

交通映像を画像処理して車や歩行者等の移動物体を検出する方式については、多数の検討が成されており実用化が進んでいる。近年は、移動物体を検出するだけでなく、移動物体の検出結果から状況を判断する研究も行われている。

H.Kollnig と H.-H.Nagel らは、自動車などの乗り物の移動を追跡して、その動きを適切に表現する動詞を生成する方式を提案した[5-1][5-2]。ただし、これら映像からの状況検出の研究は、映像から画像処理的に得た移動物体の位置・姿勢などが、予め定義した数値的な条件を満たしているかどうかを調べている方式が主であり、意味を考慮していない。

一方、G.Herzog らは VITRA (VISual TRANslator) プロジェクトにおいて、言語処理と文脈を考慮した交通事象の検出方式を提案した[5-3][5-4]。これらの手法では、冗長な表現を避けつつ、映像の変化に応じて連続的に文を生成する。本手法を用いて、道路を横断している人間の動作をテキスト化した。ただし、これらの研究はテンプレートをを用いた簡易な処理でテキストを生成しており、意味を考慮するには至っていない。

意味を考慮した映像解析の研究として、岡田らは、線図が描かれた一連の映像から、その動作をテキストで表現する図形理解の研究を行った[5-5]。また、小島らは、意味論における方法論の一つである格解析を利用して、室内の実映像から人間の行動を言語化する方式を提案した[5-6]。

ここでは、意味論を用いた従来の映像解析研究をベースに、映像から状況を言語化する方式を提案する。従来方式では、予め登録した事象を検出したが、交通事象の解析では想定していない異常事態も検出する要求がある。提案方式では知識を拘束知識として用い、知識と合致しない事象が検出された場合は異常事象と判定する。

また、従来の映像解析では、映像からある事象を検出したい場合、画像処理の専門家が事象毎に数値条件の調整やプログラムの修正を行う必要があった。本方式では、人間が意味を理解しやすい格文法に基づいてテキスト形式で知識を記述する。そのため専門家でも検出したい事象を容易に設定でき、幅広い用途や環境に適応可能となる。

## 5.2 意味論を利用した映像と交通事象との写像獲得

本節では、意味論を利用した映像から交通事象への写像獲得方法について述べる。まず 5.2.1 節では交通映像監視システムの概要を示す。次に 5.2.2 節で映像からの移動物体とその移動軌跡の抽出方式を述べ、5.2.3 節において移動軌跡からの事象抽出方式と、異常事象の検出方法を記す。

### 5.2.1 交通映像監視システム

ここで図 5.2 は提案する交通映像監視システムの構成を示す。ここでは交差点脇のポール上に据付けた固定カメラから交差点付近の交通映像を取得する。画像解析部では、この映像中から移動物体と移動軌跡を抽出する。その後、意味解析部は、知識データベースを参照して、移動物体と移動軌跡から該当する交通事象を抽出する。意味解析部では知識データベースに記述された交通事象に加え、知識データベースの内容と合致しない状況を異常事象として検出する。最後に抽出した事象を言語化して出力する。

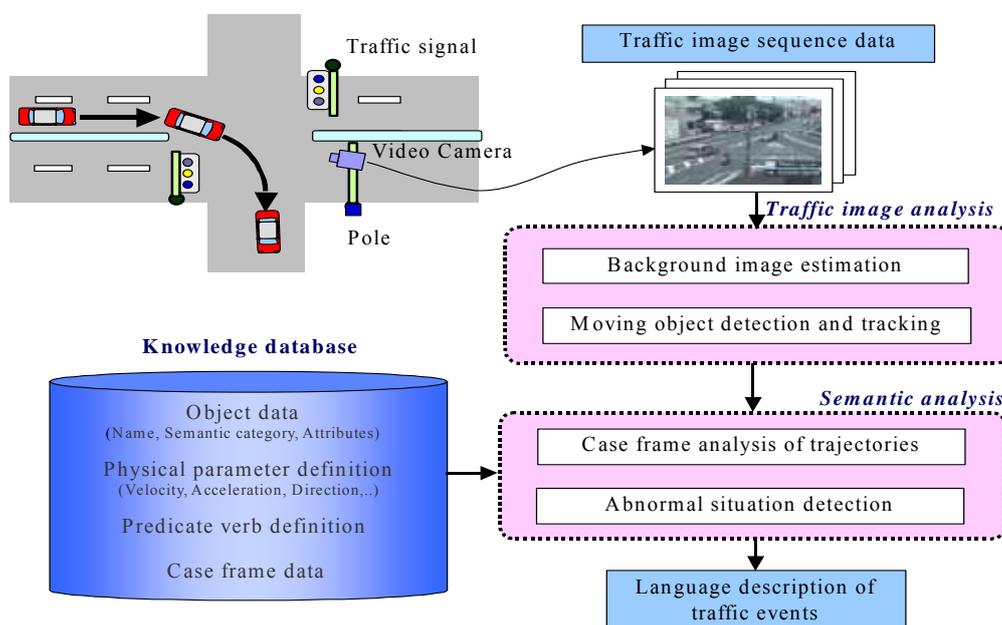


図 5.2 交通映像監視システムの構成

### 5.2.2 画像解析部の処理

図 5.3 は画像解析部の処理フローを示す．ここでは複数フレームの画像から背景画像を推定し，背景画像と現在のフレームとの差分を求めることで移動物体を抽出する．次に連続したフレーム間で移動物体をトラッキングすることにより，移動物体の軌跡を抽出する．次に，画像上における移動物体の座標値( $x,y$ )を 3 次元の仮想空間上における座標値( $X,Y,Z$ )に変換する．最後に，移動物体に関する物理的なパラメータ（加速度や速度など）を計算し，かつ移動物体の大きさに基づいて，その物体の種別（車，自転車，歩行者，犬）を判定する．

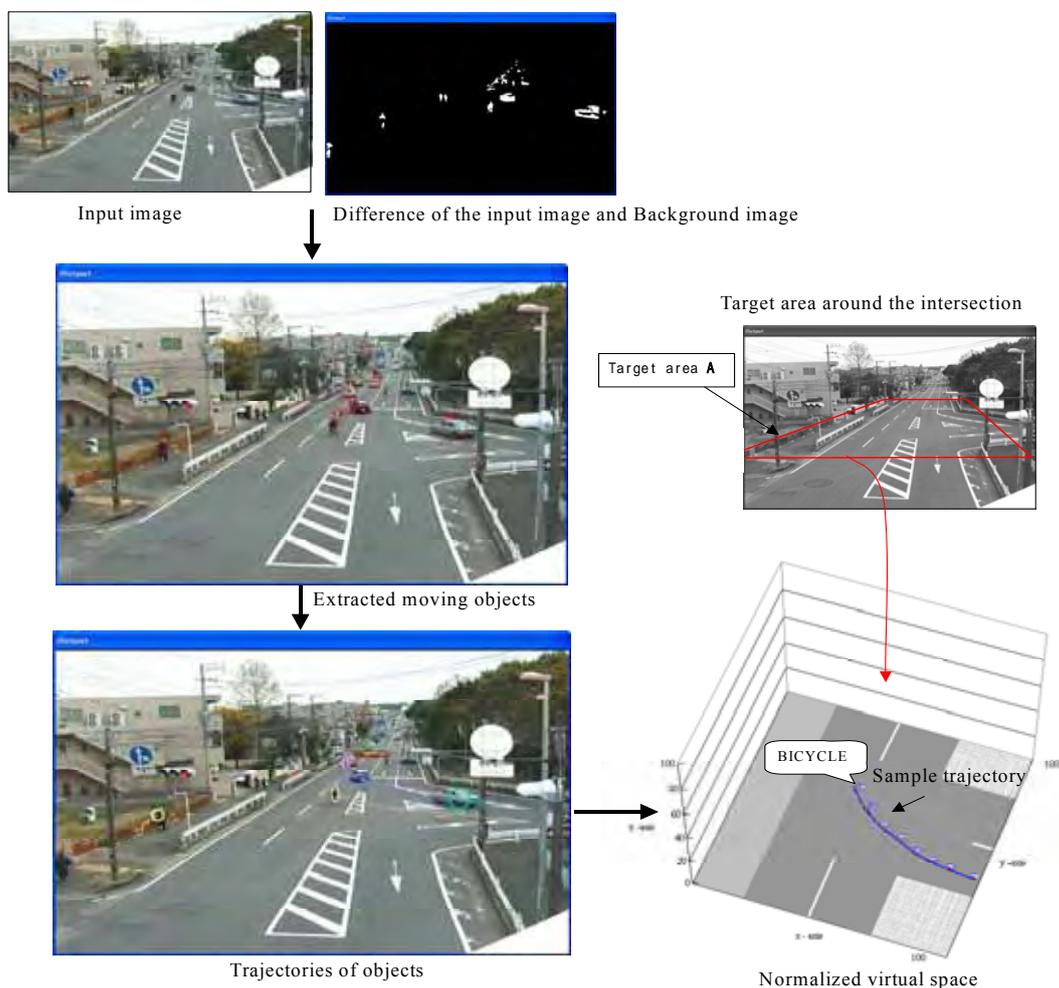


図 5.3 画像解析部の処理フロー

#### (1) 背景画像の推定

Tuzel らは動的なシーンからベイズ学習に基づいて背景画像を推定する方式を提案した [5-7][5-8] . この方式では，個々のピクセルが，互いに競合する仮説（背景か否か）を持ち，

これらの仮説が正規分布の確率を持つ．そして，この正規分布の確率を再帰的なベイズ学習の仕組みを用いて推定する．提案方式では，このアルゴリズムを用いて背景画像を推定する．

#### (2) 移動物体のトラッキング

ここでは物体の移動速度が速い，低フレームレートのビデオに適用可能なトラッキング方式を適用する[5-9]．移動量が大きい場合，フレーム間での移動物体の重なりが無くなるため，従来のシフト追跡法ではトラッキングに失敗する場合がある．このトラッキング方式では複数のカーネルを用いてトラッキングを行うことで，この課題に対処する．

#### (3) 座標変換と物理パラメータの推定

座標変換処理では，交差点の周辺領域（図 5.3 のターゲット領域 A）を 3 次元の仮想空間に変換する．この仮想空間は 100x100x100 に正規化された大きさを持つ．この仮想空間内の正規化された移動軌跡から，移動物体の持つ物理パラメータを推定する．ここでは図 5.4b の 6 個の物理パラメータを用いる．

#### (4) 物体のタイプの検出

最後に物体のタイプ（車，自転車，人，犬）を区別する．ここでは単純に仮想空間上における正規化された物体の大きさから，これらのタイプを区別する．

### 5.2.3 意味解析部の処理

意味解析部は、画像処理部で得た移動物体の軌跡とタイプ（車、歩行者など）を元に、知識データベースを参照することで事象を検出・言語化する。以降、その処理内容を記す。

#### (1) 格フレーム解析を用いた交通事象の記述

ここでは自然言語処理における意味論の方法論の一つである格文法を用いて交通事象を記述する。この格文法は Fillmore により提唱された[5-10]。格文法では動詞を中心として、他の要素（通常は名詞群）との関係を記述することで文のパターンを作成する。ここで1つの文パターンは1つの命題を表す。機械翻訳の分野では、このように動詞を中心に文を記述することで、他の記述方式よりも文パターン数を抑制できることが知られている。表5.2は Fillmore の格文法で利用される格の種類を示す。例えば、“John broke the stick” という文章は、格文法を用いることで以下のように記述できる。

[broke] 述語動詞格(PEV) – [John] 動作主格(AG) – [the stick] 対象格(OBJ)

ここで動作主格や対象格といった各々の格に適用できる単語は、述語動詞格が決まると制限される。この制限には単語の意味カテゴリが一般的に利用される（表5.3）。例えば、述語動詞格が[broke]の場合、その動作主格には意味カテゴリが Human や Living である単語を適用することができ、同様に対象格には意味カテゴリが Physical object である単語のみを適用できる。このような制限により、[broke]に関連した存在しうる多様な文を少ないパターン数で網羅することができる。

本方式では、交通事象を記述するための格として(AG, CAG, LOC)を用い、意味カテゴリとして (PHYSOBJ, LIVING, HUM, PHYSLOC) を利用する。

交通事象をこのような文法規則に従い記述する従来研究として Yuri A. Ivanov と F. Bobick の *simolified grammer* がある。Ivanov らは *simolofied grammer* を駐車場での状態を検出するために利用した[5-11]。このような他の文法を用いた方式に比べ、Fillmore の格文法を利用することは以下の有利な点がある。

- (1) 格文法は自然言語処理の分野において深層構造（表層的な表現の違いによらず、文が持つ真の意味を示す構造）を表すために提案された方式である。そのため非言語的なコンテンツから抽出された構造を表現するのにも適している。
- (2) 自然言語処理を基本とした文法であり、交通事象の記述やメンテナンスが人間にとって容易である。そのため、画像処理の専門家でなくても記述できる。
- (3) 動詞を中心としているため文パターン数が少ない。そのため少ない記述量で網羅度の高い知識を作成できる。

表 5.2 Fillmore の格文法

Case name (label)	Definition
Agent (AG)	A person or entity causing a verb's action to be performed
Counter-Agent (CAG)	The force of resistance against which a verb's action is carried out
Object (OBJ)	An entity affected directly by a transitive verb's action
Instrument (INSTR)	An inanimate entity causally involved in a verb's action
Source (SO)	The place from which something moves
Goal (GO)	The place to which something moves
Location (LOC)	The location or spatial orientation of the state or action
Time (TIME)	The time of the event
Experiencer (EXPER)	A person or thing affected by a verb's action, replacing dative

表 5.3 単語の意味カテゴリ

Semantic category	definition	Semantic category	definition
PHYSOBJ	Physical object	PHYSACT	Physical action
LIVING	Living	MENTACT	Mental action
HUM	Human	PTRANS	Physical movement
TIME	Time	ATTRTRANS	Attribute change
PHYSLOC	Physical Location	BODACT	Body action

## (2) 知識データベース

知識データベースは移動物体のタイプと移動軌跡, および推定した物理パラメータから交通事象を検出するために利用する. この知識データベースは, 拡張性と可読性を考慮して4種類のデータから構成する(図 5.4). 各種類のデータには以下の情報を記述する. この知識データベースは人間が予め記述する.

- (a) オブジェクトデータ: 車や歩行者, 犬, 自転車といった動作主格と対象行為格に適用する物体の情報, および車道や歩道といった場所格に適用する物体の情報を記載する. この情報は, 物体の名前, 意味カテゴリ, および属性(大きさや領域)で構成する.
- (b) 物理パラメータ定義: 意味解析部で利用可能な物理パラメータのリストを記載する. これらの物理パラメータは画像解析部で推定された移動物体の加速度や速度, 進行方向などである.

- (c) 述語動詞格の定義：物理パラメータと述語動詞との関係を定義する．例えば，[COLLIDE](衝突する)という動詞は，次式に示すように，二つの物体間の距離  $dist$  が 0 であり，かつそれらの物体が持つ加速度の変化量がある閾値  $\eta$  を超えた場合とする (図 5.5)．

$$\begin{aligned}
 & dist(P_{AG}, P_{CAG}) = 0 \\
 & \quad \cap \\
 & \frac{d(\alpha_{AG})}{dt} + \frac{d(\alpha_{CAG})}{dt} > \eta
 \end{aligned} \tag{5.1}$$

- (d) 格フレームデータ：個々の交通事象を定義した格フレームの情報を記載する．

(a) Object data

						
ObjectName	[CAR]	[PEDESTRIAN]	[BICYCLE]	[DOG]	[Footway]	[Roadway]
SemanticCategory	PHYSOBJ	HUM	PHYSOBJ	LIVING	PHYSLOC	PHYSLOC
Attributes	Width=5 Height=5 Length=10	Width=1 Height=5 Length=1	Width=1 Height=5 Length=5	Width=1 Height=1 Length=1	Area in virtual space	Area in virtual space

(b) Physical parameter definition

Verocity	Acceleration	Position	Direction of movement	Size	Distance of two objects
$v$	$H$	$X, Y, Z$	$H$	$W, H$	$dist(O1, O2)$

(c) Predicate verb (PREV) definition

[RUN]	[STOP]	[ACCELERATE]	[DECELERATE]	[TURN RIGHT]	[TURN LEFT]	[COLLIDE]	[PASS]
$v_{AG} > 0$	$v_{AG} = 0$	$\frac{d(v_{AG})}{dt} > 0$	$\frac{d(v_{AG})}{dt} < 0$	$\theta_{AG}^t - \theta_{AG}^{t-N} > 20$	$\theta_{AG}^t - \theta_{AG}^{t-N} < -20$	$dist(P_{AG}, P_{CAG}) = 0$ $\cap$ $\frac{d(\alpha_{AG})}{dt} + \frac{d(\alpha_{CAG})}{dt} > \eta$	$ \theta_{AG}^t - \theta_{CAG}^t  > 140$ $\cap$ $dist(P_{AG}^{t-N}, P_{CAG}^{t-N}) < \tau$ $\cap$ $dist(P_{AG}^t, P_{CAG}^t) > \tau$
continue $N$ frames	continue $N$ frames	continue $N$ frames	continue $N$ frames				

(d) Case frame data

Case name	[RUN]	[STOP]	[ACCELERATE]	[DECELERATE]	[TURN RIGHT]	[TURN LEFT]	[COLLIDE]	[PASS]
Agent (AG)	AG-PHYSOBJ or AG-LIVING or AG-HUM	AG-PHYSOBJ or AG-LIVING or AG-HUM	AG-PHYSOBJ	AG-PHYSOBJ	AG-PHYSOBJ or AG-LIVING or AG-HUM	AG-PHYSOBJ or AG-LIVING or AG-HUM	AG-PHYSOBJ or AG-LIVING or AG-HUM	AG-HUM or AG-LIVING
Location (LOC)	LOC-PHYSLOC	LOC-PHYSLOC			LOC-PHYSLOC	LOC-PHYSLOC	LOC-PHYSLOC	LOC-PHYSLOC
Counter-Agent (CAG)							AG-PHYSOBJ or AG-LIVING or AG-HUM	AG-HUM or AG-LIVING

図 5.4 知識データベースの内容

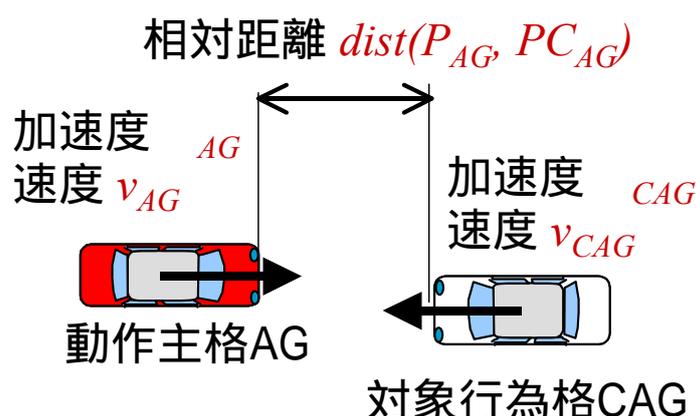


図 5.5 [COLLIDE](衝突する)の定義

(3) 映像と格フレームとの照合

ここでは時刻  $t$  における交通事象を以下の手順に従い計算する。

- (i) 図 5.4d の格フレームデータから，1つの格フレームを選択する。
- (ii) 時刻  $t$  における移動物体  $i$  と AG との関係性  $P_{AG}$  を計算する。ここで AG の意味カテゴリが，移動物体  $i$  の持つ意味カテゴリと合致する場合， $P_{AG}=1.0$  とする。もし，移動物体  $i$  の持つ意味カテゴリが不明な場合は  $P_{AG}=0.5$  とする。その他の場合は  $P_{AG}=0.0$  を持つ。例えば，移動物体  $i$  が [CAR] であると認識された場合，移動物体の意味カテゴリ  $i$  は PHYSOBJ である。また述語動詞 [RUN] の AG には意味カテゴリ PHYSOBJ を当てはめることができる。そのため，この場合の  $P_{AG}$  は 1.0 となる。
- (iii) 次に，述語動詞と移動物体  $i$  との関係  $P_{PREV}$  を計算する。移動物体  $i$  の物理パラメータが図 5.4c に示す条件式を満足する場合， $P_{PREV}=1.0$  とする。その他の場合は  $P_{PREV}=0.0$  を取る。
- (iv) 次に， $P_{LOC}$  の値を計算する。ここでは，まず図 5.4a に記載したオブジェクトデータから意味カテゴリ PHYSLOC をもつオブジェクト群を選択する。そして移動物体  $i$  が，これらのオブジェクトが持つ領域内に存在する場合， $P_{LOC}=1.0$  とする。その他の場合（オブジェクトが不明な領域に存在する場合）は， $P_{LOC}=0.5$  とする。
- (v) 対象とする格フレームが CAG を持つ場合，移動物体  $j$  ( $i \neq j$ ) と CAG の関係を示す  $P_{CAG}$  を計算する。この計算方法は上記(ii)と同じとする。
- (vi) 最後に，評価値  $E$  を式 5.2 に従い求める。全ての格において高い関係性を持つ場合， $E$  の値は 0 に近づく。関係性が低い場合， $E$  は負の大きな値をとる。

上記のステップ(i)~(vi)を全ての格フレームと、全ての移動物体の組み合わせに対して計算する。ここで評価値  $E > -1$  を取る場合、その交通事故が検出されたと判断する。

$$E = \delta_{PREV} \cdot \log(P_{PREV}) + \delta_{AG} \cdot \log(P_{AG}) + \delta_{CAG} \cdot \log(P_{CAG}) + \delta_{LOC} \cdot \log(P_{LOC}) \quad (5.2)$$

$$\delta_c = \begin{cases} 1: \text{If the CaseFrame has the case 'C'} \\ 0: \text{Otherwise} \end{cases} \quad (5.3)$$

#### (4) 異常事象の検出

交通映像監視システムでは異常事象(車が他の車と衝突したなど)を検出する要求があり、幾つかの手法が提案されている。しかし、多くの手法は予め異常事象をルールとして記述しておく必要がある。そのため想定していない事象が発生した場合、これを検出することが困難である。著者らは、知識データベースを一種の拘束条件と考え、そこに合致しない事象が検出された場合は、これを異常事象と判定する。

例えば、知識データベースに

[RUN]述語動詞格 – [PHYSOBJ]動作主格 – [PHYSLOC]場所格

と記載した場合、車が車道をはみ出して未定義の領域を走った場合、前記の映像と格フレームの照合方式で説明したように  $P_{LOC} = 0.5$  となる。評価値  $E$  は-1より大きくなり事象が検出されるが、 $E$  の値が0でないため想定外の異常事象であると判断される。

## 5.3 実験結果

### 5.3.1 仮想的なシナリオを用いたシミュレーション

異常事象を撮影した映像は入手が困難なため、仮想的な3つのシナリオを作成して本方式の動作を検証した。これらのシナリオは移動物体と、その軌跡を人間が定義したものである。各シナリオでは移動物体が右折や加速などの動作を行う事象や、人と自転車の衝突などの異常事象を模擬した。知識データベースは図 5.4 に示す内容を用いた。図 5.6 は異常事象を含むシナリオと、そこからの事象抽出結果の例を示す。

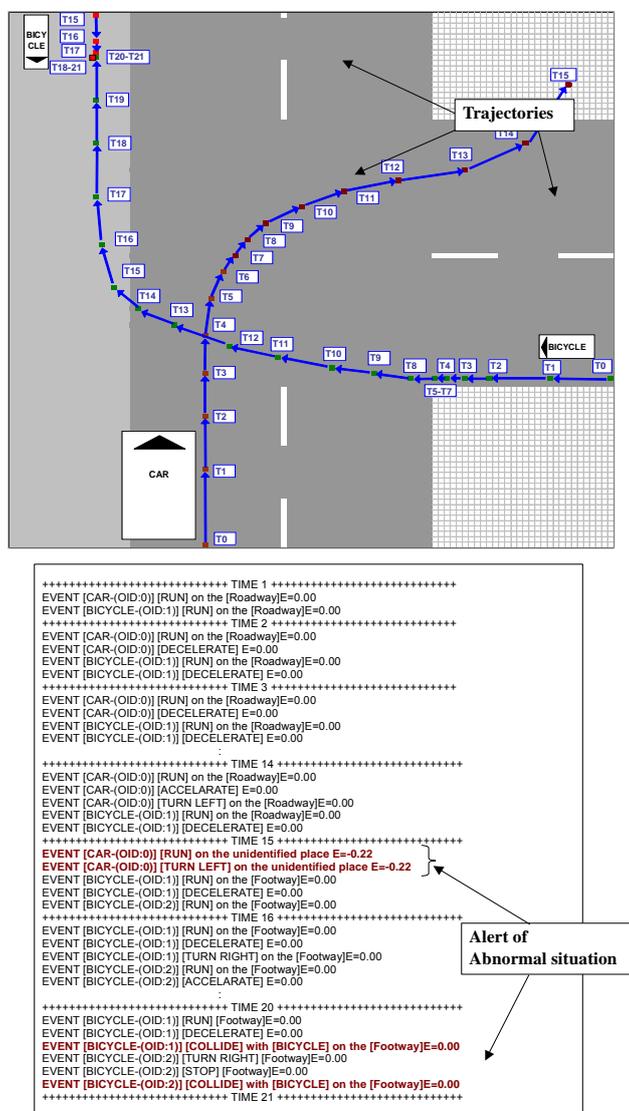


図 5.6 異常事象を含むシナリオと、事象抽出結果例

図に示すように、知識データベースに記載した全ての一般的な事象は正しく抽出できた（例えば、"A car turns right on the roadway."など）。また、知識データベースに定義した異常事象 "A bicycle collides with a bicycle on the footway."も正しく抽出できた。さらに、未定義な事象として"A car runs on the unidentified place"を抽出することができた。

### 5.3.2 交差点の映像に対する事象抽出結果

ここでは実際の交差点の映像を用いて提案方式の精度を評価した結果を記す。映像は高さ5mの歩道橋に設置した固定カメラから、交差点を斜め下方向に見下ろす形で撮影した。表5.4は撮影に用いたカメラと設定を示す。撮影した交差点は比較的事故が多い変則的な形状の交差点である（図5.7）。映像は5409フレーム、約6分間の長さを持つ。映像中には81台の車と自転車（オートバイを含む）、および5人の歩行者が写っている。

表5.4 撮影に用いたカメラと設定

Camera device	DCR-TRV900 (SONY)
Scanning system	Progressive Scan
Mode	Automatic (Brightness, Shutter speed, white balance), Zoom:Off
File format	AVI (DVI compression), 15 frames / sec
Size / Color	720 x 480 pixels, 24bit color



図5.7 撮影した交差点（本映像は Google Earth から取得）

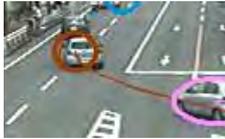
撮影した映像に対して、交通事象検出処理を適用した。知識データベースには図5.4に示す内容を用いた。ここでは全てのフレームの画像に対して画像解析処理を適用し、意味解析処理は15フレーム毎に実施した（1秒に1回）。その結果、画像解析処理で97個の移動物体を検出し、意味解析処理で1359個の交通事象を映像中から検出した。検出した交通

事象を人間が目視で確認したところ、73.8%が正しい交通事象であった（表 5.5）。検出誤りの 15.2%は、画像解析部における移動物体の抽出誤りが原因であった。例えば、1つの車を複数個の移動物体に誤分割したり、複数台の車を1台の車として誤結合するものが含まれた。このような画像処理部の誤りにより "A car collided with a car on the roadway" といった異常事象が誤検出された。また、検出誤りの残り 11.0%は意味解析部に起因した誤りであった。これは、ゆっくりと歩く歩行者を“人が歩道で止まっている”という事象として誤検出する動詞の推定誤りが大半を占めた。これは仮想空間を 100x100x100 に正規化したことによる まるめ誤差が原因である。図 5.8 は実際に抽出された交通事象の正解・誤り例を示す。

表 5.5 交通事象抽出精度

Correct	73.8 %	
Error	26.2 %	15.2%: Moving object extraction error in the image analysis.
		11.0%: Verb estimation error in the semantic analysis. (e.g. [PREDESTRIAL] [RUN] -> [PREDESTRIAN][STOP])

(a) Correct traffic events



The [CAR] - [TURN RIGHT] s - on the [Roadway] E=0.0  
(AG) (PREV) (LOC)



The [CAR] - [STOP] s - on the [Roadway] E=0.0  
(AG) (PREV) (LOC)



The [BICYCLE] - [RUN] s - on the [Roadway] E=0.0  
(AG) (PREV) (LOC)



The [CAR] - [TURN LEFT] s - on the [Roadway] E=0.0  
(AG) (PREV) (LOC)



The [PEDESTRIAN] - [PASS] s - the [PEDESTRIAN] - on the [Footway] E=0.0  
(AG) (PREV) (CAG) (LOC)



The [PEDESTRIAN] - [RUN] s - on the [Footway] E=0.0  
(AG) (PREV) (LOC)

(b) Error traffic events



Traffic signal changed



Several cars are detected as one moving object



A car is divided into two moving objects

図 5.8 抽出された交通事象の例 (a) 正解, (b) 誤り

### 5.3.3 知識データベースのサイズ

表 5.6 は、実験に用いた知識データベースのサイズと記述形式を示す。この知識データベースは図 5.4 に示す内容を含む。図に示すように、215 行（テキストと C 言語）で記述できた。交通事故の記述は格文法に従い行うため、文パターン数が少なくなりデータサイズ・作成コストが削減できる。また、自然言語に基づいてテキスト形式で記載されるため、記述が容易であり、専門家でなくても理解しやすい。

表 5.6 知識データベースのサイズと記述形式

Data	Size	Type
Object data	54 lines (about 9 lines / object)	text
Physical parameter definition & Predicate verb definition	116 lines (about 14.5 lines / verb)	C language program
Case frame data	45 lines (about 5.6 lines / verb)	text

## 5.4 事象概念への写像を用いた人間支援システム

本技術は、交通事故を未然に抑制する安全運転支援システム（DSSS：Driving Safety Support Systems）への適用が可能である。死亡事故と重傷事故の 7 割は、安全不確認と前方不注意に起因することが知られている。提案技術を用いて運転者に「見えない危険」をカーナビ等の車載端末から情報提供し、認知行為を支援することにより、これら事故の抑制に寄与できると考える。ただし、現在の事象検出精度は 73.8%であり、誤った事象を検出する割合が高い。今後、実証実験を通じてデータを集めて方式改良を進めると共に、他センサとの併用を検討して高精度化を行う必要がある。また、“追い越してきて衝突した”などのより複雑な事象を表現する方法についても検討を進める必要がある。

その他の応用例として、交差点に本システムを設置して長期間の監視を行い、事故が多い交差点での事故要因分析に利用することも考えられる。

## 5.5 第5章のまとめ

本章では、自然言語処理の方法論である格文法と交通事象に関するアプリアリな知識を元に、交差点の映像情報から、現在起きている交通事象（状況）への写像を求め、これを言語化する方式を提案した。

従来の映像解析処理では、映像からある事象を検出したい場合、画像処理の専門家が事象毎に数値条件の調整やプログラムの修正を行う必要があった。本方式では、人間が意味を理解しやすい格文法に基づいてテキスト形式で事象や知識を記述できる。そのため専門家でなくても検出したい事象を容易に設定でき、幅広い用途や環境に適応可能となる。また従来方式では、予め登録した事象を検出したが、提案方式では知識を拘束知識として用い、知識と合致しない事象を未知の異常事象として検出することができる。

実際に交差点の映像に対して本方式を適用し、73.8%の精度で正しく事象が検出できた。ただし、本技術を安全運転支援システムなどの人間支援システムに適用する場合、より高い精度と柔軟な事象記述性が求められる。今後は事象検出結果を画像解析部にフィードバックして画像認識誤りを抑制する等の対策や、より複雑な事象の記述方式について検討を行う予定である。また、別なアプローチによる異常事象の検出法として、車や歩行者の移動軌跡を収集・分析しておき、一般的な移動軌跡と異なる軌跡が検出された場合にアラートを発するような方式についても検討が必要と考える。

## 第6章 結論

### 6.1 本論文のまとめ

本論文では、写像という切り口で人間の知的情報処理を考え、写像をモデル化することで実社会に役立つ人間支援システムを開発した。具体的には、人間の脳内における写像に「集合論的写像」と「意味論的写像」が存在することを示し、各2例ずつをモデル化した。モデル化では脳機能や認知言語学における知見に基づいて、集合論的写像は幾何学的な位相情報を保持する制約を設けた。また意味論的写像は自然言語処理と、アприオリな知識を用いた推論処理を有する制約を設けてモデル化した。

#### (1)集合論的写像

第2章で「アーム先端の空間座標と、アームの身体座標（関節角）」間の写像を検討した。ここでは双方向性を持ち、学習的に写像関係を獲得することができる双方向 Feature Map モデルを提案した。写像が離散的である課題に対して、外界に適応的に素子を自己増殖する機能を適用する拡張を行い対処した。本成果を用いてアームの制御ソフトウェアを試作した。

第3章では「変形した帳票画像と、帳票の形状モデル」間の写像を検討した。ここでは、大脳皮質の視覚野で抽出されているものと同様な特徴点を帳票画像と形状モデルから抽出し、この対応付けに基づいて、広義ニュートン法に従いロバストに帳票モデルと帳票画像間の写像（変形量）を推定する方式を提案した。提案方式を FAX-OCR 装置に適用して製品化を行った。

#### (2)意味論的写像

第4章では「文書と知識概念」間の写像獲得を検討した。ここでは、まず文書中からメタ情報を抽出して文書とカテゴリ間の写像を獲得する方式を検討した。本検討では、多様なファイル形式の文書から漏れなく内容情報と論理情報を抽出する新しい文書解析方式を提案し、この解析結果からメタ情報を抽出した。本技術は「特殊用途向け文書・図面管理システム」や「企業向け文書管理パッケージ」などの幅広い製品に適用した。次に、設計書と設計知識間の写像を獲得する方式を検討した。ここでは、用語の関係性とオントロジ（部品の知識）を用いることで、抽象度の異なる設計書と設計知識間の類似性を高精度に判定する方式を提案し、「製造業向けの設計品質向上ソリューション」として製品化した。

第5章では「交差点の映像から、交通事象（概念）」への写像を検討した。ここでは自然言語処理の格文法と交通事象に関するアприオリな知識を元に、交差点映像情報から交通事象（状況）を検出し、言語化する方式を提案した。本方式は、今後高精度化を検討し、安全運転支援システムなどの人間支援システムへの適用を進める予定である。

## 6.2 今後の展望と課題

本研究では、集合論的写像の獲得に始まり、より複雑な意味論的写像の獲得について検討を進めてきた。意味論的写像の検討では、格文法を用いて交差点映像から交通事象を検出する方式を実現できた。しかし、本技術を安全運転支援システムやロボットビジョンなどに実適用する場合、提案した単純な格文法では精度や表現できる事象に限界があると考えられる。また、本研究で実現できたのは認知言語学における意味論的写像の一部（図 6.1 の点線で示す範囲）であり、メンタル・スペース部の検討が不十分である。加えて、意味論的写像に対応した脳の高次機能については、言語と認知処理との関係がどのように実現されているかなど未知な部分が多く残っている。

そこで今後は、事象検出精度の向上や、複雑な事象の記述方式の検討、交通事象以外への適用、メンタル・スペース部の実現方式検討などを進める予定である。また、バイオミメティックな情報処理の立場から、意味論的写像を神経回路モデルにより表現する方法についても検討を進めたい。長期的には、一般のユーザが自由に知識を記述して外界の状況を言語化できるモデルを構築して Prolog のような公開フレームワークを実現すると共に、今後得られるであろう脳内高次処理の知見に基づいてモデルの深堀を進める予定である。

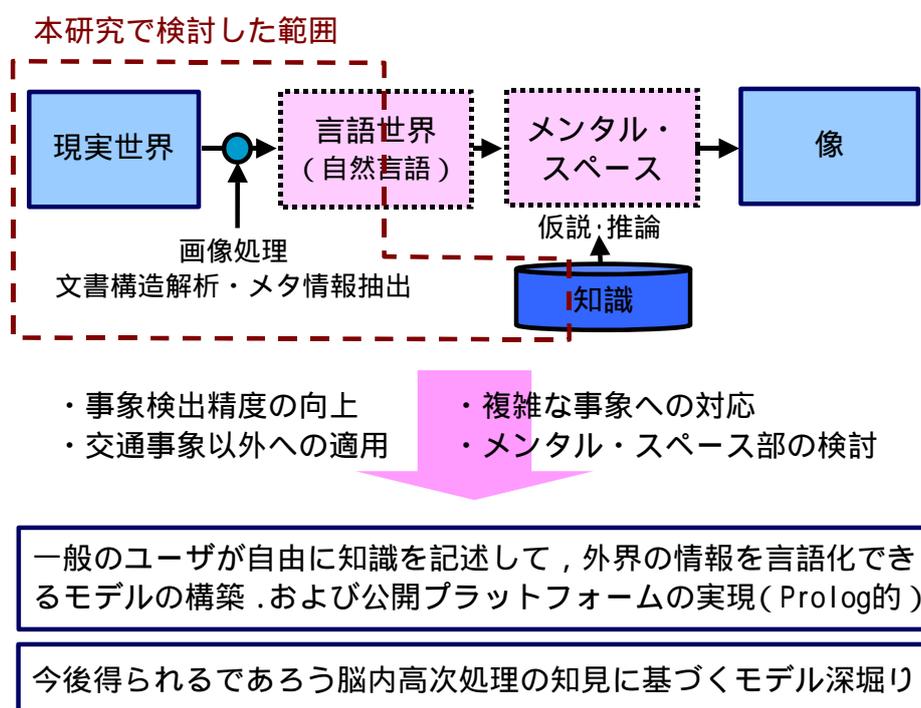


図 6.1 本研究で検討した範囲と今後の展望・課題

# 参考文献

## 第1章

- [1-1] Ludwig Wittgenstein, “*Logisch-Philosophische Abhandlung*”, In Wilhelm Ostwald (editor), *Annalen der Naturphilosophie*, (1921)
- [1-2] Alain Colmerauer, “An Introduction to Prolog III”, *Communications of the ACM*, vol 33, no 7, (1990)
- [1-3] Carl Hewitt, “The repeated demise of logic programming and why it will be reincarnated What Went Wrong and Why: Lessons from AI Research and Applications”, Technical Report SS-06-08, AAAI Press, (2006)
- [1-4] Gilles Fauconnier, ” *Mental Spaces: Aspects of Meaning Construction in Natural Language*”, Cambridge University, (1995)
- [1-5] Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T., Gaillard, W., and Theodore, W., “Regional cerebral blood flow during object naming word reading”, *Human Brain Mapping*, 3, pp.93-106, (1995)
- [1-6] Kosslyn, S.M., Alpert, N.M., and Thompson, W.L., “Identifying objects at different levels of hierarchy: A positron emission tomography study”, *Human Brain Mapping*, 3, pp.107-132, (1995)
- [1-7] Fletcher, P.C., Happe, F., Frith, U., Baker, S.C., Dolan, R. J., Frackowiak, R.S.J., and Frith, C. D., ”Other minds in the brain: A functional imaging study of theory of mind in story comprehension”, *Cognition*, 57, pp.109-128, (1995)
- [1-8] Damasio, A.R.,”Aphasia”, *The New England Journal of Medicine*, 326, pp.531-539, (1992)
- [1-9] Stromswold, K., Caplan, D., Alpert, N., and Rauch, S., “Localization of syntactic comprehension by positron emission tomography”, *Brain and Language*, 52, pp.452-473, (1996)
- [1-10] 乾 敏郎, 大津由紀雄, 岡田知久, 田中茂樹, 西澤貞彦, 小西淳二, “fMRI による構文解析に関わる領野の特定”, 第21回神経心理学会予稿集, (1997)
- [1-11] Minsky M. L. and S. Papert, “*Perceptrons, an easy in computational geometry*”, The MIT Press, (1969)
- [1-12] Rumelhart, D. E., and J. L. McClelland, “*Parallel distributed processing: Explorations in the microstructure of cognition*”, The MIT Press/Bradford Books, (1986)
- [1-13] Marr. D., ”A Theory of Cerebellar Cortex”, *Journal of Physiology*, 202, pp.437-470, (1969)

- [1-14] Kohonen. T., "*Self-Organization and associative memory*", Springer-Verlag, (1984)
- [1-15] 片岡圭太郎, 萩原将文, "意味記憶及びエピソード記憶ニューラルネットワーク", 信学技報, NC2002-188, pp. 91-96, (2003)
- [1-16] 片岡圭太郎, 萩原将文, "推論可能な言語処理ニューラルネットワーク", 信学技報, NC2004-168, pp.185-190, (2005)
- [1-17] 齋藤雅裕, 萩原将文, "概念の自動階層化による知識処理ニューラルネットワーク", 信学技報, NC2006-211, pp.137-142, (2007)
- [1-18] Chomsky N., "*Knowledge of language: its nature, origin, and use*", Praeger, (1986)
- [1-19] 乾 敏郎, 今井むつみ, 山下博志, "認知発達と生得性-心はどこから来るのか", 共立出版, (1998)
- [1-20] 今井むつみ, "ことばの学習パラドックス", 共立出版, (1997)

## 第2章

- [2-1] Miller W. T., Sutton R. S. and Werbos P. J., "*Neural Network for Control*", MIT Press, (1990)
- [2-2] 川人光男, "随意運動制御における適応と学習", 日本ロボット学会誌, 4, pp.184-193, (1986)
- [2-3] Baum E. B. and Haussler D., "What Size Net Gives Valid Generalization", *Neural Computational*, 1, pp.151-160, (1989)
- [2-4] Jordan M. I., "Constrained Supervised Learning", *Journal of mathematical psychology*, 36, pp.396-425, (1992)
- [2-5] Kohonen T., "*Self-Organization and Associative Memory*", Springer-Verlag, (1984)
- [2-6] Martinez T. M., Ritter H. J. and Schlten K. J., "Three-Dimensional Neural Net for Learning Visuomotor Coordination of a Robot Arm", *IEEE Trans. Neural Networks*, 1, pp.131-136, (1990)
- [2-7] Kuperstein M., "Adaptive visual-motor coordination in multijoint robots using parallel architecture", *IEEE Int. Conf. Automat. Robotics*, Raleigh, pp.1595-1602, (1987)
- [2-8] Dayhoff J., "*Neural Network Architectures*", Van Nostrand Reinhold, (1990)
- [2-9] 平野 敬, 佐瀬幹哉, 小杉幸夫, "Counter Propagation を拡張した双方向 Feature Map の正式", 信学技報, NC91-43, (1991)
- [2-10] Flash T. and Hogan N., "The coordination of arm movements. An experimentally confirmed mathematical model", *Journal of Neurosci*, 5, pp.1688-1703, (1985)

- [2-11] Uno Y., Kawato M. and Suzuki R., “Formation and control of optimal trajectory in human multijoint arm movement – minimum torque-change model”, *Biological Cybernetics*, 61, pp.89-101, (1989)

### 第3章

- [3-1] 鈴木克考, “文字認識装置”, *東芝技術公開集*, vol.14, no.3, pp.89-94, (1996)
- [3-2] 石谷康人, “モデルマッチングによる表形式文書の理解”, *信学技報*, PRMU94-34, pp.57-64, (1994)
- [3-3] Y. Hirayama, “A Framework for Forms Processing Using an Enhanced-Line-Shared-Adjacent Format”, *Int. Conf. on Document Analysis and Recognition*, pp.103-106, (1999)
- [3-4] U. Bohnacker, J. Schacht, and T. Yucel, “Matching Form Lines Based on a Heuristic Search”, *Int. Conf. on Document Analysis and Recognition*, vol.1, pp.86-90, (1997)
- [3-5] 平野 敬, 岡田康裕, 依田文夫, “表の構造解析による FAX 送信帳票からの文字認識フィールド抽出法”, *信学技報*, PRMU2000-71, pp.73-79, (2000)
- [3-6] 長谷川史裕, 別所吾郎, 山形秀明, “アフィン変換の係数を用いた定型帳票内の文字認識”, *信学技報*, PRMU97-186, pp.7-14, (1997)
- [3-7] 中野康明, 藤澤浩道, 国崎 修, 岡田邦弘, 花野井歳弘, “文字認識と協調した表形式文書の理解”, *信学論(D)*, vol.J69-D, no.3, pp.400-409, (1986)
- [3-8] H. Arai, K. Okada, “Form Processing Based on Background Region Analysis”, *Int. Conf. on Document Analysis and Recognition*, vol.1, pp.164-169, (1997)
- [3-9] Y. Y. Tang, C. Y. Suen, et.al, “Financial Document Processing Based on Staff Line and Description Language”, *IEEE Trans. on Systems Man and Cybernetics*, Vol.25, No.5, (1995)
- [3-10] R. Casey, D Ferguson, K. Mohiuddin, and E. Walach, “Intelligent Forms Processing System”, *Machine Vision and Applications*, vol.5, no.3, pp.143-155, (1992)
- [3-11] S. L. Taylor, R. Fritzson, and J. A. Pastor, “Extraction of Data from Preprinted Forms”, *Machine Vision and Applications*, vol.5, No.3, pp.211-222, (1992)
- [3-12] 早川秀樹, 川人光男, 乾 敏郎, “陰影による形状知覚の階層的計算モデル”, *信学論(D-II)*, vol.J75-D-II, no.11, pp.1940-1947, (1992)

### 第4章

- [4-1] 石谷康人, 住田一男, “紙文書を対象としたピボット XML 文書に基づく XML 文書変換システム”, *電子情報通信学会 信学技報*, TL2003-30, pp.7-12, (2004)
- [4-2] M. Berardi, M. Lapi, D. Malerba, “An integrated approach for automatic semantic structure extraction in document images”, *6th International Workshop, DAS 2004, LNCS 3163*, pp.179-190, (2004)

- [4-3] R. Casey, D. Ferguson, K. Mohiuddin, and E. Walach, "Intelligent Forms Processing System", *Machine Vision and Applications*, Vol.5, No.3, pp.143-155, (1992)
- [4-4] 中野康明, 藤澤浩道, 国崎 修, 岡田邦弘, 花野井歳弘, "文字認識と協調した表形式文書の理解", *信学論(D)*, vol.J69-D, no.3, pp.400-409, (1986)
- [4-5] H. Arai, K. Okada, "Form Processing Based on Background Region Analysis", *Int. Conf. on Document Analysis and Recognition*, vol.1, pp.164-169, (1997)
- [4-6] Microsoft Developer Network, "IFilter",  
<http://msdn2.microsoft.com/en-us/library/ms691105.aspx>
- [4-7] Jonathan J. Hull, Peter E. Hart, "Toward Zero-Effort Personal Document Management", *Computer*, pp.30-35, (2001)
- [4-8] Adobe Systems Inc., "*Pdf Reference: Adobe Portable Document Forman, Version 1.3*", Addison-Wesley, (2000)
- [4-9] Léon Bottou, Patrick Haffner, Yann LeCun., "Efficient Compression of Digital Documents to Multilayer Raster Formats", *Int. Conf. on Document Analysis and Recognition*, pp.444-447, (2001)
- [4-10] W3C Recommendation, "Scalable Vector Graphics (SVG) 1.1 Specification", World Wide Web Consortium, <http://www.w3.org/TR/SVG11/>
- [4-11] "XML Paper Specification", Microsoft Corp.,  
<http://www.microsoft.com/xps>
- [4-12] 平野敬, 亀代泰三, 岡田康裕, 依田文夫, "外郭ゼロ交差特徴と正準判別分析法による低品質文字認識", *電子情報通信学会 信学技報*, PRMU98-159, (1998)
- [4-13] T Watanabe, Qin Luo, N Sugie, "Layout Recognition of Multi-Kinds of Table-Form Documents," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 4, pp. 432-445, (1995)
- [4-14] 金堀 利洋, 鈴木 昌和, "PDF中のテキスト情報を利用した視覚障害者のための英文PDF科学技術文書読取りシステム", *信学論(D)*, vol.J90-D, No.3, pp.706-714, (2007)
- [4-15] Tamir Hassan, Robert Baumgartner, "Table Recognition and Understanding from PDF Files", *Int. Conf. on Document Analysis and Recognition*, vol.2, pp.1143-1147, (2007)
- [4-16] M. Armon Rahgozar, "Document Table Recognition by Graph Rewriting", *AGTIVE'99*, LNCS vol.1779, pp.185-197, (2000)
- [4-17] 橋本新一郎(編), "文字認識概論", pp.59-64, オーム社, (1982)
- [4-18] Michael A. Cusumano, "*The Business of Software: What Every Manager, Programmer, and Entrepreneur Must Know to Thrive and Survive in Good Times and Bad*", Free Press, (2004)
- [4-19] Koga T. and Aoyama K., "Quality Management System by Causal Chain Network based on Failure Propagation Model in Manufacturing Process", *Proceedings of the Design Engineering Workshop 2005*, pp.54-64, (2005)

- [4-20] Tamura, Y. and Iizuka, Y., "A Study on the Method to Manage Design Knowledge on Failures—Construction of the Knowledge Structure of a Causal Chain of Failures", In The Japanese Society for Quality Control, vol.32, no.1, (2002)
- [4-21] Suto, H., Kawakami, H. et al., "Development of Support System for Designing Interactive Artifacts", In Trans. Human Interface Society, vol.6, pp.83-89, (2004)
- [4-22] Schilit, B.N., Price, M.N., Golovchinsky, G., "Digital Library Information Appliances", In Proc. Digital Libraries '98, pp.217-226. (1998)
- [4-23] Yunwen Ye, "Programming with an Intelligent Agent", IEEE Intelligent Systems, Vol. 18, No. 3, pp.43-47, (2003)

## 第 5 章

- [5-1] H. Kollnig, H.-H. Nagel, and M. Otte, "Association of motion verbs with vehicle movements extracted from dense optical flow fields", Proc. of 3rd European Conf. on Computer Vision '94, vol.II, pp.338-347, (1994)
- [5-2] H.-H. Nagel., "A vision of 'vision and language' comprises action: An example from road traffic", Artificial Intelligence Review, Vol. 8, pp.189-214, (1994)
- [5-3] G. Herzog and P. Wazinski. Visual translator: Linking perceptions and natural language descriptions. Artificial Intelligence Review, Vol. 8, pp. 175-187, (1994)
- [5-4] G. Herzog and K. Rohr. Integrating vision and language: Towards automatic description of human movements. In Proc. 19th Annual German Conf. on Artificial Intelligence, pp. 257-268, (1995)
- [5-5] N. Okada, "Integrating vision, motion, and language through mind", Artificial Intelligence Review, Vol.9, pp.209-234, (1996)
- [5-6] 小島篤博, 田原典枝, 田村武志, 福永邦雄, "動画像における人物行動の自然言語による説明の生成", IEICE, Vol.J81-D-II, No.8, pp.1867-1875, (1998)
- [5-7] Porikli F., Tuzel O., "Bayesian Background Modeling for Foreground Detection", ACM International Workshop on Video Surveillance and Sensor Networks (VSSN), pp. 55-28, (2005)
- [5-8] Tuzel O., Porikli F. and Meer P., "A Bayesian Approach to Background Modeling", IEEE Workshop on Machine Vision for Intelligent Vehicles (MVIV), Vol. 3, pp.58-65, (2005)
- [5-9] Porikli F., Tuzel O., "Multi-Kernel Object Tracking", IEEE International Conference on Multimedia and Expo (ICME), pp. 1234-1237, (2005)

- [5-10] C. J. Fillmore, "*The case for case*", In E. Bach and R. Harms (editors), *Universals in Linguistic Theory*. Rinehart and Wiston, (1968)
- [5-11] Yuri. A. Ivanov and Aaron. F. Bobick, "Recognition of Visula Activities and Interactions by Stochastic Parsing", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.22, No.8, pp.852-872, (2000)

# 本研究に関する発表論文等

## 原著論文

1. 平野敬, 佐瀬幹哉, 小杉幸夫, “双方向 FeatureMap によるアーム制御”, 信学論 D-II, Vol.J76, No.4, pp.881-888, (1993)
2. 平野敬, 岡田康裕, 依田文夫, “ロバストなモデル照合に基づくFAX送信された一般帳票の読取り”, 信学論 D-II, Vol.J85, No.9, pp.1371-1381, (2002)
3. 平野敬, 岡野祐一, 岡田康裕, 依田文夫, “ページ記述言語の解析に基づく多様な文書からの構造化内容情報の抽出”, 信学論 D-II, 2008 年度 5 月号掲載予定

## 国際会議

1. Takashi Hirano, Yasuhiro Okada and Fumio Yoda, “Field Extraction Method from Existing Forms Transmitted by Facsimile”, ICDAR 01 (2001)
2. Takashi Hirano, Yuichi Okano, Yasuhiro Okada, Fumio Yoda, “Text and Layout Information Extraction from Document Files of Various Formats Based on the Analysis of Page Description Language”, ICDAR2007, (2007)
3. Takashi Hirano, Koichi Tanigaki, Yasuhiro Okada, Shinya Takahashi, “Push-Style Design Support System: Automatic Check of Technical Document with Design Knowledge”, ICKM2007 (2007)
4. Takashi Hirano, Shogo Yoneyama, Yasuhiro Okada, Yukio Kosugi, “Integrating Vision and Language: Semantic Description of Traffic Events from Image Sequences”, ISVC07 (2007)

## 研究会

1. 平野敬, 佐瀬幹哉, 小杉幸夫, “Counter Propagation を用いた双方向 FeatureMap の性質”, 電子情報通信学会技術報告, NC91-43 (1991)
2. 平野敬, 亀代泰三, 岡田康裕, 依田文夫, “外郭ゼロ交差特徴と正準判別分析法による低品質文字認識”, 電子情報通信学会技術報告, PRMU 98-159 (1998)
3. 平野敬, 岡田康裕, 依田文夫, “表の構造解析による FAX 送信帳票からも文字認識フィールド抽出法”, 電子情報通信学会技術報告, PRMU 2000-71 (2000)
4. 平野敬, 岡田康裕, “FAX 送信された既存帳票画像からの認識フィールド抽出方式検討”, 電子情報通信学会技術報告, PRMU 2000-224 (2001)

5. 平野敬, 亀代泰三, 岡田康裕, 依田文夫, “PDL データの解析による多様な形式の文書からの情報抽出方式の検討”, 電子情報通信学会技術報告, PRMU 2004-66 (2004)

#### 特許

1. 特許第 3179280 号, 「表を含む帳票処理装置」, 権利維持, 1994/4/15 出願
2. 特許第 3551004 号, 「罫線抽出方法および装置」, 権利維持, 1998/2/12 出願
3. 特開 2000-163516, 「文字認識装置」, 審査請求済, 1998/12/1 出願
4. 特許第 3750406 号, 「文書ファイリング装置」, 権利維持, 1999/3/19 出願
5. 特開 2001-229341, 「文字認識装置及び文字認識方法」, 審査請求済, 2000/2/14 出願
6. 特開 2001-325258, 「書類管理システム」, 審査請求済, 2000/5/15 出願
7. 特許第 3824209 号, 「文書自動分割装置」, 権利維持, 2001/4/18 出願
8. 特開 2004-227255, 「文書解析装置」, 審査請求済, 2003/1/22 出願
9. 特許第 3844765 号, 「文字認識装置」, 権利維持, 2004/6/25 出願
10. 特開 2007-048148, 「ガイダンス情報提示システム、文書・図面作成端末および知識サーバ」, 出願済, 2005/8/11 出願
11. 特開 2007-193500, 「文書または図面作成支援装置」, 出願済, 2006/1/18 出願

# 付録1 メンタル・スペース理論

従来、自然言語の意味論は論理学をベースに検討されたが、Fauconnier は数学的な視点ではなく、認知的な視点に立ちメンタル・スペース理論を提唱した。ここでは、自然言語表現は曖昧であり意味を記述することができず、メンタルスペースに存在する要素や、要素間の関係を設定するための指令であると考え（図 A.1）。

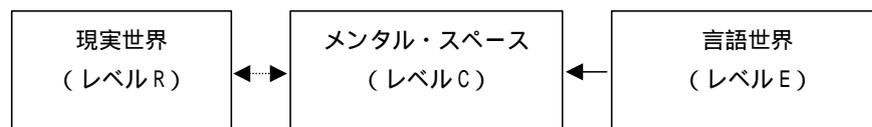


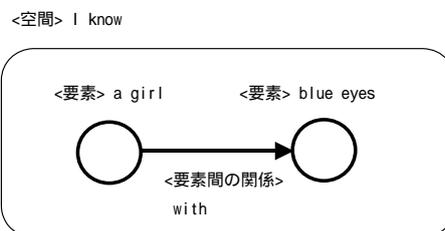
図 A.1 メンタルスペース理論

以下に、メンタルスペースの概念を、サンプルを用いて示す（「メンタル・スペース<新版>-自然言語理解の認知インターフェイス-」 フォコニエ・G 著, 坂原茂/水光雅則/田窪行則/三藤博 訳, 白水社, 1996 を参考）

## サンプル1

“私は青い目をした少女を知っている。”  
 “I know a girl with blue eyes.”

この文は、以下のようなメンタルスペースとして、以下の<空間>、<要素>、<要素間の関係>を設定する。



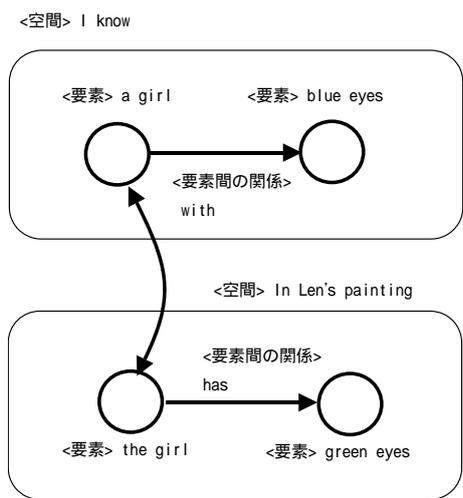
- <空間> : 私の知っている (I know)
- <要素> : 少女 (a girl), 青い目 (blue eyes)
- <要素間の関係> 持つ (with)

図 A.2 サンプル1 に対するメンタルスペース例

次に、以下のサンプル2に示す文が現れた場合、メンタルスペースは図A.3のように拡張される。

サンプル2

“ In Len's painting, **the girl** has green eyes.”  
 “ レンの絵では、**その少女**は緑色の目をしている。”



図A.3 サンプル2 に対するメンタルスペース例

図A.3では新たにレンの絵という空間が作成され、その中で閉じた要素間の関係を示すと共に、既に存在した（I know）空間内の要素との関係性も追加される。このように新しい空間を設定したり、談話において既に導入されている空間との関係を表すことを“スペース導入表現”と呼ぶ。スペース導入表現には、例えば以下の表現を利用できる。

前置詞： in Len's picture , in 1993 , in the company

副詞： really , probably , definitely , possibly , theoretically

命題結合子： if A then , either A or B

主語と動詞の結合： He believes , I hopes

ここでは空間の中に、別な空間を設けるような多重のスペース導入表現も許す。このように数理論理学で表現が困難な非常に複雑な概念や、矛盾や曖昧性を含む関係を表現できる点が、メンタル・スペースの大きな利点である。

## 謝辞

本研究を進めるにあたり、あたたかくご指導頂き、正しい方向に導いてくださった東京工業大学 総合理工学研究科 メカノマイクロ工学専攻 小杉幸夫教授に心より感謝するとともに、深く御礼申し上げます。また、同専攻の北條春夫教授、香川利春教授、初澤毅教授、小俣透教授には本論文をまとめるにあたり、有意義なご助言を賜り深く感謝致します。

本研究は、三菱電機(株)の関係各位のご協力とご支援の賜物です。FAX-OCR システムの技術開発に関し、ご支援とご協力を頂いた三菱電機インフォメーションシステムズ(株) ネットワークソリューション部の福田隆部長、小野健一課長、滝田健司氏に感謝します。

また企業文書管理パッケージや設計不具合防止支援ソリューションの開発に関し、ご指導とご支援を頂いた 同三菱電機インフォメーションシステムズ(株) ERP ソリューション部の岡村博之課長、中村伊知郎博士、亀代泰三氏、小島栄之氏、中谷壮志氏、長山馨氏、稲葉豊氏を始めとした関係者の皆様に感謝致します。

さらに特殊用途向け文書・図面管理システムの開発にあたり、適切なご指導と技術支援を頂いた 三菱電機(株) 神戸製作所 社会システム第三部の西村靖課長、谷原武郎氏、吉田剛氏、福島繁樹氏、成相剛士氏を始めとした関係者の皆様に深く感謝致します。

また FA 用表示器の開発にあたりご助言、ご指導頂きました三菱電機(株) 名古屋製作所 HMI システム部 小荒健吾博士に感謝致します。

さらに文書管理システムや設計不具合防止支援ソリューションの社内適用・評価についてご協力頂いた、三菱電機(株) 設計システム技術センター ハードウェアエンジニアリング部の長江雅史氏、川北泰之氏、同社生産技術センター 量産化技術推進部同設計システム技術部の牧野恵三氏、同社名古屋製作所 生産システム推進部 石井義一課長、戸川幸一氏、同製作所 FA システム部 高橋俊哉次長、阿部薫センタ長、新藤研二氏、山本順司氏 を始めとした関係者の皆様に感謝致します。

最後に、本研究の場を与えて下さった 三菱電機(株) 情報技術総合研究所 西井龍五所長、同研究所 前所長 肥塚裕至博士と、入社以来 ご指導いただいている同研究所 マルチメディア部門 依田文夫部門長、同研究所 音声言語処理技術部 高橋真哉部長、岡田康裕チームリーダーに深く感謝すると共に、同技術部 文書処理チームと IM チームのメンバ、および多忙な状態での支えとなってくれた家族に心から感謝致します。