

画像データベースにおけるモデル形成に関する研究

佐 藤 真 一

①

画像データベースにおける
モデル形成に関する研究

佐藤真一

博士論文

画像データベースにおける
モデル形成に関する研究

1991年度

東京大学 工学系研究科
情報工学専攻

佐藤真一

目次

序	1
第一章 概論	1
1.1 研究の背景	1
1.2 研究の目的	2
1.3 研究の範囲	3
1.4 研究の方法	4
1.5 研究の意義	5
第二章 基礎理論	6
2.1 基礎理論の概要	6
2.2 基礎理論の発展	7
2.3 基礎理論の応用	8
2.4 基礎理論の限界	9
2.5 基礎理論の展望	10
第三章 実験的検証	11
3.1 実験的検証の概要	11
3.2 実験的検証の方法	12
3.3 実験的検証の結果	13
3.4 実験的検証の考察	14
3.5 実験的検証の結論	15
第四章 結論	16
4.1 結論の概要	16
4.2 結論の考察	17
4.3 結論の展望	18
4.4 結論の意義	19
4.5 結論の限界	20
4.6 結論の展望	21
4.7 結論の展望	22
4.8 結論の展望	23
4.9 結論の展望	24
4.10 結論の展望	25
4.11 結論の展望	26
4.12 結論の展望	27
4.13 結論の展望	28
4.14 結論の展望	29
4.15 結論の展望	30
4.16 結論の展望	31
4.17 結論の展望	32
4.18 結論の展望	33
4.19 結論の展望	34
4.20 結論の展望	35
4.21 結論の展望	36
4.22 結論の展望	37
4.23 結論の展望	38
4.24 結論の展望	39
4.25 結論の展望	40
4.26 結論の展望	41
4.27 結論の展望	42
4.28 結論の展望	43
4.29 結論の展望	44
4.30 結論の展望	45
4.31 結論の展望	46
4.32 結論の展望	47
4.33 結論の展望	48
4.34 結論の展望	49
4.35 結論の展望	50
4.36 結論の展望	51
4.37 結論の展望	52
4.38 結論の展望	53
4.39 結論の展望	54
4.40 結論の展望	55
4.41 結論の展望	56
4.42 結論の展望	57
4.43 結論の展望	58
4.44 結論の展望	59
4.45 結論の展望	60
4.46 結論の展望	61
4.47 結論の展望	62
4.48 結論の展望	63
4.49 結論の展望	64
4.50 結論の展望	65
4.51 結論の展望	66
4.52 結論の展望	67
4.53 結論の展望	68
4.54 結論の展望	69
4.55 結論の展望	70
4.56 結論の展望	71
4.57 結論の展望	72
4.58 結論の展望	73
4.59 結論の展望	74
4.60 結論の展望	75
4.61 結論の展望	76
4.62 結論の展望	77
4.63 結論の展望	78
4.64 結論の展望	79
4.65 結論の展望	80
4.66 結論の展望	81
4.67 結論の展望	82
4.68 結論の展望	83
4.69 結論の展望	84
4.70 結論の展望	85
4.71 結論の展望	86
4.72 結論の展望	87
4.73 結論の展望	88
4.74 結論の展望	89
4.75 結論の展望	90
4.76 結論の展望	91
4.77 結論の展望	92
4.78 結論の展望	93
4.79 結論の展望	94
4.80 結論の展望	95
4.81 結論の展望	96
4.82 結論の展望	97
4.83 結論の展望	98
4.84 結論の展望	99
4.85 結論の展望	100
4.86 結論の展望	101
4.87 結論の展望	102
4.88 結論の展望	103
4.89 結論の展望	104
4.90 結論の展望	105
4.91 結論の展望	106
4.92 結論の展望	107
4.93 結論の展望	108
4.94 結論の展望	109
4.95 結論の展望	110
4.96 結論の展望	111
4.97 結論の展望	112
4.98 結論の展望	113
4.99 結論の展望	114
4.100 結論の展望	115

1 序章	1
1.1 研究の背景と目的	2
1.2 論文の構成	7
2 画像データベースシステム	9
2.1 概要	10
2.2 画像データベースシステムの構成	11
2.3 背景技術	14
2.3.1 画像認識・理解技術によるデータモデル獲得	14
2.3.2 マルチメディアアプリケーションシステムによるデータモデル利用	21
2.3.3 オブジェクト指向データモデルによるデータモデル管理	24
2.3.4 画像データベースシステムにおけるコンピュータ学習の利用	30
2.4 画像データベースシステムにおける機能部分への要件	33
2.4.1 画像データモデル獲得の要件	33
2.4.2 画像データモデル利用の要件	41
2.4.3 画像データモデル管理の要件	45
3 データモデル獲得を考慮した画像データモデル	49
3.1 概要	50
3.2 状態遷移モデル	50
3.2.1 状態遷移による画像認識・理解プロセス	52
3.2.2 状態間の依存関係の明示	54
3.2.3 状態遷移モデルに基づく画像理解機構の構成	54
3.3 概念的なデータモデルとの親和性	56
3.4 不完全な認識結果の扱い	57
3.5 考察	59
4 状態遷移モデルに基づいた図面理解システム	61

4.1	概要	62
4.2	システムの特徴	63
4.3	システムの構成	66
4.3.1	トークンエクストラクタ	68
4.3.2	認識カーネル	70
4.4	本枠組に基づく図面認識システムの実現	72
4.4.1	システムの実装	72
4.4.2	手書き線図形の認識	73
4.4.3	地図図面(国土基本図)の認識	77
4.4.4	システムの評価	82
4.5	考察	83
5	状態遷移モデルに基づく一般画像の認識とその利用	85
5.1	概要	86
5.2	システムの構成	86
5.3	状態遷移モデルによる一般画像理解システム	89
5.3.1	初期トークンの生成	89
5.3.2	画像理解プロセス	91
5.4	スポーツ画像検索システム	93
5.5	考察	95
6	階層構造を持つ認識・理解機構による画像理解	96
6.1	概要	97
6.2	多階層状態遷移モデル	98
6.2.1	多階層認識理解モデル	98
6.2.2	多階層状態遷移モデルに基づく画像理解システムの構成	104
6.3	実験システム	106
6.3.1	線画の解釈システム	106

6.3.2	一般画像理解システム	115
6.3.3	考察	124
7	画像データベース形成におけるコンピュータ学習の利用	125
7.1	概要	126
7.2	状態遷移型図面／画像理解システムのルール作成支援システム	127
7.2.1	システムの構成	127
7.2.2	ルール作成支援の実験	129
7.3	図形事例からのモデル獲得システム	132
7.3.1	システムの構成	132
7.3.2	実行例	134
7.4	考察	136
8	終章	137
	参考文献	141
	付録	149
A	BD木ライブラリを用いた空間 Prolog 処理系	149
A.1	概要	150
A.1.1	処理系の特徴	150
A.1.2	処理系の適用分野	151
A.2	処理系の構成	152
A.2.1	全体構成	152
A.2.2	BD木ライブラリ	153
A.2.3	図形データの内部構造	155
A.3	拡張した述語	160
A.4	考察	161

B	状態遷移型図面理解システムにおける人間機械協調によるルール作成支援システム	162
B.1	概要	163
B.2	システムの構成	163
B.3	状態遷移ルールの学習	165
B.3.1	状態遷移モデルにおけるボトムアッププロセス	165
B.3.2	ボトムアップルール作成支援	167
B.3.3	帰納推論の手法	169
B.3.4	状態遷移ルール作成支援システム	172
B.4	ルール作成支援システムの高度化の検討	174
B.4.1	パラメータ計測ルールの生成	174
B.4.2	人間機械協調型図面理解システムへの適用の検討	177
B.5	考察	180
C	図形で与えられた事例からの構造記述の学習	181
C.1	概要	182
C.2	構造記述の学習	183
C.3	構造記述学習システム	186
C.3.1	システムの構成	186
C.3.2	システムに与えられる「例」の記述	188
C.3.3	システムで扱う構造記述	192
C.3.4	学習アルゴリズム	195
C.4	システムの実装と実行例	198
C.5	考察	202
	索引	203

目 次

第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究目的与内容	2
1.3 研究方法	3
1.4 研究创新点	4
第二章 相关理论与技术	5
2.1 神经网络基础	5
2.2 深度学习概述	6
2.3 卷积神经网络	7
2.4 循环神经网络	8
2.5 生成对抗网络	9
2.6 变分自编码器	10
2.7 自监督学习	11
2.8 迁移学习	12
2.9 多任务学习	13
2.10 知识蒸馏	14
2.11 模型压缩	15
2.12 边缘计算	16
2.13 云计算	17
2.14 物联网	18
2.15 大数据	19
2.16 人工智能	20
2.17 机器学习	21
2.18 深度学习	22
2.19 神经网络	23
2.20 卷积神经网络	24
2.21 循环神经网络	25
2.22 生成对抗网络	26
2.23 变分自编码器	27
2.24 自监督学习	28
2.25 迁移学习	29
2.26 多任务学习	30
2.27 知识蒸馏	31
2.28 模型压缩	32
2.29 边缘计算	33
2.30 云计算	34
2.31 物联网	35
2.32 大数据	36
2.33 人工智能	37
2.34 机器学习	38
2.35 深度学习	39
2.36 神经网络	40
2.37 卷积神经网络	41
2.38 循环神经网络	42
2.39 生成对抗网络	43
2.40 变分自编码器	44
2.41 自监督学习	45
2.42 迁移学习	46
2.43 多任务学习	47
2.44 知识蒸馏	48
2.45 模型压缩	49
2.46 边缘计算	50
2.47 云计算	51
2.48 物联网	52
2.49 大数据	53
2.50 人工智能	54
2.51 机器学习	55
2.52 深度学习	56
2.53 神经网络	57
2.54 卷积神经网络	58
2.55 循环神经网络	59
2.56 生成对抗网络	60
2.57 变分自编码器	61
2.58 自监督学习	62
2.59 迁移学习	63
2.60 多任务学习	64
2.61 知识蒸馏	65
2.62 模型压缩	66
2.63 边缘计算	67
2.64 云计算	68
2.65 物联网	69
2.66 大数据	70
2.67 人工智能	71
2.68 机器学习	72
2.69 深度学习	73
2.70 神经网络	74
2.71 卷积神经网络	75
2.72 循环神经网络	76
2.73 生成对抗网络	77
2.74 变分自编码器	78
2.75 自监督学习	79
2.76 迁移学习	80
2.77 多任务学习	81
2.78 知识蒸馏	82
2.79 模型压缩	83
2.80 边缘计算	84
2.81 云计算	85
2.82 物联网	86
2.83 大数据	87
2.84 人工智能	88
2.85 机器学习	89
2.86 深度学习	90
2.87 神经网络	91
2.88 卷积神经网络	92
2.89 循环神经网络	93
2.90 生成对抗网络	94
2.91 变分自编码器	95
2.92 自监督学习	96
2.93 迁移学习	97
2.94 多任务学习	98
2.95 知识蒸馏	99
2.96 模型压缩	100
2.97 边缘计算	101
2.98 云计算	102
2.99 物联网	103
2.100 大数据	104

1.1 データベースシステムの構成	3
2.1 画像データベースシステムの構成	12
2.2 複合オブジェクトによる階層構造の例	26
2.3 クラス階層の例	28
2.4 内部状態を持つ計算モデルによる画像認識システムの表現	38
3.1 状態遷移モデルによるデータモデルの概要	51
3.2 トークンの状態遷移の様子	53
3.3 典型的な状態遷移図の例	55
3.4 画像理解プロセスにおけるトークンの状態遷移ダイアグラムの例	58
4.1 図面イメージの例	64
4.2 状態遷移モデルに基づく図面理解システムの概略	67
4.3 AI-Mudams の概略	69
4.4 ルールの形式	70
4.5 三角形抽出のためのルール	74
4.6 手書き図面からの三角形の抽出過程	75
4.7 三角形抽出のための状態遷移ダイアグラム	76
4.8 地図認識における状態遷移ダイアグラム	78
4.9 地図認識における状態遷移規則 (一部分)	79
4.10 地図の認識の例	80
4.11 機械設計図面の例	84
5.1 画像理解システムと画像検索システムとを含むシステムの全体構成	87
5.2 初期トークンの生成の例	90
5.3 スポーツ画像認識のための状態遷移図	92
5.4 状態遷移モデルによる不完全な認識結果を利用した検索	94
5.5 画像検索の例	94

6.1	異なる抽象レベル間の協調処理	101
6.2	多階層認識理解モデルにおける階層間の協調	102
6.3	多階層状態遷移モデルによる画像理解システムの構成	105
6.4	線画によるシーンの例	107
6.5	バスの線画による初期階層	110
6.6	バスのモデル	112
6.7	大まかな構造に着目した照合	113
6.8	照合におけるサブシンボルの利用	114
6.9	シーン階層の状態遷移	115
6.10	対象となる一般画像	116
6.11	$\sigma = 5$ 画素の Marr-Hildreth のオペレータにより生成したゼロクロスと線分 トークン	118
6.12	生成した長方形トークンの例	120
6.13	長方形により構成された人間のモデル	121
6.14	抽出された人物の例	122
6.15	スポーツ画像を対象としたシーン階層の状態遷移	123
7.1	状態遷移型図面／画像理解システムのルール作成支援システム	128
7.2	図形事例からのモデル獲得システムと線図形認識システム	133
7.3	学習システムに与えられる図形事例	135
A.1	本処理系の構成	152
A.2	BD 木による 2 次元空間データの表現	154
A.3	サンプルプログラム	156
A.4	bd.h ヘッダーファイル	157
A.5	図形データの Prolog データベース上での表現	158
A.6	空間データベースにおける空間データの定義	159

B.1	状態遷移型図面理解システムのルール作成支援システム	164
B.2	一般的な状態遷移の1ステップ	167
B.3	大規模な状態遷移ルールの作成支援	175
B.4	パラメータ計測ルールのテンプレート	176
B.5	状態遷移ルール学習システムを利用した人間機械協調型図面認識システムの構成	179
C.1	「家」の学習のための「例」	184
C.2	構造記述学習システムの構成	187
C.3	システムに与える「例」の記述	189
C.4	空間データ構造BD木による空間性の扱い	190
C.5	部分的に構造化されたオブジェクトの表現	191
C.6	実際の図形データから生成した部分的に構造化された記述	199
C.7	「バス」の学習のための例	201

表目次

1.1 文字・数値型データベースと画像データベースの比較	5
A.1 Prolog データベース上での幾何形状の表現	159

第 1 章

第 1 章

序章

1.1 研究の背景と目的

近年、映像情報を中心とするマルチメディアデータを扱うためのより高度な機器の開発、およびその普及によるパーソナル化を基盤として、こうしたデータを高度に、かつ大規模に扱い得るシステムのための枠組が望まれている。その実現のための最も有力と考えられる手段が、大規模なマルチメディアデータベース／画像データベースを中心としたマルチメディアアプリケーションシステムである。

今、データモデルの獲得からその利用までも含めた、広く考えたデータベースシステムについて考える。一般的なデータベースシステムについて、これを模式的に表した概念図を図 1.1 に示す。データモデル獲得では、情報を外部世界からセンサーなどを用いて取り込んだり、人手で入力するなどして、データベースとして運用可能なデータモデルという形で抽出する。データモデル管理は、狭い意味でのデータベースシステムであり、与えられたデータモデルに従って外部世界の情報を取り込み、データベースの利用者に高度な操作手段を提供する。データモデル利用では、このように獲得され、管理されている高度なデータモデルを用い、これをさまざまなアプリケーションに供する。一般的にデータベースシステムは、これら3つの機能部分からなると考えられる。

従来のデータベースシステムに対するアプローチでは、この各機能部分を全く個別に扱ってきた。データモデル獲得では、とにかく外部世界の実データからいかに高度な応用が可能なデータモデルを抽出するかに力点が置かれている。データモデル管理では、このデータモデルを概念的な高度な利用が可能な形で管理する方法、またこれを操作する際に用いるインターフェースのためのプログラミング言語、あるいはこれを実際の機械に実装するための手法などに重点が置かれている。データモデル利用では、高度に概念的な形で与えられたデータモデルを用いることにより、いかに分かりやすく、いかに効率良くプレゼンテーションするか、あるいは、今後情報のマルチメディア化に伴いどのような高度なアプリケーションがあり得るかなどの点に主に焦点がしぼられている。すなわち、まずデータモデルとしては高度な応用が可能なように十分概念的なレベルで与えることを前提にして、3つの機能部分はそれぞれ互いの状況に関知することなく、データモデル獲得ではこのデータモデルの

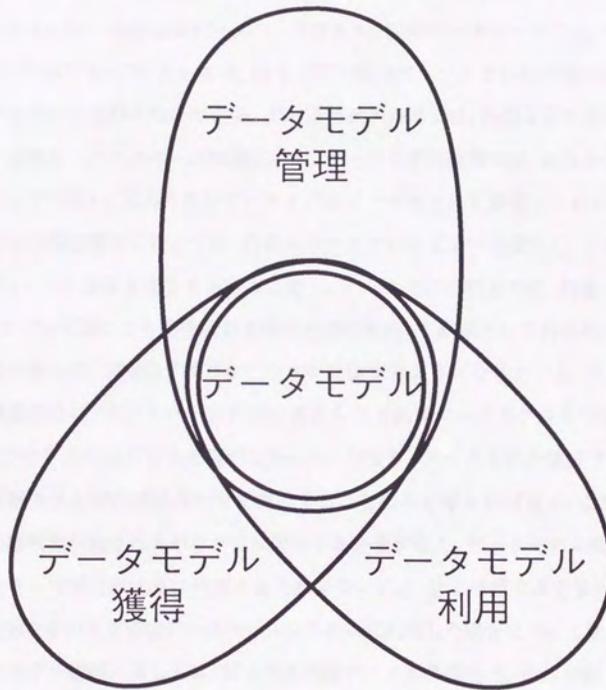


図 1.1: データベースシステムの構成

完全な抽出に、データモデル管理ではより高度で効率の良い管理手法に、そしてデータモデル利用ではこの高度なデータモデルを用いてのより高度なアプリケーションの作成に、それぞれ関心が持たれてきた。

文字・数値型のデータベースシステムの場合について考える。このデータベースとして完成しているものの一つが関係データベースである。関係データベースでは、データモデルの内部表現(内部スキーマ)としては、扱う文字・数値データとそれらの間の関係とで構成されるタプルという形態を用いている。データモデル獲得では、外部世界の事象を文字・数値データに変換してデータベースに投入する。データモデル管理では、あらかじめ設定された概念スキーマに従い、入力されたデータを内部スキーマとして蓄積しておいて、データモデル利用からの操作要求に対しては、外部スキーマというビューを設定し、これに従った関係演算を許し、この演算を実行するように働く。データモデル利用では、外部スキーマを用いて SQL などの言語により関係演算を用いた操作を行い、結果として得られた文字・数値データを受け取って、実際のアプリケーションに供するようになっている。この場合、基本的には各機能部分は、データベース内部に蓄えられる関係データモデルとの関連だけを考え、他の部分の存在は気にする必要がなかった。つまり、データモデル獲得では、外部世界から情報を取り込んで内部スキーマを構成することのみを考えれば良く、これに対してどのような関係演算が施されるかなどには関与する必要がなく、データモデル利用でも、利用しているスキーマがどのように構成されたかについては、全く考慮する必要がなかった。

この考え方をそのまま画像データベースシステムに利用した場合について考える(表 1.1)。まずデータモデル管理に関しては、扱うべき画像データの多様性や、それが表す概念の複雑さなどの理由により、従来の文字・数値型データベースにおいて有効であった関係データモデルなどによるデータモデル化では、もはや対応できなくなっている。これに対しては、オブジェクト指向データモデルなどの、より記述性の高いモデルの利用により対処できると考えられる。データモデル利用においては、扱うデータの多様性、またそのデータの特性を十二分に生かした、高度 GIS やハイパーメディアといった、高度で多彩なアプリケーションシステムが考えられている。こうしたシステムのデータモデルへの操作要求に対しては関係演算のように画一的な操作のみでは対応し切れないと考えられるが、これもオブジェク

表 1.1: 文字・数値型データベースと画像データベースの比較

	文字・数値型データベース	画像データベース
データモデル獲得	文字・数値データをそのまま入力する。	多様な素材データを高度に抽象化・概念化する必要がある。
データモデル管理	関係データベースの場合、文字・数値データによるタプルという表現で管理する。	多様なデータに対し、より概念的で複雑なモデルを表現する必要がある。データモデルに対する多彩な操作にも対応する必要がある。
データモデル利用	関係演算によりデータモデル(外部スキーマ)を操作する。結果も文字・数値データとして得られ、その扱いは簡単である。	アプリケーションごとに多彩な操作があり得る。結果のデータも多様で、魅力的に、かつ分かりやすくプレゼンテーションする必要がある。

ト指向データモデルにおける各クラスのメソッドなどの形で対応できることが期待される。

ここで、画像データベースにおけるデータモデル獲得について考える。画像データベースにおけるデータモデル(画像データモデル)では、上述のような高度で多彩な応用が可能のように、データモデル自体を高度な概念的な記述に抽象化しておく必要がある。データモデル獲得では、この抽象化作業を行わなければならない。画像データベースにおけるデータモデル獲得にはいろいろな手法が考えられるが、大きく分けて、完全に人手によりもとのデータからデータモデルを作成する方法と、コンピュータにより自動的にデータモデルを生成する方法とがある。

人手による方法は、現在実用に供されている画像データモデルの獲得方法として、おそらく最も多く見られる方法である。これは、各画像を人手によって一つずつ抽象化し、画像検索のためのキーワードを付加したり、ハイパーメディアのためのリンク情報を作成したりするという方法である。こうして生成されたデータモデルは、非常に質が高く、しかも均質なものとなるが、反面各画像ごとに人間が介入する必要があるため、大変な手間がかかるという問題がある。文字・数値型データベースの場合でも、実際の文字・数値データを人手によりキーボードなどから入力するという作業が必要であったが、これらのデータはそもそも抽象的なデータであり、データベースにはこれをそのまま入力すれば良かったので、この人手

による方法でも対応可能であった。画像データベースの場合、この人手による方法は、データベース構築のため大規模にデータモデルを作成するという用途には向かないといえる。

そこで、画像データベースの場合には、コンピュータを用いて自動的にデータモデル獲得を行う方法に期待がかかる。画像データモデル獲得のための中心技術は画像認識・理解技術であるが、現状の技術水準では、画像データベースのために望まれているような多様な対象に対する多彩な視点からの高度な認識要求に対し、十分に応えることができない。このため、実用的な画像データベースの実現に際し、この画像データモデル獲得がその大きなボトルネックとなってしまうことが危惧される。

これは、従来のデータベースシステムの構成において、データモデルの抽象度の設定レベルが、その利用者の都合にのみ合わせたものであり、獲得側のことを全く考慮しなかったためであると考えられる。画像データベースにおいては、データモデルの抽象度を、その利用側のみでなく、獲得側、あるいはさらにその管理側のことも考えて、その3つの機能部分を統合してとらえた上で設定するという視点が重要となろう。

また、このように現状の画像認識・理解の技術水準に歩み寄った形に画像データベースを構成するという視点はもちろん重要であるが、今後より先進的なマルチメディアアプリケーションが作成されるに伴い、多様・多彩で高度な画像データモデルへの要求はますます高まって行くと考えられる。この要求を満たすためには、やはり最終的には画像データモデル獲得自体のさらなる高度化が必要不可欠である。従来の画像認識・理解手法では考えられていなかった、認識結果が画像データベースにおけるデータモデルとして使用されるという点、また画像データベースにおける他の機能部分からの補助が利用できるという点に着目した、新たなアプローチに基づく画像データモデル獲得の高度化が考えられる。

本論文では、これらの状況をふまえた上で、現在その実現が強く望まれている実用的な次世代画像データベースを構成する方法について論じる。ここでのアプローチでは、特にデータモデル獲得を重視し、データモデルとしてはコンピュータにより自動化された方法で獲得可能な情報のみを用いることを前提とする。このように、データモデル獲得を重視したデータモデルの抽象度の設定を行うことにより、画像データベースを構築する3つの機能部分がデータモデル獲得の問題点を補助する形になり、バランスの良いシステムが構成できる

ことが期待できる。その上で、画像データベースにおける3つの機能部分が互いに補助し合うというこの視点から、画像データモデル獲得のさらなる高度化の方策について論じる。

以下本論文では、まず画像データベースの構成とその3つの機能部分について述べ、各機能部分に求められる要件についてまとめる。ついで各機能部分について、実際の関連技術の動向を交えて、画像データベース構築に向けてのさらに具体的な方策について論じる。その上で、このデータモデル獲得を重視したデータモデル設定の視点に立った実際の画像データモデルの例について述べ、それをういたデータモデル獲得と利用の例について紹介する。また、より概念的なデータモデルや外部システムとのインタラクションに対応した新しい画像認識・理解機構について述べる。画像データモデル獲得の高度化としては、データモデル獲得過程とデータモデル利用システム、あるいは外部システムとのインタラクションにおけるコンピュータ学習手法の役割について論じる。これらの議論およびそれらを実証するための実験システムによる知見から、画像データベースシステムにおける本論文での視点の有効性を示し、また今後の画像データモデル獲得への一つの有力な指針を示したい。

1.2 論文の構成

本章は本論文の序章であり、研究の背景と目的、および本論文の構成について述べている。

2章では、実現が望まれている高度な画像データベースの実現に向けての重要な指針について述べる。そのために、まず一般的な画像データベースシステムの構成について述べ、その機能部分に関連する重要な背景技術をあげ、その上で各機能部分に望まれる要件を明らかにする。

3章では、この要件の多くを満足する新しい視点に立った画像データモデルとして状態遷移モデルに基づく画像データモデルについて述べる。次いで4章では、この状態遷移モデルに基づく図面理解システムについて述べる。5章では、状態遷移モデルに基づく画像理解システムにより自動獲得された画像データベースを用い、これを高度に検索する画像検索システムについて述べる。6章では、さらなる画像理解手法の高度化を求めて、状態遷移モデルの拡張として多階層認識・理解モデルに基づく多階層状態遷移モデルについて述べる。

7章では、画像データモデル獲得におけるコンピュータ学習の手法の利用方について考察する。まず状態遷移モデルに基づく画像理解システム自体の作成を支援するシステムについて述べる。次いでユーザから与えられた事例から画像理解のためのモデルを自動獲得する手法について述べる。

8章は終章であり、全体のまとめと考察を行う。

第 2 章

画像データベースシステム

2.1 概要

より高度な応用が可能な次世代画像データベースシステムを実現するためには、次にあげるような点が重要となってくる。

1. 画像データベースとして運用可能な高度なデータモデルを素材画像データから大規模・体系的に獲得する。
2. アプリケーションシステムからの要求に従い、より高度で多彩な応用が可能なデータモデルを提供する。

前者はデータベース構築、すなわちデータモデルの獲得に関する要件であり、後者はデータベース利用、すなわちデータモデルの利用に関する要件である。データモデル獲得における大規模・体系的なデータモデル獲得とは、データベースとして利用できるような大規模なデータモデルを獲得することができ、しかも手法としては素材データの対象依存のアドホックなものではなく、汎用性があり拡張性も持っているようなデータモデル獲得が望まれていることを言っている。

データベース利用からの要求に従って高度で多彩な応用を行おうとした場合、利用するデータモデルは多様な視点に立って高度に概念化した形で蓄えておく必要がある。このためには、データモデル獲得の際に素材データからこのような応用が可能なような高度に概念化したデータモデルを抽出しなければならない。データモデル獲得には、大きく分けて人手によりデータモデルを作成する方法とコンピュータにより自動的に生成する方法とがある。人手による方法では、人間の持つ高度で柔軟な認知能力を用いることにより極めて品質の高いデータモデルが得られるが、データモデルの作成に非常に手間がかかり、画像データベースとして運用するための大規模なデータモデル獲得としては利用し難い。コンピュータによる方法とは、画像認識・理解技術を用いてもとの画像データからこれを概念化した記述を自動的に生成し、これを画像データベースのデータモデルとして利用しようという方法である。この方法は、大規模なデータモデル獲得としては適当であるが、画像認識・理解技術の未熟さから高度な応用が可能なデータモデルとして満足のいく概念化が行えない場合

があり得る。

本論文では、先にあげた大規模なデータモデル獲得ができ、高度な応用が可能な画像データベースシステムの構築のため、基本的にデータモデルの獲得は自動的に行うことを前提とする立場をとる。その結果データモデルとしては完全な概念化が行えないことが考えられるが、これをカバーし、高度な応用が行い得るような画像データベースシステムの構築を目標とする。本章では、このようなデータベースシステムの構築のための基本的な方針について考察する。まず2.2節では、データベースシステムをデータモデル獲得・データモデル管理・データモデル応用という3つの機能部分からなるとして考え、これらの間の密接な相互作用という新たな視点から画像データベースシステムの構成について考える。その上で本論文における画像データベース構築に対する基本的な立場を明確にする。2.3節では、これら3つの機能部分の構成と関係が深いと考えられる既存の技術についてまとめる。2.4節では、これらの考察を踏まえ、望まれている画像データベースの構築のためにシステムを構成する各機能部分に求められると考えられる要件について述べる。

2.2 画像データベースシステムの構成

扱う対象のデータモデルを中心にしてデータベースシステムを考えると、データベースシステムはデータモデル獲得・データモデル管理・データモデル利用という3つの機能部分から成り立っていると考えることができる。データベースシステムとしての処理の流れを追うと、まずデータモデル獲得が素材データからデータベースとして高度な利用が可能な形態のデータモデルを抽出する。データモデル管理では、こうしてできたデータモデルを受けとり、これを内部に蓄えてデータベース利用者に高度で統一的な利用手段を提供する。データモデル利用では、データベースシステムの利用者がこうしたデータモデルを参照し、これにさまざまな操作を加えるなどして実際のアプリケーションシステムを実現する。

データベースシステムのこれらの機能部分による構成は、画像データベースの場合にも適用できる。画像データベースシステムにおけるこれらの一連の機能部分を模式的につなぎ合わせた概念図を図2.1に示す。画像データベースシステムをこのような構成としてとらえ

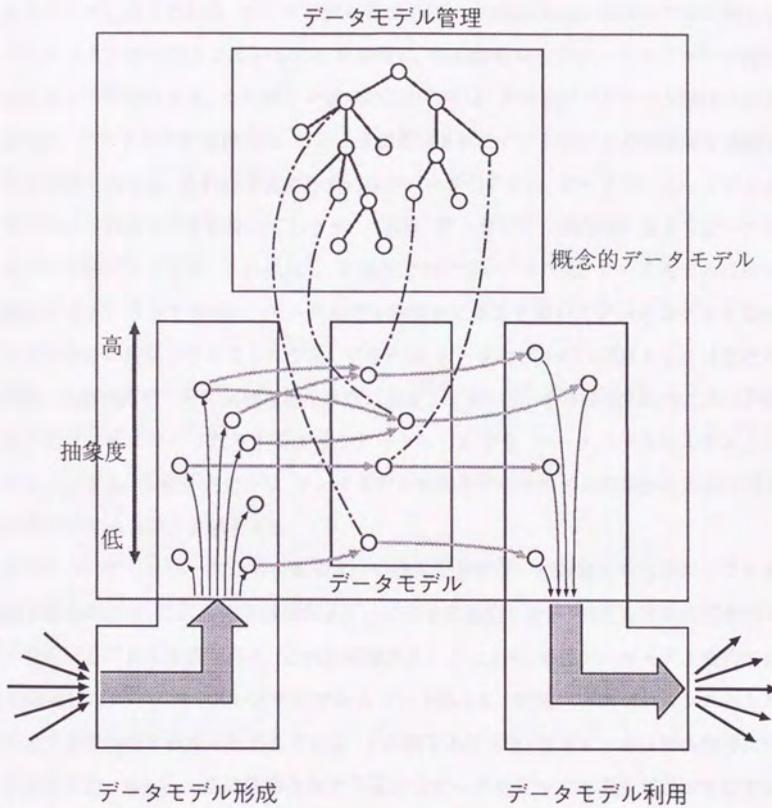


図 2.1: 画像データベースシステムの構成

た上で、本論文における画像データベース実現のための基本的な立場を明らかにしておく。

データベースシステムの実現における従来の主なアプローチでは、まずデータモデルの抽象度をデータモデル利用からの要求によって定め、他の各部分はこれを達成するように作成することが目標とされた。データモデル管理では、この利用部分の要求を十分に満たし得るデータモデルが実現されていることを前提に、利用部分からのデータモデルへの操作を実現するように動作する。この厳しい要求のしわ寄せは、すべてデータモデル獲得に来ることになる。データモデル獲得では、このように設定されたデータモデルの抽象度を達成することが目標とされる。これが達成できていないデータモデルは、データベースシステムとして受け付けられないことになってしまう。これは、データモデル利用部に重きを置いたデータモデルの設定といえる。これに対し、本論文でのアプローチでは、データモデルは自動的に獲得するという点を重視し、データモデル獲得部に重きを置いてデータモデルを設定することを第一の前提とすることにする。すなわち、データモデルの抽象度としては現状の画像認識・理解技術で十分に実現可能な水準に設定しておいて、その技術水準のために不完全であったり一様でなかったりするようなデータモデルでも、データベースシステムとして許すようにする。この不完全さは、データモデル管理やデータモデル利用からの歩み寄りによりカバーするという立場をとる。

このようにデータベースシステムの3つの機能部分が互いに補助し合う形にシステムを構成するためには、これらの機能部分を互いに密接に連携し合い、高度な相互作用を行えるような形にしておく必要がある。これを実現することにより、単にデータモデル獲得の自動化を指向したデータベースシステムであるという以上に、画像データベースシステムとしてさまざまな有用な視点が生まれてくる。その例をあげると、まずデータモデル獲得に関する点としては、もちろん自動獲得された不完全なデータモデルでも高度利用が可能である点があげられるが、これ以外にデータモデル利用からの働きかけによりさらに高度なデータモデル獲得を行い得る可能性が出てくる。データモデル利用から見ると、アプリケーション実行時の要求に従い、動的に新たな視点に立ったデータモデルを要求することができることになり、きめの細かいデータモデルを期待することができる。

このように、各機能部分間の高度な相互作用を可能にすることにより、データモデルの獲

得と利用の両面をより一層高度化できる局面があり得る。こうした視点は、画像データベースにおいては特に重要であると考えられる。本論文では、このような構成でより高度な次世代画像データベースを構成するという立場をとる。

2.3 背景技術

本節では、画像データベースシステムにおけるデータモデル獲得・管理・利用の3つの機能部分について、それらと関係の深い背景となる基盤技術について述べる。

まずデータモデル獲得については、画像データモデルの自動獲得の基盤技術はすなわち画像認識・理解技術であると考えられる。そこで現状の技術水準における画像認識・理解について、特に本論文でのアプローチに関係が深いと考えられるものについて述べる。データモデル利用については、画像データ(あるいはより広く考えるとマルチメディアデータ)を扱うアプリケーションシステムの例をあげ、その実現のためにはどのようなデータモデルが要求されるのかについて考察する。データモデル管理については、データベースにおけるデータモデルの既存の枠組の中で、画像データを中心とするマルチメディアデータを扱い得る、マルチメディアデータベースのためのデータモデルとして最も期待されているオブジェクト指向データモデルについて述べる。またこの他、より高度で知的な画像データベースシステムの構築のために必要と考えられる画像データベースシステムにおけるコンピュータ学習の利用について考察する。

2.3.1 画像認識・理解技術によるデータモデル獲得

初期の画像認識・理解のアプローチでは、扱う対象の画像を強く限定し、そこから取り出したい情報の種類もあらかじめ想定したものに限ってにおいて、その問題に完全に依存した形で認識システムを作成していた。このアプローチにより作成されたシステムは、限られた状況にしか対応できないが、その範囲内において非常にうまく働くことになる。こうした例は数多く存在する。しかしながら、その対象および目的があまりに限定され過ぎているた

め、これを画像データベースにおけるデータモデル獲得にそのまま用いるのは適当ではない。ここでは、画像データモデル獲得として利用可能なように、より広い範囲の対象を扱うことができ、より多目的性を持った画像認識・理解手法について考えることとする。

まず今日までの画像認識・理解の研究を、以下のような視点に立って分類する。

- 一様な知識表現を用いた手法。
- 対象の幾何的モデルを用いた手法。
- ユーザとの対話を用いた手法。
- 異なる抽象レベルの処理間での協調処理に基づく手法。

以下、それぞれの分類に基づいた画像認識・理解手法に関する考察を行う。

一様な知識表現を用いた手法

画像理解のための初期のアプローチの問題点の一つとして、画像処理の手法をそのまま手続的に組み合わせることによって目標の画像認識を行おうとした点が考えられる。こうして作成された画像理解システムは、その認識対象の細かい特性に深く依存したものになってしまう。これに対し、画像認識の枠組は対象によらない汎用性のある形で与えてしまい、対象に依存した部分はその枠組にのっとった体系的な知識表現で与えるという方法がある。このような方法を用いることにより、対象に依存せず多目的性を持った形での画像認識システムの構成と、体系的で大規模な画像データモデルの獲得の実現が期待できる。この例として、プロダクションシステムを用いたものと、フレーム型の記述に基づくものとをあげる。

プロダクションシステムは、IF-THEN の形式で表現された知識ベースを持ち、前向き推論を用いることにより問題解決を図る汎用の推論システムである [67]。プロダクションシステムでは、問題解決のための知識を知識ベースに蓄えることにより、後はシステムの推論エンジンがこれらの知識を適宜参照しながら問題解決を図ってくれる。この知識ベースに

蓄える知識は IF-THEN という形式のルールで一様に表現され、その意味付けも明確であるため、比較的容易に記述できるという特徴を持つ。

このプロダクションシステムを利用したデータモデル獲得の例としては、McKeown による航空写真の画像の認識システムがあげられる [30]。これは、空港の画像を認識するシステムで、その認識機構として汎用のプロダクションシステムを用いている。このため、実際に認識システムとして対象の空港の画像に依存した部分は知識ベースの部分のみであり、これも一様で記述しやすいルールの形式で与えることができる。実際の認識プロセスは、プロダクションシステムの推論エンジンに任せれば良いことになる。

このシステムでは、ルールとして認識ルールと検証ルールの2種類を与えるようになっており、基本的にはまず認識ルールで仮説を生成し、検証ルールではそれらの検証を行う。一般にプロダクションシステムは前向き推論をする推論エンジンを用いているため、仮説の検証を行う部分をこのように明示的に記述しなければならないのであるが、これがかなり複雑であるという問題点がある。

次に、フレーム [33] に基づいた知識表現を用いた方法について述べる。フレームによる知識表現は、汎化/特化関係 (is-a 関係) や部分関係 (is-part-of 関係) に基づく階層的な知識表現が可能であり、高い記述力を持つ。

この知識表現をもとにした画像認識システムとしては Mapsee2[19] があげられる。これは手書きの線画で表された地図を解釈するシステムであり、その解釈に必要な地図の知識の表現にフレームベースのスキーマと呼ぶ知識表現を用いている。道路・川・海岸・町などといった要素が、フレーム型の知識表現によりすっきりと階層的に表され、認識機構自体はある程度汎用性を残した形で構成されている。フレーム型の知識表現は、非常に高い記述力を持ち、より複雑な対象を扱う際の知識表現として有力である。知識表現という観点からはプロダクションシステムは不利であり、制御手法と対象の構造とが混然としてプロダクションルールに記述されることになるため、扱う対象の知識表現としての記述力が弱いといえる。

幾何的モデルを用いた手法

フレームによる知識表現は、認識対象としてのモデルを表現する方法というよりも、むしろ一般の知識表現の枠組である。ここではこのような知識表現の枠組を幾何学的な対象に特化して認識対象のモデルの表現のために用い、それに基づいた認識 (model-based vision) を行う方法について述べる。

まず対象を幾何学的なモデルで表現するため、対象そのものを基本的な要素に分解して表現する必要がある。この要素としては、何が記述されているか分からないような認識対象の画像からも抽出できるような、十分に一般性を持ったもの (シンボル) でなければならない [72]。Lowe は、画像からエッジを抽出し、その並行線分に着目して矩形をなす部分を抽出する例を示している [22]。これらのうち、記述として強力であり、一般性も高いものとしては一般化円錐が考えられる。これを用いた2次元画像からの3次元物体の認識のための指針については Marr が論じており [24]、またこの一般化円錐を用いたモデル記述に基づいた画像認識システムとしては ACRONYM [4] がある。

ACRONYM では、対象のモデルは一般化円錐の組合せで表現され、それらの大きさやそれらの間の関係などによる制約とで記述される。また、このモデルはフレーム型の記述のように階層化した形で与えられ、記述力が高い。認識システムとしては、まず対象の画像からモデルを構成する一般化円錐 (ただし画像上では2次元に投影された「リボン」となる) を抽出し、モデルに記述された制約を満たすかどうかをチェックして、モデルにより表される対象が画像中に存在するかを調べる。このモデルベースの方法により、認識システム自体は非常に単純な構成をとることができ、かなり複雑な構造を持つ対象でも認識できることが期待されるが、反面モデルの構造が複雑になってしまい、この対象を表すモデルをいかに作成するかが問題となる。

ユーザとの対話を用いた手法

画像情報の曖昧さや情報量の多さ、あるいは画像認識・理解技術の不完全さなどの理由から、画像データ中の複雑な状況においては、どうしてもコンピュータによる自動認識では対

応し切れない部分が出てくる。その一方で、人間の認知能力は極めて柔軟・強力であり、さまざまな対象に対するさまざまな視点に基づく認識にも容易に対応することができる。この点ではコンピュータによる画像認識・理解は遠く及ばない。今、対象の画像をどうしてもある一定の高いレベルにまで抽象化したいという要求があった場合、この人間の認知能力の助けを借りて対象をより高度に認識するという方法が考えられる。ここでは、基本的に画像データモデル獲得はコンピュータにより自動的に行うという立場を取るため、自動認識の方をベースにして、足りない部分を人間の認知能力により補うというアプローチについて考える。

大沢らは、地図図面の自動認識システムにおいて、ユーザとの対話を用いて認識結果をより高度化するシステムについて述べている [37, 64]。このシステムは、自動認識の結果完全な認識ができなかった部分をクローズアップしてグラフィック画面に表示し、ユーザの編集処理を促す。またその際、その部分の状況から考え得る認識結果の例を提案として提示してくるようになっており、ユーザの負担をある程度軽減するようになっている。

次に、人間と認識システムとの対話の別の形態の例として、人間から提示された情報から認識システムそのものを構成するというアプローチについて述べる。鳥脇らは、このアプローチを用い、X線画像から意図した画像処理システムを生成するシステム IMPRESS について述べている [17, 18, 63]。IMPRESS では、まず処理したい画像としていくつかのサンプルを用意し、人間がこれらの画像の認識結果として妥当と思われる例図形を作成する。システムはこのサンプルと例図形との組を受けとり、そのような処理を行う画像処理手順を生成してくれる。またこれを複数のサンプル画像について行い、その結果を集約することにより、同種類の画像一般に対し適用できる手順を生成するようになっている。

この他に、ユーザの例示から認識システムを生成するアプローチとしては、Winston の研究 [68, 69]、松原らの MIRACLE-IV システム [25, 26, 43]、Connel らのシステム [7] などがあげられる。これらのシステムは、いくつかの事例から一つの対象のモデルを生成するため、コンピュータ学習の手法を用いている。

こうしたユーザとの対話を用いた画像認識・理解へのアプローチでは、ユーザとシステムとの対話の「界面」が重要なポイントとなろう。人間であるユーザからの情報は極めて高

品質だが、人間に問題を提示し、人間の判断を待ち、その結果を入力させなくてはならないという手間を考えると、この情報は極めて「高価」である。ありとあらゆる部分について人間の指示を仰いでいたのでは、全く処理効率が上がらない。大沢らのシステムでは、界面を処理結果の直前の認識処理が行えなかった部分に限定し、さらにシステムの提案作成機能により、この界面での情報のやりとりができるだけ少なくなるようにしている。IMPRESSでは、界面を処理結果そのものの位置に設定することにより、人間の負担を軽くしている。また、人間から得られた情報について再び問い合わせってしまうことを避けるためには、コンピュータ学習の手法を利用するという方法が考えられる。

異なる抽象レベルの処理間での協調処理に基づく手法

一般的な画像認識・理解処理の流れは、まず与えられた画像データに対して何らかの処理を行ってある程度抽象度の上がった記述を作り、次いでその記述に対して処理を行ってさらに少し抽象度の上がった記述を作り、この記述に対してまた処理を…、というように、基本的にボトムアップに処理が進んでいく。良く見られる手順としては、まずもとの画像データに対してセグメンテーション処理を行い、画像データをセグメントに分割する。次いでこの処理結果を再統合するなどの処理により、シンボルを表す記述を作成する。そしてこのシンボルの記述を用い、実際に意味のある解釈を抽出する処理を行う。セグメンテーションなどの初期の処理は画像に記述されている意味内容には全く関係なく、画像全面に対して一様な処理を行うものが一般的である。しかしながら、この時点での処理の誤りは、最後の意味のある解釈の生成に甚大な影響を及ぼすことになる。そもそも人間の視覚について考えると、見ている画像に対し全く一様な処理を行っているとは言い難く、はじめは大雑把に画像をながめ、その後特徴的な部分だけを注意深く見る、というように、低レベルの視覚処理と高次の認知処理との間で密接に関連した処理を行っていると考えられる。コンピュータによる画像認識・理解でも、このような抽象度の低次のレベルと高次のレベルとの間での協調処理が重要であろうと考えられる。ここでは、この考え方と関係が深いと思われるいくつかのアプローチを紹介する。

まず、意味解釈のレベルとの協調によりセグメンテーションを高度に行う手法 IGS

(Interpretation-Guided Segmentation) について述べる [66]. IGS では、まず対象画像を通常の方法でセグメンテーションする。次いでこの結果に対して低レベルの解釈処理を行い、セグメンテーションが意味解釈レベルで設定されている制約条件を満たしているかどうかチェックされる。満たしていない場合は、その部分のセグメンテーションをやり直すように動作し、この制約条件を満たすセグメンテーションが生成されるまでこれを繰り返すことになる。制約条件としては、例えば風景の画像に対してであれば、「空」のセグメントは「地面」のセグメントの上に存在し、「木」のセグメントは「空」のセグメントより下で「地面」のセグメントよりは上にある、というような形で与えられる。この結果、従来のセグメンテーション処理よりも誤りが少なく、より高度な結果が得られたことが示されている。

次に、3つの異なる抽象レベルで動作するエキスパートシステム間の協調動作により、航空画像を認識処理するシステム SIGMA について述べる [27-29]. SIGMA は、「家」などといった、セマンティックスを持ったレベルでの処理を行う GRE(Geometric Reasoning Expert) と、「長方形」などのシンボルのレベルでの処理を行う MSE(Model Selection Expert) と、実際の画像データのレベルでの処理を行う LLVE(Low Level Vision Expert) という、異なる抽象レベルで動作する3つのエキスパートシステムの協調動作により、与えられた航空画像から家や道路などを認識するシステムである。まず GRE から MSE へは、「家を探せ」というような形で要求が渡される。MSE は「家」とは通常長方形に見えるといった知識を持っており、さらに LLVE に対して「長方形を探せ」という要求がなされる。LLVE では、与えられた画像から長方形を抽出する画像処理手順を自動生成し、この処理手順による処理結果を MSE に返す。MSE ではこの結果を吟味し、「家」として妥当であればこれを GRE に返す。このようにして、抽象レベルの高い部分と画像データを扱う抽象レベルの低い部分との協調により、概念的な要求に対して各画像に応じた画像処理が自動的に行われることになる。

次に、ある物体のシルエット画像を複数受けとり、これからその物体の内部構造を学習し、同時にその画像に対する画像処理をも学習するシステム MIRACLE-IV について述べる [25, 26, 43]. MIRACLE-IV は、物体の内部構造を推定するモデル獲得部と、画像処理戦略を立てて適当な視覚特徴を抽出する処理戦略部とからなる。モデル獲得部は高次のシンボル

のレベルで動作し、シルエット画像が提示されるに従い順次対象物体のモデルを更新していく。その際、その時点でのモデルに従い、処理戦略部に対し現在推定しているモデルの特徴を表す視覚特徴の抽出を要求する。処理戦略部では、試行錯誤によりこうした視覚特徴を抽出しようとし、成功例を蓄えることによりこの特徴の抽出戦略を学習して行く。このようにして、モデル獲得部というシンボルのレベルで動作する部分と処理戦略部という画像データのレベルで動作する部分との協調動作により、対象の物体の内部構造とその視覚特徴の抽出のための画像処理戦略とを同時に学習して行くわけである。処理戦略部について考えると、はじめはただ試行錯誤的に画像処理を行うが、モデル獲得部にモデルができてくるに従い、このモデルに沿った画像処理の選択ができるようになり、段々とより信頼性の高い視覚特徴が得られるようになってくる。

従来の画像処理は、セマンティックスからの情報が全くない状態で、あたかも目隠しをしたような状態で手探りの処理を行っていた。それに対しここであげたアプローチでは、セマンティックスレベルから提供された情報の利用により、より高度で信頼性の高い画像処理を行うことができる。画像認識・理解のより一層の高度化のためには、この異なるレベル間での協調処理が重要となつてこよう。

2.3.2 マルチメディアアプリケーションシステムによるデータモデル利用

マルチメディアデータを扱うアプリケーションシステムとしては、今日でもさまざまなものが考えられ、今後ますます多様なものが作成されていくと考えられる。ここでは、そうしたものの中でも典型的と思われるアプリケーションとして、画像データベース検索システムとハイパーテキスト／ハイパーメディアシステムを考える。まずこれらのシステムの概要について述べ、その上でシステムの実現のために必要なデータモデルの形態について考察する。

画像データベース検索システム

まず画像データベース検索システムについて述べる。このようなシステムにもいろいろなものが考えられるが[40]、ここでは大規模に蓄えられた画像データベースに対し、キーワードを与えることによって、その記述されている内容にまで立ち入って該当する画像データを検索するようなシステムを考える。ここではこのキーワードとしても、特に簡単のため、画像に記述されている実体の名前や、その画像を特徴付ける形容詞などの文字情報を用いるものとする。このような単純な構成のものでも、画像のさまざまな側面に基づく非常に多数のキーワードが利用できるならば、この画像へのさまざまな見方に対応した柔軟な検索のできるデータベースが構築できることが期待できる。

このようなデータベースの検索システムのためのデータモデルの構築方法であるが、まず一番直接的な方法としてはあらかじめデータモデル自体にありとあらゆる場面を想定したキーワードを付加しておくというものがある。この方法では、もちろんキーワードとしてははじめに付加したキーワードしか使えないわけであるが、実戦的な方法であり、実用例も多く存在する。

これを一歩進めた方法として、データモデルとしてはある抽象的なデータモデリング技術を用いた概念的なデータモデルとしておき、検索時には与えられたキーワードについて動的にこのデータモデルを評価し、その結果により検索のための画像データのふるい分けを行うというものがある。こちらの方がよりさまざまなキーワードに対して柔軟に対応できることになる。

このような検索が可能な画像データベースシステムの構成では、想定されるキーワードをすべて付与しておく場合にしても、ある概念的な記述を作成しておいて必要に応じてキーワードの評価が可能なおきしておく場合にしても、データモデル作成段階においては画像データに記述されている意味内容がある程度理解できている必要があるのは明らかである。

こうしたデータモデルの獲得について考えると、人間は画像を見てその意味内容を理解することができるので、実際的なアプローチとしてはこのようなデータモデルをすべて人手を使って作成してしまうというものがある。利用頻度の非常に高い画像データベースの

場合は、このようにコストをかけても高度な利用を可能とするためにこのようなデータモデル作成方法を取ることもあり得る。ここではあくまでデータモデルはコンピュータにより自動獲得するものという立場を取るが、望まれているようなさまざまな視点に立ったキーワードによる評価が可能なデータモデルの自動獲得は、非常に難しいといえる。

ハイパーテキスト／ハイパーメディア

次のマルチメディアアプリケーションシステムの例として、ハイパーテキスト／ハイパーメディア [6, 65] について述べる。

一般のテキストデータは、一次元的な文字情報の連なりであるという意味で線状のデータである。コンピュータシステム上のファイルとしてのテキストデータも同様に線状のまま管理されていた。従来の大部分の利用目的ではこのような単純な管理方法で十分であった。しかし、今後より多くの応用が生まれてくるにつれてこのような単純な扱いでは済まなくなり、より複雑で高次の構造を表すような方法が必要となってくる。

その一方で、近年コンピュータシステムはより高性能なものが安価に手に入るようになってきた。これにより、テキストデータに対する複雑で高次の構造による扱いが可能となり、しかもそれを手軽に参照することが可能となってきた。このような高次構造のテキストデータは、テキスト中のある部分から参照可能な形で、別の意味的につながる箇所に「リンク」を作成することにより実現できる。このようにして、非線形的な探索が可能な形に作成されているテキストがハイパーテキストである。

実際の典型的なハイパーテキストの実装では、もとのテキストデータを意味的なまとまりを持つ小さい部分に分割し、これをハイパーテキストのオブジェクトとする。そして、これらのオブジェクトの間に参照可能なようにラベル付けされた有向リンクを設けて、非線形的な探索(ブラウジング)ができるようにしている。このオブジェクトとして、テキストデータのみではなく、音声や画像データのようなマルチメディアデータも許すようになっているものをハイパーメディアと呼んでいる。

ハイパーメディアにおける画像データの扱いについて考える。ハイパーテキストの延長としてのハイパーメディアでは、画像データはあくまで単一のオブジェクトでしかなく、そ

の内部に働きかけて画像に記述された実体そのものを扱うようなことはできなかった。この場合、ある画像データについては、他のオブジェクトからのラベル付きのリンクのみが唯一の参照方法ということになり、画像データの扱いとしては先のキーワードによる画像検索とほとんど変わらないといえる。ハイパーテキストにおけるテキストデータによるオブジェクトでは、各オブジェクトはほとんど自明の単純な意味合いしか持たないため、このような扱いで十分であったが、画像データはテキストデータに比べ各画像の内部にさらに複雑な内部構造を持つ実体を含んでおり、このような画像データに対して前述のような扱いしかできないのではその操作方法としては極めて不十分である。ハイパーメディアにおいては、各画像データを構成する実体を参照するといった、画像データの内部構造にまで及ぶ操作が行えてしかるべきであると考ええる。

このような画像データの内部にまで立ち入った操作が行えるハイパーメディアとしては、例えば画像中のあるものを指定すると、それに関する情報が検索され、出力されてくるといえるものが考えられる。こうした操作が可能なハイパーメディア実現のためには、その内部に蓄えられる画像データとしては、各画像の特徴を表すキーワードなどが利用できることはもちろんであるが、画像の内部の実体やそれらの構造、およびそれらの特徴などの、画像内部にまで及ぶより細かい操作が可能な形態である必要がある。画像データベースを用いてハイパーメディアを構築したり、画像データベース検索のためのツールとしてハイパーメディアを利用したりする際には、画像データモデルはこのような形で構築しなければならない。

2.3.3 オブジェクト指向データモデルによるデータモデル管理

従来の文字・数値型のデータベースでは、表形式のデータを効率良く扱うことに主眼が置かれていた。こうしたデータを扱うためのデータモデルとして完成されたのが関係データモデル、すなわち関係データベースである。

一方で、画像データをはじめとするマルチメディアデータなどのように、表現したいデータに入れ子構造などのような非常に複雑な内部構造が存在し、表形式のように単純なデータモデルでは表現し得ないようなデータをデータベースシステムで扱いたいという要求が

出てきた。これらに対応するための新たなデータモデルとしては、関数型データモデル [61]、意味データモデル [1]、そしてオブジェクト指向データモデル [2, 8] 等があげられる。

これらの中でも、オブジェクト指向データモデルはマルチメディアデータベース向きの多くの特徴を持ったデータモデルであると考えられる [73]。オブジェクト指向データモデルは数多くの特徴を持つが、ここではその主な特徴のうち特にここで述べているデータモデル管理と関係の深いものを以下にあげる。

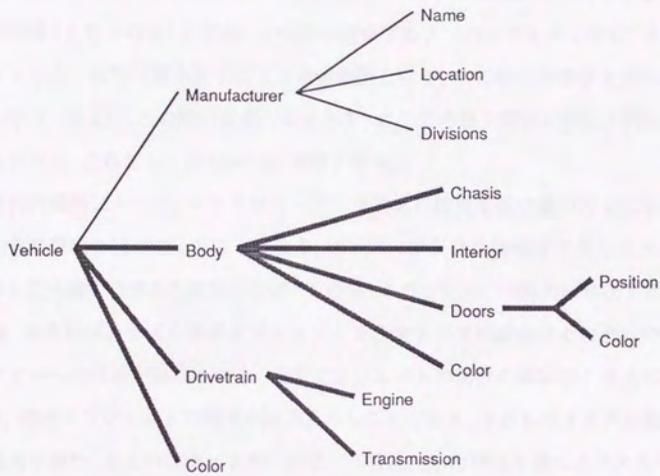
- 複合オブジェクトによるデータの記述性。
- 型階層と型継承。
- データの実体とその操作との一体化による抽象データ型。

画像データベースにおけるデータモデル管理という位置付けでは、これらをデータモデルの表現に関するものとデータモデルの操作に関するものとに分類して論ずるのが適当と思われる。本節ではこの分類に従い、オブジェクト指向データモデルの特徴について、特にデータモデルの表現とその操作に着目して、その概要を述べる。

データモデルの表現

先にあげたオブジェクト指向データモデルの特徴のうち、はじめの2点がデータモデルによる素材データの表現と特に関係が深いと考えられる。本節では、オブジェクト指向データモデルを用いた場合のデータモデルの表現に関し、その表現力や記述の際の効率などについて考察する。

まず、複合オブジェクトによるデータの記述性について述べる。オブジェクト指向データモデルでは、データモデルで表現したい対象データをより細かい独立した部分に分解し、それらの複合体として表現することができる。この複合体としてのオブジェクトを複合オブジェクトと呼ぶ。複合オブジェクトを構成する各部分も同様に複合オブジェクトで表現することができるので、入れ子構造を持つような複雑な対象でも表現できることになる。複合オブジェクトは図 2.2 に示すような親子関係で表現されるが、この関係は部分と全体の関係を表すものであり、is-part-of 関係と呼ばれる。



Thick lines: is-part-of relation

Thin lines: reference

図 2.2: 複合オブジェクトによる階層構造の例

(文献 [2] より)

次に型階層と型継承について述べる。オブジェクト指向データモデルでの表現の単位はオブジェクトであるが、これらの各オブジェクトはある特定の型(クラス)に属するようになっている。クラスにより、オブジェクトの内部変数や複合オブジェクトによるオブジェクトの構造などが規定される。同じクラスに属するオブジェクトは同じ内部変数を持ち、同じ構造を持つことになる。これらのクラスは複合オブジェクトと似た親子関係をなしていて、これを型階層(クラス階層)と呼ぶ。この時の親を子のスーパークラスと呼び、子は親のサブクラスとなる。複数の親を許すことを多重継承と呼び、この時の型階層を特にクラスラティスと呼ぶ。図 2.3 にこの階層構造の例を示す。ここでの親子関係は特化/汎化の関係を表すことになり、これを is-a (a-kind-of) 関係と呼ぶ。

この特化の関係において、クラスはスーパークラスの情報を受け継いだ上にこれにいくつかの独自の情報を付け加えることにより、スーパークラスの特化したクラスとなる。この情報を受け継ぐ機構を型継承と呼ぶ。この時、そのクラスに内部的に存在する変数やその初期値、あるいはクラスが複合オブジェクトを表すクラスの場合はその部分のクラスの種類などといった情報が継承される。複合オブジェクトの場合の部分のクラスの継承により、複合オブジェクトの構造が継承されることになり、子どものクラスは親のクラスと同じ構造を持つ、あるいはそれを少し変更して特殊化した構造を持つクラスとなる。

この型階層により、似たような構造を持つクラスをいちいちゼロから作成することなくある既存のクラスのサブクラスとして作成することによりより効率の良いクラスの作成ができる。

データモデルの表現力という観点から言うと、まず複合オブジェクトによりかなり複雑な構造を持つような対象でも表現することができ、ほとんどありとあらゆる対象をこのオブジェクト指向データモデルにより記述できると考えられる。対象の構造のうちでも極めて詳細な部分まで含むようなデータモデルでもオブジェクト指向データモデルにより表現できるが、このようなデータモデルを形成するのはかなり大変な作業を伴うことになる。データモデルの記述力については、このデータモデルの構造のテンプレートであるクラスの作成の際、型階層の利用によりクラスの記述そのものを効率良く行うことができ、データモデルを効率良く記述することができるといえる。

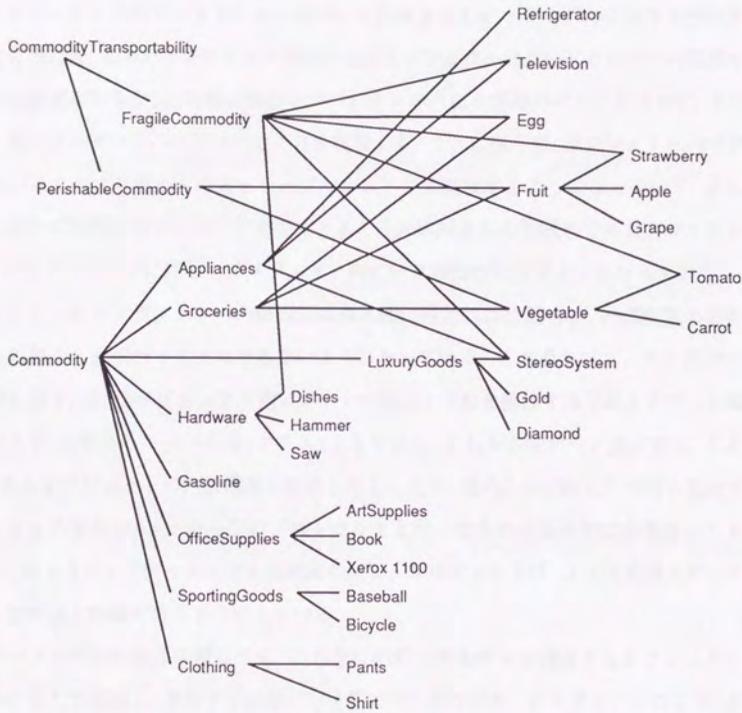


図 2.3: クラス階層の例

(文献 [62] より)

データモデルの操作

データモデルの操作としては、先にあげた特徴のうちの最後の点との関係が深い。ここではこのオブジェクト指向データモデルの特徴の一つである抽象データ型について述べる。

オブジェクト指向データモデルに基づいて作成されたデータモデルに対する操作を実行するに当たり、このデータモデルを構成する各オブジェクトは互いにそれぞれの情報を参照し合う必要があるが、この時に他のオブジェクトの内部を直接のぞくことはせず、オブジェクト間のメッセージパッシングでこれを実現する。これに限らず、オブジェクト内のデータの扱いはすべて外部からのメッセージパッシングで実現するようになっていて、あるオブジェクトは外部からメッセージを受けるとこれに対応した手続きであるメソッドを起動し、このメッセージに従ったオブジェクト内データの操作を行うようになっている。

このようにオブジェクトの内部構造は基本的に外部には公開せず、内部変数の参照や操作も外部から直接行うことはできないようになっている。これをオブジェクト間での情報隠蔽と言う。また、オブジェクト内にデータの実体とそれを操作する手続きすべてを取り込んだ上で、全体として一つの型(クラス)としている。これを抽象データ型と言う。これにより、あるオブジェクトの内部構造を変更したとしても、他のこれに対して作用を及ぼすオブジェクトの変更は、メッセージのプロトコルさえ同一であれば基本的に必要ないことになる。このようにしてデータモデル作成時のモジュラリティを上げ、より大規模なデータモデルを効率良く作成できるようにしている。

データモデルの操作に対しても、これをこのデータモデルを構成するオブジェクトのメソッドとして実現し、そのまま抽象データ型の中に取り込む。こうすることにより、多様なデータモデルへの多様な操作の実現に対して、対象のデータモデルに依存した部分はデータモデル内部でメソッドとして処理してしまい、外部から見るとメッセージという統一的な処理手順で実現されることになる。

2.3.4 画像データベースシステムにおけるコンピュータ学習の利用

画像データベースシステムのさまざまな局面においてコンピュータ学習の利用が考えられる。例えば、データモデル獲得において画像理解システムの自動獲得に利用する場合や、データベースシステムのユーザとのセッションにおいてより円滑な処理のために利用する場合などが考えられる。以下に特に重要と考えられる画像データベースシステムにおけるコンピュータ学習利用の局面をあげる。

1. 人間機械協調型の画像データモデル獲得システムにおける利用。
2. 画像認識のためのモデルの自動獲得における利用。

ここでは、まずコンピュータ学習の手法の分類について述べた上で、これらの各局面について述べる。

コンピュータ学習の手法の分類

通常のコンピュータプログラムは、はじめにプログラムとして与えられた手順に従いその範囲内であらかじめ決まった仕事のみを実行することができる。コンピュータ学習とは、この手順自体を何らかの方法で自ら変更することにより、その振舞いの一層の高度化を実現するための手法である。

まず文献 [5] に従い、学習者の推論の量に応じたコンピュータ学習の分類を行う。

暗記による学習 学習するべきものをそのままの形で学習者に与える。学習者はこれをただ内部に取り込むだけという形態の学習手法である。

指示による学習 教師と呼ばれる外部システムからの指示に従い学習を行う。教師は、学習プロセスがうまく働くように学習者に適切な指示を与える必要がある。学習者の負担は軽い。

類推による学習 新しい状況に対し、すでに持っている知識からの類推によって得られる知識により対応するという学習手法である。

例からの学習 ある概念について、学習者にはこれを表現している事例である正の例とその反例である負の例が与えられ、これからその概念の記述を帰納的に推論するという学習手法である。これは帰納学習の特殊な場合であると考えられる。

観察および発見からの学習 これは教師の不要な学習手法であり、学習者は外部環境との相互作用を観察することにより一連の概念を学習する。これは帰納学習の一般的な場合であると考えられる。

コンピュータ学習を利用しようとしている局面や得られる情報の性質などから、実際にはこれらのうちから適当な手法を用いることになる。

学習する知識のタイプにもさまざまなものがあり得る。一般的には、獲得する知識としてはパーセプトロンにおけるパラメータやプロダクションシステムにおけるルールなどといった定型的な扱い易いものが望ましい。

人間機械協調型の画像データモデル獲得における利用

画像認識・理解技術を用いた画像データモデル自動獲得では、さまざまな対象に対応できる柔軟なデータモデル獲得はいまだに実現できない。認識対象の局面を限り、例えば一定の記述ルールの範囲内で書かれた地図図面や回路図面などの認識システムでは9割方自動獲得が可能となってきているが、それでも残りの数パーセントはまだ自動獲得できないで残ってしまう。この部分まで完全に獲得されたデータモデルが要求されている場合、人間機械協調型のアプローチがあり得る。

人間の認知能力は極めて強力であり、あらゆる局面について柔軟に解釈を行うことができる。その反面、人間に問題となっている箇所を提示してその指示を仰ぎ人間の判断の結果を入力させるという一連の処理は、人間にとってもシステムとしても負担が大きく、極めて「高価な」処理であるといえる。また、人間はコンピュータと違い疲労する。特に似たような局面について何回も人間に判断を仰いでしまうような場合、この疲労の進む度合は激しくなる。

大沢らのシステム [37, 64] や IMPRESS [17, 18, 63] では、人間とシステムの接する界面を

うまく選び、この間でやりとりする情報量が少なくなるようにすることによりこの問題に対処しようとしている。ここでコンピュータ学習の利用により、この人間とシステムとのセッションから得られた知識によってシステム自体を適応させていくという手法を用いれば、人間機械協調がより高度化できることが期待できる。

この場合の学習方法としては、情報が学習者向きには与えられないため、例からの学習が観察と発見からの学習のような帰納学習の方法を用いることになるろう。

画像認識のためのモデルの自動獲得における利用

先にモデルベースの画像理解手法は強力ではあるが、そのモデルの作成にコストがかかってしまうという問題点があると述べた。このモデルの作成をコンピュータ学習の利用により自動的に行うというアプローチがある。

まずは Winston の ANALOGY システム [68] があげられる。これは対象の世界を積木の世界に限定し、いくつかの事例を受けとることによって積木によって構成される「アーチ」などの概念を獲得するシステムである。システムに与える事例としては画像イメージで与えることを考えているが、対象が積木の世界に限られているためその概念的な記述がある程度容易に得られる。学習システム本体に与えられる事例としては、こうして得られた概念的な記述を表す意味ネットワークの形で記述される。

ANALOGY のような事例からの学習を行うシステムでは、学習したい概念の典型的な正あるいは負の事例ばかりを与えてしまった場合、意図したような正しい概念の記述が得られない。対象の概念を表しているかどうか非常に微妙であるが、確かに正の（あるいは負の）事例であるといういわゆるニアミスを与えることが重要である。このニアミスの発見が困難である場合があり、事例からの学習の一つの問題点でもある。MARVIN システム [44] では、この点の解決方法として、学習プロセスに有効であるニアミスを仮説として生成し、この真偽をユーザに問い合わせるといったアプローチをとっている。このシステムも ANALOGY と同様積木の世界に限って概念の学習を行うシステムであるが、上記のアプローチにより効率の良いモデル獲得が実現できる。これは人間機械協調の視点からも有効なアプローチである。

Connellらのシステム[7]は、ANALOGYシステムと同様事例からモデルを獲得するシステムであるが、対象としてはより一般的なものを考えており、ハンマーの画像や飛行機の画像からそのモデルを抽出することができる。基本的な動作はANALOGYシステムとほぼ同様であり、与えられた画像からかなり複雑な意味ネットワークを生成し、これからモデルを抽出するように動作する。

MIRACLE-IVシステム[25, 26, 43]は、画像処理部とモデル獲得部から構成されており、これらの相互作用によりモデル獲得を行うようになっている。先の2つのシステムでは学習システムは外部から与えられた意味ネットワークなどによる記述をそのまま受けとり、これを用いてモデルを抽出しようとするものであった。従ってモデルの抽出に必要な特徴などはこの意味ネットワークにあらかじめ含まれていることが必要であった。MIRACLE-IVでは、画像処理部はモデル獲得部からの指示に従いモデル獲得に必要な画像特徴を必要に応じて抽出しようとし、その結果を学習していく。このようにして、生の画像からのモデル獲得を実現しているといえる。

2.4 画像データベースシステムにおける機能部分への要件

2.4.1 画像データモデル獲得の要件

実用的で高度な応用が可能な次世代画像データベースの構築のためには、そのデータモデル獲得として高度に概念化されたデータモデルを、大規模に形成し得る手法が望まれる。それも画像認識・理解技術などを利用して、自動化された形で実現する必要がある。

この画像データモデル獲得に対する要件としては以下のような点が考えられる。

1. 大規模・体系的なデータモデル獲得が可能である。
2. 現状の画像認識・理解の技術水準に従い、可能な限り高い抽象化が可能である。また、今後の技術の進展にも対応できる。
3. 不完全にしか認識できなかったような部分でも、ある程度の利用には耐える形で提供

できる。

4. 利用部からの動的な要求にも応え得る。

ここでは、これらの要件の画像認識・理解の枠組の中での実現方法について考察する。

大規模・体系的な画像データモデル獲得

大規模・体系的なデータモデル獲得のための方策について考える。

画像データモデル獲得では、生成するデータモデルがデータベース環境で使用されることを考えると、大規模で、その種類も多様な画像データに対し、多彩な視点に立った画像データモデルが生成できることが望まれる。これを実現するためには、データモデル獲得の方法としては対象に依存したアドホックな方法ではなく、多目的性のあり拡張性にも富んでいる見通しの良い体系立った方法をとる必要がある。

画像認識・理解システムを手続き的に記述してしまう方法では、作成したシステムはその対象の画像の種類に特化したものになってしまい、これを再利用して他の種類の画像にも適用できるようにするのは大変困難であった。これに対し、知識表現を用いることにより、認識・理解の機構と対象に依存した知識部分とを分離した画像認識・理解手法では、基本的には処理の対象の種類が変わっても認識・理解機構は共通に用いることができ、知識表現の部分のみの変更で対応することができる。この知識表現の部分も、意味付けが明確な見通しの良い形で作成可能であったり、階層的に作成されていてその再利用が容易であったりするならば、種類の多様な画像データに対しても容易に対応できることが期待される。

このような特徴を持つアプローチとしては、先に述べたプロダクションシステムを用いたものや、フレームに基づく知識表現を用いているもの、あるいはモデルベースのものなどがあげられる。

より高度な画像認識・理解手法への対応

より高度な画像認識・理解手法については、先に述べたようなアプローチをはじめとして、さまざまな方法論があり得る。こうした中でも特に異なる抽象度間での協調に基づく方

法は、期待が持てる手法であると考え、また、後に述べる「利用面からの要求」により状況や目的を限定してより高度な認識を行うという方法も、実戦的という点で評価できよう。こうした高度な画像認識・理解手法に画像データモデル獲得を対応させる問題、また今後のこれらの技術の進展にも対応して行くための方策について考える。

画像データモデル獲得をこのようにさまざまな画像認識・理解手法に対応させるための一つの方策として、画像獲得のための枠組としてはさまざまな画像認識・理解手法を実現できるようなシステム作成の枠組を与えるようにし、今後の技術の進展についてもこの枠組の範囲内でシステムを作成することにより対応するという方法が考えられる。プロダクションシステムを用いた画像データモデル獲得はこれに近い方法である。これは汎用の問題解決のための枠組なので、広範囲の画像認識・理解手法を記述することができる。このような枠組を用意することにより、現状の高度な画像認識・理解技術にも対応することができ、今後の技術の進展にも対応するような画像データモデル獲得が実現できる。

このような画像認識・理解のための専用の枠組を利用するという方法を用いる場合、特に重要なのはこの枠組が他の画像データモデル獲得のための要件を満足するものであるということである。このように設定することにより、その枠組に沿って画像認識・理解手法を記述することで、特に意識することなく他の要件を満たした画像データモデル獲得の実現が期待される。

不完全な認識結果の利用

不完全な認識結果をある程度の利用は可能な形で提供するための方策について考察する。

一般に画像データは極めて曖昧なデータであり、状況や利用目的などが特定されない場合、そのデータの解釈を一意に決定するのは難しい。しかしながらこのような場合でも、あるデータの解釈を最終的な形に決定することができないとしても、ある程度曖昧さを含んだ情報ならば得ることができる。利用のされ方によっては、このような曖昧さを含んだ情報でも十分利用可能な場合が往々にしてある。例えば、机の上のある位置にある角度でコーヒーカップが置いてあって、その他に雑然と重なり合った資料が置かれている状況で、この資料の置かれ方があまりに雑然としていて、どの部分がどの資料と対応しているのかが決

定できないような場合でも、コーヒーカップを手取ることはできるし、資料の大体の分量を推測することもおそらく可能である。

画像データベースシステムにおいても、このような状況は十分に考えられるし、おそらく大部分の画像データからは曖昧さを含んだ解釈しか取り出せないと考えられる。画像データベースシステムでこのような曖昧さを含んだ解釈を扱うためには、データモデル獲得で曖昧さを含んだ状態のデータモデルを作成できなければならない。

曖昧さを含む解釈結果を扱う方法はいくつか考えられるが、ここでは ATMS を用いた方法と、内部状態を持つ計算モデルに基づく方法とをあげる。

ATMS を用いた方法 ATMS[9-11] は非単調推論に基づく一種の探索システムである。ATMS を用いた推論システムは ATMS 本体と問題解決システムからなる。問題解決システムは、ある時点で得られている事実と、あらかじめ用意されている問題領域の知識とから、与えられた問題の解決を図るために一連の推論を行う。その際、与えられている事実からは判断できないような曖昧な状況については、その都度適当な仮説 (assumption) を立てて行く。ATMS では、問題解決システムで推論を行うたびに仮定した事実の間の依存関係を受け取り、これを束 (lattice) の形に保存する。そしてこの束を見て、現在までの状態で矛盾のない仮定を選んで問題解決システムにこの仮定を渡し、その事実関係を調べさせる。このようにこれまでの推論結果から矛盾しない仮定を順次生成することで、最終的には問題解決システムが満足するような仮説を決定する。こうして効率良く解空間を探索する推論システムが構成できることになる。

ATMS での曖昧なデータの扱いは、仮説という形で行う。この仮説は、それまでに得られている依存関係に矛盾しない限り、すべての可能な仮説を同時に保存する。例えば、あるデータの解釈が A であっても最終的に全体として矛盾のない解釈が得られ、またその部分の解釈が B であってもやはり全体としては矛盾のない、また別の解釈が得られる場合、このデータには曖昧さがあることになり、ATMS ではこれらを同時に保存する。このため、曖昧なデータの「曖昧さ」まで保存した形で扱っているといえる。

ATMS あるいは TMS[13] を用いた画像理解システムとしては、Kanade らの都市部の航

空画像の解釈システム [20] や、柴藤らの一般化円錐に基づいた3次元シーン解釈システム [14, 15] などがある。TMS では ATMS と異なり、複数の仮説を同時に持つことはできず、常に一つの仮説のみしか扱えないが、Kanade らのシステムではまず画像からエッジの線分を取りだし、これらの線分が主に直方体の建物のどの部分に相当し得るかについて矛盾のない解釈を生成するために TMS を用いている。柴藤らのシステムでは、一般化円錐を基にした仮説の管理に ATMS を用いている。

これらの例でもそうであるが、TMS や ATMS で扱えるのは基本的にはシンボルデータであり、そのデータの曖昧さについても、あるシンボルがいくつかの解釈候補のうちのどれになるかが決定できないという、いわば離散的な曖昧さしか扱うことができない。実際の画像データにおける曖昧さとしては、ある部分の解釈がどうなるのか皆目見当がつかないような場合や、ある部分のシンボル化そのものが曖昧さのためにできない場合などのように、かなり質の違う曖昧さがありうる。こうした曖昧さをいかに扱うかについて考慮する必要がある。また、TMS や ATMS で扱う曖昧さはあくまで処理の途中のみに存在し、これを曖昧な認識結果として利用することを考慮していない。実際にこれを画像データモデル獲得として積極的に利用しようという場合には、この点も考察する必要がある。

内部状態を持つ計算モデルに基づく方法 従来の画像認識・理解の枠組では、画像中のある部分の解釈が、その曖昧さのため一意に決定できない場合、認識結果としてはその部分は「認識できなかった」という情報しか得られなかった。一般の画像認識システムを内部状態を持つ計算モデルでとらえることによりこれを図式的に説明すると以下ようになる(図 2.4)。初期状態において画像データが入力された後、この画像データに対してさまざまな処理を行い、画像の各部分について状況判断を行いながら、状態遷移を伴って少しずつ画像の解釈を作成して行くというように画像認識過程を実現する。ここでいう状態とは、解釈の途中の対象画像の記述や、画像認識システムの内部変数などをすべて含む総体としての状態を指している。もし与えられた画像データが完全に解釈できたならば、システムは結果を出力する状態に遷移し、解釈した結果をすべて出力して最終状態に到る。しかしながら通常はこのようになることは稀で、大抵は画像データの曖昧さや認識システム自体の不備か

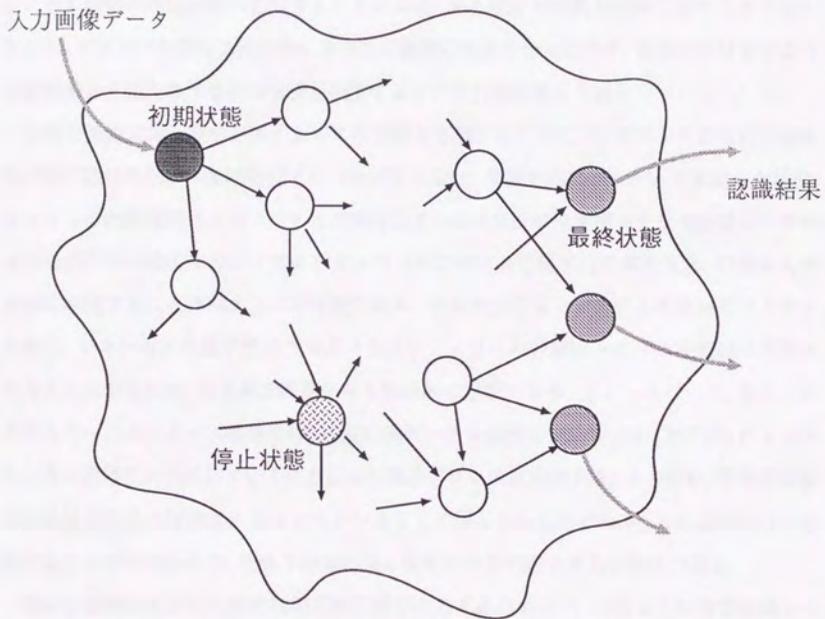


図 2.4: 内部状態を持つ計算モデルによる画像認識システムの表現

ら、状況判断ができないまま最終状態に到る途中状態で停止してしまうことになる。このとき、システムからは結局何の出力も得られなかったため、「認識できなかった」という結果しか得られないことになるわけである。

この計算モデルでは、対象の曖昧な認識結果は途中で停止してしまっているシステムの内部状態に反映される。不完全な認識結果を積極的に利用しようという場合、この画像認識システム自体の内部状態を参照することにより、ある部分の解釈が実際にはどこまで進んでおり、どのような曖昧さのためにその先の解釈に進まなかったのか、最終的にはどのような解釈がありえたのかなどの情報を抽出するという方法が考えられる。

実際の画像認識・理解システムでこの方法を実現するためには、まずシステムの内部状態が明示的に外部から参照可能でなければならない。手続き的にシステムを記述した場合、システムの内部状態はそのシステムが動作しているタスクの作業用メモリや処理シーケンス内の実行中の場所を表すプログラムカウンタなどにより暗示的に表現され、外部から明示的に参照することはほとんど不可能である。プロダクションシステムを用いたシステムの場合、システムの内部状態はプロダクションシステムの作業用メモリに定型的に表現されることになるため、ある程度外部からも明示的に参照できる。フレームベース、あるいはモデルベースのシステムの場合は、一般に実データを参照してフレーム(モデル)のインスタンスを段階的に作成していくことにより認識プロセスを実現する。そのため、不完全な認識結果は不完全に構成されたインスタンスとして得られることになり、これは明示的に参照することが可能なので、曖昧な認識結果を曖昧なまま利用できる可能性がある。

曖昧な認識結果をある程度利用可能な形で出力するためには、このように画像認識システムの内部状態を明示的に参照できることは必須であるが、さらのこの内部状態の意味付けが明確でなければならない。つまり、最終的にはどのような解釈を表す状態に到り得るのか、どのような解釈はあり得ないのかということが内部状態により明確に表されている必要がある。一般にはこのような情報は明示されてはおらず、システムの内部状態間の依存関係から推論することになる。このために、この内部状態間の関係が明確に表現されている必要がある。プロダクションシステムの場合、内部状態間の関係は知識ベースのプロダクションルールの解析により比較的容易に得られると考えられる。フレームに基づく手法の場合、

不完全な認識結果は不完全に生成されたインスタンスを表すフレームで表現されるが、これは複数の最終的解釈候補を表すことは考慮されてはならず、結果の曖昧さを完全には表せない可能性がある。

利用面からの要求への対応

一般に画像データには極めて大量のデータが含まれている。先ほどあげた机の上にコーヒーカップと資料とが雑然と置いてあるような画像を考えた場合、コーヒーカップと資料とが置かれているという情報よりさらに、例えばコーヒーカップの詳細な3次元形状に関する情報や、その表面の模様、あるいは資料に書かれた文章までも読み取ることが必要とされる場合もあり得る。もちろんこうした情報をあらかじめすべて抽出してしまえば、後にありとあらゆる利用方法に対応することができるはずであるが、上記のようなあまりに詳細な情報を、必要がないかも知れないのにむやみに抽出しようとするのは好ましくない。さらにこうして作成したデータモデルは、画像データベース環境で利用されるということを考えると、データベースとして必要な大量のデータに対して巨大なコストをかけてこうした情報を抽出することは、全く非現実的である。

しかしながら、全く抽象化を行わない生の画像データのままで画像データベースを構築するのでは、高度な利用の不可能な使い道のないデータベースになってしまう。そこそこのコストをかけてある程度の抽象化を行った状態でデータベースを構築しておいて、後のより詳細な情報の抽出は、そのデータベースの利用段階になってから必要に応じて行うというアプローチが現実的であろう。この「ある程度の抽象化」のレベルは、データベースの規模とデータモデルに期待される抽象度との間のトレードオフで決まることになる。

また、画像認識・理解が一般に困難な問題であることを述べたが、対象の画像に関して状況や認識の目的などが明確になった場合、認識自体が容易になる場合があり得る。例えば、ある画像を見せられて「これは何の絵か」という質問がなされた場合、一般には画像理解システムがこれに答えられることは全く期待できない。しかしこの質問が「これは動物の絵か」というようなものであり、これにyes/noで答えることが求められているような場合は、かなり答えられる可能性が出てくる。こうした状況や目的などは、一般にはそのデータ

モデルの利用時に確定されることになる。

こうした事情を考えると、画像データモデル獲得には、このように抽象化がほどほどの状態でとどまっている画像データベースを、利用時に必要に応じてより抽象化したり、利用時に明確になった状況や目的などを利用して認識できなかった画像データの抽象化を行ったりできる機能が望まれる。これは画像データモデル獲得を、利用時の要求に応じて制御できるような形に構成しておくことにより実現できる。

このように制御性に優れた画像認識・理解システムとしては、認識対象に依存した知識と制御すべき認識機構とが明確に分離できるようなものが望ましい。すなわち、プロダクションシステムやフレームベースのシステム、またモデルベースのシステムなどは制御性も比較的優れている。このほか、認識の途中結果によって認識機構の停止や続行などを制御するために、「不完全な認識結果の扱い」も考慮する必要がある。

2.4.2 画像データモデル利用の要件

画像データベースを利用する局面においては、より高度で多彩なアプリケーションが実現可能であるようにしておくことが望まれる。画像データモデル利用では、こうしたアプリケーションの要求を満たすようにデータモデルの利用、すなわちデータベースへの操作を提供する必要がある。

これらの実現のために、画像データモデル利用に望まれる要件としては以下の点が考えられる。

1. データモデル利用部からデータモデルへの操作要求は、統一的な「演算」という形態で与えられる。
2. データモデルの信頼性を扱うことができる。

本節では、これらの要件についてより詳しく述べ、その実現のための方法論を述べる。

「演算」によるデータモデルへの操作の抽象化

画像データベースにおいて、利用システムからデータモデルへの操作としてどのようなものがあるかについて考える。ここでは先に述べた2つのアプリケーションシステムを例にあげて考える。

まず画像データベース検索システムでのデータモデルへの操作は、与えられたキーワードについてこれに相当する画像データを検索することである。これは適当なデータモデルさえ作成しておけば比較的簡単に実現でき、しかもかなり一般性の高い操作である。

次に、ハイパーメディアにおけるデータモデルへの操作について考える。画像オブジェクトが他の複数のオブジェクトからラベル付きリンクで参照されている場合、参照元のオブジェクトとリンクのラベルとによる一種のキーワードで画像オブジェクトが修飾されていると考えることができる。このリンクの参照は、画像オブジェクトのキーワードによる検索と考えることができるので、この時の操作は画像検索の場合と同様、かなり一般的な操作であり、データモデルさえ適当な形に構成しておけば比較的簡単に実現できると考えられる。

しかしながら、ハイパーメディアのデータモデルへの操作として特に重要と考えられる画像データの内部にまで立ち込んだ操作について考えると、この操作としては非常に多様なものがあり得る。例えば、画像中の「ボタン」であれば「押す」ことができたり、「コック」であれば「ひねる」ことができたり、「人物」がいれば「話す」ことができたり、などといった操作が無数にあり得る。このため、これを一般的・統一的に扱うことは極めて困難であり、基本的にはアプリケーションごとに場当たりに対応するしかない。画像データベースのアプリケーションシステムとしては他にもまだまだ多様な種類のものがあり、それぞれ多彩なデータモデルの操作を行うようになってきている。これらすべてについていちいち対応するのはほとんど不可能である。

その一方で、画像データベースシステムというらえ方をすると、データモデルの操作はデータモデル利用側からデータモデル管理側への要求という形でなされるので、ここは一貫性のある一般的な形で行うのが望ましい。そこで、このデータモデルへの操作を「演算」という形で抽象化して扱うことを考える。演算は、データモデルの操作のうち一般性が高く

比較的単純なものであり、これら演算の組合せによりデータモデルへのさまざまな操作を実現できるようなのと考える。有限個数の演算を用意することにより、かなりの種類の複雑な操作でも表現できるようなのが望ましい。これらの演算の組合せでさまざまな操作を統一的に実現できる枠組を用意しておくことにより、アプリケーション側はアプリケーションに特化した複雑な操作でも与えられた演算の組合せでこれを実現することができる。このように、データモデルへの操作を演算の組合せで表現できるような枠組を用意することにより対応するという方法が考えられる。

先に述べたアプリケーションシステムについて、そのデータモデルの操作をデータモデル利用システムの「演算」というとらえ方をしてみる。まず画像データベース検索システムでのデータモデルへの操作は、キーワードによる画像の評価、キーワードによる画像のクラス分け、あるいはキーワードによる画像の概念クラスタリングといった、かなり一般性のある演算であるといえる。

次いでハイパーメディアについて考える。ハイパーメディアを実現するためのデータモデルは、これを構成するリンク情報などのようにハイパーメディアの構成そのものに深く依存したものにより構成されるため、一般的な議論はしにくい。ハイパーメディアのデータモデルに対する演算要求はおおまかには次のようなものとなる。まずある画像オブジェクトが他の複数のオブジェクトからラベル付きリンクで参照されている場合、参照元のオブジェクトとリンクのラベルとによるキーワードで画像オブジェクトが修飾されていると考えることができる。このリンクの参照は、画像オブジェクトのキーワードによる検索と考えることができるので、この演算はキーワードによる画像オブジェクトの評価であると考えることができる。画像データの内部にまで立ち入った操作について考えると、この操作としてはあまりに多様なものがあり得るため、一般的に述べることはほとんど不可能である。こうした操作は一般性を持った演算に分解するのが難しいが、基本的な演算そのものを対象のデータモデルと一緒に管理し、データモデルへの操作はそれに付随して与えられている演算を組み合わせることにより実現するという方法が考えられる。

一般のマルチメディアアプリケーションでは、データモデルに対して多種多様な操作を要求するが、このようにデータモデル操作の記述の枠組を用意することにより、この多種多

様な無数の操作を一般的なデータモデルへの演算によって実現するというアプローチが望ましいと考える。

データモデルの信頼性

データモデル獲得を完全自動で行うとすると、どうしても得られるデータモデルの品質は一定ではなくなり、さまざまな信頼度を持った記述が入り混じった形で得られると考えられる。一方で、従来のアプローチではデータベースアプリケーションシステムは扱っているデータモデルの信頼性については一般的に全く考慮しておらず、言うならばデータモデルには100%間違いがないと信じて処理を行うようになっていた。この両者をそのまま接続すると、当然不整合が生じてくる。これをうまくつなぎ合わせるためには、アプリケーション側でもデータモデルの信頼性が扱えるようになっている必要がある。

データモデル獲得側では、得られたデータモデルについて信頼度などを付加するのはそれほど難しい処理ではないと考えられる。処理コストを下げるために敢えて信頼性の低いデータモデルを生成するようなことも可能であろう。データモデル利用側では、データモデルに対して行う演算に対し、その結果にどの程度の信頼性を要求するかという情報を付加してデータモデル管理部に渡す。これに対してデータモデル管理部では、例えば高い信頼性を要求されれば、蓄えているデータモデルのうち信頼性の高いものだけを用いて確実な結果のみを返したり、信頼性を上げるためにデータモデル獲得部に対しデータモデルをさらに詳しく調べるように動的に要求するなどして信頼性の高いデータを作成したりすることになる。低い信頼性を要求されれば、誤ったデータが混じることもあり得るが、とにかく拾い落しのない結果が得られることが期待できる。

実際のデータモデル利用は、いくつかの操作(あるいは演算)を組み合わせることにより実現される。信頼性については各演算を適用するごとに考慮しなくてはならず、これらを組み合わせる時には信頼性も合成されていくことになる。この信頼性の合成については、Dempster-Shaferの確率モデル[58]などが利用できると考えられる。データモデル利用における信頼性の扱いに関しては、今後より一般的に考察していく必要がある。

2.4.3 画像データモデル管理の要件

画像データベースとして、自動化された大規模なデータモデル獲得が可能で、高度で多彩な応用が可能であるようなものが望まれており、これを実現するためのデータモデル獲得とデータモデル利用における要件について述べてきた。データモデル獲得では、素材データから可能な限り高度なデータモデルを抽出しようとするが、どうしても抽象化できなかった部分についてもある程度の利用ならばできるような形でデータモデルを提供するための方策について述べた。データモデル利用では、多種多様なデータモデルに対して多彩で高度な操作を実現するための方策について検討した。これら2つの機能部分の要求には互いに相容れない部分も少なからず存在し、これをそのまま接続することはできない。

データモデル管理は、これら2つの機能部分の間の媒介となり、全体として3つの機能部分からなる1つの統合されたシステムとするための極めて重要な位置にある。画像データベースシステムとして構成するためには、データモデル管理ではデータモデル獲得/利用の2つの機能部分への要件を実現するようにし、しかもこれらからの相容れないような要求に対しても何とか答えられるようにうまく両者を接続する必要がある。

これを実現するための画像データモデル管理の要件としては以下のような点があげられる。

1. 記述力の高い概念的なデータモデルにより管理する。
2. データモデル獲得部により生成された、信頼性の面で一様ではない不完全な認識結果を、可用性の高い形で取り入れる。
3. 必要に応じて、データモデル獲得部に動的・局所的な要求を発することができる。
4. データモデル利用部には、データモデルの不完全さをなるべく意識させない。
5. データモデル利用部に対し、汎用で可用性の高い「演算」を提供できる。

これらは大きく分けると、データモデル管理の内部的な問題と、データモデル獲得の要件と関連した問題、およびデータモデル利用の要件と関連した問題の3つに分けることができる。ここではこのそれぞれについて述べる。

データモデル管理の内部の問題

データモデル管理では、実際にデータモデルを扱ってこれにさまざまな操作を行うことによりデータベース利用側にサービスを提供する。すなわち、この部分は狭い意味でのデータベースシステムに相当する。このため、データモデル管理にはデータの永続性の実現、トランザクション処理、データの機密処理、コンシステンシの保存などといった、データベース固有のさまざまな問題点も考慮する必要が出てくる。これらの問題点は、既存のデータベースではすでに論ぜられているので、ここではデータモデルの基本的な部分は既存のものを用いるのが良い。

この既存のデータモデルとしても、ここで考えている画像データベースの実現になるべく即したものにすることが必要である。すなわち、記述力の高い概念的なデータモデルが望ましい。このような要求に答え得るデータモデルとしては、マルチメディアデータモデル、その中でも特にオブジェクト指向データモデルが最も期待が持てるといえる。

データモデル獲得との関連の問題

データモデル獲得と関連した問題点としては以下の2点があげられる。

1. データモデル獲得部により生成された、信頼性の面で一様ではない不完全な認識結果を、可用性の高い形で取り入れる。
2. 必要に応じて、データモデル獲得部に動的・局所的な要求を発することができる。

まず、データモデル管理における不完全な認識結果の扱いについて考察する。データモデル獲得から得られる不完全な認識結果は、最終的にどのような解釈になるかが分からないまま、曖昧な状態で得られる。この時、データモデル獲得の要件が十二分に満たされているならば、なぜ、どの程度曖昧なのかが明示されている状態でこの結果が得られることが期待できる。このような曖昧なデータモデルを、他の内部構造まで詳細に得られているデータモデルと同様に扱っていくことが要求される。

内部構造の曖昧なデータモデルに対してその詳細な内部構造を必要とする操作が与えられた時に、これを行うための指示をデータモデル獲得部に対して発するような機構も望ま

しい。これは、データモデル獲得での利用面からの要求への対応という要件を実現するためにも必要である。この機構の実現のためには、不完全な内部構造への操作に対して動的に対応することができる必要がある。また、その要求の結果データモデル獲得部からその曖昧なデータモデルを少し詳細化した結果が得られ、これが段々と最終的な認識結果に近付いていくことにも対応できる必要がある。

オブジェクト指向データモデルは、これらの条件を満たし得ると考えられるデータモデルである。内部構造の良く分からない曖昧なデータモデルであっても、外部からはその構造を直接はのぞけず、必ずメッセージを介してしかそのオブジェクトに操作を加えられないため、外部から見れば他の完全に認識されているオブジェクトと区別がつかず、全く同様の扱いが可能である。どの程度曖昧なのか、などの曖昧さを表す情報もオブジェクトの内部に取り込み、メソッドの実現に際しこの情報を参照して、より実情に即したデータモデルへの操作を実現することも可能である。

不完全な内部構造への操作に対する動的な対応も、抽象データ型を積極的に用いて、データモデル内のメソッドとして不完全な内部構造への操作が行われた時の動作を手続的にデータモデル内に記述してしまえるため、実現可能であろう。その結果内部構造が段々詳細化していく過程は、情報隠蔽されたデータモデル内部で徐々に内部変数が具体化していくことにより実現できる。またオブジェクトの型(クラス)自体が段々と変化していく場合、例えば物体→乗物→4輪→トラック→…というような場合、これはこのクラスが段々と特殊化していくことになり、オブジェクトのクラスがクラス階層を特化の方向に下っていく過程により表現することができる。

データモデル利用との関連の問題

データモデル利用と関連した問題としては、以下の2点が考えられる。

1. データモデル利用部には、データモデルの不完全さをなるべく意識させない。
2. データモデル利用部に対し、汎用で可用性の高い「演算」を提供できる。

データモデル利用側にデータモデルの不完全さを意識させないという点は、不完全なデータモデルの扱いに深く関係している。これについては、不完全に認識された曖昧なデータモデルであっても他の完全に認識されたデータモデルと同様に扱う必要があることを述べたが、これをデータモデルの利用側にも拡張することによりこの要件を満足することができると考えられる。

汎用で可用性の高い演算を提供するという点に関しては、データモデル利用の章で述べたように必要なデータモデルの操作を実現するために最小限の演算を用意し、データモデル管理ではデータモデルごとにこの演算を用意するという方法が考えられる。データモデル利用から見ると、この演算を用いて必要な操作を記述するようになり、いわばこの演算はデータモデル操作の記述のためのプリミティブということになる。データモデル管理から見ると、あくまで有限個の演算について内部に曖昧さを含んでいたり十分に抽象化されていないような状態でも何とか実現することにより、データモデル利用にはデータモデルの不完全さを意識させず、その上高度な操作が提供できることになる。

こうした要求もオブジェクト指向データモデルで実現できると考えられる。データモデル利用部からのデータモデルの操作に対し、必ず抽象データ型のメソッドで対応するようになっていれば、データモデル利用部にはデータモデル内部の曖昧な状態などを意識させることなく、利用部からの操作要求に答えることができる。またこの要求を演算としておき、さまざまなクラス(特に特化の関係にある親子クラス)に共通に同じ演算が利用できるようにしておくことにより、利用部でこの演算をプリミティブとして作成した高度なデータモデル操作要求に答えることができる。データモデル管理側でも無数の操作要求にいちいち答える必要がなく、演算の実現についても型継承をうまく用いることによりオブジェクトのモジュラリティ高く効率良く行うことができる。

第 3 章

データモデル獲得を考慮した画像データモデル

3.1 概要

これまでに、データモデル獲得の自動化を指向した画像データベースの実現に向けて、画像データベースシステムを構成する3つの機能部分であるデータモデル獲得/管理/利用の各部分に対して望まれる要件について考察してきた。本章では、これらの要件をある程度満たした画像データモデルの一例として、状態遷移モデル [45, 53] に基づく画像データモデルについて述べる。状態遷移モデルは画像認識・理解のための機構のモデルであり、これに基づく画像データモデルには以下のような特徴がある。

- 画像認識・理解機構を、内部状態を持つ計算モデルで実現している。
- 見通しの良い認識・理解機構が実現できる。
- 不完全な認識結果をデータモデルとして利用可能な形で生成することができる。
- より概念的なデータモデル (オブジェクト指向データモデル) との親和性が高い。

これらの特徴は、これまで述べてきた画像データベース実現のための要件に対して大変有利に働くことになる。

本章では、この状態遷移モデルの概要について述べ、これに基づいたデータモデルについて述べる。状態遷移モデルに基づいた認識理解機構である状態遷移型認識理解機構による実際の認識プロセスについては、章をあらためてより詳しく述べる。

3.2 状態遷移モデル

状態遷移モデルに基づく画像認識・理解機構では、認識プロセスを先に述べた内部状態を持つ計算モデルでとらえる。データモデルとしては、この認識プロセスの内部状態をできるだけそのまま利用できるようなっている。このため、データモデル獲得における認識プロセスの内部状態をデータモデルとして利用できるように留意して構成している。状態遷移モデルに基づいて構成したデータモデル獲得部と、それにより生成したデータモデル、およびその利用部との関係を模式的に表した概念図を図 3.1 に示す。

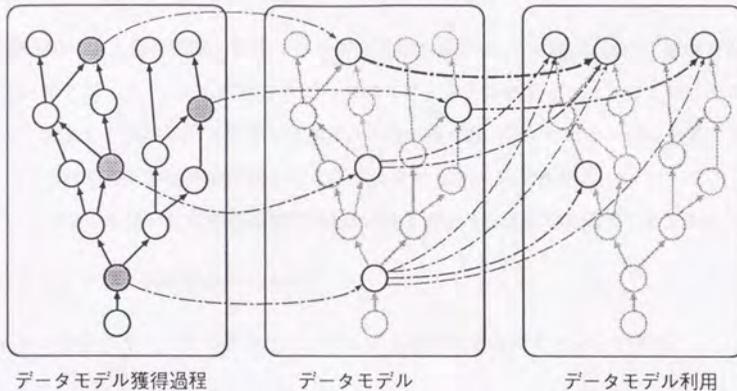


図 3.1: 状態遷移モデルによるデータモデルの概要

一般に、現状の画像認識・理解技術を用いた画像データモデル獲得では、完全な認識結果が得られるとは限らない。これは状態遷移モデルを用いたデータモデル獲得についても同様であり、対象画像のある部分は高い抽象度の概念的なレベルにまで認識プロセスが進んでいるが、他の部分では画像の若干の変形やノイズなどにより概念レベルに至る途中状態で停止している。また他のある部分では、全く認識プロセスが進んでいないような部分もある(図 3.1左部)。

状態遷移モデルに基づく画像データモデル獲得では、このように不完全な認識結果を含み、かつその抽象度も高いものや低いもの入り混じったまま、その内部状態をそのままデータモデル獲得結果として出力するようになっている。それを受けて生成された画像データモデルは、図 3.1中央部に示すように、認識結果の各段階、および最終的に得たい概念レベルに至る高い抽象度の状態との、状態遷移による関係が明白な形で得られることになる。

本節では、これを実現するための状態遷移モデルに基づく画像認識・理解プロセスの概要について述べ、これを用いた画像データモデルについて述べる。

3.2.1 状態遷移による画像認識・理解プロセス

状態遷移により画像認識・理解プロセスを実現するため、まず認識対象を、状態を持ち得る実体(オブジェクト)から構成されているようにとらえる必要がある。ここではこの実体として、画像上で幾何学的な実体を持っている線・棒・箱・点などのシンボルを表すものとする。これらの対象画像を構成するオブジェクトをトークンと呼ぶ。

このトークンを基にした画像認識プロセスは以下のような過程で実現できると考えられる。

- 各トークンへの適切なラベル付け。
- いくつかのトークンのグルーピングによる構造化されたトークンの生成。

基本的には、認識対象の画像の解釈を表す各トークンの適切なラベリングによって画像認識プロセスが行われる。複雑な構造を持つような対象については、この構造を表すようにいくつかのトークンをグルーピングすることによって構造化し、かつその構成要素の各トークンはそれぞれ適切にラベリングすることにより、こうした複雑な構造を持つような対象でも扱うことができる。

状態遷移モデルでは、このように表現される画像理解プロセスをトークンの内部状態の状態遷移により実現する。各トークンに付与されるラベルがすなわちトークンの内部状態であるとする。こうすることによって、状態遷移によりトークンへの適切なラベリングプロセスが実現できる。また、いくつかのトークンをグルーピングして新たなトークンを発生させることにより、グルーピングによる構造化に対応するトークンが生成できる。

このようにして、トークンの状態遷移により画像理解プロセスを実現する。この様子を図3.2に示す。この状態遷移モデルによる画像理解プロセスは、一般性の高いものであり、多くの画像理解のための手法はこの状態遷移モデルによりモデル化できると考えられる。

状態遷移モデルによる画像理解では、トークンの最終状態がすなわち画像理解の最終結果を表すことになる。2章で述べた内部状態を持つ計算モデルにおける「状態」とは、いわば認識対象を構成するすべてのトークンのすべての内部状態の総体として表現されるもの

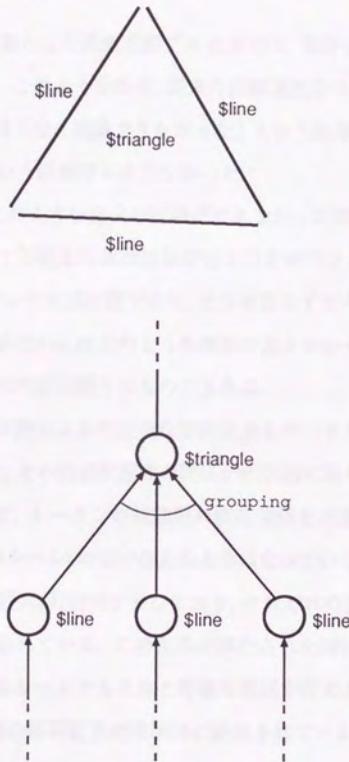


図 3.2: トークンの状態遷移の様子

である。ここでいうトークンの状態とは、先の章で述べた「状態」を構成する一つの要素と
いうことになる。

3.2.2 状態間の依存関係の明示

一般に複雑な画像を対象とした画像理解プロセスでは、期待したような完全な認識結果
が得られるとは限らない。このような場合、従来の画像理解システムでは、完全な認識がで
きなかった部分に関しては「全く認識できなかった」という結果しか得られず、その結果を
何らかの形で利用するという余地はあまりなかった。

状態遷移モデルでは、このような完全な認識ができなかった部分についても、その部分の
トークンの内部状態として不完全な認識結果を得ることができる。状態遷移モデルにおい
ては、認識結果は各トークンの内部状態であり、とりも直さずどのような経路で認識が進み、
完全な認識結果が得られるためにはどのような情報が足りなかったのかなどをすべて表し
ている画像理解システムの内部状態そのものでもある。

この各トークンの内部状態による不完全な認識結果をデータモデルとして利用するため
には、2章で述べたように、その内部状態間の関係が明示的に得られるような形になってい
る必要がある。このために、トークンの状態間の依存関係を状態遷移グラフ(状態遷移図)
や状態遷移規則(状態遷移ルール)の形で与えるようになっている。状態遷移図の例を図3.3
に示す。矢印が状態間の遷移の可能性を表しており、それぞれの矢印ごとにその遷移が起
り得るための条件が付与されている。この条件が満たされた時に限りその状態遷移が行わ
れることになる。状態遷移ルールでもこれと等価な表現が行われることになる。このよ
うにして、トークンの状態間の依存関係が明示的に表現されている。

3.2.3 状態遷移モデルに基づく画像理解機構の構成

従来の画像理解システムのアプローチでは、システム自体が認識対象に深く依存してし
まうことが多く、新しい対象を認識するシステムを構築する場合、対象ごとにシステムを始
めから作成し直す必要があった。状態遷移モデルに基づくアプローチでは、状態遷移を表す

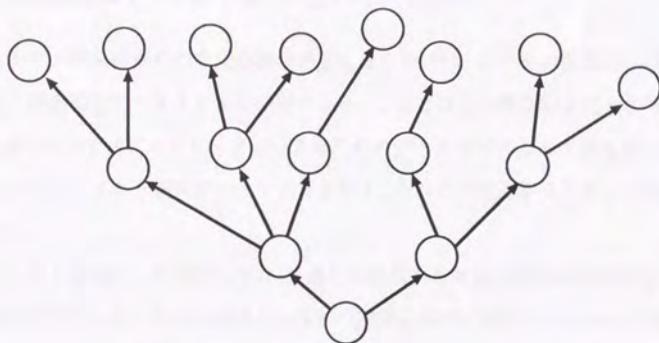


図 3.3: 典型的な状態遷移図の例

状態遷移規則などの部分を明示的に表すことにより、画像理解システムの構成が比較的見通しの良いものとなり得る。認識対象に依存した部分は状態遷移を表す部分に集中しており、この状態遷移を実現する画像理解機構の部分は、認識対象とは独立した対象によらず共通に利用できるような構成となる。

このような構成をとることにより、認識対象が変わった場合でも認識機構は同じものを利用でき、この対象に対応したシステムの作成が比較的容易である。また認識機構が対象に依存しない汎用の構成であることにより、その機構をさまざまな要求に従って制御することも比較的容易にできると考えられる。こうした意味で、状態遷移モデルによる画像理解システムの機構は見通しが良い構成であるといえる。

この状態遷移モデルに基づく画像理解システムの機構については、4章で実際のシステムに沿ってより詳しく述べる。

3.3 概念的なデータモデルとの親和性

データモデル利用面からの高度な操作に応えるためには、データの概念的な側面を表すようなより概念的なデータモデルが必要である。ここではこの概念的なデータモデルとして、現時点でマルチメディアデータベースのためのデータモデルとして最も適当であると考えられるオブジェクト指向データモデルを考え、これと状態遷移モデルとの親和性について論じる。

オブジェクト指向データモデルでは、2通りの視点から対象を階層的に抽象化して扱う。一方が複合オブジェクトによる部分と全体との関係に基づく階層で、is-part-of 関係による階層である。もう一方が型階層(クラス階層)に基づく is-a 関係による階層である。これら2通りの階層関係に基づき、is-part-of 関係による階層では複雑な構造を持つ対象をより単純な小さい部分の複合体としてとらえ、is-a 関係による階層では特化・汎化関係により対象をより抽象的なオブジェクトとしてとらえることを可能にしている。

ここで、状態遷移モデルに基づく画像認識・理解プロセスについて考える。まず認識対象はトークンという形に分解される。これらのトークンの内部状態は、はじめは最も一般的な状態になっており、今後いかなる状態にも遷移し得るような状態が初期状態として与えられる。この後の認識プロセスにおいて行われる状態遷移としては、大別すると2通りのものがあり得る。一つはあるトークンの属性やその周囲状況などを調べ、これがある条件を満たしている場合このトークンをもとの状態から少し特殊化した状態へと遷移させるというものである。もう一つはあるトークンとその周囲のいくつかのトークンの内部状態を調べ、これがある条件を満たしている場合これらのトークンをグルーピングして別の新たなトークンを生成するというものである。

これらの状態遷移について考えると、あるトークンを特殊化するという状態遷移は、すなわちオブジェクト指向データモデルでの is-a 階層をさかのぼり、トークンをもとより少し特化したクラスに属するように状態遷移したものととらえることができる。また、いくつかのトークンをグルーピングして新たなトークンを生成するという状態遷移は、もとのトークンにより構成されるトークンが生成されることになるので、オブジェクト指向データモ

デルでの複合オブジェクトが生成されることに対応し、is-part-of 階層に沿った状態遷移であるということができる。

このことから、ある対象の認識のための状態遷移規則が与えられているとすると、これによりトークンの状態遷移の過程が決定でき、これに沿ったオブジェクト指向データモデルのためのデータモデルの階層も自ずから決定されることになる。あるいはオブジェクト指向データモデルの階層が与えられている場合、これに沿った形で状態遷移規則を作成することも容易に実現できると考えられる。このようにして、状態遷移モデルに基づいた画像認識・理解機構により生成したトークンの内部状態としての認識結果は、容易にオブジェクト指向データモデルの枠組で扱うことができるといえる。

3.4 不完全な認識結果の扱い

状態遷移モデルにおける不完全な認識結果の利用の可能性について考察する。認識結果としてのデータモデルの利用法としては極めて多種多様なものが考えられるが、これを実際に実現するのはより概念的なデータモデルを擁したデータモデル管理での問題であると考えられる。データモデル獲得としては、認識結果をこのような高度な利用が可能な形態とするため、対象の画像を概念的データモデルにより、利用側が望んでいる抽象度で表現することが求められる。すなわち、画像を構成する各トークンがオブジェクト指向データモデルなどの概念的データモデルにおいて任意の抽象度のオブジェクトとして識別できることが望まれる。

ここでは、状態遷移モデルによって得られた不完全な認識結果としてのトークンの内部状態について、これを任意の抽象度の状態として扱うための方策について考察する。

図 3.4 は、画像理解プロセスにおけるトークンの状態遷移の様子を表す状態遷移ダイアグラムの例である。図中で、A、B、C はまだ意味的なレベルに達していない中間状態を表しており、X、Y は利用可能な意味を持つ、いわば概念的なレベルに達している状態を表している。いま、画像理解プロセスにおいて、あるトークンが中間状態 A であるという結果が得られたとする。図を見ても明らかな通り、状態 A にはまだ極めて数多くの状態遷移の可能性

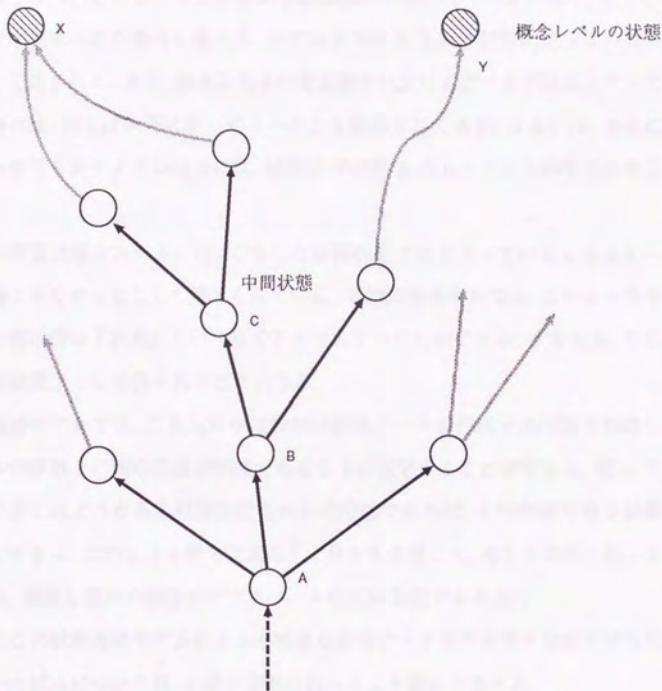


図 3.4: 画像理解プロセスにおけるトークンの状態遷移ダイアグラムの例

が残されており、このままではこのトークンが最終的にどういう意味を持ちうるかについて、ほとんど何の情報も得られない。ここで状態遷移のための条件がそろって、このトークンが状態 B に遷移できた場合を考える。この場合、このトークンは、知られている概念レベルの状態になるとすれば、X か Y という状態にしかなり得ないことが分かる。

このような A, B, C といった不完全な認識結果が得られている状況で、X という状態のトークンが要求された場合を考える。まずはまさに X という状態に至っているトークンが結果として返される。また、結果にある程度正確さに欠けるデータが混在していても良いという場合には、例えば中間状態 C のトークンも結果として返す。あるいは、さらに正確さの低いデータでも許すような場合には、状態 B や状態 A のトークンも利用されることもある。

従来の画像理解システムでは、こうした中間状態でとどまっているようなトークンはすべて認識できなかったとして捨てられていた。状態遷移モデルでは、このような中間状態は外部から明示的な「状態」という形でアクセスすることができる。すなわち、これらもすべて「認識結果」として得られることになる。

状態遷移モデルでは、これらの中間状態は認識ルールを作成する段階で期待している概念レベルの状態との間の関係が明確であるように設定することができる。従って、図 3.4 で概念的に示したようなある程度限定された利用法であれば、充分利用可能な結果として得ることができる。このような形で認識の不十分さを考慮して、何とか利用に供しようという考え方は、複雑な図面や画像のデータベース化には有用である [41]。

実際にこの状態遷移モデルによる不完全な画像データモデルを不完全ながら利用に供しようという試みについては、5 章で実例に沿ってより詳しく述べる。

3.5 考察

データモデル獲得過程を考慮した画像データモデルとして、状態遷移モデルについて述べた。その特徴として、以下のような点をあげた。

1. 画像認識・理解機構を、内部状態を持つ計算モデルで実現している。

2. 見通しの良い認識・理解機構が実現できる.
3. 不完全な認識結果をデータモデルとして利用可能な形で生成することができる.
4. より概念的なデータモデル(オブジェクト指向データモデル)との親和性が高い.

1と3はデータモデル獲得と管理に対する要件である不完全なデータモデルの利用を実現している. 2はデータモデル獲得の要件の大規模・体系的なデータモデル獲得の実現,さまざまな画像認識・理解技術への対応,および利用部からの動的な要求への対応に深い関わりがある. 4はデータモデル管理の要件での概念的なデータモデルによる管理の実現に必要な不可欠である. このように,状態遷移モデルによる画像データモデルは画像データモデルとして望ましい多くの特徴を持っているといえる.

続く2つの章では,実際のシステムによりこの状態遷移モデルによる画像データモデルの有効性について検証する.

第 4 章

状態遷移モデルに基づいた図面理解システム

4.1 概要

状態遷移モデルに基づく画像データモデルは、3章で述べたように、本論文で考えている画像データモデルとして望ましい多くの特徴を持つ。これに基づく画像理解システムは、画像データモデル獲得の要件をほぼ満足することになる。本章では、この状態遷移モデルに基づく画像認識・理解機構の一具現例として、状態遷移モデルに基づく図面理解システムについて紹介する [45-48, 53, 55, 56]。元来状態遷移モデルは画像認識・理解の機構のためのモデルであったが、ここでは一般の画像を扱うよりも実装が容易であると考えられる図面データに対する認識システムに適用することとした。

本システムの特徴としては以下のようなものがあげられる。

- 不完全な認識結果でもデータモデルとして利用可能な形で生成できる。
- 新たな対象図面に対応した図面理解システムを構築するのに際し、その作成が容易である。
- ボトムアップ処理とトップダウン処理の併用により、強力な認識プロセスが実現できる。
- 認識機構自体が見通しの良い構成となっており、外部からの制御が行いやすい。

状態遷移モデルの概略および特徴についてはすでに前の章で述べたが、本章では特にその認識機構としての側面について着目し、これについて重点的に述べ、その有効性について論じる。

以下、状態遷移モデルに基づく図面理解システムの特徴について述べ、その構成について述べる。ついで本システムの特徴である対象依存の部分が簡単に作成できることを実証するため、実用的なレベルの図面理解システム(ここでは地図図面を対象とする図面認識システム)に適用した例について紹介する。そして、この状態遷移モデルに基づく図面理解システムについての考察を行う。

4.2 システムの特徴

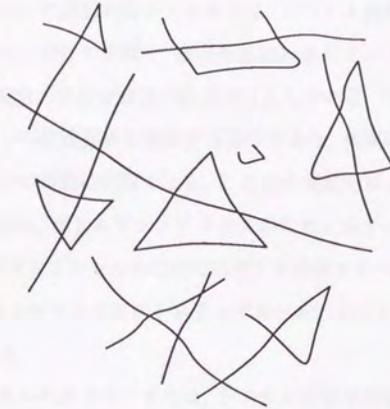
図 4.1 に、ここで述べる図面理解システムの処理対象として想定しているいくつかの種類
の図面の例を示す。図 4.1(a) は手書きされた線図形の例で、図 4.1(b) は地図図面の一部を
表している。本システムでは、状態遷移モデルの利用により、この図で示した例のように独
立したシンボルで構成されており、各シンボルの適切な意味づけ(ラベリング)によって解
釈ができるような図面を想定している。ここでのシンボルとは、線・棒・箱・点などとい
った幾何学的な実体を持ったものをいう。状態遷移モデルではこれらのシンボルをトークン
と呼び、図面理解プロセスはこれらのトークンの内部状態の遷移として実現する。

状態遷移モデルに基づく図面理解システムの機構としての特徴は以下のようなものがあ
げられる。

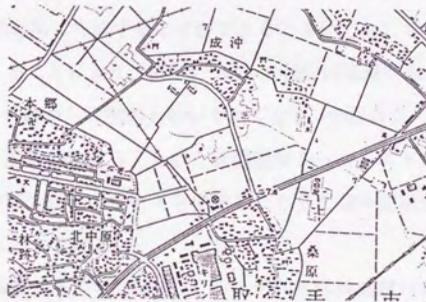
- ボトムアップ処理とトップダウン処理の併用により、強力な認識プロセスが実現で
きる。
- 新たな対象図面に対応した図面理解システムを構築するのに際し、その作成が容易で
ある。
- 認識機構自体が見通しの良い構成となっており、外部からの制御が行いやすい。

これらの点について説明する。

図面理解を含め、一般に画像理解プロセスでは、ボトムアップ的な処理とトップダウン的
な処理とを併用する必要がある。状態遷移モデルでもこれを実現するため、トークンの状態
遷移をボトムアッププロセスとトップダウンプロセスの2つに分けて行う。ボトムアップ
プロセスでは、各トークンが自律的に自分の比較的近傍の周囲状況を調査して、それに従っ
てその内部状態を変更することにより、ボトムアップ的な状態遷移を実現する。トップダウ
ンプロセスでは、ボトムアップ処理の結果をもとにして、認識対象を表すモデルとのトップ
ダウンのマッチングを行う。この両者の併用により、まずボトムアッププロセスでは局所的
な視野で明白な局面について判断することによって、認識プロセスの効率化を図り、トップ



(a) 手書きの線図形の例



(b) 地図図面の例

図 4.1: 図面イメージの例

ダウンプロセスでは大局的な視野に立った状況判断によって、より複雑な状況でも認識することが可能となる。

状態遷移モデルに基づいた図面理解システムでは、システム自体を認識対象に依存した部分と依存しない部分とに分けて実現し、新たな認識対象について構築し直さなければならない対象に依存した部分の作成が容易であるようにしている。対象に依存しない部分は、図面を構成するトークンの状態遷移を実現する部分であり、基本的には認識対象にはよらずさまざまな対象について共通に利用できる。ここでの実装ではこれを認識カーネルと呼ぶ。対象に依存した部分は、ボトムアッププロセスのためにはトークンに与える状態遷移ルールとして、トップダウンプロセスのためにはモデルを表すルールとして、それぞれシステムに実装される。それらをそれぞれボトムアップルールとトップダウンルールと呼び、合わせて認識ルールと呼ぶ。

従来の図面理解システムのアプローチでは、システム自体が対象に深く依存してしまうことが多く、新しい対象を認識するシステムを構築する場合、対象ごとにシステムを始めから作成し直す必要があった。しかし、状態遷移モデルに基づくアプローチでは、認識カーネルの部分は対象によらず共通に利用することができ、対象ごとに用意しなければならないルールの部分も比較的容易に作成することができるように工夫している。

このような構成をとることにより、認識カーネルの部分は認識対象によらない汎用的な形態をとることになり、外部からの制御も行いやすくなる。今回のシステムの場合、認識カーネルの動作としては、基本的には与えられた状態遷移ルールを各トークンに適用していくのみであり、任意の段階で認識プロセスを中断したりまた途中段階から再開したりという制御は極めて行い易くなっている。

認識のための手続きをルールの形にして、システム本体は処理対象に依存しないようにしようというアプローチには、他にプロダクションシステムを用いたものがある [30]。しかしながら一般に、プロダクションシステム自体にはトップダウン処理をする機能がなく、あくまでボトムアップ的にルールを対象に適用するのみである。そのためこれを図面認識システムに応用する場合には、ルールの方をうまく構成してトップダウン的な効果が出るようにしなければならず、ルール作成の負担が大きいという問題がある。これは、あくまで

ログクシオンシステムは汎用の問題解決システムであるので、プロダクションルールをボトムアップ的に効率良く実行する方法に力点が置かれているためである。

また、対象のモデルとのトップダウンマッチングを行い、その認識を行おうとするアプローチとして、フレームを利用したものがある [19, 34]。モデルはフレームの形で階層的に表現される。これを対象とトップダウンにマッチングすることによりその解釈を行う。フレームは記述性が良く、かなり複雑なモデルでも記述できる。しかしながら、一般にトップダウンのマッチングは効率が悪く、特に図面データのように対象を構成する要素が大量にある場合、その認識プロセス実行の効率の低下は著しい。

本章で述べている状態遷移モデルに基づく図面理解システムでは、これらの問題点をある程度解決しており、画像データモデル獲得として望ましい特徴を持つシステムであるといえる。

4.3 システムの構成

図 4.2 にシステムの構成図を示す。

対象の図面は、スキャナーによりビットマップデータに変換され、システムに入力される。このデータはまずトークンエクストラクタによって処理され、基本的なシンボルを表すトークンによる記述に変換される。この時、各トークンには初期状態が付与されており、初期トークン記述となっている。

ついで各トークンは、それぞれ状態遷移を行うように認識カーネルにより処理される。認識カーネルは与えられた認識ルールに従って状態遷移を行うが、この時まずボトムアッププロセスのモードとなり、ボトムアップルールを適用して各トークンをボトムアップ的に処理する。こうしてすべてのトークンがこれ以上状態遷移できないという状況になった時、自動的にトップダウンプロセスのモードとなり、今度はトップダウンルールを適用してトップダウン的な認識プロセスを行うように動作する。各トークンは認識カーネルにより順次その内部状態を遷移していく。その任意の途中状態は、いつでも認識結果として利用できる形式であるが、トップダウンプロセスでもこれ以上状態遷移できないとなった時点のトー

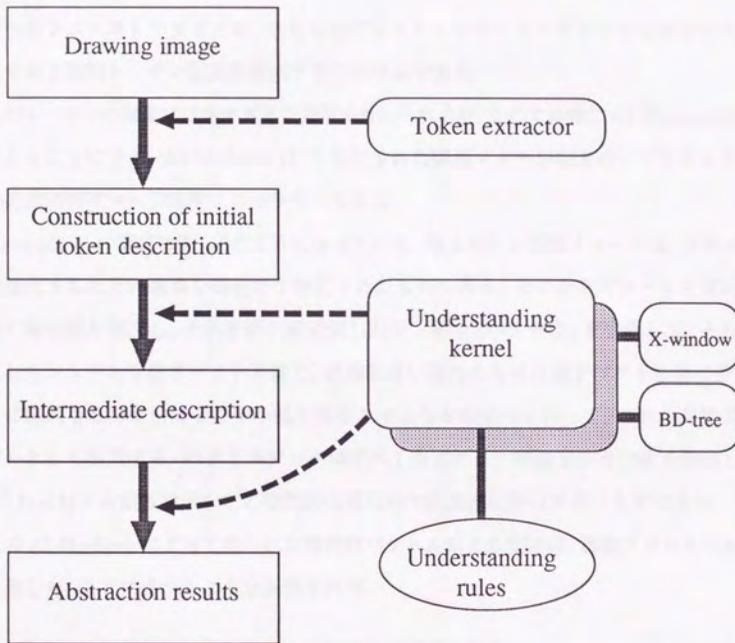


図 4.2: 状態遷移モデルに基づく図面理解システムの概略

クンの内部状態が、システムの認識プロセスの最終結果となる。

以下で、トークンエクストラクタ、認識カーネルのボトムアッププロセス、トップダウンプロセスについて順次述べる。

4.3.1 トークンエクストラクタ

トークンエクストラクタとは、与えられたビットマップイメージからシンボルレベルの記述である初期トークン記述を生成するシステムである。

このトークンの抽出にはさまざまな方法が考えられるが、ここでは特に AI-Mudams[23,42] を考えることにする。AI-Mudams は、2 値化された図面イメージ向けのソフトウェアベースの高速図面イメージ処理プロセッサである。

AI-Mudams の構成は図 4.3 のようになっている。与えられる図面イメージは、スキャナーで数値化された上に適当な閾値で 2 値化されたものである。そこからイメージを構成する図形の輪郭線を抽出し、それを折れ線近似したシンボル (ベクトル) を生成して、それらを孤立したシンボルを表すベクトル群と、比較的長い線的なものを表すベクトル群に分ける。さらに線的なものからはその中心線を発生させるなどの処理を行って、これらを結果のシンボルとして出力する。ベクトルデータはすべて空間データ構造 BD 木 [39] で管理しており、それに対する領域探索などの空間的な操作は十分高速に実行することができる。

この AI-Mudams によって得られた輪郭線ベクトルによる記述は、認識プロセスの初期記述に適した、以下に述べるような特色を持つ。

- イメージが線分というシンボルのレベルで記述される。
- 元の 2 値化されたイメージと比較して、折れ線近似の時の誤差しか入り込まず、線の太さなどの情報も含めてほぼ完全な形で元のデータを取り出すことができる。
- シンボルレベルの記述であるため、画像記述に比べてデータの量は大きく減少している。

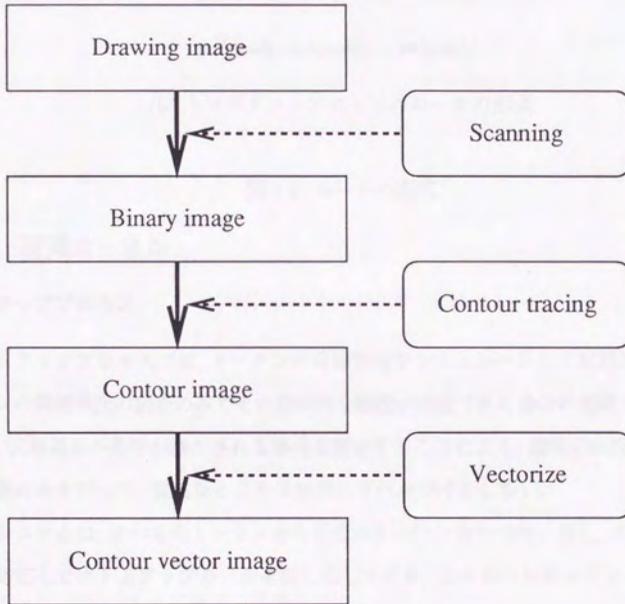


図 4.3: AI-Mudams の概略

$$cur_state \leftarrow cond_1, cond_2, \dots, cond_n,$$

$$next_state_pred.$$

(a) ボトムアッププロセスのルールの形式

$$goal \leftarrow cond_1, cond_2, \dots, cond_n,$$

$$subgoal_1, subgoal_2, \dots, subgoal_n.$$

(b) トップダウンプロセスのルールの形式

図 4.4: ルールの形式

4.3.2 認識カーネル

ボトムアッププロセス

ボトムアッププロセスでは、トークンの自律動作をシミュレートして状態遷移を行い、各トークンの周囲状況の調査のみでその意味的な解釈が決定できる場合の意味づけを行う。すなわち、状態遷移の条件が満たされる場合を限定することにより、確実に状況判断できる局面の解釈のみを行って、複雑なところは無理にラベルづけをしない。

まずシステムは、すべてのトークンから任意のトークン一つを拾い出し、そのトークンの状態に対応したボトムアップルールを探し出してくる。このルールはエジンバラ Prolog の記述法を用いて表され、トークンの現在の状態をヘッド部分に持ち、満たすべき制約条件および次の状態への遷移の指示をボディ部に持つ(図 4.4(a))。システムは、ルールのボディ部にあるすべての条件が満たされるかをテストし、もし満たされればその後続く状態遷移を行う。途中で条件が満たされることがわかったら、そのルールは捨て、別のルールを取り出して、始めから条件の試行をやり直す。

ある時点において複数のルールが同時に満たされる場合には、どのルールが正しい選択であるかその時点ではわからないので、一般にはバックトラックなどによって、満たされる

すべてのルールを試行してみなければ、正しい認識プロセスは実現できない。しかし、このバックトラックを可能にした場合、処理の効率が著しく落ちてしまう。そこで本システムのボトムアッププロセスでは、特に初期のバックトラックが必要ではないような、意味づけが明白な状況についての認識プロセスを効率良く行うことを考えている。従って、ルールの充足性が排他的になるようにルールを構成しなければならない。この場合のトークンの状態遷移は決定的に行うことができる。後のトップダウンプロセスではバックトラックが可能のように動作するが、ボトムアッププロセスでは、そのトップダウンプロセスの際の解空間の大幅な刈り込みを行い、解釈の可能性を絞り込んでいると考えることができる。

すべての条件が満たされなくなると、そのトークンは停止状態となり、回りのトークンが状態を変化させることによって周囲状況が変わり、状態遷移の条件が満たされるようになるまで待つことになる。

一つのトークンの処理が終ると、システムは次のトークンを選び出してこのプロセスを繰り返す。この繰り返しは、すべてのトークンが停止状態になるまで行われる。こうしてシステムは、すべてのトークンの状態遷移を疑似並列的にシミュレートすることになる。

このボトムアッププロセスの具体的な動作例は、4.4.2 節で実例に沿って詳しく説明する。

トップダウンプロセス

ボトムアッププロセスですべてのトークンが停止してしまうと、認識カーネルはトップダウンプロセスを起動する。

このモードでは、ボトムアッププロセスで生成された記述から、与えられた認識対象のモデルが抽出できるかどうかをトップダウンに探索する。もしあるモデルが抽出されると、それがこのトップダウンプロセスの結果として出力される。もし失敗するとこのモデルは捨てられて、バックトラックにより次の候補のモデルが試行される。

ボトムアッププロセスとは異なり、トップダウンプロセスでは、トークンが複数通りに満足されることを許している。もしあるモデルが満たされた時に、別の解釈の可能性(チョイスポイント)が残っている場合、後のバックトラックによる解釈のやり直しを可能にするため、これらは保存される。後に解釈に矛盾が生じ、この時点での解釈が誤りであることがわ

かった場合、ここのチョイスポイントが再び呼び出され、他の解釈がなされることになる。こうしてシステムは非決定的な認識を実現している。

モデルはルールの形でシステムに搭載される。これはボトムアッププロセスの場合と同様エンジンバラ Prolog の記法で表され、図 4.4(b) のようになる。実際の動作では、認識したい対象のモデルとマッチするヘッド部を持つルールが選ばれ、ボディ部の条件が試行されて、これがすべて満足されると、ボディ部に記述される、そのモデルを構成する要素を表すいくつかのモデルを、さらに再帰的に抽出するように動作する。もしどこかで失敗すると、システムはもっとも最近のチョイスポイントまでバックトラックして、そこからモデルマッチングを再開することができる。このようにして、最終的にあるモデルがある導出過程をもって満たされると、その過程がそれに関与した各トークンのトップダウンプロセスにおける状態遷移となる。

マッチングに成功したモデルを1つの状態と考えると、この状態は、モデルを構成する各構成要素の状態から、それらをグルーピングして状態遷移したものと考えることができる。

トップダウンプロセスの具体的な動作例については、ボトムアッププロセスとともに 4.4.2 節で事例に沿って詳しく説明する。

4.4 本枠組に基づく図面認識システムの実現

4.4.1 システムの実装

これまで述べた枠組に基づいて、実際の図面の解釈を行う図面認識システムを作成した。システムはトークンエクストラクタと認識カーネルとから構成される。トークンエクストラクタとしては C 言語で作成された AI-Mudams システムを使用した。この処理系は C 言語でおよそ 2500 行で記述されている。認識カーネルは C 言語と Prolog とで記述され、いくつかのライブラリがリンクされる。このライブラリとしては、まず空間的な評価を効率よく行うために、C 言語で記述された BD 木ライブラリを使用している。また、処理の途中経過や結果をグラフィックディスプレイに表示するために、X-Window システムの Version

11 Release 4 の Athena Widget ライブラリを Prolog システムから呼び出すようなインターフェイスを用意し、これを使用している。4.3.2 で述べたボトムアッププロセス、トップダウンプロセスを含む実際の認識プロセスを行う本体は Prolog で実現している。この認識プロセスの支援を行う部分、すなわちトークンの形状、それらの空間的な関係や存在する位置などといった基本的な幾何的な状況をテストするために空間的な評価を行う部分、および外部ライブラリとのインターフェイス部分は、C 言語で記述されている。処理系のサイズとしては、BD 木ライブラリの部分が C 言語で約 1000 行、認識カーネルの Prolog 部分が約 2000 行、その他の C 言語の部分が約 1000 行となっている。

実際のシステムは 4.5 MIPS の Sun-3/260 ワークステーション上で、Prolog の処理系としては SICStus Prolog を用いて作成している。C 言語で書かれたさまざまなプリミティブやライブラリなどを Prolog システムにとり込むため、SICStus Prolog の foreign 述語などを利用している。

以下、手書きの線画と地図という 2 種類の異なった図面を対象にして、本システムの動作について説明する。

4.4.2 手書き線図形の認識

図面の認識の例として、まず図 4.1(a) に示した手書きの線画から得た図 4.6(a)(一部分) のトークン記述から、三角形をなす部分を取り出すという認識問題を考える。

この問題に対するボトムアップルールは図 4.5(a) で表される。このルールによるボトムアッププロセスの戦略は、具体的には以下のようなようになる。

1. すべての線分トークンをチェックして、自分の 2 つの端点の近くに端点を持つ線分が、それぞれ 1 本ずつあるような線分トークンを選び出す。
2. その結果見つかった隣合う線分が、比較的まっすぐにつながっている場合(すなわち、それらの線分のなす角度がほぼ 2 直角である場合)、線分のまっすぐなつながりを表す新しいトークンが生成される。

```

$line :-
  $get_self(S),
  $get_terminal(S, X, Y),
  $near_line_terminal(X, [L1]),
  $near_line_terminal(Y, [L2]),
  $create($line_line(S, L1)),
  $create($line_line(S, L2)).

```

```

$line_line(A, B) :-
  $is_straight(A, B),
  $transform($connect(A, B)).

```

```

$line_line(A, B) :-
  $is_corner(A, B),
  $transform($corner(A, B)).

```

(a) ボトムアップルール

```

$triangle :-
  $corner(A, B),
  $edge(A, C), $edge(B, D),
  $corner(C, E), $corner(D, F),
  $edge(E, F).

```

```

$corner(A, B) :-
  $token($corner(A, B)).

```

```

$corner(A, B) :-
  $continue_line(A, B),
  $is_corner(A, B),
  $assert($corner(A, B)).

```

```

$edge(A, A).

```

```

$edge(A, B) :-
  $token($connect(A, C)),
  $edge(C, B).

```

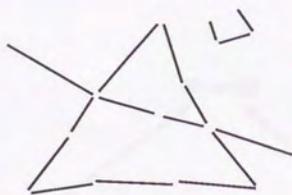
```

$edge(A, B) :-
  $continue_line(A, C),
  $straight(A, C),
  $assert($connect(A, C)),
  $edge(C, B).

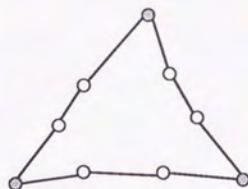
```

(b) トップダウンルール

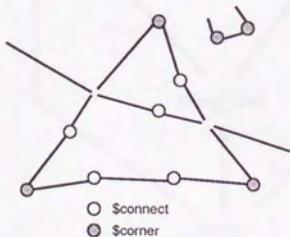
図 4.5: 三角形抽出のためのルール



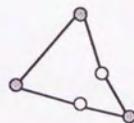
(a) トークンによる記述(一部分)



(c) トップダウンプロセスの結果(1)



(b) ボトムアッププロセスの結果



(d) トップダウンプロセスの結果(2)

図 4.6: 手書き図面からの三角形の抽出過程

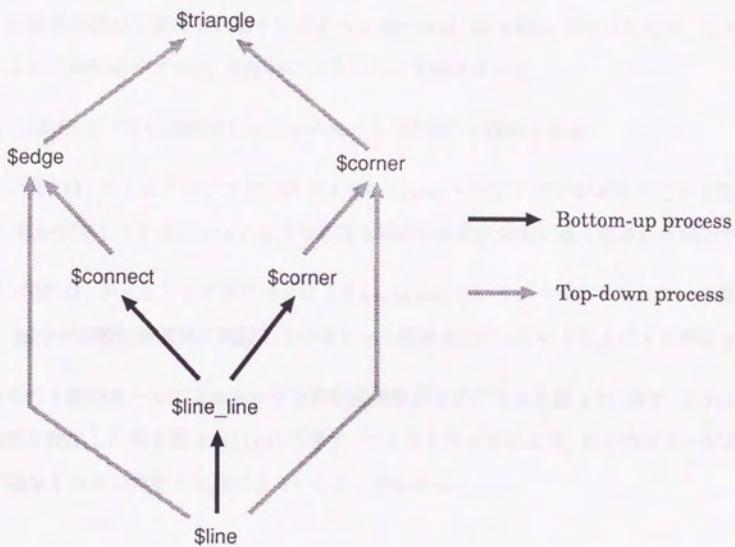


図 4.7: 三角形抽出のための状態遷移ダイアグラム

3. 隣合う線分が折れ曲がってつながっている場合 (すなわち、それらの線分のなす角度が2直角より十分に小さい場合)、角を表す新しいトークンが生成される。

このルールを図 4.6(a) に適用すると、その結果は図 4.6(b) のようになる。イメージの上できわめて状況が明白な場合についてのみ、\$connect や\$corner というラベルがふられる。直線が交差しているといったような複雑な状況は、ラベル付けされないのでおかれているが、これらはトップダウンプロセスで処理するために、そのまま出力される。

その結果を受けて実行されるトップダウンルールは、図 4.5(b) のようになる。このルールによる三角形のモデルは、具体的には以下のように表される。

1. 三角形は、“互いに接続した3つの角と3つの辺”で構成される。
2. “角”は、ボトムアッププロセスにより\$corner というラベルがふられている部分か、線が交差しているといったような複雑な場所から非決定的に取り出される部分である。
3. “辺”は、ボトムアッププロセスにより\$connect というラベルがふられている部分か、線分が比較的直線的に接続しているという部分をたどっていくことにより構成される。

これらの2組のルールによるトークンの状態遷移ダイアグラムを図 4.7に示す。これにより三角形を抽出した例を図 4.6(c),(d)に示す。バックトラックにより、もとのイメージから抽出可能な2つの三角形を認識できていることがわかる。

4.4.3 地図図面 (国土基本図) の認識

次に、ある程度実用的な図面理解システムの例として、図 4.1(b) に示した、国土地理院から出されている 1/25,000 国土基本図の図面理解システムを考える。これは、前述の AI-Mudams によって得られた記述を入力とし、その中のトークンの状況に従って、植生界、植生を表すシンボル、特定地区界、狭い道といった状況を認識することができるシステムである。これらの対象はいずれも、比較的空間的な広がりを持っており、局所的な状況調査と判断のみでは認識し難い対象である。すなわち、ある程度グローバルな状況判断が必要になっ


```

%
%      Bottom-up rules
%
'$medium' :-
    '$get_corners'(Cs),
    '$transform'('$medium'(Cs)).

'$medium'([P1, P2]) :-
    '$transform'('$linear_symbol'(P1, P2)).
'$medium'([P1, P2, P3]) :-
    '$transform'('$triangle_symbol'(P1, P2, P3)).

'$triangle_symbol'(P1, P2, P3) :-
    '$override'(P1, P2),
    '$override'(P1, P3),
    '$no_override'(P2, P3),
    '$transform'('$v_symbol'(P1, P2, P3)).
'$triangle_symbol'(P1, P2, P3) :-
    '$override'(P2, P1),
    '$override'(P2, P3),
    '$no_override'(P1, P3),
    '$transform'('$v_symbol'(P2, P1, P3)).

```

(a) ボトムアップルール

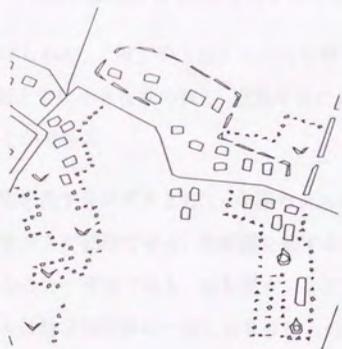
```

%
%      Top-down rules
%
'$sb' :-
    '$token'('$linear_symbol'(A, B, A1, B1)),
    '$no_token'(A, '$SB'(_, _)),
    '$sb'('$ls2'(A, A1), A, B, [A, B], M1),
    '$sb'('$ls2'(B, B1), B, A, [B], M2),
    '$sb_line'(M1),
    '$sb_line'(M2).
'$sb' :-
    '$token'('$v_symbol'(C, A, B, A1, B1)),
    '$no_token'(A, '$SB'(_, _)),
    '$sb'('$vs3'(C, A, A1), A, B, [A, B], M1),
    '$sb'('$vs3'(C, B, B1), B, A, [B], M2),
    '$sb_line'(M1),
    '$sb_line'(M2).

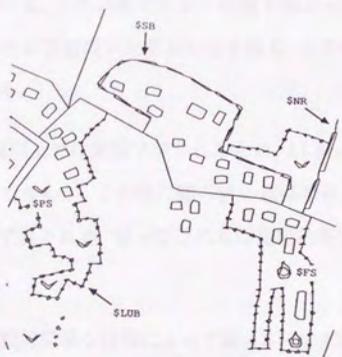
```

(b) トップダウンルール

図 4.9: 地図認識における状態遷移規則 (一部分)



(a) 特徴的な地図の一部



(b) システムによる認識の結果

図 4.10: 地図の認識の例

てくるが、これに対しては、本システムのトップダウンプロセスにおけるバックトラックの機構が有効に働くことになる。

認識する各対象について個別に説明すると、以下のようになる。

植生界 地面に対して付けられた、「畑」や「田」といった植生について、その境界を表す閉領域である。地図上では小さな点の列と、道路や後に述べる特定地区界も植生界の一部を構成することができる。

植生を表すシンボル 植生を表すシンボルとして、今回のシステムでは果樹園を表すシンボルと畑を表すシンボルとが認識できる。果樹園を表すシンボルは、その形状からボトムアップに決定することが可能である。畑を表すシンボルはv字型をしているが、このv字型のシンボルは特定地区界の一部ともなり得るので、トップダウンプロセスによらないと決定できない。

特定地区界 特定地区界は特定地区の境界を表すものである。地図上では比較的ストロークの短い破線で表される。これが直角に近い角度で曲がったりした場合に、畑のシンボルと区別のつかないv字型のシンボルとなり得る。この破線の他、道路も特定地区界となることができる。

狭い道 一般の道路は、地図上では実線で表されるため、AI-Mudamsによって線分トークンとして得ることができるが、この他に幅の狭い道路が存在し、地図上では比較的ストロークの長い破線で表される。従ってこれらは比較的長い棒シンボルの列として得られる。

これらの解析から、比較的簡単な作業によって図4.8に示す状態遷移図を得ることができる。システムに与える初期トークンとしては、AI-Mudamsにより比較的長い線のものはその中心部を表す「芯線」を、小さくまとまっているものはその輪郭線をそのまま、それぞれ線分トークンおよびシンボルトークンとして抽出し、これを用いた、状態遷移図で一番下にあるのが初期トークンの状態で、線分トークンに当たるのが\$lineという状態であり、シンボルトークンが\$loopである。上にあるのがゴールの状態であり、それぞれ\$LUBが植生界、

\$SB が特定地区界, \$NR が狭い道, \$PS が畑シンボル, \$FS が果樹園シンボルを表す。図 4.8 の状態遷移の様子に従って作成したルールの一部を図 4.9 に示す。全体では、ボトムアップルールとトップダウンルール合わせて 83 のルールを実装した。このシステムを図 4.1(b) に対して動作させた例の一部を図 4.10 に示す。これにより、システムが上で述べた構成要素を正しく認識していることがわかる(なお、地図上の孤立したシンボルを扱うこのシステムは、大沢らの AI-Mudams をベースにした道路や鉄道などの連続した成分を認識するシステム [37, 64] と補間し合って、国土基本図そのものの認識システムを形成している)。

4.4.4 システムの評価

状態遷移モデルに基づいた図面理解システムの例として、手書きの線図形の認識システムと、地図の認識システムを紹介した。システムに載せたルールの数としては、線図形の認識の方では 9 つ、地図認識システムの方では 83 のルールを要した。実際にルールを作成する時間としては、線図形の方では 1 時間程度で十分であり、地図の方では基本的な動作を行わせる部分の作成に数時間、これらを実行させて最終的に正しく動作するようにするために数日程度と、比較的短時間で実現することができた。

一般のプロダクションシステムと異なり、ルールを実行する認識カーネルにトップダウン推論の機構を持つため、トップダウン的な効果を実現するためのルールを用意する必要がなく、比較的容易にルールを作成できた。実際のルールの作成は、慣れれば認識プロセスの基本的な部分については数時間程度と、かなり短い時間で可能であると思われる。

動作時間については、地図の認識システムについて述べると、AI-Mudams の出力した結果から図 4.1(b) の全体について処理するのに、4.5 MIPS の Sun-3 ワークステーションでの実測時間で十数分程度かかった。これは、Prolog によって記述されているために認識カーネルの動作が遅いためと考えられるが、それでもそれほど非現実的な時間がかかるというわけではない。また、トークンの空間的探索は BD 木によってトークンの全数に対して対数時間で処理できるため、問題の大きさに対してもたかだか多項式時間で処理ができる。

4.5 考察

図面理解のための新しい機構として、状態遷移モデルを用いた実際の図面理解システムの例を紹介して、その評価を行った。本機構に従った図面理解システムでは、その本体である認識カーネルは、一定の範囲のさまざまな対象図面について共通に使用することができ、新たな対象についても、状態遷移ルールの再構築だけで認識システムを作成することができる。また、このルールの作成も比較的容易である。そのほかの画像データモデル獲得に対する要件についても、その多くを満足する図面理解システムとして実現できており、この状態遷移モデルが画像データモデル獲得のための一つの有力な手法であることが検証された。

現状の機構では、ここで述べたような図面や、フローチャートや論理回路図面などのように、局所的な状況判断である程度の解釈ができ、その結果を周囲の解釈との間で折り合いをつけていくことにより全体の認識が可能であるような対象には対応できる。しかし、複雑な機械設計図面など、構成要素の種類が多く、出現する形も不定型なイメージに対しては対応し切れない面がありうる。図 4.11 に機械設計図面の例を表す。(a) のような簡単な例であればこの認識システムでも処理することが確認されているが、(b) のような例では、グルーピングと高次のモデルマッチングとを併用するような、より大局的な状況判断が必要となつてこよう [56]。

第 5 章

状態遷移モデルに基づく一般画像の認識と その利用

5.1 システムの構成

この章では、状態遷移モデルに基づく一般画像の認識と、その利用について述べる。まず、状態遷移モデルの概要を説明し、次に、状態遷移モデルを用いた一般画像の認識の仕組みを説明する。最後に、状態遷移モデルを用いた一般画像の利用について述べる。

5.1 概要

状態遷移モデルを用いた画像理解システムにより作成した画像データモデルは、たとえその獲得段階で満足が行く抽象化が行えなかったとしても、それらの不完全に抽象化されたデータを不完全なりに利用することが可能である。本章ではこれを実証するため、状態遷移モデルを一般画像の認識に用い、その不完全な結果を利用する一手法について述べる。

一般画像は図面に比べて抽象度がさらに低く、曖昧さも大きいので、その認識処理はさらに困難なものとなる。そこで認識結果としては完全なものが得られることは期待せず、抽象度の低いものも混在した形で生成し、利用システム側でこれを吸収して、ユーザにはこの不完全な認識結果をなるべく意識させないようにすることを考える。

ここでは、多少認識する世界を限定するため、与えられる画像データとしてはスポーツシーンのカラー画像に限られるものとし、これを状態遷移モデルに基づく認識システムにより、完全に自動的な手法である程度抽象化したデータモデルをもってデータベース化する [50]。その利用システムとしては、このデータベースからスポーツシーンの種類で検索できるシステムを考える [71]。

この2つのシステムにより、コストが高く実現の困難な認識プロセスを現状の技術水準である程度まで行い、その不完全な処理結果を利用システム側が補間するという例を示している。本章では、この例を紹介し、状態遷移モデルにおける不完全な認識結果の利用について論じる。

5.2 システムの構成

図 5.1 に画像理解システムと画像検索システムとを含むシステムの全体構成を示す。画像理解システムがすなわち画像データモデル獲得にあたり、画像検索システムは画像データモデル利用にあたる。ここでのシステムではデータモデル獲得によって得られたデータモデルをそのままデータモデル利用側が参照するという形になっており、データモデル管理の役割は検索システムが兼ねていると考えられる。ここで、データモデル利用によりなされ

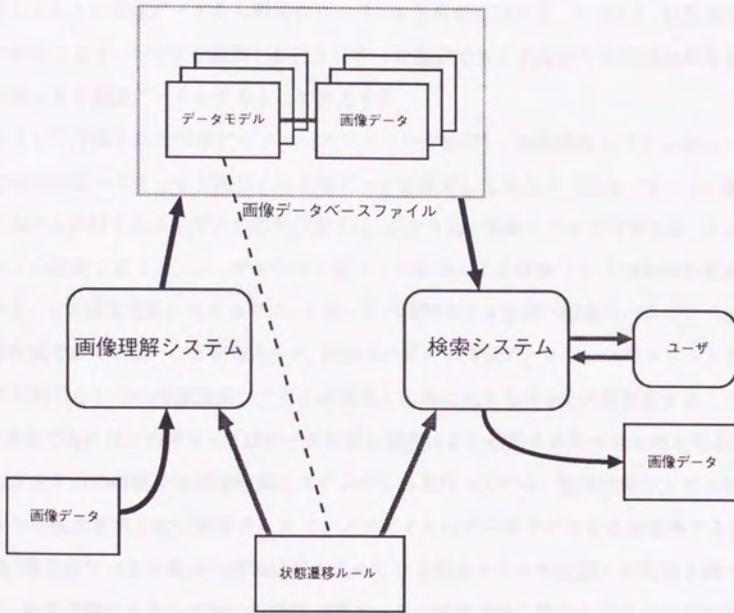


図 5.1: 画像理解システムと画像検索システムとを含むシステムの全体構成

るデータモデルに対する「演算」とは、画像データのキーワードによる検索だけに絞られることになる。

扱う対象画像は、画像理解システムにより処理されて、認識結果としてのデータモデルと画像データとの組により構成される画像データベースファイルへと変換される。画像理解システムは状態遷移モデルを利用した構成となっており、認識のための状態遷移ルールを参照してもとの画像データから画像データモデルを自動生成する。このとき、状態遷移モデルの利用により、不完全な認識しか行えなかった部分でも不完全なりの認識結果を表す中間状態のまま画像データモデルとして出力する。

こうして作成された画像データベースファイルを参照し、画像検索システムはユーザからの指示に従ってユーザが所望した画像データを検索してくるようになっていく。画像検索システムに対するユーザからの検索要求は、かなり高い抽象レベルで行われる。スポーツシーンの画像で言うと、ユーザからはスポーツの種類による検索という抽象的な要求がなされる。一方画像理解システムでは、スポーツの種類のような高い抽象レベルのデータモデルは生成できていないことが考えられ、作成されている画像データベースファイルとデータモデル利用としての画像検索システムの要求との間に大きなギャップが存在することになる。本来であればこのギャップはデータモデル管理により補間されるべきものであるが、今回のシステムの構成では画像検索システムがこれを行っている。画像検索システムはユーザからの検索要求に対し画像データベースファイルのデータモデルを直接参照するが、その際、得られている不完全な認識結果と求めている概念レベルの記述との関係を調べるために、画像理解システムが用いた状態遷移ルールを状態遷移の様子を表す状態遷移図として参照する。

このようにして、状態遷移モデルに基づく画像理解システムの処理結果の不完全な画像データモデルを、データモデル利用としての画像検索システムで不完全なりに利用することを実現する。

5.3 状態遷移モデルによる一般画像理解システム

5.3.1 初期トークンの生成

一般画像を状態遷移モデルで扱うためには、与えられたカラー画像をトークンというシンボルの記述に変換する必要がある。以下にその方法を示す。なお、入力となるカラー画像は 512×432 というサイズで、各ピクセルごとに RGB それぞれ 256 の階調で表されたフルカラー画像である。

1. 色の数を 16 に制限する。
2. 色の数が 16 に制限された画像を、16 種類のラベルが振られたクラスタリング結果と見て各クラスタの輪郭を抜き出す。
3. 輪郭を折れ線近似して線分に変換し、線分で囲まれた閉領域とその内部の色情報というシンボルとする。

色の数を 16 に制限する部分では、画像を構成する各画素の色情報の RGB 空間における分布状況に従い、もとの画像をなるべく崩さない形で 16 個の色を 3 次元の色空間に最適配置するアルゴリズムを用いている [21]。

このようにして、色情報が付加された、線分による閉領域というシンボルを生成する。これを状態遷移モデルのための初期トークンとする。このトークンの初期状態としては、 $\$chain(Cmap, R, G, B, H, I, S)$ というフォーマットで与えられる。Cmap は 0 から 15 までの 16 種類の色のラベルを表す。R, G, B は色情報であり、いわゆる R, G, B の値を 256 で正規化して表している。H, I, S も色情報であり、H は -180 度から 180 度までで正規化した色相、I は $768 (= 256 \times 3)$ で正規化した明度、S は 10000 で正規化した彩度を表す。

図 5.2 に処理結果の線分によるイメージの例を示す。

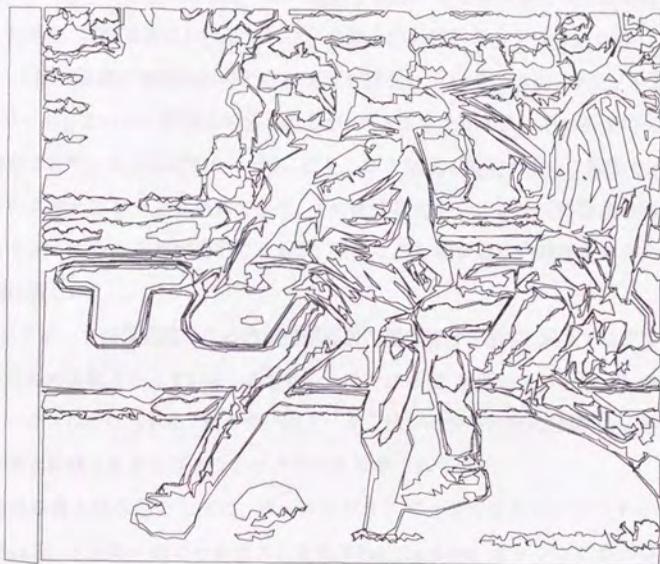


図 5.2: 初期トークンの生成の例

5.3.2 画像理解プロセス

このようにしてスポーツシーンの画像から生成された初期トークンに対し、状態遷移モデルに従い各トークンの状態遷移により画像理解プロセスを実現する。ここでは説明のため、図 5.2 のサッカーシーンの画像を例にとる。

認識システムとしては、もちろんサッカー画像専用の認識システムとってしまうわけにはいかないで、ある程度の範囲の一般の画像を処理できるようなものにならなければならない。この場合、状態遷移におけるトークンの最上位の状態としては「白い棒」、「黒いかたまり」、「緑の領域」程度のもとならざるを得ない。「(テニスの) ラケット」、「(サッカーの) ゴール」といった状態も考えられるが、そこまで上げるにはその画像のスポーツの種類が特定できている必要があり、一般にはそこまで状況を特定するのは困難である。人物はすべてのスポーツシーンに現れてきて、その服装などからスポーツの種類の特定のための大きなヒントとなることが期待できるが、残念ながら現状の状態遷移モデルの枠組ではその識別は難しい。

図 5.3 にスポーツ画像認識のための状態遷移図の基本部分を示す。現状ではボトムアッププロセスのみの実装となっている。まず各トークンは明度、彩度によって白、黒、および色のついたトークンに分けられる。色のついたトークンはついで各色別に分類され、さらに形状や大きさなどに従ってさまざまなトークンへと分類されていく。

図における最上位状態としては、サッカーやラグビーでの芝生のグラウンドに相当する `$green_field`、サッカーのゴールポストを表す `$white_post`、ボクシングやプロレスのリングロープを表す `$ring_rop` の 3 種類が示されている。人物については、服のしわなどによって体が複数の細かいトークンに分かれていたり、顔が背景の色に溶け込んでしまっているなどしているため、その同定はかなり困難であろうと思われる。

このように、現状の認識システムの認識結果は極めて抽象度が低く、単に各トークンの色と形状によって分類しただけに留まっている。このようにして作成された不完全な認識結果は、利用システムとしての検索システムから参照できるように、例えばサッカーという状況を特定するにいたる途中状態として、各トークンの内部状態そのままの形で出力される。

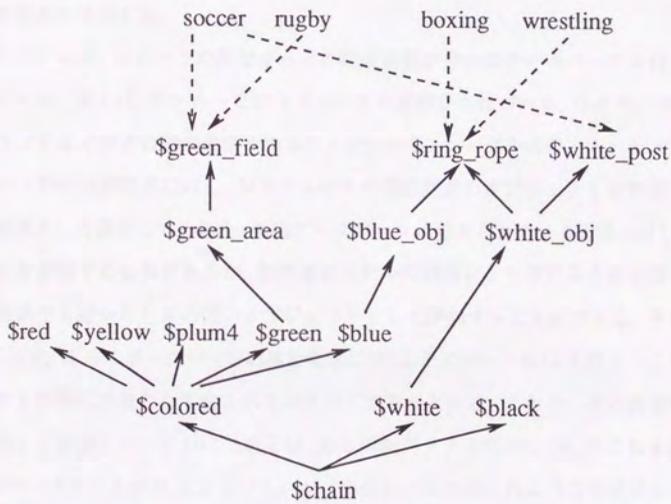


図 5.3: スポーツ画像認識のための状態遷移図

5.4 スポーツ画像検索システム

検索システムには、スポーツの種類という極めて抽象化された形で検索要求がなされる。これに対し、認識システムから渡された画像データベースとしての記述は、スポーツの種類のような抽象度には全く及ばない抽象度の低い不完全なデータである。ここで述べるシステムは、このような不完全な認識結果を用いながらユーザの要求を何とかある程度まで満足するため、検索結果の確からしさを多少落すことにより得られている認識結果を最大限生かした検索を実現する。

検索システムは、スポーツの種類からその構成要素が分かるデータベースを持った構成となっている。例えば、サッカーというスポーツの種類からはゴール、サッカーボール、芝生のグラウンドなどがその構成要素であることが分かる。ユーザから与えられたスポーツの種類という形の検索要求に対し、システムはその構成要素のオブジェクトが存在する画像を検索結果として提示してくれる。画像データベースに蓄えられている画像に対し、構成要素の存在を評価する必要があるが、状態遷移モデルの利用により不完全ながら得られている認識結果でも確からしさを低いオブジェクトとして評価することができる。その様子を図 5.4 に示す。あるスポーツシーンの検索要求に対し、そのシーンには A, B という種類のオブジェクトが構成要素として含まれているはずであるとする。これを一連の画像データモデルに対して評価していくわけであるが、ある画像データモデルについてこれを評価した時、そのデータモデルには A, B といういわば概念レベルに達したようなオブジェクトが生成されていなかったとする。ここで、状態遷移図の利用により状態間の依存関係が明示的に得られるので、この画像には A, B というオブジェクトになり得る状態のトークンが存在することがわかる。こうしてこの画像は比較的高い確からしさでこのスポーツシーンを記述している画像であると評価されることになる。

数十程度の画像を持つ実験的な画像データベースを作成し、「サッカー」というキーワードで検索を行った結果を図 5.5 に示す。正しいサッカーの画像も検索されているが、不正確なデータとしてラグビーの画像が一枚検索されてしまっているのが分かる。

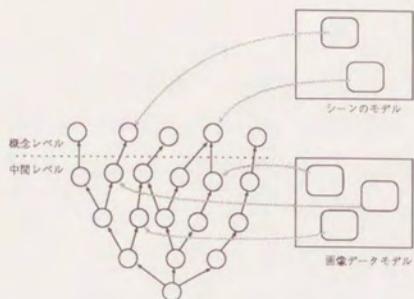


図 5.4: 状態遷移モデルによる不完全な認識結果を利用した検索



図 5.5: 画像検索の例

5.5 考察

状態遷移モデルに基づく画像理解システムによって生成された不完全な認識結果を利用し、不完全ながらある程度の利用に供するシステムの例を示した。状態遷移モデルによる認識結果は、たとえそれが概念レベルに到っていない不完全な結果であっても、状態遷移図という形で各中間状態と概念レベルとの対応が明示的に与えられるため、これを参照することによりこのような不完全な結果を利用することができることになる。

このデータモデルの利用として、ここではこれを抽象化して表した「演算」としては画像データのキーワードによる検索のみに留まっていたが、他のデータモデルへの操作も実現できるような「演算」についても検討する必要がある。また、今回の検索システムでは不確かな認識結果でも検索できるようになっているが、「どの程度」不確かなものを検索するかについては対応していない。これに対応するため、データモデルの信頼度を扱うという点も考察の余地があろう。

また、ここで紹介した画像理解システムはあまりに不完全であり、その認識結果をより高次の応用に供するのには無理がある。少なくとも、スポーツシーンの認識システムとしては、人物を識別することができ、その服装を大体認識できるようにする必要があると考えられる。