実世界画像に対する画像認識の研究

柳　井　啓　司

目次

第 1 章

序論

1.1 はじめに

視覚は，人間が実世界から情報を獲得する手段として，最も大きな役割を果たしている．例えば，歩いている時に目の前の障害物を認識したり，人と会ったときにその人が誰であるか識別したり，本に書かれている文字を読み取ったりなど，様々な場面で視覚による認識が行われる．視覚は人間が日常生活を送る上でなくてはならないものとなっている．実世界を視覚によって認識する時，人間は必要に様々な応じて様々なレベルで「認識」を行っている．歩いている時に障害物を認識する時は，それを避けるために障害物のおおよその位置と形を認識する．また，人の認識の場合は，ただ「人」がいると認識すればよい場合と,「佐藤さん」「鈴木さん」というように人物の同定まで必要になる場合とが考えられる．障害物の認識の場合と違って，人の存在の認識の場合は，予め一般的な「人」の視覚的特徴，人物の同定の場合は「佐藤さん」「鈴木さん」という個人の視覚的特徴をそれぞれ予め知識として持っている必要がある．一方，文字の認識の場合は，文字という記号を識別することが目的となるが，認識するためには文字に関する知識を持っていることが必要不可欠である．この様に「視覚による認識」には，対象個別の知識が不要で実世界の空間構造だけが分かればよい場合,「人」や「文字」という一般的な対象に関する知識が必要な場合，個人という特定の対象に関する知識が必要な場合の 3 つのレベルが存在する．これら 3 種類の認識は，それぞれ，3 次元の復元 (reconstruction)，対象の識別 (classification) もしくは分類 (categorization)，対象の同定 (identification) と呼ばれる．

本研究においては，その中でも特に，人を「人」として認識する場合の様に，画像中の物体を一般名称のレベルで認識することに注目する．一般名称とは,「机」「椅子」「自動車」「空」などの人間の認識にとっての基本的な物体の概念を表す名称である．人間はこうした認識を普段から無意識のうちに行っているが，同じことをコンピュータ上で実現することは容易なことではない．なぜなら，実世界に存在する物体の種類がとても多く，さらに，同一の名称を持つ物体物体についてもとても多くの視覚的な多様性が存在するために，個々の物体に関する知識の表現や獲得，視覚入力との照合の方法などに難しい問題を含んでいるためである．また，そうした問題以前の問題として，人間が創造した概念体系の一部である物体の一般名称と，実世界の物体を結び付けることは，そも

そも哲学的な問題である「認識論」に関係した問題を含んでおり,「椅子」「机」とは何であるかという明確な定義をコンピュータに与えることが困難であるという根本的な問題も存在している．こうした問題のために，3 次元情報の復元に関してはステレオカメラやレンジファインダの利用によって，かなり実用に近いレベルまで研究が進んできている一方，物体の一般名称レベルでの認識は実用とは程遠いレベルにあるのが現状である．例えば，我々にとって身近な物の代表である「椅子」「机」を認識することすら現状では難しい．けれども，現在飛躍的なハードウェアの進歩をとげているヒューマノイド型のロボットの視覚機能や，爆発的に増加するデジタル画像の意味内容による自動処理などの実現のためには，一般名称レベルでの画像認識の実現が不可欠であり，実用的なレベルでの認識の実現が大いに求められているといえる．

この様な一般名称レベルでの物体の認識という人間の視覚を人工的に実現しようとする試みは，コンピュータが現れる以前の昔から現在に至るまで，長い時間に渡って行なわれ続けている．そして，その成果として，視覚システム (vision system) が数多く実現された．ところが，それらのシステムの多くは，“人間らしさ” に欠けていて，いかにも “機械らしい” ものがほとんどであった．つまり，今までの画像認識システムは，ある特定の状況を想定していたり，必要な情報が完全に得られることを仮定しており，“人間らしさ” である一般性 (generality)，柔軟性 (flexibility) を真の意味で実現しているものは存在しなかった．これに対して，人間の視覚は実世界を対象としており，認識の対象や場面を限定することを必要としない一般性 (generality)，そして必要な情報が不完全であったり，曖昧にしか得られない状況下でも，うまく認識できる柔軟性 (flexibility) を兼ね備えている．コンピュータ上での視覚システムが少しでも人間の視覚に近付くためには，これらの 2 つの一般性と柔軟性の性質を高いレベルで実現することが必要不可欠であるといえる．

このような一般性と柔軟性を持つ視覚システムの実現は画像認識の研究が始まった当初より目標とされていたものの，実世界を対象とする人間の様な多目的の視覚システム (general-purpose vision system) の実現を最初から目指すことは極めて困難であったために，当初は世界を限定して研究が行なわれた．そして，最初に一定の成功を見たものは,『積木の世界』を対象とするものであった．その後も現在に至るまで，多目的な視覚システムを目指して研究は行なわれているのものの，初めから何らかの条件を仮定していたり，結果的にある条件下でしかうまくいかないという研究がほとんどであり，人間の視覚に匹敵する一般性と柔軟性を実現しているシステムは皆無であるといえる．

こうした一般性と柔軟性を目指した視覚システムの研究の一方で，初めから目的が限定されているために一般性は必要なく，その目的の範囲内での柔軟性だけ実現されていればよいという視覚システムの研究も存在する．例えば，顔画像の検出システムや，文字認識システムなどである．このような特定の用途向けの視覚システムなら，現在でも実用化され，広く普及している視覚システムが存在する．しかし，本研究では人間の視覚による情報獲得機構をコンピュータ上で実現するという科学的な興味を追求することを前提としたいと考えるので，ここでは一般性と柔軟性を視覚認識システムに与えることに興味がある．我々の目標は，特定の状況にのみ 99%の認識率を持つシステムよりも，多様な状況に対して平均的に 70〜80%程度は認識できるシステムの実現である．

このような多目的の認識システムの研究は大きく分けて，モデル表現と認識手法の研究と，システムの構成法の研究の 2 つに分けることができる．モデル表現と認識手法の研究では，同一種類で

も様々な形状や見え方を持ったり，他の物体によって一部が隠されてしまうこともあるような対象をいかに正確に認識するかということが大きな問題になっている．一方，システム構成の研究では，いかに多くの認識手法を統合して，それらを柔軟に組み合わせて認識を実現するかということが課題である．これらの２つの研究は，どちらも多目的認識システムを作る上で不可欠なもので，両方を並行して研究していく必要があり，両方の研究の成果を高いレベルで統合することが一般性と柔軟性を持ったシステムの実現の必要条件であると考える．

1.2 本研究の目的

本研究は，実世界画像中の様々な物体を一般名称で認識する汎用画像認識システム (general-purpose image recognition system) の実現を目的とする．本研究における「物体」とは，画像が表しているシーン中に含まれる，人間にとって何らかの意味を持った概念に対応する物理的実体のことである．画像中に含まれる物体の認識を行なうことにより，その画像が何を表現しているかおおよそ知ることができる．勿論，画像中のすべての物体を認識したからといって，画像の伝えている情報をすべて認識したことにはならないが，物体が認識ができれば，画像の伝える情報のかなりの部分を知ることができる．したがって，物体の認識は認識の基本ともいえるものであり，「**物体認識 (object recognition)**」と呼ばれている．

具体的には本研究では，実世界画像の持つ２つの困難な問題 (1) 多数の種類の物体が存在しており，物体の種類によって適する認識方法，モデル表現が異なる．(2) 単一種類の物体でも様々な個体が存在し，さらに，画像中での見え方も多様である． に対して，それぞれシステム構成法と認識方法の観点から研究を行った．(1) の問題に対しては，従来の画像認識の研究で多数提案されている特定種類の物体に対する認識手法と知識表現を統合して対処することとし，そのために，多数の認識手法と知識表現を統合するための認識システム構築法を提案する．(2) の問題については，単一種類の物体の画像中での様々な現れ方に対応するために，多数の学習画像を WWW (World-Wide Web) から自動収集し，自動的に画像認識のための知識ベースを構築する方法を提案する．

本研究においては，実世界に対する「認識」を取り扱うが，一般に「認識」と言った場合に，何をどうすれば「認識」できたとみなすかという問題が存在する．これは重要な問題であり，予め定義しておく必要がある．物体認識には，初めに厳密なモデルを定義して，画像中の要素がどのモデルに該当するか調べることが「認識」である場合と，初めに対象を表す概念があって，それをモデル化して，画像中の要素がどの概念が表している対象に相当するのかを調べるのが「認識」であるとする場合の２種類があるが，本研究においては後者の認識を目的とする．より具体的に言えば，本研究では，対象を表す概念として，我々が普段用いている物の一般名称を用いることにする．つまり，一般名称が表す対象をモデル化し，そのモデルによって画像中から物体を探し出し，その一般名称を答えることができれば，「認識」ができたとみなすことにする．

本研究での対象を表す概念としての一般名称とは，例えば，「机」「椅子」などのような，人がぱっと見た時にすぐに思いついたり，幼児が最初に覚えるような，**基本認識レベル (basic-level category)**[1] の名詞のことである．例えば，「乗用車」を認識しようと考えた場合，「乗用車」は「飛

行機」「電車」も含んだ概念である「乗り物」でもある一方，さらに対象を特定した名称である「セダン」や「トヨタ ヴィッツ」でもあるかも知れないというように，乗用車という物理的実体を表すには多くの名称が存在して,「乗用車」という名称はその中の一名称でしかないことが分かる．一般に，対象を表す名称 (もしくは概念) は，単に「乗り物」と言うような広い範囲を含む一般的な名称から，対象を限定した固有名詞の「トヨタ ヴィッツ」，さらには「山田さんの所有するメタリックブルーのトヨタ ヴィッツ」というようにある単一の個体を指す名称まで広く存在し，それらが階層構造を成している．この階層構造の上下関係は instance-of 関係であるが，これ以外にも，物体の構成要素の全体部分関係 (part-of 関係) や，物体の素材の関係 (made-of 関係) を考えることが可能で,「乗用車」は「タイヤ」でも「車体」でも「窓」でもあるとも言え，また,「乗用車」は「鉄板」や「ゴム」「ガラス」などであるとも言える．こうした概念の階層構造の中において，E. Rosch ら [1] は，どのレベルでの認識が人間にとって最も基本的な認識であるかを心理実験によって明らかにしている．例えば,「動物」「秋田犬」よりも「犬」の方が人間に認識にとっては基本的な名称で，普通は人間が見たときに最初に思いつくのは「犬」という名称である．この人間の認識にとって基本的な概念レベルを基本認識レベル (basic-level category) と呼ぶ．E. Rosch らは，基本認識レベルでは同一名称の対象は多くの共通の性質を持っていて，特に (a) 形状の類似性．(b) 運動，動作，操作の類似性を持っている．ということを述べている．(a) の形状の類似性は,「動物」のレベルではいろいろな形の種類が存在するの対して,「魚」「鳥」「犬」などの基本認識レベルでは，それぞれの分類毎に分類内で共通の形状，大きさが存在している．そのため，基本認識レベルでは，その分類内での平均的な形状 (prototype shape) を用意することで認識が可能となると述べている．この性質は，静止画像からの物体認識を行う上では，好ましい性質である．(b) の運動，動作，操作の類似性は,「動物」では種類によって，泳ぐ，飛ぶ，歩くといろいろな運動を行う動物が存在するが,「魚」「鳥」「犬」はそれぞれ固有の運動を持っている．これは，動画像から物体の認識を行う場合に有効な性質である．このような理由から，本研究では，主に基本認識レベルの名称を対象を表す概念として用いることとする．もちろん，基本認識レベルには例外があったり，個人の文化的背景によって異なることがあるが，多くの場合は適用できる考え方である．

以上をまとめると，本研究における認識システムとは,『実世界画像中に含まれる物体の存在と基本認識レベルでのその一般名称を認識するシステム』であると定義する．我々の研究の目的は，このようなシステムおいて，一般性と柔軟性を少しでも向上させるための新しい提案を行なうことである．なお,「実世界画像」とは，我々が生活している 3 次元空間をカメラで撮影した静止画像のことで，特に断りがなければ，通常は人間が目にするシーンを撮影したものである．つまり，特に制約はなく (unconstrained)，人間にとって一般的な画像を認識の対象とする．

1.3 本研究の位置付け

本研究は，大きく分けて 2 つの研究から成っている．1 つ目はマルチエージェントによる画像認識システムの研究で，2 つ目は WWW (World-Wide Web) からの収集画像を学習画像とする画像認識システムの研究である．

1つ目の研究は，マルチエージェントによる画像認識システムに関する研究である．この研究では，初めに，画像認識システムを特定の一種類の物体のみを認識するエージェントの集合体として構築するためのシステム構成法 MORE(Multi-agent Object REcognition Architecture) の提案を行った [2, 3]．MORE では，中央制御機構に相当するものはなく，それぞれのエージェントは独自にそれぞれの対象に適した手法で画像を解析し，自分の担当する対象のみを認識する．そして，その認識結果を通信によって交換し合い，対象間の関係の知識を主に用いることによって協調を行ない，全体の整合性が保たれるような結果のみが最終的な結果となる．このようなマルチエージェントによるシステムは，アルゴリズムの統合，柔軟な制御構造の実現，システムの拡張などに優れている特徴がある．従来のシステムでは，屋外画像なら屋外画像のみ，航空写真なら航空写真のみの様に，単一種類の画像しか同時に認識対象とすることができなかった．これは，システムの構成が画像の種類に依存したものになっていたり，認識のための知識が相互に密接な関係を持っていたためである．それに対して，MORE に基づくシステムでは，各エージェントが，自分の認識対象とする物体の認識のための物体固有の知識の表現方法および認識手法を完全に独自に定めることができる．統一した表現を用いるのは，エージェント間でのそれぞれの認識結果の調整を行なうのに必要な物体候補の表現法と，物体相互の間の関係知識だけである．そのために，システム構成としては画像の種類とは完全に独立しているので，屋外と屋内の様なまったく異なる複数認識領域に対する認識が同時に可能となっている．

また，認識手法の面からは,「画像中の領域のみに基づくのではなく，物体の本来の形状を推測することによって認識を行なう手法」を提案した．従来の多目的な物体認識システムにおいては，領域分割による画像からの領域の切り出しを認識の手法の中心として来たが，この方法であると，物体の一部が他の物体によって隠されているような状況では領域分割自体が難しいことがある．つまり，手前にある物体と後ろに隠れている物体の画像上での境界は，手前の物体が作り出しているものなので，後ろの物体の領域の切り出しを行なう正確に行なうには，両方の物体に対する知識が必要になる．ところが，1つのエージェントは自分の担当する物体に関する知識しか持っていないので，後ろにある物体の場合は正確な領域の切り出しが不可能である．そこで，他の物体によって隠されている部分を無視して，物体の本来の形状を推測することによって認識を行ない，隠れている部分も含めてその物体の占める領域とする．しかし，こうすると，手前にある物体が認識された時に矛盾が起こる．その矛盾は予め与えてある関係知識を利用することによって解決する．例えば，もし，机の上に本が認識されて，机の領域を競合を起こしても，机と本の間に「本は机の上に存在しうる」という関係知識さえあれば，本が机の上にあるとみなされて，両者は両立する．これは，従来のシステムになかった特徴であるといえる．

さらに，提案したマルチエージェント物体認識構成法 MORE に基づいて，物体が物体の上に載っているという“支持関係”を考慮した，室内画像認識システムを構築した [4, 5, 6]．支持関係を考慮することによって，複雑なオクルージョンを含む室内シーンの画像に対する認識が可能となった．従来のシステムでは，物体が十分に画像中に表れていないと認識ができず，室内シーンのような複雑なオクルージョンを含む画像に対して対処できなかった．それに対して，我々の提案するシステムでは，物体が物体の上に載っているという関係である物体間の支持関係を定性的に推論すること

によって，他の物体によって隠されている物体の認識を可能としている．具体的には，最初に画像中に明確に表れている対象に対して 3 次元構造モデルを当てはめることによって物体の 3 次元構造を推定し，次に推定された物体の 3 次元構造を利用して，物体間の支持関係をチェックすることによって，部分的にしか見えていない物体の存在を推定したり，実在しない物体の候補を消去し，最終的に全体として整合性のとれた認識結果を得ることができる．

次に，画像認識システムが物体を認識可能であるためには，その物体がある一定以上の大きさを持って画像中に現れている必要があるという問題を克服するために，近年 CCD 技術の発展によって容易に得ることが可能となった 100 万画素以上の高解像度画像を入力画像とすることのできる多重解像度に対応したマルチエージェント物体認識システムを構築した [7, 8]．高解像度画像を認識に用いる場合，10〜30 万画素程度の画像を想定した従来のシステムにそのまま入力すると計算時間や必要記憶容量の著しい増大という問題が発生する．そのため，従来のシステムでは，元々高解像度の画像であっても 10〜30 万画素程度に縮小して，システムに入力することが多かった．しかし，単純に縮小して入力することは，高解像度画像からしか得られない詳細なシーン情報までを捨ててしまっていることになる．そこで，本研究では，我々が提案しているマルチエージェント画像理解システム構成法 MORE を用いて，エージェントの協調作用によって，画像中の認識すべき部分を選び出し，予め数段階に縮小された画像から適切な解像度の画像を対象に応じて選択する機構を実現した．そして，その結果，処理時間をあまり増大させることなく，効率的に高解像度の画像を認識に利用することを実現した．

2 つ目の研究では，WWW から多数の画像を自動収集し，それを学習画像として，画像認識を行うシステムを実現した．人間は生活の中で,「物」に関する知識を逐次的に学習している．しかし，計算機システム上のソフトウェアによって実現された認識システムには実世界中を自由に移動できる様な「身体」がないので，人間のように実世界での生活を通して，自ら新しい「物」を見て学習することは出来ない．もちろん，ロボットならそれが可能であるが，現時点では実世界で人間と同様に生活できるロボットは存在しない．しかし，計算機システムには，World-Wide Web というコンピュータネットワーク世界が存在する．この世界には，実世界に関する情報が大量に存在していて，計算機システムはネットワークを通して，自由にその中をアクセスして，情報を獲得することが出来る．これは，まさに，計算機システムが WWW という擬似的な実世界である計算機ネットワーク世界の中で「生活」することができると言うことができる．そこで，人間が実世界での生活を通して学習していく「物」に関する知識を，計算機システムが WWW を利用して獲得し学習するということを試みる．これは，実世界画像に関する知識の獲得を WWW から自動的に行うという新しい試みであり，従来の実世界画像認識の研究における知識獲得の困難性の問題点を解決するための新しい提案であると言える．

そのための研究として，初めに WWW からの大量の画像収集方法を提案した [9, 10, 11]．この方法では，WWW 上の複数のテキストサーチエンジンを利用して，収集したい画像のキーワードを含む Web ページを収集して，さらにその中に含まれる画像を収集する．そして，収集した画像をクラスタリングすることによって，互いに類似している画像のみを選び出し，それをキーワードが表す画像として出力する．従来の方法では，初めに画像をキーワード情報で WWW から収集し，そ

の後，システムの検索結果の提示に対してユーザが選択を行い対話的に望ましい画像を徐々に絞り込んでいくというインタラクティブな方法によって，収集を行っていた．そのような方法は収集枚数が少なくてよい場合には有効であるが，大量に収集するには不向きであり，困難であった．本研究では，画像をクラスタリング結果の大クラスタの選択によって自動的に絞り込み，処理途中でのユーザの介入を不要としたために，大量の画像の自動収集が可能となっている．実験により，キーワード入力のみで6～8割程度の適合率で，数百枚単位で画像がWWWから自動的に収集出来ることが示された．

次に，WWWから収集した画像を学習画像として，類似画像検索に基づく画像自動分類を試みた[12, 13, 14]．画像特徴を用いた類似画像検索は，画像内容に基づく画像検索 (Content-Based Image Retrieval, CBIR) として，様々な方法が提案されており，ここでは，カラーヒストグラムによる方法[15], カラー情報に基づく Earth Mover's Distance(EMD)[16], 領域分割を用いた Integrated Region Matching(IRM)[17] による方法の3つを用いた．その結果，EMD と IRM による方法がよい結果を納め，20クラスの場合に約50%の精度で，各クラス50枚ずつの一般実画像のテスト画像集合に対する自動分類を行うことができた．これは単語入力のみで，画像に関する知識をまったく与えることなく，画像分類が可能になったということであり，従来のWWW画像検索から，WWWからの画像知識の獲得という一段階進んだ新しい研究を提案している．また，本実験では，高速化の方法として，高速インデクシング手法である SR-Tree[18] を EMD や IRM による類似度計算に適用し画像分類の高速化を図った．

最後に，本研究における2つの研究の関係をまとめる．1つ目のマルチエージェントによる画像認識システムの研究では，多数の認識モジュールの統合の手法を提案しているが，認識モジュール自体は人手で構築する必要があるので,「机」「椅子」などの簡単な形状の人工物以外に対応した認識モジュールを構築するのが困難であった．そこで，2つ目の研究では，形状の複雑な物体の認識を学習によって実現するために，WWWから多数の多様な画像を収集し，それを学習画像として，実世界画像の認識を行うことを試みる．これら2つの研究は，相補的な関係にあり，将来的にはこれらの研究を融合させることによって，真に実世界に対応して認識システムの実現が可能となると考える．

1.4 本論文の構成

本論文は，全部で10章からなる．

第1章では，研究の背景，目的，位置付けについて述べる．

第2章では，実世界画像に対する従来の研究についてまとめる．研究の流れを大きく分けると，システム構築に関する研究と，認識手法に関する研究の2つがあることを示す．

前半部の第3章から第6章では，実世界画像に対応したシステムを構築するために，異なる多数の認識手法や知識表現の統合をマルチエージェントの考え方に基づいて実現する方法について提案する．

第3章では，従来の実世界画像に対する認識システムの構成法についてまとめる．従来の認識システムでは，例えば，屋外画像のみ，航空写真のみ，という様に対象を予め想定してシステム構築が行われてきた．そのため，システムの構成が認識対象の画像の種類に依存したものになっていたり，各対象物の認識のための知識が相互に密接な関係を持っていために，異なる種類の画像に対する知識を混在させることが困難で，様々な種類の画像が存在する実世界画像の認識には適用が難しいという問題点があったことを指摘する．

第4章では，マルチエージェントによる画像認識システムの構築法を提案する．多様な認識対象に対応するために，本研究ではマルチエージェントによってシステムを構成する．各エージェントは1種類の物体のみを認識する独立した認識システムであり，物体毎にそれぞれに異なる知識の表現および認識手法を用いることができるため，システム構築の自由度が高い．システムの全体の最終的な認識結果は，エージェント間の相互作用によって求める．実験により，提案手法によって室内画像と屋外画像の両方に対応できる認識システムを構築出来ることを示す．

第5章では，マルチエージェントによる画像認識システムに物体間の定性的な位置関係に関する推論機構を導入することを提案し，より複雑な画像の認識を可能とする．実世界画像においては，物体が物体の上に載ったり，手前に位置したりして，物体が物体を隠すオクルージョンが発生する．オクルージョンのために一部分しか画像中に現れていない物体を認識可能とするためには，物体間の位置関係を利用することが不可欠であるが，従来は主に画像上での物体領域同士の2次元的な位置関係しか利用されていなかった．ここでは，物体の定性的な3次元情報を利用して，定性的な3次元位置関係の推論を行うことにより，実世界画像で問題となるオクルージョンに対処する方法を提案する．実験では，室内画像に対するシステムを実現し，その効果を示す．

第6章では，画像中に小さくしか現われていない物体の認識を高解像度画像を利用することによって認識可能とする方法を提案する．単純に高解像度画像を用いることは，認識時間の著しい増大を招くが，ここでは，多重解像度解析を導入することによって，効率的な認識を実現することを提案し，より複雑な実世界画像が認識可能となることを実験にて示す．

後半部の第7章から第9章では，画像内容を表すテキスト情報を伴った多種多様な画像をWWWから自動収集することによって，画像認識のための知識ベースを自動構築し，同一種類でも多様な個体が含まれる実世界画像を認識可能なシステムを実現する方法について提案する．第6章までででは，認識対象のモデルを人手で与えていたために，それぞれの対象毎に適切な認識方法およびモデル表現を採用することが出来たが，その一方で「机」「椅子」などの簡単な形状の人工物以外に対応した認識モジュールを構築するのは困難であるという問題点があった．そこで，第7章以降では，学習による認識システム構築を試みる．

第7章では，多数の学習画像を用いた実世界画像の認識について従来の研究をまとめる．そして，従来の研究では，学習画像を収集することが困難であったために，顔画像や自動車の画像などの限定された対象にしか実験が行われていなかったという問題点を指摘する．

第8章では，実世界画像を大量にしかも手軽に収集する方法として，WWWから自動的に大量の

実世界画像を収集する方法について提案する．WWW 空間中に存在する画像は現在数億枚と言われ，様々な画像が存在している．WWW 空間中に存在する画像はその多くが画像内容を表すテキスト情報を伴っているので，テキスト情報を解析することによって，ユーザの望むあらゆる画像をWWW から収集することが可能である．

第 9章では，提案した画像収集法を用いて様々な実世界画像を自動収集し，それらを学習画像として，実世界画像に対する認識を行うことを提案する．最初に学習の段階として，WWW から認識したい対象，例えば,「ライオン」「りんご」などの画像を各種類 (クラス) 毎に数百枚から数千枚程度収集する．そして，それらから色情報，テクスチャ情報などを画像特徴として抽出し，各クラス毎に画像特徴に関する知識ベースを構築する．次に，認識の段階では，認識対象の画像から同様に画像特徴を抽出し，知識ベースと照合を行い，最も可能性の高いクラスに分類し，認識を行う．実験では，この提案手法により，単語入力のみで画像に関する知識をまったく与えることなく，画像分類が可能になることを示す．

第 10章は，本論文の内容をまとめ，今後の実画像認識の研究についての課題，展望を述べる．

なお，本論文では,「画像認識 (image recognition)」 と「物体認識 (object recgnition)」という用語を併用するが，物体認識を画像に対するものに限定した場合，前者は後者のより広い概念であり，本研究で行う様な物体の存在を認識する物体認識のみでなく，画像の 3 次元的な構造の認識や，物体の動きの認識なども含まれる．以後は，画像認識全般に共通するような事については画像認識という用語を用い，物体の認識に限定される時は物体認識という用語を用いることとする．

第 2 章

実世界画像の物体認識

2.1 はじめに

物体認識では，3次元である実世界のシーンをカメラによって撮影することにより生成された2次元画像を対象に認識を行なう．通常，認識は画像中の対象とシステムに予め与えられている対象モデルの照合を行うことで行なわれ，画像中の対象がどのモデルに対応するかラベル付けを行なうことで「認識」できたと見なす．一般に物体認識の入力画像は静止画に限定される訳ではないが，本研究では静止画を対象としており，本論文では静止画の認識についてのみ触れる．

本章では，初めに簡単に物体認識の一般的な処理手順について述べ，次に物体認識には classification と identification の2種類があることを説明する．さらに，従来の物体認識の研究の流れを簡単に説明し，その後，主に classification のための物体認識の手法について簡単に説明する．

2.2 物体認識の一般的手順

一般的な物体認識において対象とする画像は，単一静止画，複数静止画像，ステレオ静止画像，動画，距離画像などのあらゆる画像であり，それらは，数値データの集合体として計算機中で表現される．画像認識によって最終的に得られる結果は，画像の表しているシーンの記述，例えば存在する物体の名前と位置の記述などで，記号による画像の説明である．

一般に物体認識の処理は，特徴抽出 (feature extraction)，特徴のグループ化 (grouping) もしくは画像の分割 (segmentation)，モデル照合 (model-matching) の3段階で行われる (図 2.1)．これらの3段階は，それぞれ低レベル処理 (low-level processing)，中レベル処理 (middle-level processing)，高レベル処理 (high-level processing) と呼ぶこともある．第一段階の特徴抽出では，画像に対して，例えば，エッジ抽出，領域抽出，特徴点抽出などを行ない，モデル照合で必要な特徴を抽出する．画像は数値データの集合体で，多くの数値データによって一つの画像が表現されている．例えば，横 320 画素，縦 240 画素の大きさの画像の場合では，76800 個もの点によって画像が表現される．そのため，画像に対して直接モデル照合を行うことはせずに，通常はその前に特徴抽出によって，モデル照合に必要な情報のみを取出しておく．なお，特徴抽出では，実際の特徴抽出処理を行う前

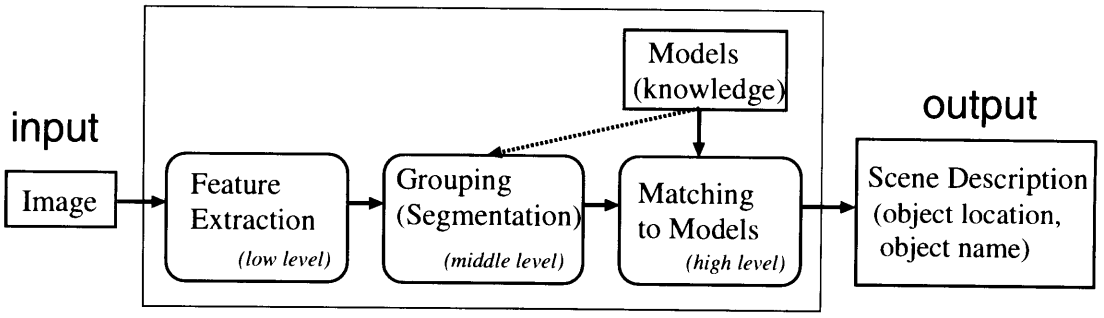

図 2.1 一般的な物体認識の流れ

に，画像の平滑化や色の減色などの前処理が行われることが普通である．また，このような第一段階での解析を行なわず，直接画像に対して 2 次元画像照合を行なう手法も提案されている [19].

第二段階の特徴のグループ化では，抽出した特徴を第三段階でモデル照合しやすいように一まとまりにグループ化する．グループ化の際には，図中で点線矢印で示すように認識対象に関する知識が用いられる場合もある．グループ化が画素単位で行われる場合は，領域分割 (segmentation) と呼ばれる．また，特徴のグループ化を行わずに，入力画像から任意の大きさの矩形をスライドさせることによって，機械的にブロック領域を多数切り出して，それらすべてについてモデル照合を行うという方法もある．

次の段階のモデル照合では，第二段階でグループ化した特徴集合と予めシステムに蓄えられているモデルとの照合することにより，画像中の対象がどのモデルに対応するかラベル付けを行なう．モデルは，抽出する特徴量との照合が行い易い形式で表現され，予め認識システムに蓄えられている．こうして，入力画像に対して，記号であるラベルを付けることによって，認識は行われる．つまり，数値データから記号的な情報を抽出するのが「物体認識」であるといえる．

以上のことから,「物体認識」の手法は，特徴抽出，グループ化もしくは領域分割，モデル表現，モデル照合の 4 つが研究課題となっていることがいえる．これらの 4 つのうち，最初の 2 つは必ずしも認識対象に依存しないので，それぞれ独立して単体のアルゴリズムとして研究が行われることが多い．一方，残りの 2 つは認識対象に依存する場合のが通常なので，画像認識システムの研究の中に組み込まれて研究されることが多い．

2.3 分類と同定

実世界画像に対する物体認識には大きく分けて 2 種類の認識がある [20].

- classification (分類)
- identification (同定)

Classification とは物体の種類 (an object) を区別する認識で，つまり，一般名称としての呼び名を認識する．Categorization とも呼ばれる．一方，identification は個々の物体 (the object) を区別する認識である．例えば，自分の机と他人の机を区別するのが identification で，机と椅子を区別するのが classification である．

従来は identification を目的とした物体認識が主流であった．なぜなら，認識対象の正確な形状を予め与えておくことができ，それらと画像を照合することによって認識が可能になるので，問題となる点が「画像と形状モデルをいかにうまく照合するか」という一点に集約されるからである．Identification は昔から様々な手法が提案され，盛んに研究されてきている．しかし，それでも，実画像に対する identification には多くの課題が残されており，難しい問題といえる．

もう一方の classification は，identification よりさらに難しい．それは，対象の正確な形状が分からないために，何を認識の手がかりとして用いればよいかという問題があるからである．つまり，一般名称で表される概念をどのようにモデルとして表現すればよいかという問題がある．例えば，椅子の classification をしようとすると，椅子には 1 本足の回転する椅子もあれば，4 本足の椅子もあり，また，ソファーの様な椅子もあれば，公園のベンチのような椅子もある．これらの様々な椅子をどうすればすべて椅子と認識できるであろうか？ これには，様々なアプローチがあり，代表的な椅子の形状モデルを与えておいてそれと照合する方法，対象の構造を抽出し構造モデルと照合する方法，物体の画像中での 2 次元的な見え方，模様，色などの特徴と物体間の関係を組み合わせて認識する方法，椅子の画像をとにかく大量に集めて画像同士の照合を行う方法など様々な方法が提案されている．

本研究においては，2 種類の物体認識のうち，classification を行なう画像認識の方に関心があり，以後，本論文における画像認識は classification を指すものとする．

ちなみに，人間は正確さを要求される identification よりも classification の方が得意である．しかし，コンピュータは逆に，厳密な定義が与えにくい classification よりも identification の方が得意である．このことは，人間とコンピュータの認識方法の違いを端的に表しているといえる．

2.4 従来の物体認識の研究

従来の物体認識の研究では，認識において最も重要である認識対象のモデル表現と，モデルと対象の照合手法を中心に研究が行なわれてきた [21, 22, 23, 24, 25, 26].

コンピュータビジョンの研究が盛んになった約 30 年前から，言うまでもなく実世界画像の認識の実現を目指して研究が行なわれていた．しかし，当初より物体認識はとても困難な問題であることは認識されており，最初に成功を見た研究は，限定された世界『積木の世界』を対象としたものであった．

その後，実世界画像に対する研究として，2 次元的な取扱いのできる画像，例えば，航空写真などの様な画像に対する理解システムがさかんに研究されるようになった．2 次元であると，視点の移動による見え方の変化という 3 次元特有の煩わしい問題に悩まされることがなく，対象としては扱い易い．認識の方法は領域分割の延長線上にあり，同じ対象を表している領域を切り出して，そ

の形状や色，模様，領域間の関係などを手がかりにしてラベリングすることによって認識を実現している．本研究の題材とするような予め物体の完全なモデルが得られない場合の実世界シーンの認識は，古くは Tenenbaum[27] らの領域分割した領域に対する緩和法によるラベリングによる認識があるが，こうした方法は非常に単純な方法であり，複雑な画像に対しては有効ではなかった．その後は Nagao[28, 29], Ohta[30], The Schema System[31], SIGMA[32] などの画像中の物体毎に認識手法を用意する知識ベース型の画像理解システムが登場し，図 2.1で示した様な画像特徴を構造化してモデル照合を行うボトムアップ処理だけでなく，ボトムアップ処理と，モデルから画像特徴を推定するというその逆の処理の流れであるトップダウン処理の融合を実現した．しかし，これらのシステムはどれも空間的情報の利用が 2 次元的であるので，航空写真や遠景の風景画像を対象としており，オクルージョンが多く発生し 3 次元的な取り扱いが不可欠である室内画像に対しては応用されなかった．

3 次元の実世界を対象とする認識では，モデルベースト (model-based) による物体認識の研究が一般的に行なわれた．モデルベーストとは，認識の対象とする物体の形状モデルを知識としてあらかじめいくつも用意しておいて，画像とモデルの照合を行うことにより，画像中にモデルの表す物体の存在を認識する方法である．この手法は，現在でも 3 次元物体に対する identification 的な物体認識の基本となるもので，形状モデルの表現方法には様々な方法がある．モデルの表現の最も一般的な方法は，物体の 3 次元幾何形状をモデルとするものである．他にも複数方向からの見え方 (appearance) を用いて 3 次元物体を認識する方法 [19] や，一般化円筒を用いた構造表現によって対象を要素に分解してネットワークやグラフなどによって構造的に表現する方法などがある．また，パラメータによって形状モデルの形に幅を持たせることも行なわれた [33].

しかし，どの表現方法も物体の形状を直接認識に利用している．つまり，認識する対象の形状が完全に既知であるか，もしくは既知の形状に多少変形を加えた程度でないと，正しい認識が不可能である．ところが，実世界の画像を認識しようとすると，実世界に存在する物体の形状は無限ともいえる程あり，そのすべての形状が既知であることはあり得ない．また，海や道路などのように，明確な形状を定義することすらできない物体も存在する．したがって，未知の形状を持ったり，形状の定義できない物体であっても，何らかの手がかりによって認識を行なわなければならない．また，同一の概念に属する対象であっても，その形状は一通りではなく，さらに，場合によっては決まった形状を持たないものすらある．例えば,「家」とか「木」とかいう名詞が表現する対象に属する形状は無数にある (図 2.2)．そのため形状のみでなく，構造，色，模様，さらには物体間の関係や，それらよりも抽象的な概念を用いて対象を認識する必要が出てくる．

このような形状以外を用いた 3 次元画像に対する認識は先に述べたように 1980 年前後に盛んに研究されていた [34]．しかし，当時の研究のほとんどが 3 次元画像を航空写真と同じように 2 次元的な画像として取り扱っており，領域分割を行なった後に，関係や構造の情報を利用してそれぞれの領域にラベリングを行い，classification 的な物体認識を実現していた．この様な方法では，初期の領域分割の結果が最後まで結果に影響してくることや，対象が 3 次元であるのにも拘らず，3 次元的な取り扱いがなされていないという問題点があった．そのため，その後，D.Marr の提案した「視覚認識への計算論的アプローチ」[35] の影響で，3 次元情報の復元が重視されるようになり，こ

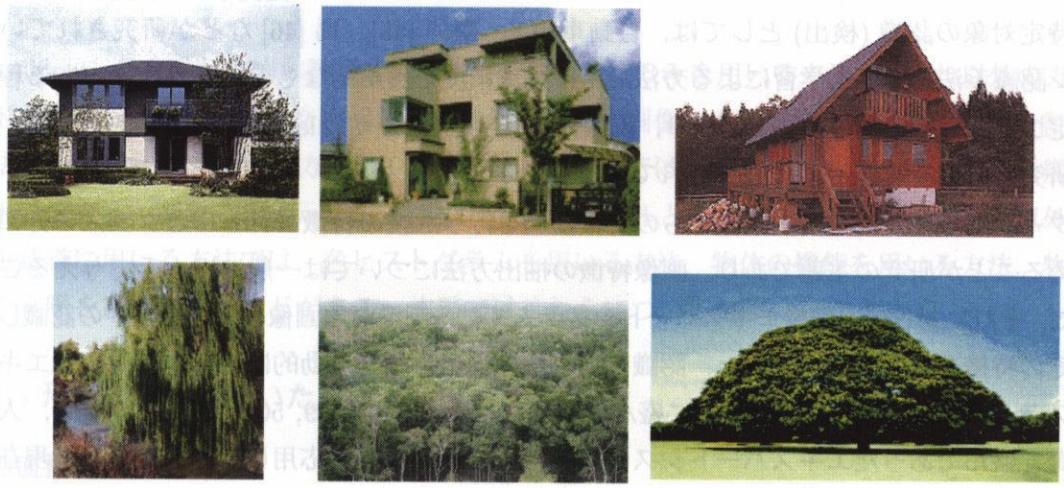

図 2.2 「家」「木」を表す画像には様々なものが存在する

うした領域分割＋ラベリングのような 2 次元的な物体認識の手法は下火となった [36].

近年，再び，形状以外の情報も用いて classification 的な物体認識を行なう方法が試みられている．これらでは，物体の機能や目的，物体間の関係などが用いられている．このうち，物体間の関係については 1980 年前後に盛んに研究されていたが，それ以外の，機能や目的などについては，物体認識に利用するアイデア自体はかなり以前から存在したものの，知識表現や認識機構の困難さのためにあまり研究が行なわれておらず，特に 3 次元の実世界を対象とする認識について，研究が始まったのは 1990 年代になってからである．形状以外の情報も用いて classification 的な物体認識を行なう手法として，物体の機能を推測して機能から物体を認識する function-based recognition [37]，物体の候補を複数出して物体間の関係によって最終的な結果の選択を行なう context-based recognition [38] などが提案されている．

今まで述べた研究は，どれも人手でモデルを記述し用意しなければならない方法であり，非常に多くの種類の物体が存在する実世界に対応するには限界がある．そのため，学習を用いた認識も広く研究されている．学習を用いた認識は大きく分けて，二通り存在する．1 つは，予め対象に適した特徴抽出方法を用意しておいて，それを用いて学習画像から画像特徴を抽出してモデルを構築し，認識対象画像から抽出した画像特徴と比較することによって，クラス分類を行うパターン認識手法による方法．もう 1 つは，認識対象画像とその画像中の認識したい要素やその特徴を併せて入力すると，自動的にその認識の手順を生成する画像処理エキスパートシステムである．

パターン認識は，画像に限らず実世界の様々なパターン情報，例えば，文字，音声，生物の行動パターンや DNA 配列などをいくつかのクラスに分類する研究分野であり，統計理論にその基礎を置いている [39, 40]．パターン認識は，画像認識の研究が始まる前から行われており，現在でも盛んに研究が行われており，「認識」の基礎とも言える研究分野である．画像認識の場合，人手でモデルを構築するのでなければ，ほとんどの場合，パターン認識手法によって認識が行われる．例えば，人の顔画像の認識はパターン認識手法による学習を用いた方法が一般的であり [41, 42, 43, 44]，H.

Schneidermanらの研究 [44] においては 9 割以上の精度で画像からの人の顔の検出に成功している．他に特定対象の認識 (検出) としては，自動車 [44]，裸体 [45]，馬 [46] などが研究されている．パターン認識手法を用いた学習による方法では，認識精度を実用的なものとするには，ある程度対象を想定して特徴抽出を行う必要があり，現状では，学習を用いた classification においては，ほとんどの研究において認識対象が限定されている．

パターン認識の研究では，画像からの特徴抽出を行った後の特徴ベクトルとクラスの分類の関係を求めることが研究の主題であり，画像特徴の抽出方法については一般には人手で与えることが多かった．それに対して，画像エキスパートシステムでは，認識対象画像とその画像中の認識したい要素やその特徴を併せて入力すると，認識手順や特徴抽出手順を自動的に生成する．画像エキスパートシステムの研究は 1980 年代後半に盛んに研究された [47, 48, 49, 50]．これは，当時，人工知能の分野で盛んであったエキスパートシステムの手法を画像処理に応用したものであり，現在まで研究は行われている [51, 52, 53, 54]．例えば，IMPRESS[48, 53] では，主に医用画像を対象として，画像とゴール画像（望ましい出力画像）の組のみを入力することによって，ゴール画像に類似した画像要素を抽出するための処理手順をそれぞれの処理要素のパラメータも含めて自動生成する．しかし，これらの画像エキスパートシステムも，医用画像から臓器の一部を抽出したり，細胞写真や航空写真から比較的単純な形状の要素を抽出する様なそれほど複雑でない処理手順の自動生成に留まっており，複雑な形状の物体や 3 次元的な物体の抽出を行うことは処理手順の組合せ爆発のために難しいと考えられる．

また，近年，コンピュータの進歩により大量の画像を高速に処理できるようになってきた．そこで，画像データベースおける検索手法を物体認識に応用するという試みがなされている [55, 56, 57]．画像内容に基づく画像検索 (content-based image retrieval, CBIR)，つまり，画像特徴を手がかりとした画像データベースの検索では，質問画像が与えられて，それに類似した画像がデータベースから画像特徴に基づいて検索される． CBIR では，従来の画像認識とは異なり，データベース中に様々なジャンルの画像が含まれるので，予め画像の種類を限定することが出来ない．そのため，どのような種類の画像でも対応可能な色情報やテクスチャ情報が画像特徴として用いられる．特に，カラーヒストグラム [15] は CBIR ではよく用いられる画像特徴である．CBIR の手法を用いた画像認識では，基本的には，含まれている主な物体の名称が予め分かっている学習画像を大量に用意して，認識したい画像に類似している学習画像を検索する．そして，学習画像と認識対象画像が類似しているので，検索された学習画像に含まれている物体が認識対象画像にも含まれていると見なして，認識を行ったこととする．こうした方法は，画像同士の類似度の計算方法を定めて，学習画像を大量に集めれば，認識が可能となるので，比較容易にシステムが実現できる反面，すべての対象に対して，単一の一般的な方法を用いるために，認識率が高くすることが難しいという問題がある．この CBIR の手法による画像認識の研究については，第 7章で詳しく述べる．

2.5 物体認識の方法

Classification を行なう物体認識には，主に対象の形状のみを用いる方法と，それ以外の方法がある．形状を用いる方法では，2 次元的な方法としては，画像を直接照合する方法，領域分割を用いる方法，局所的特徴を用いる方法，3 次元的な方法としては，構造を用いる方法，3 次元幾何モデルを用いる方法，複数の 2 次元ビュー (2-D view) をモデルとする方法などがある．形状以外の手がかりを認識に用いる方法では，色ヒストグラムを用いる方法，物体の機能を用いる方法，物体間の関係や文脈を用いる方法などがある．本節ではこれらについて簡単に説明する．

2.5.1 形状のみに基づいた方法

画像の直接照合

画像の一部を切り出した後，モデル画像との直接照合を行う．ただし，直接照合と言っても，画像から画像特徴を抽出せずに画像自体を特徴として扱うという意味の「直接」であって，単純に差分をとるわけでなく，複数枚のモデル画像を主成分分析 (principal component analysis, PCA) によって圧縮し，圧縮された部分空間内で照合を行うのが普通である．この方法は顔画像の認識において最初に用いられた [41]．その後，H. Murase らのパラメトリック固有空間法 [19] によって，3 次元物体の identification にも用いられるようになった．この方法では，3 次元世界の画像であっても，画像を 2 次元的なものとして取り扱っており，3 次元物体を対象にする場合は，多数のモデル画像を用意する必要がある．

領域分割

前節で述べたように，領域分割を行なった後に，領域の自体の色，形状，テクスチャなどの特徴と，領域同士の関係や構造の情報を利用して各領域にラベリングを行うことによって認識を行う方法である．

局所的特徴の利用

画像の直接照合や，領域分割による方法では，物体が部分的にしか見えていない場合や，形状が複雑で領域分割がうまく行かない場合には，対処することが難しい．そこで，C. Schmid ら [58] は局所的な特徴の組み合わせによって，画像の照合を行う方法を提案した．具体的には，最初に Harris interest point detector[59] によって，画像中から 100 点程度の特徴点を選び出す．次に，各点の画素値や微分値等を特徴ベクトルとし，それらの集合によって 1 枚の画像を特徴付けることにする．照合は，未知の画像に対して，同様に特徴ベクトルの集合を求めて，モデル画像 (または学習画像) の特徴ベクトルの中から，それぞれ近い特徴ベクトルを探して，ある程度類似しているモデル画像に対して投票を行う．この際，特徴点間の相対的位置関係を考慮することによって，無駄な投票を防ぐことを行う．最終的に最も多くの投票を集めたモデル画像にマッチしたと見なす．D. Lowe[60] も同様の方法によって，オクルージョンのあるシーンにおける物体認識を実現している．ただし，

これらの研究は identification の物体認識である．

一方，M. Weber らの [61, 62] 研究では，多数 (300 枚程度) の正例，負例の両方の学習画像から interest point detector を用いて局所パターンを抽出し，それらをクラスタリングすることによって対象に特徴的な局所パターンを選び出し，人間の顔や自動車の classification を学習によって実現している．

構造

全体の形を認識するのではなく，物体を部分に分解して認識する方法である．部分は通常，円筒や直方体などのプリミティブな形状で表現し，それらのつながりをネットワークやグラフなどによって構造的に表現する．構造を認識するのでモデルの情報量が減り，形状が多少違っても構造が同じであれば同一対象であると見なされるので，classification に適した方法である．しかし，大局的な構造が似ているものは区別できなかったり，部分に分解しにくいものは認識できないという欠点がある．

Recognition by components [63] では，“geon” と呼ばれる一般化円筒表現によって物体の部分を表現し，その部分の関係から物体の構造を記述し，認識することを提案している．また，D. Marr による著作 *Vision* [35] の中でも同様のことが記述されている．

3 次元幾何モデルとの照合

Identification でよく用いられる 3 次元幾何モデルと対象の照合を基本とした手法である．すべてのモデルに対して，画像中の対象にできるだけ近付くように変換を施してみて，その中で変換された後のモデルと画像中の対象との差が最も小さかった時のモデルに対象がマッチしたとみなす．モデルは，認識した対象のクラスの代表するような 3 次元幾何モデルを 1 つまたは複数個用いて，その中のモデルにマッチすればそのクラスに属するとみなす．この方法では，すべてのモデルについて変換して照合する必要があるために計算量が多くなる欠点がある．

D. G. Lowe による研究 [64] では，perceptual organization によって，画像から直線を抽出した直線のグループ化を行い，それに対して，3 次元モデルを当てはめることによって，物体の認識を行った．他にも多数の同様の研究が存在していたが，そのほとんどは identification を指向していた．過去においては,「物体認識」と言った場合，この様な 3 次元モデルを画像に当てはめる方法による認識を指していた時期もあった．また，依然として現在においても，狭義の意味での「物体認識」はこのことを差すことが多い．

複数の 2 次元ビューとの照合

この方法では，3 次元モデルの代りに，物体のモデルとして複数方向からのビューの画像をモデルとする．Linear combination [20, 65] では，正射影を仮定した場合，1 つの物体について任意の 3 方向からのビューの画像があれば，任意の方向からのビューの画像を作り出すことが出来ることを示している．

Recognition by prototypes [20, 66, 67] では，椅子や自動車の複数のビューの画像をプロトタイプモデルとして，linear combination を用いた 2 次元ビューの画像照合によって，classification を実現している．

2.5.2 形状以外に基づいた方法

色ヒストグラム

この方法は，M. J. Swain[15] によって最初に用いられた方法で，画像認識の方法としてではなく，画像データベース検索の手法として提案された．そのため，対象に依存しない方法で，どのような種類の画像に適用可能である．ただし，画像中から物体を見付ける手法ではなく，画像同士の類似度を計算する方法である．ヒストグラムの作り方は様々な方法があるが，最も簡単な方法としては，RGB の各軸を等間隔に分割することによって，画像中の各画素の色を減色し，各色の頻度を計数してヒストグラムを作る (図 2.3)．これを画像の特徴量として，類似画像の検索を行う．例えば，図 2.3では，RGB の各軸を 5 つに分割し，256^3 色の色空間を $5^3(= 125)$ 色に減食している．そして，それら各色の頻度を計数し，125 次元の色特徴ベクトルを求め，ベクトル同士の距離を例えば，ユークリッド距離などで定めれば，類似画像の検索を行うことができる．

画像データベースに予め一般名称の分かっている物体の画像を用意しておくことによって，精度はあまり高くはないものの，類似検索による物体に認識が可能となる [55, 57]．また，近年，identification を指向した研究ではあるが，色ヒストグラムの方法を用いて，画像中からの物体の探索を高速に行う手法が提案されている [68, 69]．

また，近年，B.Schiele ら [70] によって，色ヒストグラムを拡張して，色以外の特徴，例えば，ガウシアン微分係数 (Gaussian derivatives) やガボール係数などのテクスチャ特徴を各画素について求めて，それをヒストグラムによって表現する手法 multidimensional receptive field histogram が提案されている．

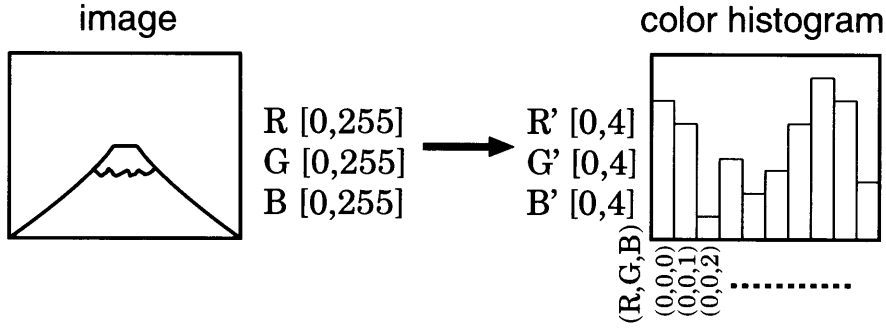

図 2.3 色ヒストグラムの作り方．

対象の機能に基づいた認識

物体の概念モデルの利用の一つの方法として，最近，物体の機能に基づく認識手法 (function-based) が提案されている [71]．これは，通常は人工物の認識に限って用いられ，対象物の機能 (function) を認識の手がかりとする認識手法である．つまり，従来の様に直接形状モデルから対象を認識しようというのではなく，まず対象の形状からその対象の持ち得る機能を認識し，認識された機能をもとに最終的に対象を認識する方式である (図 2.4)．

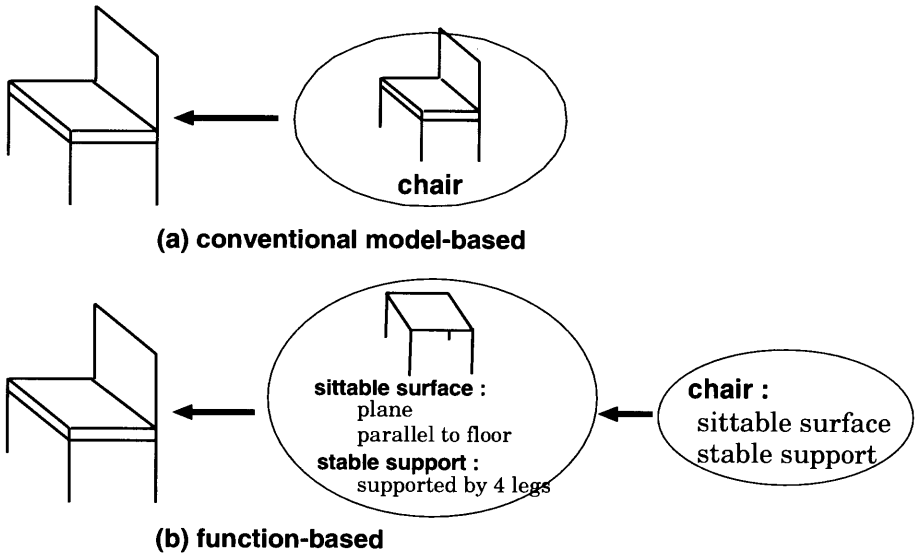

図 2.4 (a) 従来のモデルベーストでは形状モデルと画像中の対象との照合によって認識を実現していた (b) ファンクションベーストでは機能を提供しているような形状を探し，機能から認識を行なう．

この方法では，認識対象の全体の形状を認識の手がかりとしていた従来の方法と異なり，対象のある一部分の形状に注目して認識していることになり，認識の手がかりとして機能を用いることによって注意の集中 (focus of attention) が実現されている．また，「人は，物の物理的な形を知覚するのではなく，物の物理的形状が人に与えている機能を知覚している」という J. J. Gibson のアフォーダンス理論 (affordance theory)[72] に近い考え方である．

L. Start ら [37, 73] は，椅子，ベンチ，テーブル，ベッド，本棚をそれぞれの function を用いて認識するシステムを提案した．ただし，このシステムは単に function を認識に用いることのみを目的としているので，与えられるデータは 2 次元画像でなく，3 次元の形状モデルである．システムはまず，対象とする物体がその物体の特有の機能を持つ形状を備えているか調べる．例えば，椅子であれば，座ることのできる面 (sittable surface) と，安定した支え (stable support) が必要である．これらの sittable surface と stable support は物体の形状から認識される．このような，各物体が備えるべき条件は，大きさ，方向，安定性，隣接関係，前方空間のゆとりをそれぞれ調べる 5 つのプリミティブ処理ルーチンによって確認される．この場合，sittable surface は，方向は床面とほぼ並行で，高さ，大きさとも人間が座るの適していることが条件であり，stable support は，sittable

surface を床面から支えている構造が存在して安定していることが条件である．これら 2 つの条件さえ備えていれば，その物体は人間に対して座る機能を提供していることになり，それ以外の形状がどうなっていようとも椅子であると認識される．つまり，図 2.5の物体は，すべて椅子であると認識される．

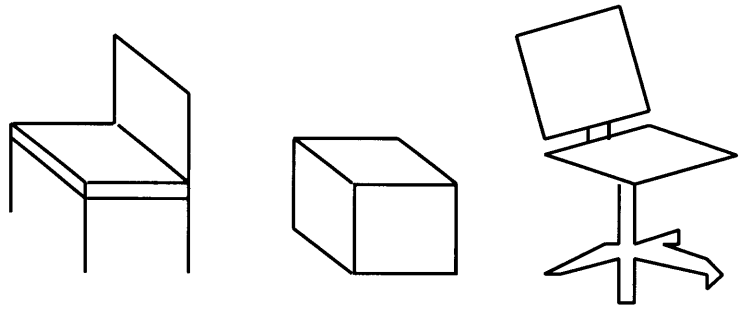

図 2.5 いろいろな椅子

関係を用いた認識

従来の認識では，物体はそれぞれ別々に存在するものであり，物体間の関係情報は，領域分割の結果にラベリングする認識方法以外では，あまり積極的には用いられて来なかった．

ところが，人間が認識を行なう場合，物体間の関係情報は非常に重要である．例えば，図 2.6(a) の絵はそれだけ見ると，何を表しているか可能性があり過ぎて分からないが，図 2.6(b) のように人間の形を描くだけで，それが壁と天井と床，つまり部屋の隅を表していることが容易に想像できるようになる．これは，図 2.6(a) のような図形に，立っている人間の存在という情報が与えられることにより，図形が制約を受け，表している対象の範囲が限定されたので，認識が可能になったということであろう．

このような物体間の関係の情報を一般に文脈 (context) とよび，このような文脈情報を利用した認識手法を context-based による認識という [74].

通常，画像中には複数の物体が含まれ，それぞれが何らかの関係を持って存在している．つまり，すべての物体は周囲の関係との中に存在している．そのため，関係情報は認識にとって有用な情報を提供してくれることが多く，これを利用することによってより柔軟な認識が可能となる．

単に物体の関係情報といっても，次のように様々なものが考えられる．

- 関連する物体の位置関係：
  「机の前には，椅子があって，上には電気スタンドがのっている」

- 力学的平衡関係：
  「物は引力を受けるので，安定した構造物の上にしか存在しない」

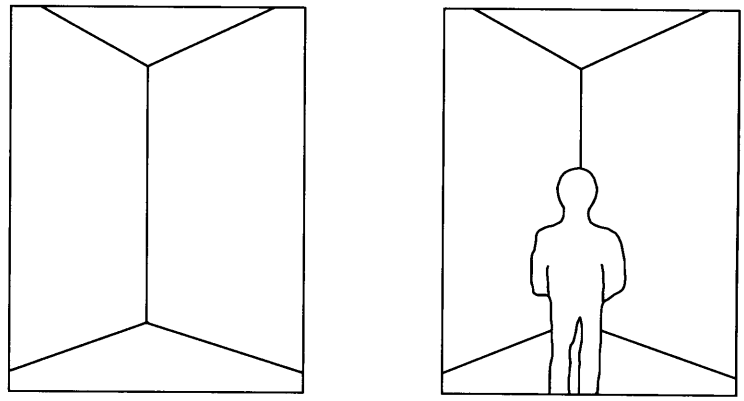

図 2.6 コンテキスト情報が (a) 得られない場合，(b) 得られる場合

- 状況による文脈関係：
  「オフィスのシーンなので，机の上の白い紙は書類だ」
- スケールの関係：
  「消しゴムと同じくらいの大きさの電車なので，模型だ」

これらの知識は無数あり，どのような情報を認識に用いると効果的であるかは難しい問題である．従来のほとんど研究においては，関係を用いる場合には，もっとも利用し易い物体の位置関係のみが利用されていて，それ以外の関係を認識に用いた研究はほとんど存在していない．

T. M. Strat ら [38] は，context-based という言葉を掲げて，コンテクスト情報が認識処理の中心になるようなシステムを提案した．このシステムは，屋外の自然風景の画像を対象としている．このシステムの特徴は，過去に行なわれた領域分割と関係によるラベリングの方法とは異なり，限定された special-purpose なオペレータで対象の候補をたくさん抽出しておき，その候補を関係によって組み合わせて選択することである．認識は，評価値の高い順番に候補を組み合わせて，いくつか組合せを求めた上で，最終的に最も良い組合せを画像全体の解釈として選択することによって行なわれる．さらに，オペレータが非同期的に動作し，生成された候補を見ることによって関係情報を利用して，さらに別の候補を生成する動作も行なうという関係によるトップダウン認識も実現している．ただし，このシステムは，1 枚のサンプル画像を認識するためだけにシステムに予め人手によって与えておく関係の知識である context set が 75 も必要で，知識の記述が細か過ぎる欠点がある．

また，同様の研究としては，J. Marti らの研究 [75] がある．こちらのシステムでは，候補を生成する認識プロセスを学習で構築する点が異なっている．

2.6 まとめ

本章では，初めに，物体認識には classification と identification の 2 種類があることを説明し，本研究においては前者の認識を目標とすることを述べた．さらに，従来の物体認識の研究の流れを簡単に説明し，その後，主に classification のための物体認識の手法について簡単に説明した．

第 3 章

認識システムの構成法

3.1 はじめに

実世界の画像に対する物体認識システムを構築するには，実世界の様々な不確定な要素に対して対応できるように設計しなければならない．そのため，単一の画像認識アルゴリズムを用いるだけでは，実世界に対して一般性，柔軟性のあるシステムを構築することが難しく，いくつもの特性や機能の異なるアルゴリズムを統合することが必要である．また，物体認識においては，画像からモデルを探すボトムアップ処理のみではなく，モデルから画像中の要素を探すトップダウン処理も効果的であり，これらをうまく組み合わせることも必要である．

本章では，まず初めに，一般的な画像理解システムの処理の流れについて簡単に触れ，その後，複数アルゴリズムの統合と，柔軟な制御構造の実現に適しているといわれている分散協調型のシステム構成について説明する．

3.2 ボトムアップとトップダウンの融合

第 2.2節で述べたように，一般的には画像認識の処理は，第一段階で画像からの特徴抽出を行い，第二段階で特徴のグループ化を行い，第三段階で，予めシステムに蓄えられている知識やモデルのデータとの照合を行って，画像中に含まれる要素の認識を行なうという構成になっている (図 2.1).

誤差やノイズのない理想的な世界に対してはこのような入力から出力まで一方向の処理のみで正確な認識が期待できるが，実世界では誤差やノイズなどの不確定な要素を多く含んでいるために，それのみでは精度のよい認識は期待できない．そこで，実際の多くの画像認識システムにおいては，入力から出力への一方的な処理の流れだけではなく，図 3.1の太い矢印で示されたような高次レベルから低次レベルへの流れも存在する．例えば，最初に左方向の処理を行ってモデルを選択し，次に，そのモデルに合わせた特徴抽出，グループ化を再度行って，再びモデル照合を行うことによって，より精度の高い認識を行ったりすることがある．この場合の左方向の処理をボトムアップ処理，一度モデル照合をしてから右方向に戻る処理をトップダウン処理と呼ぶ．ボトムアップ処理はデータ駆動型であり，画像から抽出した情報を構造化，抽象化して，モデルとの比較を行なうのに対して，

トップダウン処理はモデル駆動型で，モデルから推定される構造を画像から抽出することによって，モデルとの適合を検証する．

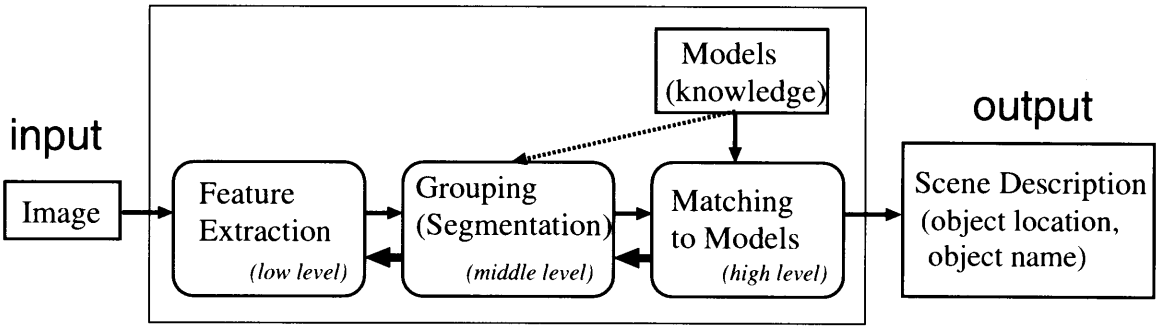

図 3.1 トップダウン解析を含む画像認識の処理の流れ

従来の画像認識システムでは，ボトムアップ処理によっておおまかな構造を抽出し，システムが持っているモデルからもっともらしいものを選択する．次に，その選択したモデルの適合性を検証するためにトップダウン処理を実行する．選択したモデルが間違っていると思われる場合は，ボトムアップ処理の適当な段階まで戻って，処理をやり直し，別のモデルを選択するというように，ボトムアップ処理とトップダウン処理を交互に行なうのが，一般的であった (図 3.2)[76].

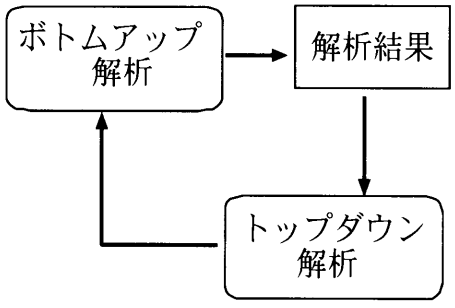

図 3.2 従来のボトムアップとトップダウンの融合方法

しかしながら，このようなボトムアップとトップダウンを交互に行なう方法では，それぞれの処理が独立して実行されるので，効率のよい処理とはいえない．

そこで，ボトムアップとトップダウンを単純に交互に行なうのではなく，より柔軟な制御構造を実現するためのシステム構成法が必要となる．そのためのシステム構成法として，処理の流れを固定しない分散協調型のシステム構成法という方法が存在する．これについては，第 3.4節で詳しく説明する．

3.3 複数アルゴリズムの組合せ

単一の視覚認識アルゴリズムは，ある特定の状況には有効であっても，そうでない場合には有効でないことがある．複雑なシーンに対応するため，複数の異なる特性を持ったアルゴリズムを組み合わせることが必要になってくる．それぞれ異なる特性を持った複数の同じ処理を行なう違うアルゴリズムを組合わせればに柔軟性が増し，領域に依存する (domain-dependent) 複数のアルゴリズムを組み合わせればより一般的なものとなる．

複数のアルゴリズムを組合わせる方式として，以下の３つが考えられる．

- 逐次型
- 並列型
- 分散型

逐次型は，最も一般的な組み合せ方で，低レベルのアルゴリズムで，ある程度シーンの特徴を抽出しておいてから，最も適切なアルゴリズムを選択し，処理を行なう方式である．しかし，低レベルのアルゴリズムでの処理がうまくいかない場合は，適切なアルゴリズムが選択できないことがある．

並列型は，複数のアルゴリズムを並列に並べておき，それらを１つの画像に対して同時に適用して，出力を最後に１つにまとめる方式である．それぞれのアルゴリズムは，完全に独立しており，並列に実行できる．しかし，処理の途中で各アルゴリズムが部分結果を提供し合うことによる相乗的な性能向上が期待できない．

分散型は，アルゴリズムを並列に並べるところまでは並列型と同じであるが，互いに協調する点が異なる．アルゴリズムの組合せ方としては，最も進んだ形である．お互いのアルゴリズムが他の処理結果や中間結果を参照し合うことによって，精度の高い解析が期待できる．また，第 3.2節でも述べたように，トップダウンとボトムアップの融合を実現するのにも有効である．

次節では，アルゴリズム統合と柔軟な制御構造の実現にそれぞれ利点を持っている分散協調型のシステム構成法について詳しく説明する．

3.4 分散協調型のシステム構成法

分散協調型のシステム構成法とは，複数の互いに協調し合うサブモジュールによって，システムを構成する手法である．ここでのサブモジュールは，単なる一つの画像処理オペレータでもいいし，もっと高機能なそれ自体が何らかの単独のシステムであってもいい．サブモジュールがある程度の独立性，自律性を持っている場合は，その意味を込めて一般に『エージェント』と呼んでいる．システムがエージェントの集合体として構成される場合は，マルチエージェントシステムと呼ばれる．

この分散協調型のシステム構成法は，音声認識システム Hearsay-II(1980)[77] において，黒板モデルとして初めて提案されたもので，黒板と呼ばれる共有メモリを介して，サブモジュールが情報交換を行ないながら，処理を行なうというシステム構成法である (図 3.4)．この Hearsay-II の提案

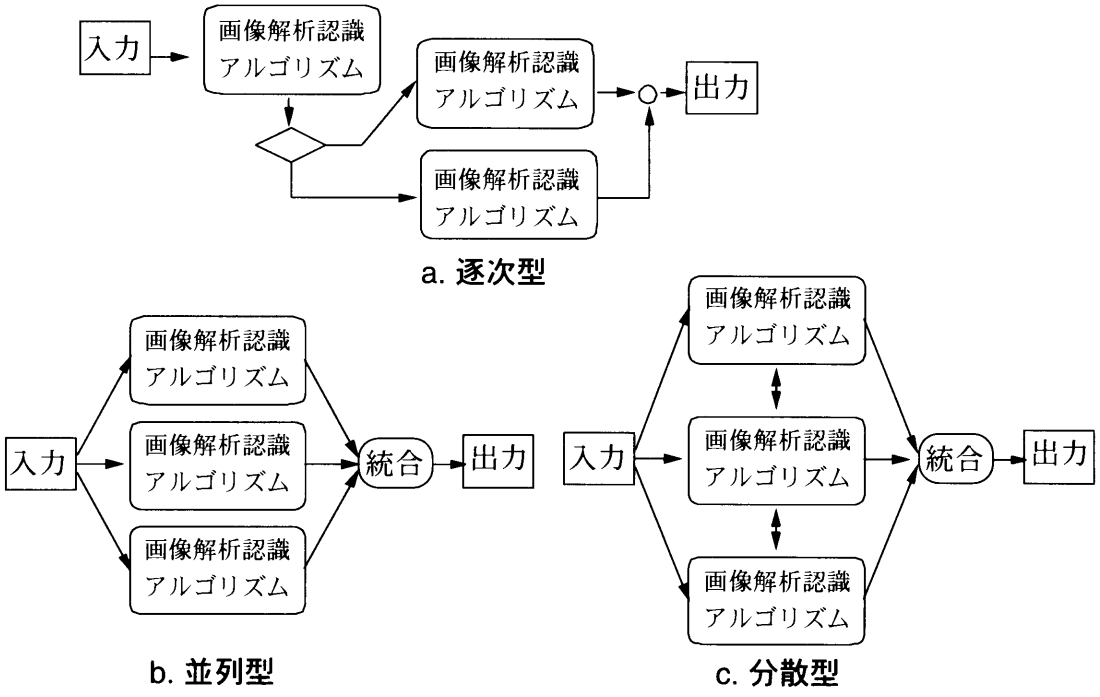

図 3.3 複数のアルゴリズムを組合せる方式

した黒板モデルは，その後の分散協調システムの基本方式として，その後の研究に大きな影響を与えた．

黒板モデルのシステムは，黒板とよばれる共有メモリと，その黒板の回りに配置される複数のモジュールによって構成される．モジュールは，システムの中においては，対象物を認識するための知識とみなすことができ，通常は簡単なオペレータ，例えば，エッジ抽出やエッジから平行エッジを抽出するオペレータなどである．これらのモジュールは，knowledge source とも呼ばれる．知識としてのモジュールは互いに独立なので，簡単に追加することが出来，モジュールを追加することによって，システムの拡張が簡単にできる．これらのモジュールが黒板を通して相互作用を行なうことによって，認識が実現される．各モジュールは，黒板を常に監視していて，自分が処理を行なうべきデータが黒板に書き込まれると，勝手に読み出して，処理結果を再び黒板に書き込む．つまり，黒板の内容に応じて，適切なモジュールが処理を行なう機構が実現されている．この各モジュールの処理過程は各モジュールが独立に行なう．しかし，モジュールが勝手黒板に書き込みを行なうと黒板内に矛盾するデータが書き込まれたりして，整合性がとれなくなるので，黒板モデルでは，黒板へのアクセス制御が必要である．一見，簡単に並列処理化出来そうであるが，並列化するとこの制御機構が複雑になる．長所としては，黒板の内容に応じてモジュールが選択的に起動し，柔軟な解析過程の制御が出来ることと，複数のモジュールの存在によって自然なアルゴリズムの統合ができること，モジュールの追加によって容易に機能が拡張できることなどである．

共有メモリを介して情報交換が行なわれる黒板モデル以外にも，直接，構成要素間で通信をする方法もあり，通信はメッセージパッシングによって実現される．この方式ではシステムの制御機構

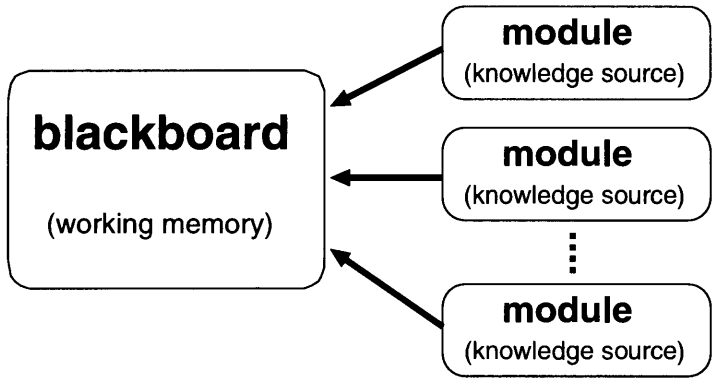

図 3.4 黒板システム

は小さいもので済み，各エージェントの実行に必要な知識や処理結果は，各エージェントに分散されて格納され，通信によって互いの処理結果にアクセスする．

なお，画像認識においてはデータ量が多いために共有メモリを介した通信が好まれ，従来のシステムではほとんどが共有メモリ方式を採用している．

分散協調型のシステム構成法は，一般に分散協調アーキテクチャと呼ばれていて，分散人工知能 [78, 79] やマルチエージェントシステム [80] のシステム構成法のことである．分散協調アーキテクチャを考える時，システム全体に注目するのか (マクロな視点)，個々のエージェントに注目するのか (ミクロな視点) によって，２通りのシステム構築のアプローチがある．

システム全体に注目する場合，初めにシステム全体としての目的があり，それに従った形で個々のエージェントも目的を持つことになる．つまり，初めにシステムの全体の構成を設計し，それをエージェントに分割することになる．

逆に，個々のエージェントに注目する場合，個々のエージェントにそれぞれ独立した目標を持たせておいて，それらが同一の世界で協調しながら個々の目標を達成しようとする時，システム全体としての振舞いがどうなるかを予想して，個々のエージェントを設計することになる．

前者を分散人工知能，もしくは分散協調問題解決，後者をマルチエージェントシステムと分けて呼ぶこともある [81]．

さて，一般に，分散協調システムの効用は，次のようなことが言われている [78, 80]．

- 多数のエージェントが協力して問題の解決に当たるため，全体の処理能力，処理効率が向上する．
- 達成可能な処理内容あるいは処理領域が拡大する．つまり，単体では，不可能であった処理が，多数のエージェントが集まって協調動作を行なうことにより可能となる．
- あるエージェントが故障したり，不良であっても，他のエージェントが肩代りすることが可能な柔軟なシステムを構築できる．

- エージェントの追加によって，容易にシステムの拡張，機能追加が行なえる．

これらのことは，現実世界の多様性，不確実性などを克服するためには都合のよいことであり，画像理解システムのような実世界の問題を扱うシステムの構成法には適しているといえる．しかし，分散協調による応用システムは，まだ研究途中であり，これらの効用が果たして本当にいえるかどうかは，今後の研究次第である．これらの効用を実証することは，第4章で述べるマルチエージェントによる画像認識システムの研究の目的の一部ともいえる．

3.5 画像理解システムにおける分散協調

画像理解システムにおいて，複数アルゴリズムの統合や，柔軟な制御構造の実現のために，黒板モデルやマルチエージェントモデルの考え方を用いた分散協調型のシステム構成法が注目を集めている．[82, 83, 84, 85, 86, 87]

画像理解における分散協調システムには，**機能分散型**と**空間分散型**の２つの分散の形態がある．機能分散型は，システムをどのようにモジュール化して，分散・協調させるかというアルゴリズム統合の観点からの分散方法であり，一方，空間分散型は，画像空間中に分散している局所情報をいかに協調させて画像全体の構造を理解するかという制御構造の観点からの分散方法である．これら２つの形態は，同じ分散協調型といっても，考え方に大きな違いがある．２つの方式は，それぞれ重点の置き方が異なる方式である．つまり，機能分散協調方式は，プログラムの分散協調なので，異なる複数のアルゴリズムの統合に重点を置いた方式であり，空間分散協調方式は，データの分散協調なので，データフローの制御構造に重点を置いた方式である．以下では，それぞれについて，詳しく説明する．

機能分散方式

機能分散方式とは，画像処理・理解アルゴリズムをエージェントの単位とする方式である．この方式は，画像理解システムに特有の方式ではなく，単にシステムをエージェントに分割するにいうことである．この方式では,『エージェント ＝ システムを構成するモジュール』ということであり，システムをモジュール分割することによって構築するという従来の方式と基本的には同じである．しかし，エージェントであるので，モジュールとは違い，コントローラに制御されることなく，自律的に動作を行ない，他のエージェントと協調動作を行なうことができるのが普通である．

この方式は，システムに関する分散協調の方式といえる．そのため，エージェントの数や種類は，システムの設計時に決まり，エージェントの構成は静的である．

空間分散方式

空間分散方式は，画像空間中に分散している局所情報をいかに協調させて画像全体の構造を理解するかという観点からの分散方法である．つまり，対象とする画像中の要素をエージェントの単位

とする方式である．この空間分散方式は，システム構築の方式というよりも，画像理解の手法の一つと言ってもよく，機能分散型とは基本的な考え方が異なる．

具体的な処理としては，初めに前処理と簡単な画像解析を行ない，画像からその構成要素を大まかに抽出し，それぞれの要素をエージェントとする．エージェントは，他のエージェントと協調することにより，要素よりも大きな構造や，初期の解析で発見できなかった要素の存在を発見したりする．エージェントは動的に生成，消滅，併合などが行なわれ，エージェントの構成は動的に変化する．

この方式では，機能分散方式と違って，システムの構築方法は定めていない．しかし，画像を分散協調で扱うので，通常は自然とシステムも分散協調型になることが多い．この方式は，画像中の要素についての分散協調の方式といえる．

それぞれの方式の特徴について，表3.1にまとめて示す．

表 3.1 機能分散方式と空間分散方式の比較

方式	分散	指向	主な効果	エージェントの構成
機能分散方式	システムの分散	プログラム指向	アルゴリズム統合	静的
空間分散方式	画像要素の分散	データ指向	データフロー統合	動的

3.5.1 機能分散協調方式のシステムの例

機能分散協調方式のシステムの例を４つ取り上げる．渡辺らによる並列ステレオ対応探索 [88] は処理内容は同じであるが特性の異なる３つのアルゴリズムを並列に並べたシステムで，典型的な機能分散方式のシステムである．角らによる並行トップダウン処理 [89] は，複数の解析モデルを並行に処理する並行トップダウン処理を提案している．また，松山らによる航空写真解析システム [29, 90, 91]，B. Draper らによる風景画像認識システム The Schema System[31, 92] は黒板モデルを用いたシステムである．

同一処理アルゴリズムの並列実行

渡辺ら [88] は，異なる特性で同一内容の処理を行なうアルゴリズムの統合の実験として，異なる画像特徴に基づく３つのステレオ対応探索アルゴリズムを並列に実行するシステムを提案している (図 3.5)．このシステムにおいては，協調処理を「複数の処理が共通の目標に向かって相互に支援し合う形態」と定義している．また，このシステムでは，結果の統合は，各エージェントが自分の処理結果に対して自ら独自の基準で付ける確信度 (confidence value) に基づいて行なうことになっている．

ステレオによる３次元空間の認識においては，左右の画像の対応点を求めることが最も重要な問題であるが，実際には，類似した特徴を持つ点が多数存在するために，対応点を求めることは困難

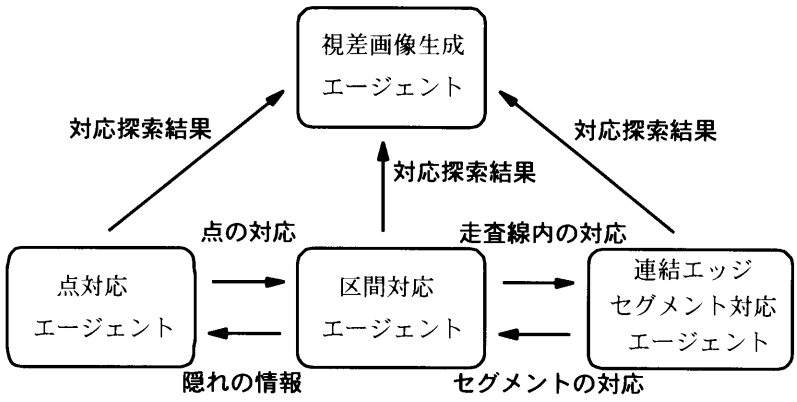

図 3.5 ステレオ対応探索エージェントの協調

な場合が多い．そのため，対応点を探索するアルゴリズムは，環境や用途に応じて，多数開発されている．しかし，1 つだけで，すべての場合に適用できるアルゴリズムは存在しない．そこで，異なる環境に適用するアルゴリズムを複数併用することが必要になる．

このシステムでは，表 3.2に示す 3 つの特性の異なるアルゴリズムを統合しており，これらの 3 つのアルゴリズムに対応するエージェントが，協調動作を行ないながら，ステレオ対応探索を行なう．協調動作の内容は，基本的に，不足情報を他のアルゴリズムの処理結果や中間結果を参照することよって補うことである．

結果の統合は，各エージェントが自分の処理結果に対して独自の基準で行なった 0 から 1 までの自己評価値の最も高いものを選ぶということで実現している．このことは，異なるエージェントが異なるアルゴリズムで処理した結果を統一的な基準を用いて客観的な方法で評価することが難しく，結局，**処理結果は各エージェントが自分自身で評価するのが適当である**ということを示している．

表 3.2 ステレオ対応探索アルゴリズムの特徴

	適用範囲	精度
点対応法	○	△
エッジ区間対応法	△	○
セグメント対応法	×	△

並行トップダウンによるシステム

角らによる顔画像認識システム [89] では，正面からの顔認識モデル，斜めからの顔認識モデルなどの複数の認識モデルを用意し，それらを同時にトップダウン的に動作させ，最も確からしい出力をしたモデルを正解として採用する (図 3.6)．

渡辺らの研究 [88] と異なる点は，渡辺らのシステムが 3 つの特性の異なるアルゴリズムの結果を統合して利用するのに対して，こちらのシステムではある特定の状況にしかうまく行かないアルゴリズムを複数組み合わせている．つまり，アルゴリズムの選択は，実際に処理を行なってみて，その結果から判断するしかないという考えに基づいている．従って，正面からの顔写真に対して，斜め顔解析アルゴリズムが解析を行なうような「無駄な」処理を敢えて行なうことによって，結果として多様な入力画像に対して柔軟に対応することができている．このシステムでは並列計算機を用いているので，無駄な処理をしても，処理時間が大幅に増加するようなことにはなっていないという特徴がある．これは，**並列計算機の高速性を認識システムの柔軟性の向上に利用**しているということである．

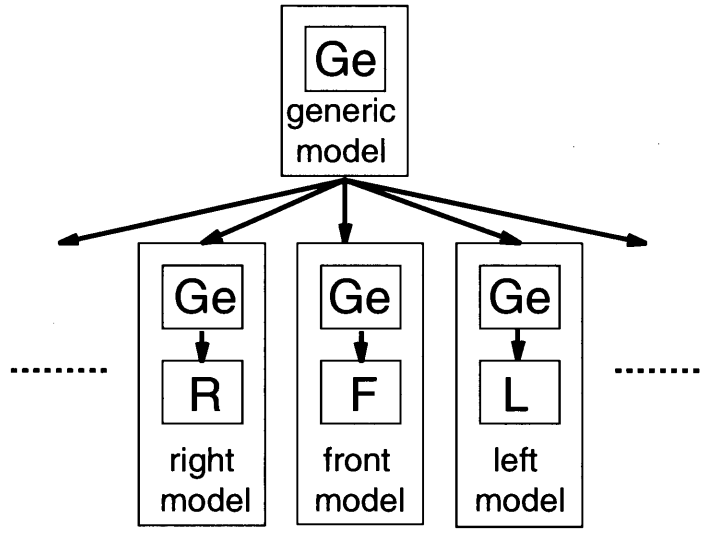

図 3.6 並行トップダウンによる処理

黒板モデルによるシステム

松山らによる航空写真認識システム [29, 90, 91] は，領域分割した航空写真に対して，複数の特徴検出モジュールと複数の対象物認識モジュールを黒板モデルによって協調的に動作させることにより，田畑，森，住宅地などを認識する (図 3.7)．特徴検出モジュールは入力画像に対して直接処理を行ない，その結果に対して対象認識モジュールが認識処理を行なう．1 つの対象物認識モジュールは，それぞれ 1 つの対象物を認識し，田畑認識モジュール，森認識モジュール，住宅地認識モジュールなどがある．

このシステムでは，特徴検出モジュールの認識結果が複数の対象物認識モジュールに利用されるという，**処理結果の共有**が実現されている．

B. Draper らによる The Schema System[31, 92](図 3.8) は，風景画像に対して物体認識を行なうシステムである．各認識モジュールはスキーマと呼ばれ，それぞれが special-purpose な認識シス

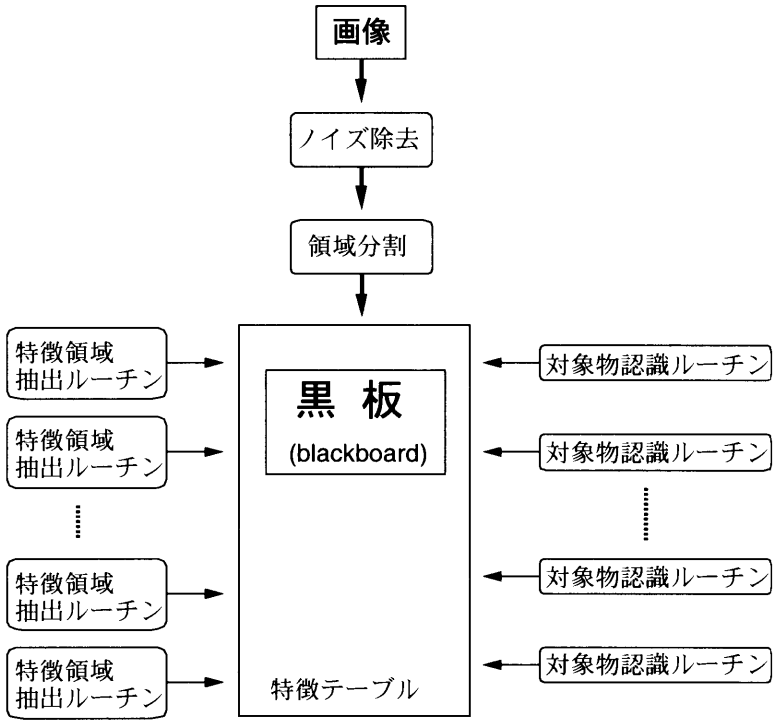

図 3.7 黒板モデルによる処理

テムになっている．つまり，松山らによるシステム [90] における特徴検出モジュールと対象物認識モジュールを 1 つにまとめたようなものである．そして，各スキーマは内部にもローカルな黒板を持ち，それ自身がさらに黒板システムになっているという，**階層的な黒板モデル**を提案している．

The Schema System では，物体の仮説が生成されるとその物体のインスタンスが動的に生成される．インスタンスは黒板を介して互いに通信を行ない物体の存在の可能性を高めようとする．このため，このシステムは空間分散協調方式であるとも言える．なお，このインスタンスには物体だけでなく，例えば，road-scene instance の様にシーン自体もインスタンスになり，これによってシーンの文脈の利用が実現されている．

他の機能分散協調方式の例としては，テーブル上の食器の認識 [93]，階層的なマルチエージェントシステム構成によるカメラ制御を含めたアクティブビジョンシステム [94, 95]，認識システムにおける高レベル，中レベル，低レベルのそれぞれの処理を独立したエージェントとし，工業部品の認識を実現したシステム [96, 97, 98] がある．ただし，これらの研究では，従来モジュールと呼んでいたものを単にエージェントと呼び替えているに過ぎず，第 3.4節で述べた分散協調型のシステムの特性を十分生かしたものとはなっていない．

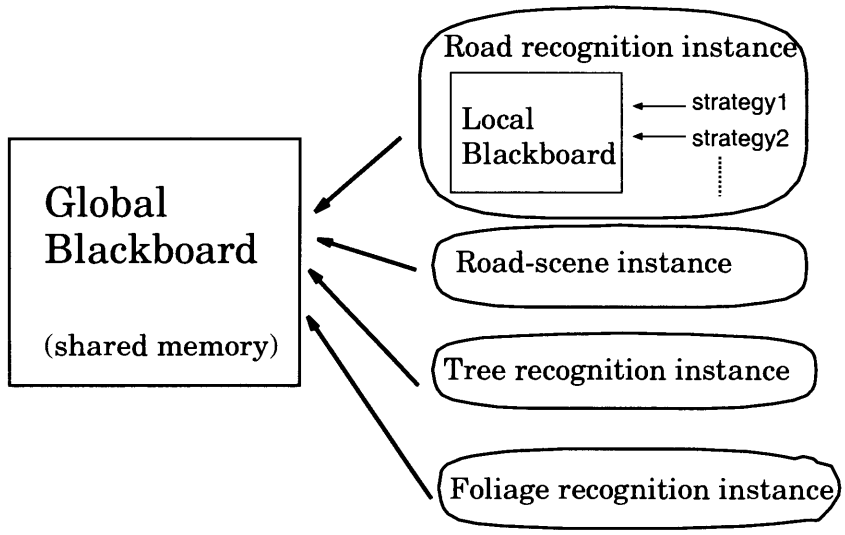

図 3.8 The Schema System

3.5.2 空間分散協調方式のシステムの例

松山らによる航空写真認識システム SIGMA[28, 32] は，空間分散協調方式の最も古典的なシステムで，空間推論を行なうことによりボトムアップとトップダウンの自然な融合を実現している．

SIGMA[28, 32] では，画像中の認識対象物の空間的な関係に関する知識に基づいて，画像の構造を認識，推論するための処理方式として空間分散協調方式を利用する画像理解システム SIGMA が提案されている．SIGMA 自体は汎用的な画像理解システムであり，実際のアプリケーションとして航空写真の認識を行なっている．

SIGMA では，画像から認識された対象物は，独自の知識に基づいて推論を行なうエージェントとしてみなされ，多数のエージェントが互いに協調しながら空間推論を行ない，シーンの構造を抽出する．

システムの構成は，図 3.9のように 3 つのエキスパートと呼ばれるモジュールからなっている．それぞれ，低レベル処理エキスパート，モデル選択エキスパート，空間推論エキスパートと呼ばれている．高次のエキスパートが 1 つ下のエキスパートを呼び出す形で，処理が行なわれる．SIGMA では，システムの構成要素に関してはエージェントという言葉は用いておらず，画像中の要素 (認識対象物) をエージェントと呼んでいる．

具体的な処理としては，最初に，低レベルエキスパート，モデル選択エキスパートによって，認識し易い対象物がいくつか認識され，対象物 1 つにつき 1 つのエージェントが生成される．このエージェントは，実際には対象物の種類に応じて用意されたクラスのデータ構造のコピー，つまり，オブジェクト指向言語でいうところのインスタンスである．このインスタンスには，認識された対象物が持つ具体的な属性値が書き込まれ，一度インスタンスが生成されると，後はエージェントとして自律的に活動を始める．

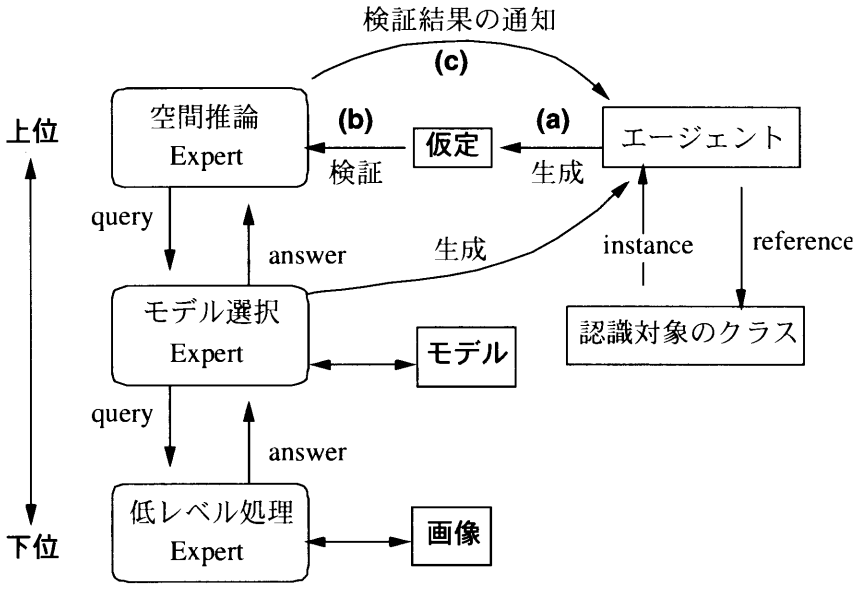

図 3.9 SIGMA のシステム構成

エージェントが推論に用いる知識は，それが属するクラスの中にルールとして含まれている．ルールは，条件，仮定，実行内容の 3 つの部分からなっている．エージェントは，条件が満たされているルールを見つけると，そのルールに従って，仮定を生成する (図 3.9(a))．生成された仮定は，空間推論エキスパートによって，検証され (b)，結果がエージェントに通知される (c)．結果を受けとったエージェントは，結果に応じてルールに書かれた実行内容を実行する．なお，空間推論エキスパートは，モデル選択エキスパートを呼ぶことによって，実際の処理を行なっている．

次に，この空間推論によってトップダウン処理とボトムアップ処理が自然に実現される例を見てみる．エージェント s が，空間的関係 REL に関するルールを起動して，f(s) という仮定をしたとする (図 3.10)．すると，空間推論エキスパートは，仮定 f(s) がエージェント t に当てはまるかどうか検証し，仮定が正しければ，エージェント s とエージェント t の間に関係 REL が認められたことになる．つまり，これで空間の構造が一つ認識できたことになる．この一連の推論過程が，ボトムアップ解析によるシーン構造の抽出ということになる．これによって，例えば，家が複数個あれば，住宅地という上位の構造が発見できたりする．

また，図 3.11では，エージェント s とエージェント u が，整合性がある 2 つの仮定 f(s),h(u) を行なっている．その場合，空間推論エキスパートは，1 つに統合して処理しようとするが，仮定に当てはまるエージェントが存在しない．そこで，空間推論エキスパートは，下位のエキスパートに，統合した仮定を満たす対象物の検出を要求する．要求を受けたエキスパートは，通常よりもスレッショルドを下げて，画像から目的の対象物に対応する画像特徴を探索する．そして，仮定を満たす対象物が検出出来た場合，エージェント t を生成し，エージェント s と t の間に関係 REL1，エージェント u と t の間に関係 REL2 が結ばれる．これが，SIGMA におけるトップダウン解析である．例えば，2 つの家がその間の未認識領域を家であると仮定することによってトップダウン解析が起

動される (図 3.12).

以上のような**物体間の位置関係から空間推論によるボトムアップとトップダウンの自然な融合の実現**が，空間推論を行なうエージェントを用いたこのシステムの最も大きな特徴である.

他の空間分散協調方式の例としては，画像の領域分割を目的とした，分散協調領域分割 [99, 100]，文書画像の文字領域の切出し [101]，顕微鏡写真からの細胞の切出し [102] などのシステムが挙げられる．領域分割以外の空間分散協調方式の利用としては，線画解釈において線や面などの物体の要素をエージェントとし，エージェント間で整合性をチェックすることによって，全体としての解釈を行う研究 [103] がある.

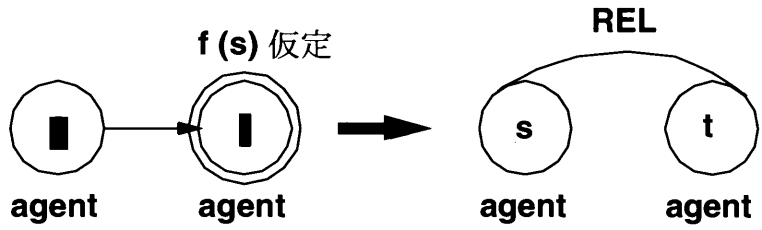

図 3.10 ボトムアップによる構造抽出

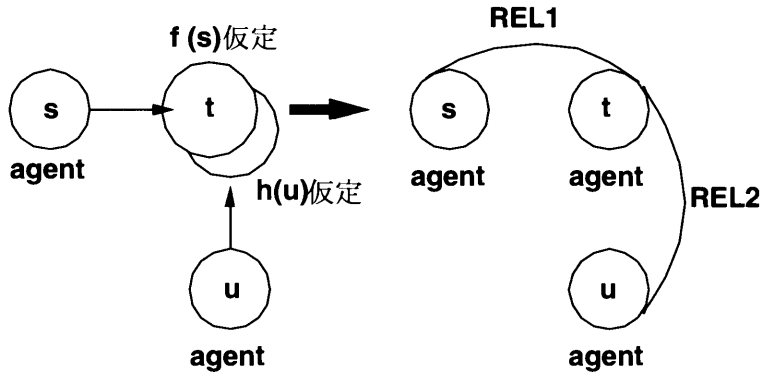

図 3.11 トップダウンによる構造抽出

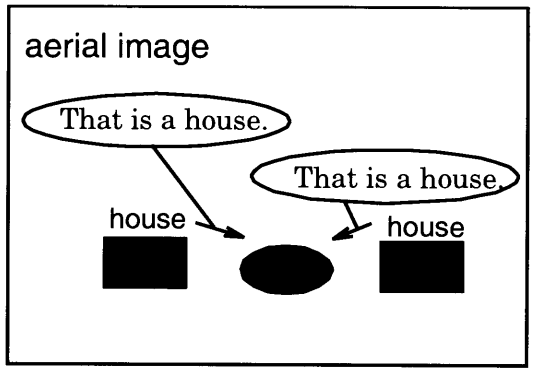

図 3.12 トップダウンによる構造抽出の例．２つの家がその間の未認識領域を家だと仮定している．

3.6 実世界画像に対する認識システム実現のための考察

本節では，第2章で述べた従来の物体認識手法，および本章で述べた物体認識システム構成方法を踏まえて，実世界画像に対する認識システム実現のための考察を行う．

3.6.1 実世界画像の認識の困難点

ここでは，実世界画像に対する画像認識システムの実現が難しいといわれている理由である３つの不完全性 [84] について説明する．

実世界に対する物体認識においては一般に次の３つの不完全性が存在するといわれている．

- 画像の不完全性
- 知識の不完全性
- 処理の不完全性

画像の不完全性とは，画像は常にシーンを完全に表現している訳ではないということで，あくまでも画像はある空間中の一点におかれたカメラによって写し取られたシーンの断片に過ぎないということである．本研究においては，１枚の画像のみについての物体認識を目的としているが，１枚からでは３次元の奥行き情報を得ることは難しく，奥行き情報の欠落という，大きな不完全性がある．また，照度不足や，量子化する際の誤差の問題なども含まれる．

知識の不完全性とは，どんな認識対象も完全に知識表現できる訳ではなく，また，すべての対象を完全に知識として与えることも不可能であるということである．特に，本研究のように一般名称で対象を認識する場合，一般名称の表す概念を明示的にモデル表現することが必要であるが，それは一般に困難な問題であり，完全に表現することはほとんど不可能と言ってよい．また，対象が限定された認識システムを作る場合でさえも，どんな認識対象も数値データによって厳密に表現できるものではなく，ある程度のずれをもってしか表現しかできないということを常に考慮しておく必要がある．

処理の不完全性とは，あらゆる対象に適用できるような完全なアルゴリズムが存在しないということで，これを補うためには複数の方法を統合していくことが必要不可欠である．

このように実世界に対する物体認識においては，対象，手段，知識のどれも不完全性を持っており，完全な情報下を前提とした “計算機” であるコンピュータで物体認識を実現することは極めて困難な問題であるといえる．このことは，物体認識に限らず，実世界を対象にした計算機システムの構築，いわゆる “リアルワールドコンピューピューティング” 全般の問題であるといえる．

以上の様にすべてが不完全であるので，どのような画像認識システムにおいても完全な認識結果は得られない．それは人間の視覚も同じことであって，“見間違え” などということはよく起こることである．つまり，物体認識で実現するべきことは，このすべてが不完全な状況の中から，いかに信頼できる，つまり最も可能性が高い結果を導き出すかということである．

また，実世界を対象にしたシステムを実現した場合，実現した認識システムの評価が困難であるという問題もある．これは，実世界の画像が無限の視覚的変化を持っているため[1]に，評価する時に用いた画像によって，認識結果が大きく影響を受けてしまうというということである．こうした問題に対して，顔画像認識の場合は，論文を発表する際に評価に用いた画像セットを公開することが行われており [42, 43]，それを用いて後から発表された論文で同じ画像セットを用いて比較がおこなわれている [44]．しかし，こうしたことは，比較的多くの研究が行われている人の顔の認識においてのみ主に行われていることであり,「机」「椅子」などの顔画像以外の認識においては，標準的な評価画像というものは存在していない．ただし，画像データベース的手法による認識の研究においては，6 万枚の著作権フリーの画像を含んでいる Corel 社の Corel Image Library が事実上の標準評価画像となっている [55, 56]．Corel Image Library は枚数は多いものの，様々な画像が含まれているので，同一のカテゴリーに含まれる画像は多くても 100 枚程度に過ぎない．しかも，プロの写真家が撮影した整った画像のみを集めていて，同じカテゴリに含まれる画像は同一のカメラマンが似たような構図で撮影した場合が多く，必ずしも実世界画像に対する認識システムの評価に適した画像であるとは言えないという問題がある [14]．

以上のような問題が存在することを踏まえた上で，様々な対象を認識可能な general-purpose な認識システムを実現するには，どうすれば良いか，多種類を認識出来る一般性 (generality)，多様な状況において対象を認識出来る柔軟性 (flexiblity) の 2 つ観点から以下では議論を行う．

3.6.2 多種類の物体の認識

一般性：多くの種類の対象を認識できること．

多くの場面で認識できることとは，つまり，多目的に使える general-purpose なシステムであるということである．そのためのシステムを構築するには，ある目的にのみ適用できるようなシステムの設計の仕方でなく，できるだけ幅広い対象領域において適用できるシステムを設計しなければならない．

認識の対象画像には様々な種類のものが考えられる．例えば，室内の画像，ビルの建ち並ぶ屋外，

[1] 320×240 の大きさの 256 階調濃淡画像には，$256^{320\times240} \approx 10^{184953}$通り ものパターンが存在する．

自然の風景，歩いている人間，レントゲン写真，回路図面など，その種類はほとんど無限にあると言える．人間もこれらのすべてが理解できる訳ではない．レントゲン写真，回路図面などは専門の知識がなければ，見ても何を意味するかさっぱり分からないだろう．こう考えると，画像認識において一般性を実現するには，多くの知識を与えて置くことが必要であるといえる．

その場合，知識表現が問題になる．つまり，すべての画像中のすべての対象を表現できるような統一的な表現方法は存在するのかという問題であるが，これに対しては，それぞれの対象によって異なる適した知識表現，認識手法を用いる方が都合がよいことが多い．例えば，常に一定の形状を持つものなら幾何モデルで表現すればよいし，様々な形状を持つものであればそれらの共通な特徴を表すような概念的なモデルを用いるべきである．

この様に知識表現を統一せずに，それらを統合して扱うには，第3章で紹介した様な分散協調型もしくはマルチエージェント型のシステム構成法を用いるのが適当であると考えられる．このシステム構成法では，サブシステムの集合体としてシステムを構成するので，それぞれ違った知識表現，認識アルゴリズムを持つ special-purpose なサブシステムを統合することができる．統合の方法についての問題は残るが，システム構築の方針としては，一般性は special-purpose な知識を大量に与えることにより，基本的には実現可能であると考える．つまり，個々に対する認識手法は対象に依存していた方が一般的なものに比べて高性能な物ができるので，それらを統合することによって一般性を出した方が結果的に高性能になる．

このことを物体認識についてさらに考えた場合，サブシステム，つまりエージェントの単位をどうするかという問題がある．システムの目標は物体の認識であるので，物体 1 つについて 1 つのエージェントとするのが自然であると考える．従来の分散協調型の研究でも，物体 1 つが 1 つのエージェントになっていることが多かった．ここで物体 1 つというのは，classification の場合，一般名称 1 つが指す範囲の物体である．もし，1 物体名称を担当できるエージェントを構築できるとするならば，エージェントを数多く，できれば，辞書に出ているくらいの数のエージェントを用意することによって，general な物体認識システムの実現が可能である．しかし，実際には 1 物体名称を担当できる柔軟なエージェントを作ることは簡単ではなく，後程述べるように学習を用いて構築を試みることが必要となってくる．なお，物体 1 つに 1 つのエージェントを割り当てる考え方は，M.Minsky の『心の社会』[104] に通ずるものがある．

各物体それぞれに関する知識はこのようにエージェント毎に異なる知識表現で与えておく方が合理的であるが，同時に複数物体に関係する知識，つまり，物体間の関係知識をそれぞれのエージェント毎に独自の表現方法で与えておくことは合理的といえない．もし，ある一つのエージェントがほかのすべての関係エージェントとの関係知識を持っていなくてはいけないならば，エージェントを一つ追加する度に，既に存在しているエージェントにもその新しいエージェントとの関係知識を追加してやる必要が出てくる．そこで，関係知識を統一した表現で与えてやることによって，エージェント間で関係知識の交換が可能になる．そうすれば，新しくエージェントを追加した場合に，新しいエージェントが既に存在しているエージェントに対して自分との関係の関係知識を教えてやることができる．つまり，物体内のみの知識はそれぞれのエージェントが独自に表現し，物体相互の間の知識は統一したものとすることが，一般的なシステムを構築するためには必要であると考える．

このような考え方に対して，special-purpose なものの集まりは所詮 special-purpose の集まりに過ぎず，general なものとはいえないという批判があるが，本研究の目的とする classification を行なう，つまり一般名詞を認識する物体認識に関していえば，認識対象の普通名詞は高々辞書に出ている程度の有限個であるので，それだけの数の認識エージェントを作ることができれば，general なシステムが構築できると我々は考えている．

3.6.3 多様な状況での物体の認識

柔軟性：多様な状況において対象を認識できること．

柔軟性の問題は実画像に対する認識における問題点そのものである．つまり，ノイズがのっていたり，対象の一部分が他の対象によって隠されていたり (オクルージョン)，また，3次元の物体が視点の方向によって見え方が変わるというような問題に対応することが，柔軟性の実現には必要である．

従来の物体認識の研究では，その対象とする画像に条件を付けることが多かった．例えば，オクルージョンがなく認識する物体が完全に見えていることを前提条件とすることがあるが，そのような前提条件は実世界の画像を認識する場合，満たされることは極めて少なく，そのような前提条件をつけている限り，柔軟性のある認識を実画像に対して実現することは不可能である．つまり，柔軟性を実現するためには，実画像に対して条件を付けることなく，認識ができなくてはいけない．以下に，実画像特有の問題点を挙げる．

- 照明条件が変化する．
- 隠れ (オクルージョン) がある．
- 画像の端で物体が切れる．
- スケールが分からない．
- 見る方向によって，見え方が変化．
- 1種類の物体でも，形が多様．

これらの問題を克服することが，柔軟性の向上につながると考えられる．次に，これらの問題について個別に触れることとする．

照明条件の変化 同一物体を同一方向から見ていた場合でも，照明条件の変化によって，見え方 (view) は変化する．そのため，物体の境界やテクスチャが不明瞭になり，物体認識のための手がかりとなるエッジや領域などの特徴の抽出がうまく行なえないことがある．

対策としては，完全なエッジ抽出，境界線抽出，領域分割を期待せず，局所的な特徴に対して，予想される形状の大域的なトップダウン的当てはめや，SNAKE[105] などのある程度形状を仮定した抽出方法を用いることが考えられる．また，濃淡画像では濃淡値がほぼ同じであっても，色調は異なることが多いので，カラー画像を用いることも有効である．

隠れ (オクルージョン) がある 実世界には多数の物体が存在しているので，1 枚の実世界画像に複数の物体が含まれている状況は普通であり，したがって物体が視点に対してその後ろにある他の物体の一部を隠してしまうオクルージョンが起こるのはよくあることである．

対策としては，手前の物体から認識し，後ろの隠れている物体は部分的な形状と，手前にある物体との関係などから認識することが考えられる．これも，ある程度形状を仮定して認識することが有効である．

画像の端で 物体が切れる これは画像がシーンの一部分を切りとっている以上仕方ないことであるといえる．

部分的な形状，他の物体との関係でできるだけ認識するが，物体の大部分が切れている場合には認識不可能である．

スケールが分からない スケールが分からないとは，近くの小さいものと，遠くの大きいものの区別が付かない．これは 3 次元を 2 次元の画像で見ているのだから仕方のないことである．

物体の絶対的な大きさが分からないということは，人間が 1 枚の写真を見る時にも起こることである．例えば，世界最小の○○の新製品といっても，その製品だけの単独の写真を見せられてもどれくらい小さいのか分からない．普段なじみの深いもので大きさがよく知られているような，例えばタバコの箱などと比較してもらえると，その相対的な大きさの関係から，その新製品の小ささがよく分かる．このように，物体の大きさ (スケール) を知るには，他に存在している物体との相対的な大きさの関係が重要な手がかりとなる．

見る方向によって 見え方が変化 これも 3 次元を 2 次元の画像で見ているために起こる問題である．

対処の方法としては，見る方向によらない特徴的な形状で認識する，何通りか典型的な見え方を与えておく，3 次元モデルを利用することなどが考えられる．人間は見慣れた方向から見る時は 2 次元的な照合を行ない，見なれない方向から見た時は頭の中で 3 次元的なイメージを回転されることによって照合を行なう，ということが心理実験より明らかにされている [106]．したがって，2 次元的な認識と 3 次元的な認識を統合することは，柔軟な認識を行なう上では必要である．

同一種類の物体でも 形が多様 これは第2章で触れた classification の問題で，今までの問題点と別の次元の問題であり，認識する対象の定義の問題に係わってくる．本研究での認識する対象の物体を一般名称で表現されることにしたために起きた問題である．認識の対象の定義が厳密な形状によってなされるならば，このような問題は起きないが，本研究において問題にする一般名称物体の定義は，例えば人工物の場合その機能や目的などによってなされている．例えば，机なら「物を載せる台で人がその上で作業を行なう.」といった具合である．したがって，本当に柔軟な認識を実現するには，この定義を満たすかどうかによって，認識を行なうべきである．

第2章でも触れたが，この問題に対処するには，その物体の本質を表すような形状，物体の構造の解析，他の物体との関係などから，機能を推測し認識する function-based の考え方を用

いることが，最も本質的な方法である．それ以外にも，考えられる形状をすべてモデル化する方法や，代表的な形状をモデル化する方法，形状以外の色やテクスチャ，構造などを認識の手がかりとして対処する方法もある．

以上をまとめると，

- ２次元的，３次元的認識の両方の利用
- 物体間の関係情報の利用
- 形状を仮定したトップダウン的な抽出方法
- 部分的な形状，特に本質を表す形状に注目

などの方法が，実画像に仮定を付けない柔軟な認識を実現するには，有効であるといえる．

このうち，１番目は複数の手法の組合せの問題で，２番目の関係の利用は物体間の問題である．この１番目，２番目はマルチエージェントアーキテクチャを採用することによって，実現が容易になる．また，３番目，４番目は物体個々の認識の手法の問題で，認識エージェントの作り方の問題になる．

また，柔軟な認識モジュールを人手で作ることは困難なことであるので，学習を導入することが考えられる．その場合，予め対象に応じて，モジュールを作り込んでおいて，それから，学習するという方法が考えられるが,「顔画像」などのごく少数の例外を除いて，その方法は確立されていない．そこで，本研究の後半部の第9章においては，対象を想定しない一般的な方法を用いて，学習によって，special-purpose な認識モジュールを構成することを行う．学習に際しては，出来るだけ柔軟な，つまり，認識対象物体についての多様なビューに対応できるように，多くの学習画像を用意し，それを学習画像として利用した．しかしながら，実際の問題として，多くの種類の物体について，それぞれ多様なビューの画像を収集することは困難である．人間は非常に長い年月を掛けて，様々な物体の視覚的特徴を学習していくのに対して，人間の視覚と同様の柔軟性を持った認識を実現するには，それと匹敵する量を視覚的情報を学習させてやる必要がある．そこで，WWW(World-Wide Web) から学習画像を自動収集することを行う．WWW は巨大なコンピュータネットワークであるインターネット上の情報空間であり，そこには実世界の情報が電子的な形で蓄積されている．その中には，視覚情報や画像情報も多量に含んでいる．WWW にはテキスト情報を伴った画像データが大量に存在しており，テキスト情報を解析することによって，かなり高い精度で大量にある特定の種類の画像を収集することが可能である．そこで，WWW から特定の物体名称に対応する学習画像を大量に集め，それを利用して認識を行うことによって，さまざまなビューに対応した画像認識の実現を目指す．

3.7 まとめ

本章では，初めに，ボトムアップ処理とトップダウン処理の融合と，アルゴリズム統合がシステムの柔軟性の向上には不可欠であることを述べ，その後，その２つを実現するためには，マルチエー

ジェントによる分散協調型のシステム構成が適していることを述べた．分散協調型のシステムには，機能分散方式と空間分散方式があり，それぞれ，アルゴリズム統合と，ボトムアップ処理とトップダウン処理の融合に適している．その後，機能分散方式分散協調型のシステムの例として４つ，空間分散方式の例として１つのシステムをそれぞれ紹介した．

また，実世界画像に対する認識システム実現のための考察をまとめた．その結果，実世界画像に対する物体認識において一般性，柔軟性を向上させるには，マルチエージェントアーキテクチャを用いることによって，多様な認識手法を組み合わせることが，有効な方法であると考える．

そこで，第4章で述べるシステムでは，第3.6.2節，第3.6.3節で述べた，以下の様なことの実現を目指す．

- １エージェントに１物体名称の認識を担当させ，物体毎に異なる表現知識，認識手法を用いる．
- 物体相互の一般的関係を知識として予め与え，認識に利用する．
- 物体の全体の形状や構造ではなく，物体の機能を提供する形状や構造などの，物体の本質的な形状や構造をモデル化し，それらを画像中から探して，認識の手がかりとする．

第 4 章

マルチエージェント型画像理解システムの提案

実世界に対応した画像理解システムを実現するためには，多種多様な画像に対して認識ができることが必要である．そのために，従来の様に認識対象となる画像の種類を予め想定してシステムを構成するべきではなく，多様な画像，認識対象に対して対応できる柔軟かつ拡張性のあるシステム構成とするべきである．そこで，本章では，そうした認識を実現するためのシステムの新しい構築方法として，単一種類の物体のみを認識する認識プログラム (**認識モジュール**) を複数用意し，それぞれに通信機構 (**通信モジュール**) を付加することによってエージェントを構築し，その集合体として認識システムを構築する方法を提案する．なお，ここで前提とするシステムの入力は単一の濃淡画像であり，出力はその濃淡画像中に含まれる一般名称を持つ物体の名称と画像中におけるその物体の占める領域である．

4.1 はじめに

従来の実画像を対象とした画像理解システムの研究では，ほとんどの場合，認識対象とする画像を予め想定していた．例えば，Nagao[91], SIGMA[32] は航空写真，Ohta[30], The Schema System[31] は風景画像をそれぞれ対象としていた．対象画像を限定しておくことよって，実画像特有の困難な問題を予め想定してシステムを構成することが可能であった．しかし，認識対象を限定することがない，より人間に近いシステムを構築するためには，システムの構成法自体に予め限定された対象の性質を反映させるべきではない．

そこで，このようなシステムの構成法として，画像を直接扱う部分と，記号的な処理を行なう部分に分けてシステムを構成することによって，対象画像に依存しないシステムを実現することを提案する．基本的な考え方としては，単一種類の物体を認識する認識プログラムを複数個用意し，それぞれに協調機構を付加することによって，プログラム同士が相互作用を行なえるようにし，その集合体としてシステムを構築するということである．これを実現するために，マルチエージェントによってシステムを構築する．

従来のマルチエージェントによる画像理解システムでは，複数アルゴリズムの統合を目的としていたり [87]，空間構造の柔軟な利用を目的としていた [83] が，エージェント間の相互作用を個々に考える必要があったり，対象の階層性を利用しているために，システムの構築が非常に複雑であった．そのため，認識対象を限定していることが多く，我々の目標とするような一般的な認識システムを構築する場合には現実的でなかった．しかし，本システム構成法では，単一種類の物体を認識する認識プログラムをシステムの他の部分とは独立に構築し，それらを共通の協調機構を付加して統合するだけなので，従来より容易に一般的な認識システムが構築可能であるという特徴を持つ．

4.2 協調に基づく認識

従来の実世界画像に対する物体認識の研究では，システムが認識する範囲を真に認識すべき範囲に近づけるように努力が行われてきた．例えば，図 4.1のように，システムが「椅子」「机」と認識する範囲を，なるべく真の「椅子」「机」の範囲に近づけようとしてきた．しかし，ある程度以上近づけようとすると，物体の形の多様性のために，認識方法が複雑になったり，対象物体に関する多くの知識をシステムに与えたりする必要が出てきて，それは容易ではなかった．

これに対して本章で提案するマルチエージェントによる協調型のシステムでは，単一クラスの物体の認識を行う認識モジュールは，物体の形の多様性に対応するため，真の「椅子」「机」の範囲を含む広い範囲を認識するような大雑把な認識をする (図 4.2)．そのため，モジュール間で競合が起こりやすくなるが，その競合は，物体の位置関係の知識などに基づいて解決する．つまり，個々のモジュールだけで認識を完全に行うのではなく，個々は簡単な認識のみをし，その能力の不十分さをエージェント同士の協調によって補うことを狙いとするシステムの構築を目指す．

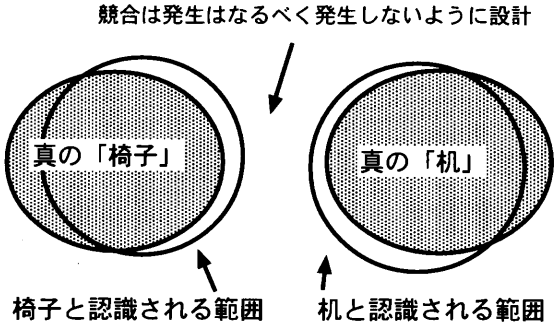

図 4.1 従来の認識．なるべく真の「椅子」「机」を認識できるようにそれぞれのモジュールを設計．

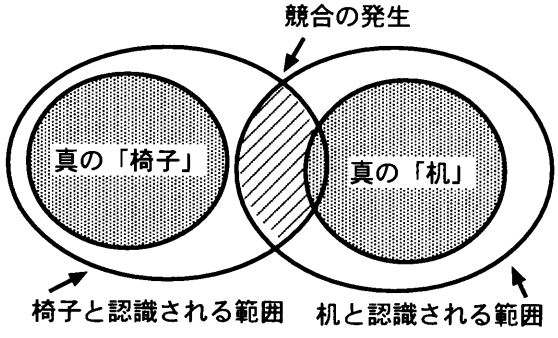

図 4.2 本システムの方針．各認識モジュールは真の「椅子」「机」より広い範囲を認識．競合が発生するが，協調によって解決．

このように本システム構成法では，「複雑な」実画像に対応するためのシステムを，複数個の「簡単な」画像認識モジュールを統合することによって，シンプルに構築することを目的としている．

4.3 システムの概要

本章では認識システムを，単一種類の物体のみを認識するエージェントの集合体として，マルチエージェントシステムとして構築する．エージェントは，画像に含まれているある一つの同一の一般名称をもつ対象，例えば,「椅子」などのような物をすべて認識する認識システムに，他のエージェントとの協調を行なう通信システムを付加したものである．我々はここで提案する画像理解システム構成法を **MORE (Multi-agent architechture for Object REcognition)** と呼ぶこととする．

システムは，物体個別の認識エージェントの集合体になっている (図 4.3)．認識エージェント以外にもホストモジュールが存在するが，これは単に，ユーザから与えられた画像をすべての認識エージェントにブロードキャスト (broadcast) する，物体間の関係に関する知識をファイルから読込んで関係する認識エージェントに送る，そして，すべての認識エージェントの処理が終了した後に結果を集め，ユーザに提示するという，ユーザインタフェース的な役割のみを果たしている．従って，中央集権的な機構は存在しない．また，本システムでは，エージェントが非同期的に動作し，通信のみによって情報のやりとりを行なっているために，非同期通信型マルチエージェントシステムになっている．非同期の通信型システムでは，システム全体に共通な「時間」が存在しないため，マルチエージェントシステムのなかでも特に解析が困難とされているシステム [107] である．本システムでは，後述する様にそれに対する対策をいくつか行なっている．

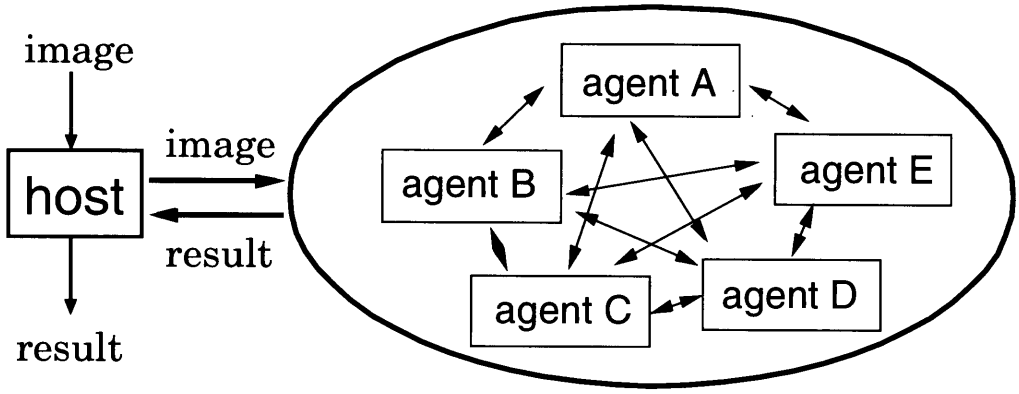

図 4.3 システムの構成

認識エージェントは，それぞれのエージェントが独立して，ホストモジュールより送られた入力画像に対して認識を行なう．各エージェントはそれぞれ異なる処理を行なうことになっており，ここでは，各エージェントの認識方法については特に規定しない．また，物体個々の認識に必要な対象に関する知識は統一知識表現を定めていないので，各認識エージェントがそれぞれの表現法で内部に保持している．

各認識エージェントは，それ自体が独立した認識システムであり，自分の対象とする物体をでき

るだけ多く認識できるように動作する．しかし，画像には「1 つの画像の領域は 1 つの物体に対応する」という原則があるために，複数の認識エージェントが同じ領域を異なる物体として認識してしまうと，結果の競合が発生し，システム内で矛盾した結果が存在することになる．そこで，競合を解消するためにエージェント間で交渉が行なわれる．このエージェント間の交渉によって，常にシステム内のすべてのエージェントの認識結果が整合性が保たれている．

このようなマルチエージェント型のシステム構成を採用した場合，全体を制御する機構がないために，認識エージェントの追加のみで認識可能な物体の数を増やすことができるという特徴がある．

本システムは以下のような特徴を持っている．

- 通信のみによるエージェント間での情報交換．
- エージェントを認識モジュールと通信モジュールによって構成．
- エージェント間の競合解消に関係知識を利用．

それぞれについて，詳しく説明する．

4.3.1 通信のみによるエージェント間での情報交換

本システムでは，共有メモリを用いた黒板システムとは異なり，各エージェント間の情報交換はすべて通信によって行なわれている．そのため，各エージェントの認識結果は，認識したエージェントがその他のエージェントへ向かってブロードキャストすることによって，他のエージェントに伝えられる．その結果を受けとったエージェントは，それがすでに認識されている自分自身の物体結果と矛盾していないかチェックし，もし矛盾していなければ，そのままにし，矛盾していれば，矛盾する結果をブロードキャストしたエージェントに対して異義メッセージを送る (図 4.4(b))．そして，2 つのエージェント間で，競合解消のための交渉がおこなわれ，どちらか取消しか両立の決定が行なわれ矛盾が解消される．こうして，認識結果をブロードキャストした方としては，異義メッセージが送られて来ない場合，もしくは競合解消によって取消されなかった場合は，承認されたとみなすことができる (図 4.4(a))．しかし，一度承認されたとしても，その後に新たな認識結果が他のエージェントからブロードキャストされ，競合が起こった場合には取り消されることもある．このようにして，常にすべてのエージェント間で認識結果の整合性が保たれるようになっている．

また，通信のみによって協調が行なわれ，共有メモリ的なものを一切用いていないので，分散メモリ型並列計算機上でのインプリメントが容易である．実際，本システムは分散メモリ型並列計算機 AP1000+上にインプリメントされている．

共有メモリを用いる場合に並列処理を導入するとなると，共有メモリのアクセスの制御に複雑な制御機構が必要であるが，本システムのようにすべて通信で行なう場合には複雑な機構は必要なく，各エージェントが統一された通信プロトコルと，情報の非同期性に対応していれば良い．なお，非同期性への対応については，第 4.8節で述べる．

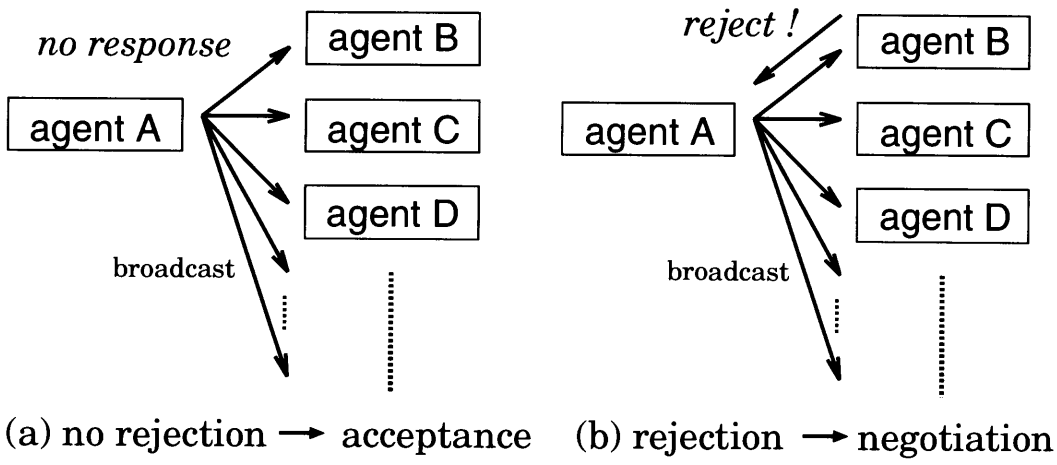

図 4.4 (a) 認識結果が他のエージェントに承認された場合 (b) 異義が出た場合

4.3.2 エージェントを認識モジュールと通信モジュールによって構成

各認識エージェントは他のエージェントとの通信を行なう通信モジュールと，入力画像に対する認識を行なう認識モジュールの 2 つから成り立っている (図 4.5)．互いの情報交換は，他のエージェント間と同じように，すべて通信によって行なわれる．

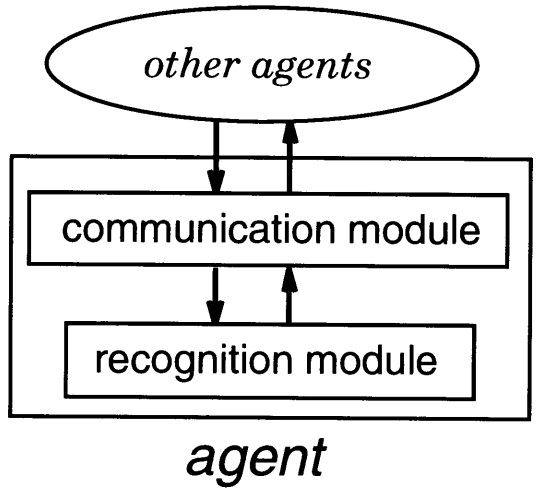

図 4.5 各認識エージェントの内部構成

このように，画像から物体を直接認識する認識モジュール部分と，認識結果と他のエージェントの認識結果の整合性をチェックする通信モジュール部分に分けることにより，認識モジュールと通信モジュールを並列に動作させることができる．そうすることにより，認識モジュールが認識を行なっている間でも，通信モジュールは他エージェントの認識結果との競合をチェックし，競合解消のための交渉を行なうことができる．

また，このように 2 つのモジュールに分かれていることによって，認識モジュールが物体個々の認識に直接必要な知識のみを取り扱い，通信モジュールが物体相互の間の関係に関する知識を取扱うというように，物体固有の知識と関係の知識が別々のモジュールで扱われている．これは，認識モジュールが物体知識を用いた認識のみに専念し，関係知識の利用に必要な他のエージェントの認識結果をブロードキャストによって受けとっている通信モジュールが関係知識の取扱いを行なうというように分業されているということを意味している．

これら 2 つのモジュールは，画像に対して直接認識を行なわずにその結果のみを取り扱う通信モジュールはすべてのエージェントで共通なものとし，認識モジュールのみをそれぞれの物体に適した方法で構築する．こうすることにより，特定物体の認識システムである認識モジュールを構築するのみで，エージェントの追加が実現できる．なお，物体間の関係知識は，システムに 1 つ存在する関係知識ファイルに納められ，これに新しく追加するエージェントに関係する関係知識を追加すればよい．このように，通信部分を各エージェントで共通なプログラムを用いることにより，容易にシステムの拡張が行なえるようになっている．

また，通信モジュールと認識モジュールの間の情報交換が通信で行なわれているため，共有メモリ使用時のような複雑な制御機構を追加せずに，1 つのエージェント内に複数の認識モジュールを用意し，状況に応じて使い分けることも可能である．例えば，物体が認識し易い状況では 2 次元的な簡単な認識モジュールで認識を行ない，複雑な状況では 3 次元的なモデルを用いた認識を行なって，さらに，完全に未知の形状に対してはその機能を推測するような高次の認識を行なうようにすることも可能である．つまり，エージェントをさらにマルチエージェントによって構築することも可能である．本システムのアーキテクチャはこのようにさらに発展させることのできる可能性を持っている．

4.3.3 エージェント間の競合解消に関係知識を利用

各認識エージェント間で認識結果に不整合が発生した場合，両エージェント間で交渉を行ない，競合を解決する．その時に，単純に結果の比較を行なう訳ではなく，物体間の関係の知識も利用して解決を行なう．例えば,「机」と「本」が競合し，両者の認識結果の形状が明らかに異なる時，“**book on desk**” という関係知識が存在していれば，両者は矛盾がなく，両立することとみなす (図 4.6)．このように，関係知識を用いることによって，一見矛盾している結果でも，実は整合性がとれていることが分かる．この場合に様に関係知識によって両立した場合は，小さい領域の方が手前にあることとする．

こうした関係知識を利用した矛盾の解消によって，小さい物体が前にあるために後ろの大きな物体の一部が隠れている時，つまり，オクルージョンが発生している時に，そのオクルージョンを無視するような認識結果が出力されても，手前にある小さい物体と両立する関係があるならば，矛盾が起きない．こうすることによって，図 4.6(a) の様に，机の領域を，上に載っている物体を無視して，本来の机の形状のまま認識することができる．

また，一方，両者の間に関係がない時は，両者を，各認識モジュールが自分自身の認識結果 (物体候補) に対する確信度を自己評価した値である**形状評価値**と，他のエージェントの認識結果との

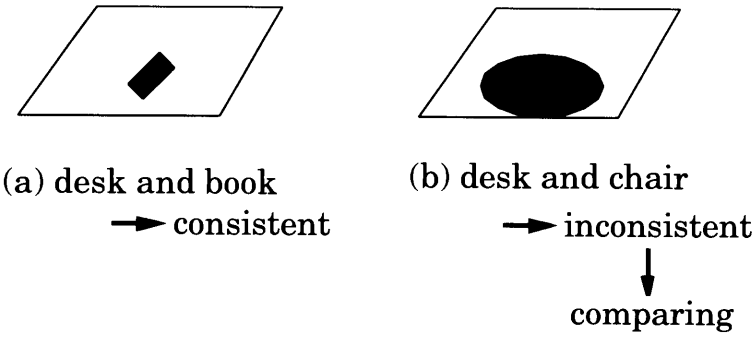

図 4.6 (a) 関係知識によって両立する場合 (b) 両立せずに比較が行なわれる場合

関係を関係知識によって評価することによって求められた**関係評価値**と，場合によっては両者の面積の大きさを利用して，比較を行なう．そして，どちらか一方を正しい結果と判定し，他方を取り消す．この比較を行なう前に，予め関係知識を用いて関係評価値を計算し，比較の材料とする．結果の比較の方法に関しての詳細は，第 4.6.4節で触れる．

他にも物体間の関係知識は，トップダウン的な認識モジュールの呼び出しや，一度取消された認識結果の復活などに利用される．それぞれ，第 4.7節，第 4.6.5節で，詳しく触れる．なお，これらの処理に利用される関係知識はすべて共通のものであり，それぞれの処理向けに用意されるということはない．

本システムで用いられる関係知識は，２物体間の相対的な関係を記したもので，２物体の通常考えられる関係を記したものである．人間は，関係知識をいわゆる「常識」として扱っていて，辞書などには通常出ていない知識であり，正確には定義できない曖昧な知識である．しかし，認識には有用な知識である．具体的には，２物体間の位置関係，大きさの関係，形状の複雑度の関係，濃淡の関係など，様々な関係がある．本システムにおいては，これらの関係はすべて相対的な定性的なものとして表現され，数値によって絶対的に定義することが難しい知識である「常識」を表現するのに都合がよくなっている．関係知識とその評価方法は，第 4.6.2節で詳しく述べる．

4.4 認識の流れ

本システムにおける物体認識の基本的な方法は，認識エージェントが物体を１つ認識すると，それを他のすべてのエージェントに対してブロードキャストし，他のエージェントから異義が出なければ，承認されたとみなし，もし，異義が出れば，そのエージェントと交渉して，結果の両立，取消を決定を行ない，システム全体で整合性を保つ，ということである．この方法は，エージェント間の協調作用によって，常にシステム全体で認識結果の整合性を保つという考えが元になっている．

システムの実際のメッセージの流れを図 4.7の例に基づいて述べる．

1. まず最初に，認識対象画像がすべてのエージェントの認識モジュールに送られ，認識モジュールの認識が始まる．通信モジュールはメッセージ待ち状態に入る．

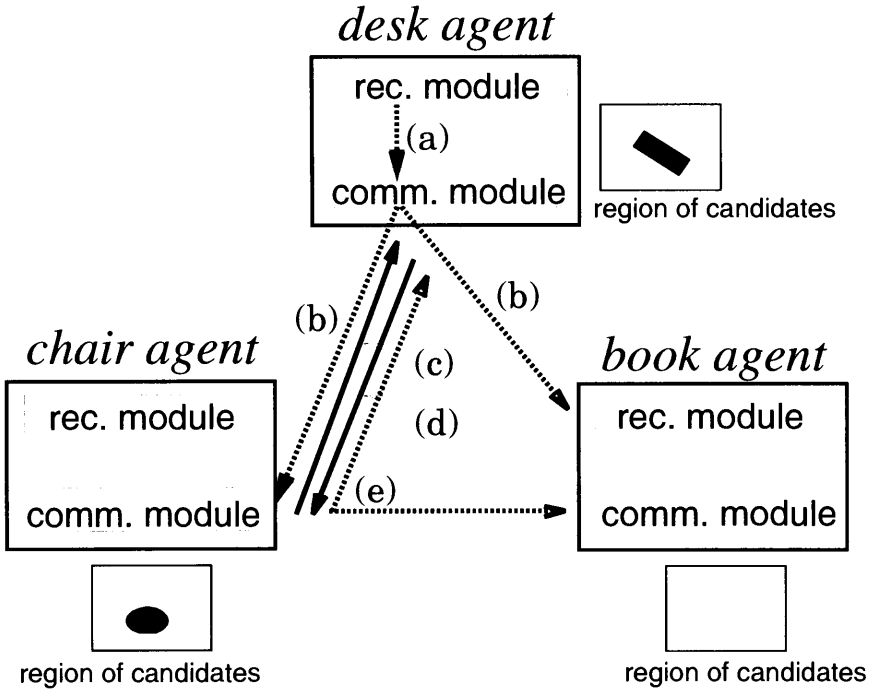

図 4.7 メッセージの流れ．(a) 物体候補の情報．(b) 物体候補の存在範囲の情報．(c) 異義メッセージ．(d) 異義メッセージを受けた物体候補の詳細情報．(e) 取り消しメッセージ．

2. 認識モジュールが物体候補を 1 つ生成すると，その情報を通信モジュールに送る (図 4.7**(a)**).

3. 通信モジュールは候補の情報を記憶し，その存在範囲 (実験システムでは矩形で表現) の情報を他エージェントへブロードキャストする **(b)**.

4. 候補の存在範囲情報を受け取ったエージェントの通信モジュールは，自分の物体候補の領域と重複していないかチェックする．

5. もし重複があれば，物体候補をブロードキャストしたエージェントに，異義メッセージを送る **(c)**．異義メッセージには，自分より相手の面積が小さい時は自分の重複する結果の詳細情報 (物体候補の領域の画素と後述する形状の評価値など) も含める．そうでない時は，相手の詳細情報要求を含める．

6. 異義メッセージに詳細情報要求が含まれている時は，重複する認識結果の詳細情報を送る **(d)**.

7. 詳細情報を受け取ったエージェントは，両立か，どちらか取り消しの判定する．取り消しの場合は，全エージェントに対して，取り消しメッセージを送る **(e)**．両立の場合は，相手のエージェントのみに結果を送る．

8. 再び，通信モジュールはメッセージ待ち状態に入る．

9. すべてのエージェントのすべてのモジュールが待ち状態になり，通信路上にメッセージが存在しなければ，認識が終了する．

4.5 システムの動作の詳細

本節ではシステムの動作の詳細について述べる．最初に，本システムにおいて重要な役割を果たしているモジュール間の通信メッセージの詳細について説明し，その後，認識モジュール，通信モジュールではそれぞれどのようにメッセージを処理しているかを解説する．なお，ここでメッセージと呼んでいるのは，モジュール間で送受信されるデータのことである．

4.5.1 通信メッセージ

すべてのモジュールの動作は，第4.4節で述べたように，すべてメッセージの通信によって引き起こされる．各モジュールは，初めにメッセージ待ちに入り，メッセージを受けとると，それについての処理を行なう．そして，再びメッセージ待ちに入る．基本的な動作は，このサイクルの繰り返しである．つまり，メッセージ駆動による動作を行なっている (図 4.8)．したがって，本システムにおいては，通信メッセージが重要な役割を果たしている．

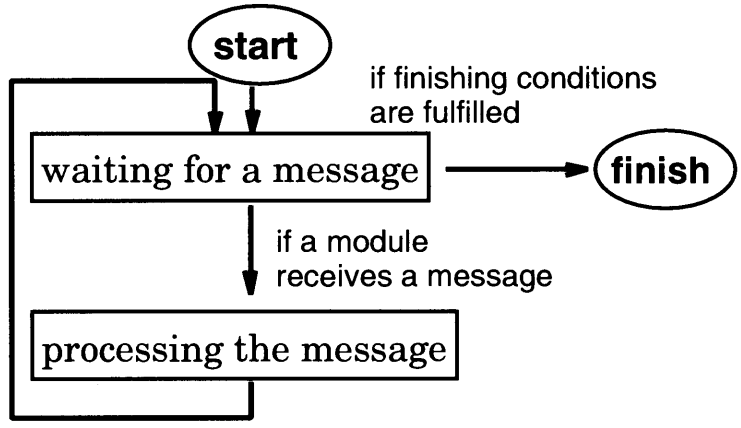

図 4.8 メッセージ駆動によるモジュールの動作．

メッセージは表 4.1に示す種類がある．この表には，メッセージ名とその内容，送信元と送信先，そのメッセージが送信されるきっかけとなるメッセージの番号，そして，そのメッセージを送信したことによってその返答として受信することになるメッセージの番号を記してある．この返答メッセージの欄には，複数の番号が記されているものがあるが，それはどちらかのメッセージが発生することを意味し，括弧で括られているものは送られる可能性があるということを意味にしている．表にはスペースの関係上，一部省略があり，送信元と先の項目の,「通信」はある 1 つの通信モジュール,「認識」は 1 つの認識モジュール,「ホスト」はホストモジュール,「全通信」は他のすべての通信モジュール,「他通信」は異なるエージェントの通信モジュールをそれぞれ意味する．また，メッ

表 4.1 メッセージの種類

送信元と先	no.	メッセージ名	内容	要求	返答
通信 → 全通信	**1**	CANDIDATE	物体候補情報	13	(4, 5, 1)
	2	CANCEL	物体候補の取消し	4, 6	(1)
	3	FINISH	エージェントの動作終了	14	—
通信 → 通信	**4**	REJECT	異義メッセージ	1	6, 7
	5	REJECT&DETAIL	異義＋詳細情報	1	2, 8
	6	DETAIL	認識済み物体の詳細情報	4	2, 8
	7	NO_EXIST1	既に取消済み (1)	4	—
	8	NO_EXIST2	既に取消済み (2)	5, 6	—
通信 → 認識	9	REQUEST_REC	初期認識要求	17	(13), 14
	10	REQUEST_COND	条件付再認識要求	1	(13), 14
	11	STOP_REC	停止要求	3	—
	12	IMAGE_REC	画像の送信	17	—
認識 → 通信	**13**	CANDIDATE_REC	物体候補	9, 10	1
	14	FINISH_REC	認識処理の終了通知	9, 10	(11)
他通信 → 認識	15	ANSWER_DETAIL	認識済み物体の詳細情報	16	—
認識 → 他通信	16	QUERY_DETAIL	直接の詳細情報要求	—	15
ホスト → 全通信	17	INIT_IMAGE	画像の送信	start	19
	18	INIT_RELATION	関係知識の送信	start	—
通信 → ホスト	19	LAST_RESULT	最終認識結果の送信	17	end

セージ番号で太字になっているものは，画像が入力されて最終的な結果が出力されるまでの 1 回の認識に間に複数回送信され，しかも，認識動作の中で重要な役目を持っているメッセージである．

なお，本システムにおいては，メッセージは必ず順序を持って到着することを想定している．つまり，モジュール A と B から同時に C に向かってメッセージが送信されたとしても，A からのメッセージか，B からのメッセージかのどちらかが先に到着することになる．また，A が C に対して続けて複数のメッセージを送信した場合に，その到着順序が入れ替わることがないことも想定する．

4.5.2 認識モジュールのメッセージ処理

認識モジュールと関係のあるメッセージは表 4.1の no.9 から 16 までである．認識モジュールのメッセージ処理は，基本的には認識要求の受信とそれに対する認識結果の送信，認識終了の確認の送信である (図 4.9).

最初に，認識モジュールに送られてくるメッセージは『no.12 画像の送信』で，入力画像がその

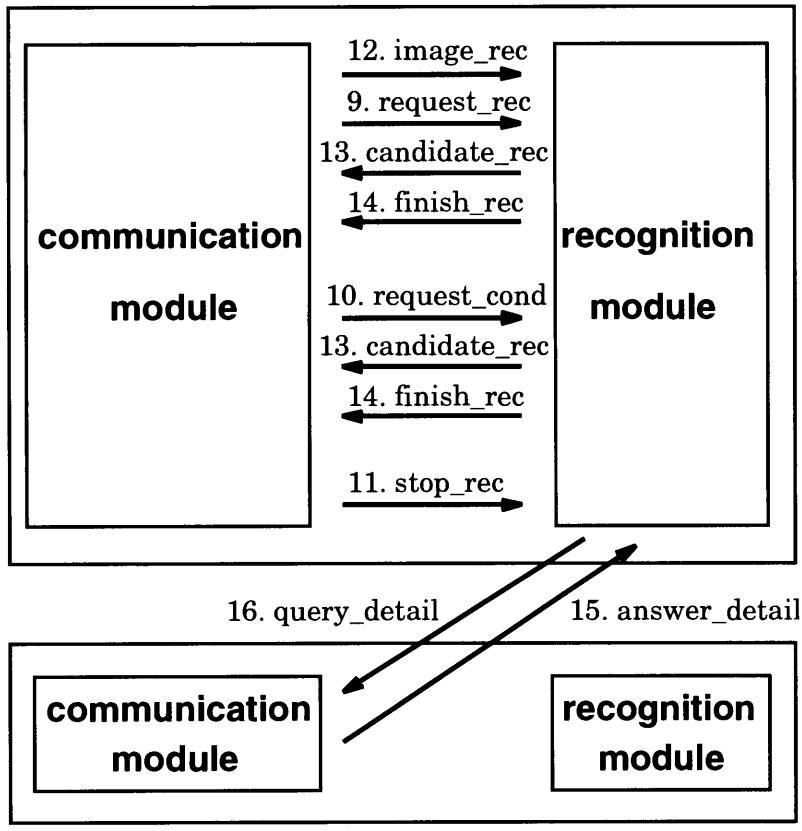

図 4.9 認識モジュール関係のメッセージ

まま送られてくる．その後，すぐに続いて『no.9 初期認識要求』が送られて来て，それを受けて認識モジュールは，画像に対してそれぞれが担当する物体が含まれていないかどうか認識を始める．認識の方法は各認識モジュール毎に異なるのでここでは触れないが，認識モジュールは物体を 1 つ検出する毎に 1 つずつ『no.13 物体候補』として通信モジュールに送信する．

この時送信するメッセージには，物体の占める領域を囲む矩形の大きさと位置，矩形の中で実際に物体が占めている領域の 2 値画像，認識モジュールが付けた認識結果に対する評価値，の 3 つの情報が含まれている．この『no.13 物体候補』メッセージは，物体が発見される度に送信されるので，物体が見付からなければ送信されないし，多く見つかればその数だけ送信されることになる．なお，ここで送信する認識結果は，結果同士の領域が重なっていても，後は通信モジュールが適切な処理を行なうので，そのまま送信して構わない．

こうして，認識モジュールがすべての対象物体を認識し終ったら，『no.14 認識処理の終了通知』を送信し，認識モジュールはメッセージ待ち状態に入り，『no.10 条件付再認識要求』か『no.11 停止要求』のどちらかのメッセージが送られて来るのを待つ．

『no.10 条件付再認識要求』が送られて来ると，認識モジュールはメッセージ中に含まれている情報を利用してトップダウン的な認識を行なう．メッセージ中の情報とは，対象とする物体が存在する可能性の高い領域の矩形，予想される大きさ，そして，『no.10 条件付再認識要求』が送信される

きっかけとなった対象物体と関係のある物体候補の物体番号である．認識モジュールは，きっかけとなった物体についての実際に占める領域などの詳しい情報が欲しければ,『no.16 直接の詳細情報要求』をその物体のエージェントの通信モジュール宛に物体番号と共に直接送って,『no.15 認識済み物体の詳細情報』を得ることができる．通常認識モジュールは同一エージェント内の通信モジュールとしかメッセージ通信を行なわないが，このメッセージのやりとりだけは例外的に他のエージェントと行なうことが認められている．

こうして，認識モジュールは『no.9 初期認識要求』の時とは違って，他の認識結果の情報を利用して，トップダウン的な認識を行なう．新たな物体の候補を検出した場合は，前述した初期認識の場合と同様に『no.13 物体候補』を通信モジュールに送り，結果を報告する．そして，新たな物体が発見されなくなったら,『no.14 認識処理の終了通知』を送信し，再び認識モジュールはメッセージ待ち状態に入る．

そして，やがてシステムのすべての認識モジュール，通信モジュールがメッセージ待ちに入ると，認識の終了が検出され，通信モジュールから『no.11 停止要求』が送られてくる．これを受信したら，認識モジュールはモジュールの実行を終了する．

4.5.3 通信モジュールのメッセージ処理

通信モジュールは，その名の通り通信が仕事であるので，表 4.1のすべてのメッセージと関係している．これらの 19 個は，大きく分けると，初期データ，最終データについてのメッセージ，他のエージェント内の認識モジュールからの問合せメッセージ，認識結果の整合性をとるための協調作用に関するメッセージの 3 種類に分類できる．

初期データ，最終データについてのメッセージは,『no.17 画像の送信』『no.12 画像の送信』『no.18 関係知識』『no.19 最終認識結果の送信』の 4 つである (図 4.10)．最初の 3 つは初期データについてのメッセージであり，それぞれ，ホストモジュールから各通信モジュールへの入力画像のブロードキャスト，各通信モジュールから認識モジュールへの入力画像の送信，ホストが関係知識ファイルから読み込んだ関係知識を関係するそれぞれのエージェントの認識モジュールに送信されるメッセージである．

一方，最終データのメッセージは『no.19 最終認識結果の送信』の 1 つで，各通信モジュールがシステム全体の認識の終了を検出したら，直ちに，ホストモジュールに対して，現在取り消されずに残っている自分の物体のすべての認識結果をまとめて 1 つのメッセージとして送信する．

2 番目の他のエージェント内の認識モジュールからの問合せメッセージは，前節で述べたが，他エージェントの認識モジュールからの『no.16 直接の詳細情報要求』に対して,『no.15 認識済み物体の詳細情報』を返信する (図 4.9)．

3 番目の認識結果の整合性をとるための協調作用に関するメッセージは，本システムの基本原理であるエージェント間の協調を実現するためのメッセージであり，最も重要なものである．前節で説明した認識モジュールに関係のあるものを除くと全部で 7 種類あり，すべての通信モジュールに同時に送られるブロードキャスト型のもの 2 種と，1 対 1 型のもの 5 種がある (図 4.11)．

『no.9 初期認識要求』『no.10 条件付再認識要求』を通信モジュールが認識モジュールに送ると，

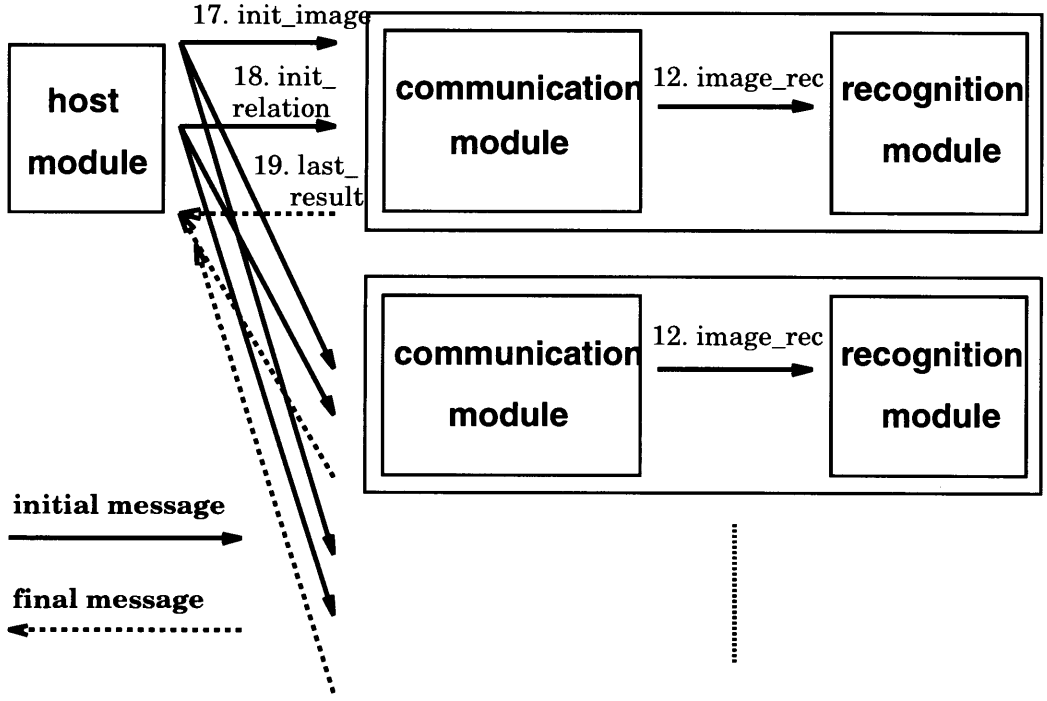

図 4.10 初期メッセージと終了時メッセージ

対象物体の認識が行なわれ，物体が検出された場合には，検出された物体候補の情報が『no.13 物体候補』として通信モジュールに送られてくる．すると，通信モジュールはその候補情報を自分で記憶し，その物体を囲む矩形の座標値と実際の面積の値を『no.1 物体候補情報』として，すべての通信モジュールに対してブロードキャストする．この時，通信メッセージの大きさを小さくするために，実際の画素情報は送らない (図 4.12)．例えば，80 × 80 pixels の矩形で囲まれる領域に物体候補が含まれる場合，矩形の座標値 (左上と右下の x, y 座標値をそれぞれ 2byte ずつで表現すると合計で 8 byte) と面積の情報 (4 byte で表現) のみだと合計 12 byte となるが，1pixel を 1bit で表現した場合，すべて送ると画素情報と矩形の座標値を合計して 808 byte となり，約 67 倍もの違いがある．

『no.1 物体候補情報』を受けとった他のエージェントの通信モジュールは，その候補情報を記憶しておいてから，その矩形の領域が既に認識されている自分の物体候補と重複していないかチェックし，もし，ある一定以上重複している場合,『no.1 物体候補情報』を送信した元の認識モジュールに対して，異義メッセージを送る．この時，重複の対象となっている自分の物体候補の面積と，相手の候補の面積を比較し，相手が大きければ『no.4 異義メッセージ』，小さければ『no.5 異義＋詳細情報』を送る．この２つのメッセージはどちらも異義を表明しているのであるが，異義の調停をどちらのエージェントが主導で行なうかが異なる.『no.4 異義メッセージ』は「こちらで判定するので，詳しい情報を送りなさい.」という意味で，受けとったエージェントは直ちに『no.6 認識済み物体の詳細情報』を返信する．一方,『no.5 異義＋詳細情報』の方は「詳しい情報も一緒に送ります

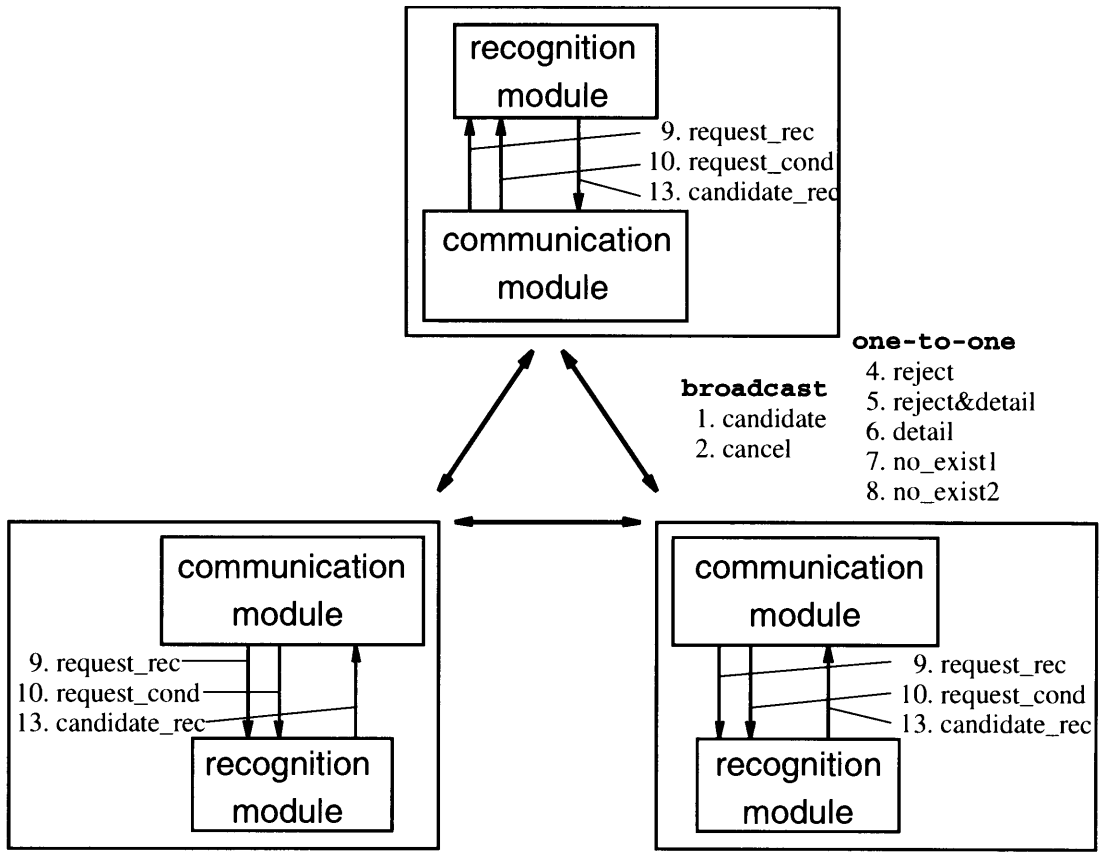

図 4.11 通信モジュール関係のメッセージ

ので，そちらで判定して下さい」という意味である．このように面積の小さい方のエージェントで比較すると決めてしまうことによって，もし 2 つのエージェントが同時に『no.1 物体候補情報』を送信し，さらに同時に『no.5 異義＋詳細情報』を送った場合に，同じ競合を別々のエージェントが判定するということが防げ，2 つの矛盾した結果が生まれることを防止することが可能である (図 4.13，図 4.14)．また，面積の大きい物体の方が競合が生まれる確率が高いので，負荷分散の効果も期待できる．

そして，どちらか一方のエージェントで判定が行なわれ，結果が出ると『no.2 物体候補の取消し』メッセージとしてブロードキャストされる．このメッセージには，取り消される物体の数と，物体候補番号，取消しの判定に形状，関係，面積のどれが使われたか，そして，どれだけの差で取消しになったかの情報が含まれている．判定結果は，両立でどちらも取り消されないということもあるがこの場合もキャンセル 0 として『no.2 物体候補の取消し』を相手のみに送る．また，判定方法については第 4.6で述べるが，判定は競合相手が複数存在する場合，1 対複数でまとめて行われる場合がある．これは，判定するエージェントの自分の物体候補複数個と，相手の 1 つの物体候補との間で判定が行なわれる場合で，場合によっては複数の自分の物体候補がまとめて取消しになることもある．こうした時でも，その結果は 1 つのメッセージにまとめられる．したがって，判定が行な

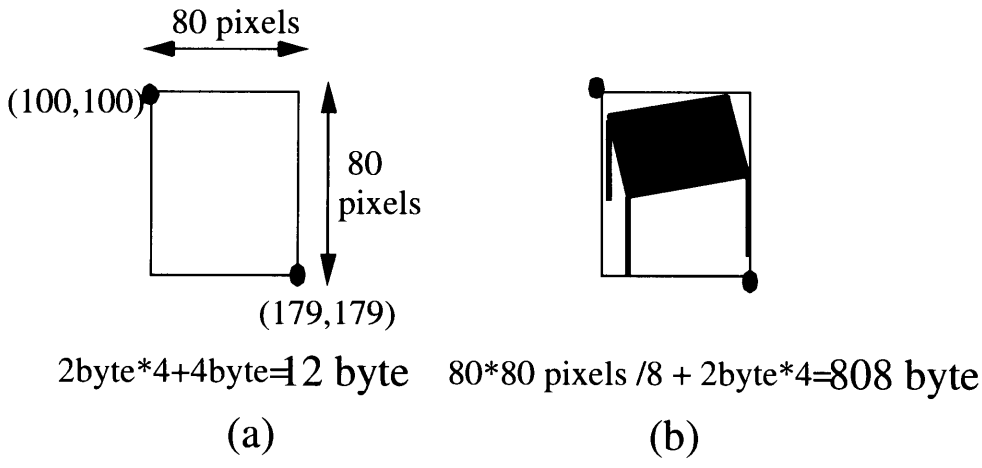

図 4.12 (a) 矩形のバイト数　(b) 画素データのバイト数

われると常に,『no.2 物体候補の取消し』が 1 つだけ全エージェントに向かって送信される.

他に,『no.7 既に取消済み (1)』『no.8 既に取消済み (2)』というのがあるが，それぞれ，前者は『no.4 異義メッセージ』に対して，後者は『no.5 異義＋詳細情報』『no.6 認識済み物体の詳細情報』に対して，対象となる両者の物体のどちらか一方が，他の物体との競合解消などの理由のために既に取り消されていた場合に送信される (例えば，図 4.14). これらは通信に時間遅れのある非同期型マルチエージェントシステム特有の現象で,「あるエージェントが知っている他のエージェントについての状態は現在のものではない」ということに起因している．ここでは 2 つの分けているが，特にその必然性はなく，システムの実装上の都合で 2 つに分けている.

なお，本節では『no.3 エージェントが認識を終了』メッセージについて触れていないが，これはシステムに全体の終了を判定するためのメッセージであり，このシステムの終了を判定するのは実はかなり複雑な問題が存在するので，これについては後ほど第 4.8.2節で触れる.

4.6 競合の解決

本システムにおいては，異なるエージェント間で競合が起こった時は，当事者であるエージェント間で交渉を行ない，競合を解決する．その時，前述の通り，実際の判定は面積の小さい方の物体のエージェントで行なわれ，判定を行なうエージェントに対して，競合する相手の物体候補が実際に占める画素の情報が送られる．判定を行なうエージェントは，その情報を元に自分の物体で認識済みのものとの重なりを調べる．もし，ある一定以上重なっているものがあれば，比較を行なう．この時，ある一定以上重なっているものが複数あれば，複数に対して，それぞれ 1 対 1 で行なわれる (図 4.15). つまり，他の物体候補 1 つに対して，複数個の自分の物体候補と比較が行なわれる可能性がある．なお，異義メッセージは送られてきた矩形情報と，自分の認識済みの物体の占める領域の比較によって送られるので，異義メッセージが送られていても，実際に画素同士を比較した場

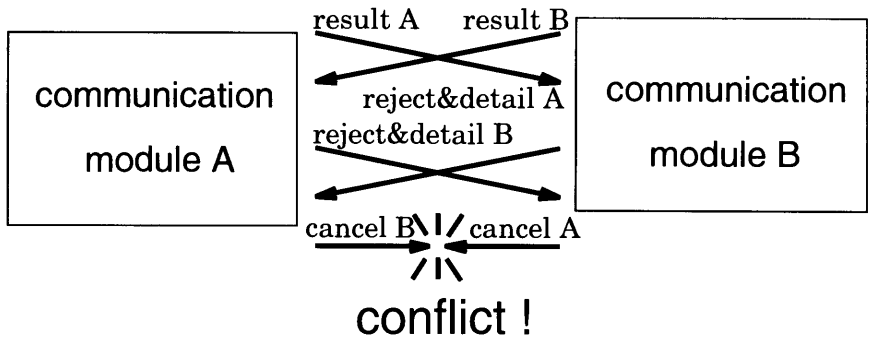

図 4.13 判定する方が統一されていないと矛盾することがある

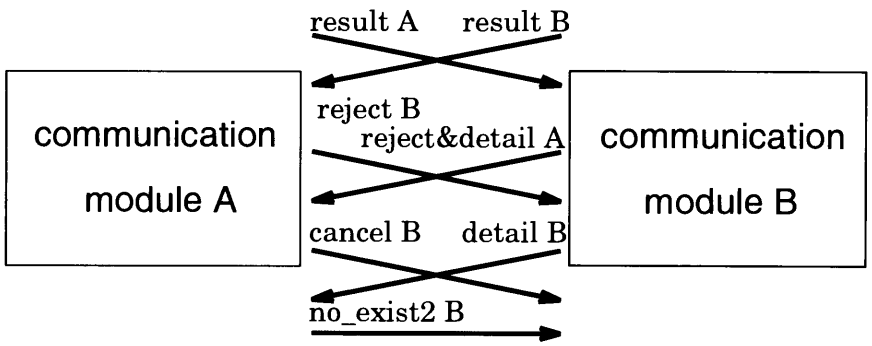

図 4.14 判定する方が統一されていると矛盾しない

合には，重なっていないこともある．その場合は，両立として扱われる．以上の様に，1 対複数の場合でも 1 対 1 が複数回行なわれるので，常に比較は 1 対 1 で行なわれることになり，1 対 1 の比較についてのみ考えればいいことになる．

この 1 対複数の競合の時に，それぞれ 1 対 1 として比較された結果をどうまとめるかという問題があるが，これに対しては，相手の物体候補が取り消される結果が 1 つでも出た場合には，相手候補を取り消すことにする．従って，複数比較した結果が全部両立の場合は全部両立，1 つでも相手候補の取消しがあれば相手候補の取消，それ以外は自分の物体候補の中で取消しになったものが取消しということになる．例えば，図 4.15において，$A > B$ を B の取消し，$A = B$ を両立とした場合，

$$\begin{aligned} A1 &> B1 \\ A2 &= B1 \\ A3 &< B1 \end{aligned}$$

であるならば，$A1 > B1$ があるので，B1 の取消しということになる．B1 が取り消されたので，$A3 < B1$ ではあるが，A3 は取消される相手がなくなったので，取消されずにそのまま残ることになる．

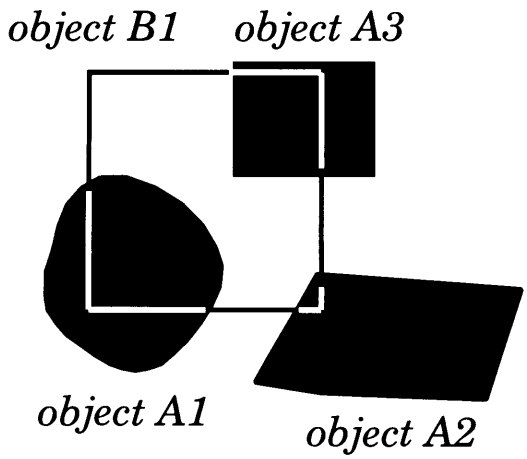

図 4.15 1 対複数の競合

次に比較で必要な形状の評価値，関係知識とその評価の仕方，関係評価値について触れ，その後比較の方法について説明する．

4.6.1 形状の評価値

形状の評価値は認識モジュールが独自に出力結果につけた評価で，本章で実装したシステムにおいては，形状の評価値を 1, 2, 3, 4, 5 の 5 段階にしてある．評価の付け方はそれぞれの認識モジュールによって異なるので，明確な基準を定めることができない．そこで，本システムでは，おおよそ表 4.2のようなおおまかな基準に基づいて値を決定することとしている．なお，これは The Schema System [31] で用いられていた方法を参考にしたものである．

表 4.2 形状評価値のおおよその基準

評価値	おおよその基準
5	すべての形状が得られている
4	ほぼ全体が得られている
3	最低限の構成要素に加えて，付加的な要素も見付かっている
2	最低限の構成要素だけが見付かっている
1	最低限の構成要素らしいものが見付かっている

例えば,「椅子」なら表 4.3のようになる．基本的に「椅子」のような構造がはっきりしている人工物の評価値は，その物体が機能を発揮するのに最も必要な要素を，最低限の構成要素とみなし，それ以外にどれだけの構成要素が見付かるかで，評価値を付けている．

形の評価値を 5 段階と少なめに設定してあるのは，形状評価値は認識モジュールの結果の自己評

表 4.3 「椅子」の形状評価値の例

評価値	検出されて構成要素
5	座面＋背もたれ＋安定した足
4	座面＋安定した足 (1 本以上 or (1 本＋足元の安定した構造))
3	座面＋足 or 座面＋背もたれ
2	正方形か円に近い形状を持った座面
1	形が正方形でも円でもない座面

価値であり，これらを比較に用いる場合に，異なる対象を異なる基準で評価している訳であり，それらを同列に比べることは，あまり厳密に行なっても意味がないという考えに基づいている．形状の評価の段階を少なくすることにし，後は関係知識を活用して決定するという方針を本システムではとっている．したがって，この段階の 5 という値は，理論的な値ではなく，経験的な値である．

4.6.2 関係知識とその評価

本システムで用いられる関係知識は，2 物体間の相対的な関係を記したもので，2 物体の通常考えられる関係を記したものである．人間は，関係知識をいわゆる「常識」として扱っていて，辞書などには通常出ていない知識であり，正確には定義できない曖昧な知識である．しかし，認識には有用な知識である．具体的には，2 物体間の位置関係，大きさの関係，形状の複雑度の関係，濃淡の関係など，様々な関係がある．本システムにおいては，これらの関係はすべて相対的な定性的なものとして表現され，数値によって絶対的に定義することが難しい，曖昧な知識である「常識」を表現するのに都合がよくなっている．

関係知識は「物体名」＋「関係」＋「物体名」という形で表現される．例えば,「本は机の上にある.」という関係は，“book on desk” というように表現される．表 4.4にいくつか例を示す．なお，タイプの意味については後ほど述べる

表 4.4 関係知識の例

```
desk near chair
desk larger_than chair
desk higher_than chair
ws on desk
book on desk
```

本システムで現在使用可能な関係は以下の通りである (表 4.5).

このような関係知識は，システムに 1 つだけ存在する関係知識ファイル中に書かれており，最初

表 4.5 本システムで使用できる関係

no.	relation name	内容	type
1	near	近くにある	A
2	on	上に載っている	A
3	larger	面積が広い	B
4	higher	位置がより高い	B
5	more_complex	複雑度が高い	B
6	more_straight	境界線が直線的	B
7	lighter	平均輝度値が高い	B
8	high_texture	輝度値の分散が大きい	B

にホストモジュールがそれを読み込み，関係のあるエージェントに送信される．関係知識は2つの対象間の関係を記したものであるので，この時必ず，同じ関係知識が2つのエージェントに対して送信される．

次に関係知識の評価の方法について触れる．関係知識は各エージェントが自分の物体候補と，他の物体候補の間に，関係が成立しているかどうかチェックすることによって行なわれる．例えば，“**book on desk**” という関係知識を「机」エージェントが持っていた場合，初めに,「机」エージェントは，過去に本エージェントから認識結果のブロードキャストがあったかどうか，自分が保持している他のエージェントの物体候補情報の中から調べる．ブロードキャストされた情報には,「本」の占める領域の矩形の情報が含まれているので，それを元にして，自分の物体の認識結果との位置関係を調べる．この場合，関係 “**on**” なので,「本」が，画像上での「机」領域に含まれれば，関係が満たされることとする．

4.6.3 関係評価値

ある物体候補についての関係評価値を評価する時は，その候補を生成したエージェントが，自分の持っている他のエージェントの物体候補に関する情報に基づいて評価を行なう．

評価は評価を行なうエージェントが持っている関係知識を1つずつチェックすることによって行なわれる．まず，関係知識を1つ用意し，それに記されている相手の物体候補がないかどうか調べる．もし，あった場合には，関係を調べる．関係も成り立っていた場合には，評価値にある値が足されることになる．ここで，関係の種類によって，加算される値が異なる．Aタイプの関係では関係の相手が複数あった場合に，1つでも成り立てば1が加算される．逆に2つ以上成り立っていても1しか加算されない．Bタイプの関係では，関係が成り立った場合，1/(関係の対象物体候補の個数) が加算され，従って，関係がすべての相手について成り立った場合には1が加算されることとなる．例えば,「椅子」物体候補が3個存在している場合に “**desk larger_than chair**” から加

算される「机」候補の評価値は，3 個の「椅子」候補中の 2 個の候補との間で “desk larger_than chair” が成り立っている場合は，合計 2/3 となる．

このような関係知識の評価をすべての関係知識について行ない，その値を合計したものが関係評価値となる．ただし，競合解消時に，直接競合している物体候補との間に何らかの関係があった場合には，その関係は関係評価値を計算する場合には除外することとする．

この関係評価値の付け方については，理論的に決めることは恐らく不可能であり，経験的に決めるしかないと考える．

4.6.4 物体候補の比較

競合が起こった時には，最終的には 2 つの候補を評価し，比較することによって，どちらが生き残るか決める必要がある．この処理は本システムの認識処理の中でも，重要な処理であると同時に，大変難しい問題でもある．本章で実装したシステムにおいては，暫定的に，非常に単純な方法をとっている．

前述したように，比較を行なう前に，両者の間に両立しうる関係があるどうかチェックし，もし関係がある場合には，両者は両立ということになる．ここで関係がない場合のみ，次に説明する比較が行なわれ，どちらか一方の結果が取消しになる．

比較は，両者を形の評価値と，他の物体候補に対する関係の評価値と，面積の関係によって行なわれる．

まず初めに 5 段階の形状評価値によって評価を行なう．ここで，どちらか一方の形状評価値が小さければ小さい方は取消しとなる．もし，形状評価値が同じだった場合は，関係評価値を用いて，同じような比較を行ない，評価値が小さい方が取消しとなる．さらに形状評価値も等しかった場合は，単純に面積の小さい方が取消しということにしてある (図 4.16)．

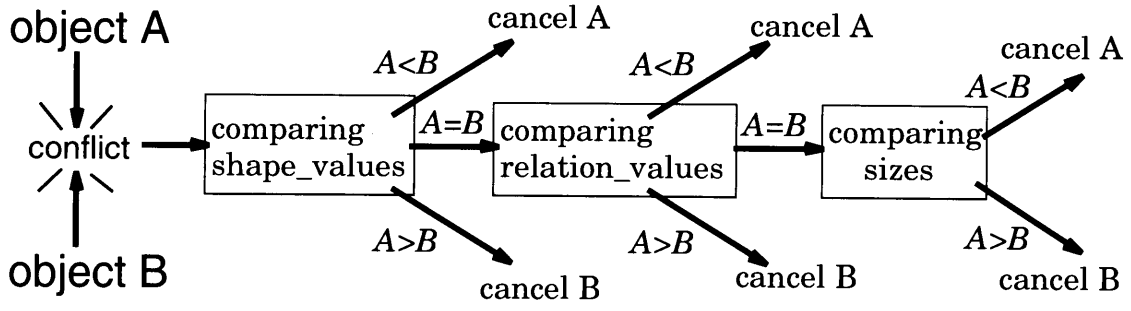

図 4.16 結果の比較

4.6.5 物体候補の取消し，復活

競合を解決するための結果の比較が終了すると，その結果をすべてのエージェントに知らせるために，取消しメッセージがブロードキャストされる．取消しメッセージには取消す認識結果の数と，

その物体の識別番号が含まれていて，それを受けとったエージェントは，それに基づいて，他物体についての認識結果情報を修正する．それと同時に，取消しメッセージを受けとった各エージェントは，取消しメッセージによって取り消された物体候補によって，以前に自分の候補が取り消されていなかったどうか調べる．候補がキャンセルされたことによって，その候補のためにキャンセルされていた自分の物体候補がある場合，キャンセルの理由がなくなった訳であるから，復活することができる (図 4.17)．

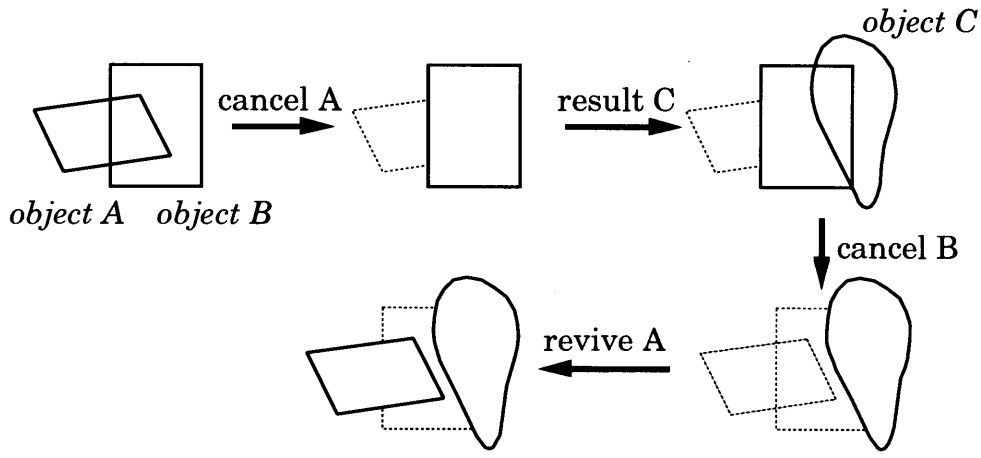

図 4.17 ある物体候補を取り消した物体候補が取り消された場合．

復活は，新たな物体候補を認識エージェントから受けとった時と同様に,『no.1 物体候補情報』メッセージをブロードキャストすることによって行なわれる．その後のエージェント間の協調は，新たな物体候補情報をブロードキャストした時と同様である．

また，復活は，一度取り消された物体候補との間に関係が存在する物体候補の情報を受信した時にも行なわれる．そのため，各エージェントは他の物体候補情報を受けとると，過去に取り消された自分の候補と関係が成り立つかチェックしている．ただし，この復活は，新たな物体候補によって，関係評価値が増えた結果，一度決定された判定を覆すことができるとエージェントが判断できる場合のみ行なわれる．この判断は取消しメッセージに含まれている情報を基に各エージェントの通信モジュールが行なう．復活の仕方は，前述の場合と同じである．

このような復活の仕組みを導入することによって，実は，取消しと復活が交互に起こり続け，ループができてしまうという現象が，起こる可能性がある．この問題に対する対策については，後ほど，第 4.8.1節で詳しく説明する．

4.7 トップダウン認識の起動

エージェントの生成した物体候補の情報はブロードキャストされて，他のすべてのエージェントに伝えられる．この時，自分の物体候補と関係のある物体候補の情報を受けとると，その両方の物体候補に間に成り立っている関係から，新たな自分の物体候補の存在領域を探すための何らかの手

がかりが得られることがある．例えば，図 4.18の例では,「机」エージェントが「机」を認識し，そのことを知った「本」エージェントは “book on desk” という関係知識を用いて,「机」の上にある四角い形の物体は「本」だと推論している．このように，関係知識とそれに含まれる物体の認識結果を合わせることによって，自分の物体の存在する可能性が高い領域や，大きさなどの，認識に役立つ手がかりを知ることができ，他のエージェントの認識結果からのトップダウン的な認識が実現できる．

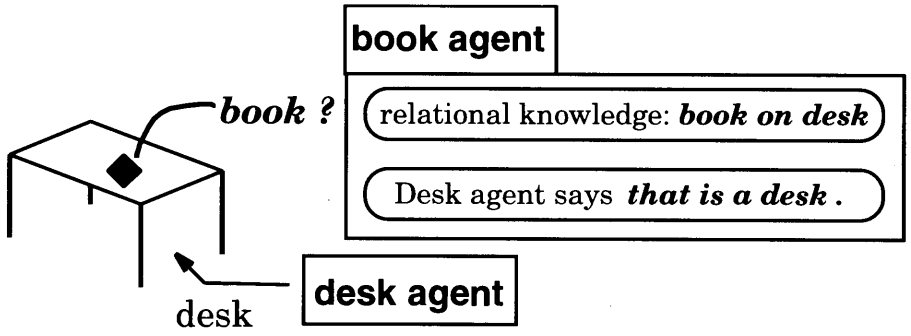

図 4.18 トップダウン認識の例．机の認識結果と「本は机の上にある」という関係知識から，机の上の本を探しにいく．

これを実際に実現するために，本システムでは，すべてのエージェントは，他のエージェントが生成した物体候補情報を受けとると，自分の関係知識と照合し，もし，なんらかの手がかりが得られる場合には，通信モジュールから認識モジュールへその手がかりの情報を条件付再認識要求メッセージとして送信する．例えば，図 4.18の例では,「本」エージェント内の通信モジュールが,「机」の位置と “on” という関係から予想され得る「本」の存在領域と，“on” という関係，その相手の物体番号を条件付き認識要求メッセージとして,「本」の認識モジュールに送信する．そうすると，認識モジュールは予想存在領域に対して，再認識を行なう．再認識の具体的な方法は認識モジュールの実装方法に依存するが，基本的には，初期認識の時に認識できなかった物体に対して再認識を行なう訳であるから，閾値を下げたり，認識できたとみなす最低基準を引き下げたり，あるいは，関係のある既認識の物体のエージェントに対して直接詳細情報要求のメッセージを送って，関係のある物体の認識結果の情報そのものを参照することによって，認識パラメータを変更したり，場合によっては，再認識専用の認識アルゴリズムを用いたりして，初期認識で認識できなかった物体を認識しようと試みる．そして，認識ができた場合は，結果を通信モジュールに送り返す．もちろん，どんなに解析しても認識できない場合は，認識終了メッセージのみを送り返す．

本章におけるシステムの実装では，関係 “on” のみが，このトップダウン的な認識の起動に利用されている．

4.8 非同期動作の問題点への対策

非同期的にエージェントを複数個動作させる場合，その時の特有の問題である「あるエージェントが知っている他のエージェントについての状態は現在のものではない」ということに起因する問題がいくつか現れてくる．1つは，2つ以上のエージェントが互いの結果をキャンセルし合うこと．もう一つは，終了の判定の問題である．本システムでは，通信モジュール間の通信は基本的に認識モジュールからの結果出力に基づいて行なわれるので，認識モジュールの結果出力の数が有限で，かつ，キャンセル ⇒ 復活 ⇒ キャンセルの無限ループが形成されない限り，システムは必ず停止する．認識モジュールの出力が有限というのは，認識モジュールを設計する際に，まったく同じ物体候補を重複して生成することがないようにして，有限個の認識結果しか出力しないように作ればよいので大きな問題とはならないが，キャンセル復活ループが作られないようにするのは非常に困難なことである．以下にこれに対して取られている対策を述べる．

また，システム全体の終了の判定も実は難しい問題である．なぜなら，非同期システムでは，通信にはある一定の時間がかかるために，各エージェントの持っている情報は時間遅れを伴っているものにしかなり得ない．つまり，各エージェントは，ある瞬間のシステム全体の状態を知るということは不可能である．一定時間通信がなく，安定状態に入ったと思ったら，実はあるエージェントの認識モジュールは動いていて，突然，結果をブロードキャストしはじめ，それに伴って，競合解消のためのエージェント間の通信が活発に行なわれることは，十分起こりうることである．そのため，一定時間通信がなければ終了という単純な方法ではシステム全体の終了を検出するのは不可能で，きちんとした終了判定の方法が必要になる．それが行われないと，すべてのエージェントが通信を終了しているのにも拘わらず，他のエージェントからのメッセージを待ち続け，デッドロック状態に陥ってしまうことも考えられる．以下で，さらにそのための方法についても触れる．

4.8.1 キャンセル ⇒ 復活ループの回避

キャンセル ⇒ 復活ループとは，複数の物体の候補の間で，AがBを取り消して，BがCを取り消して，CがAを取り消して，それによって，Bが復活し，Cも復活し，Aも復活し，再び，3つとも取り消される，という無限サイクルが構成されてしまうことである (図 4.19)．

これを回避するには，図 4.19(a) のようなループを構成するような比較の結果が出ないようにすればよい．また，逆に，3つ以上のエージェント間での競合解消のメカニズムを導入しない限り，そうするしか解決方法が存在しない．この方法は，つまり，A, B, C に対して，必ず，$A > B$, $A > C$, $B > C$ のように，比較の結果の順序が一意に決まるようにするということである．つまり，A, B, C の間の順序に推移律が成り立っていて，順序関係が存在していることが必要である．

これは，実は常に守られるべき原則である．もし，比較の結果の順序が一意でないと，例えば，A, B, Cが互いに競合するとして，初めに2つの間で比較を行ない，次にその2つのうちの残ったほうと，まだ比較されていない1つを比較するとする時に，

$$A > B$$

$$B > C$$

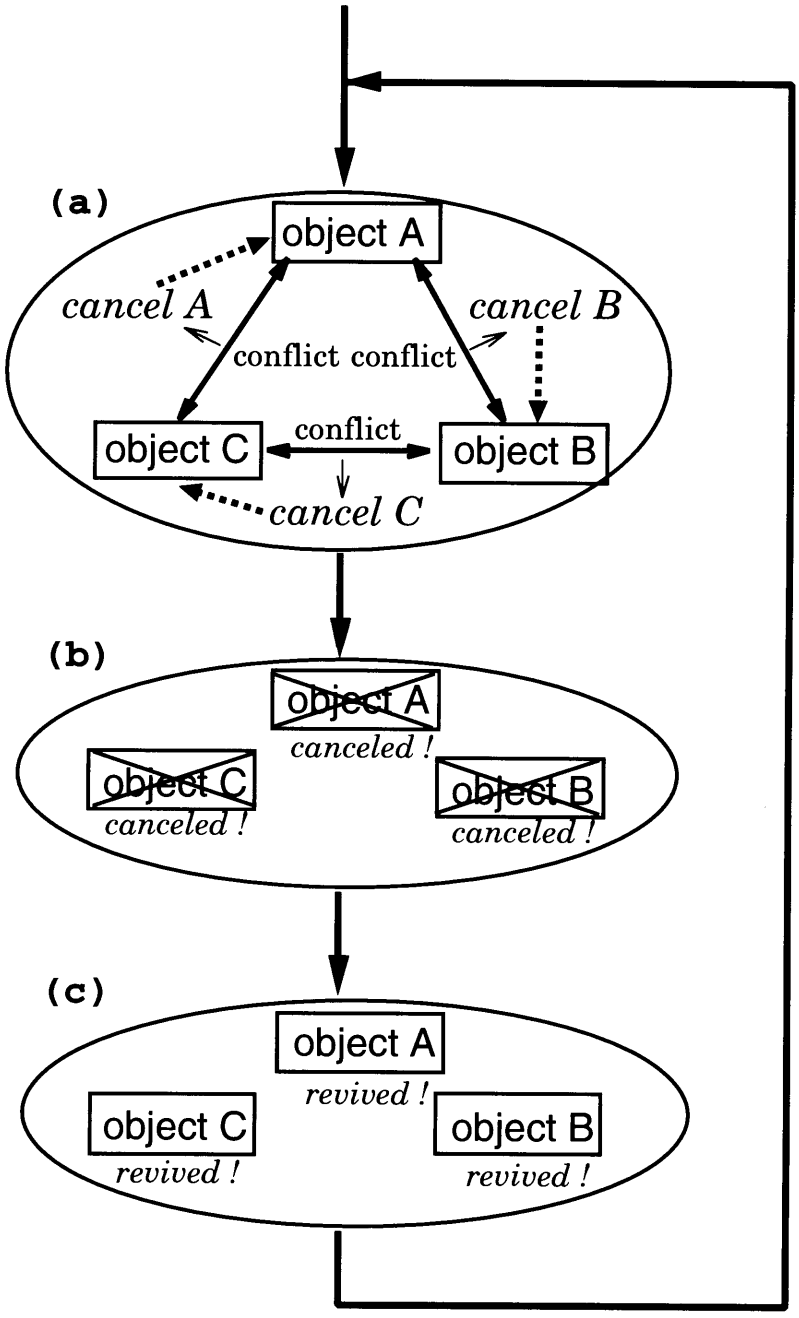

図 4.19 キャンセル ⇒ 復活の無限ループ

$$C > A$$

であるとする．この時，

$A > B$, $C > A$ の順に比較を行なうと $C > A > B$ となり，Cが最終結果として残る．
$B > C$, $A > B$ の順に比較を行なうと $A > B > C$ となり，Aが最終結果として残る．
$C > A$, $B > C$ の順に比較を行なうと $B > C > A$ となり，Cが最終結果として残る．

つまり，どんな結果でも出てしまうことになる．

従って，$B > C$, $A > B \Longrightarrow A > C$ の原則が常に守られるようにしなくてはいけない．そうすることによって，無限ループが構成されることを防ぐことができ，また，比較の順番によって結果が変わるという事態も回避される．

本システムでの競合時の比較の方法は，値の比較であるので，基本的にはこのような原則は守られている．しかし，同じ認識結果に対して，複数回比較が行なわれる場合，関係の評価値がその2回で同一でないことがある．その場合は，$B > C$, $A > B \Longrightarrow A > C$ の原則が必ずしも守られない．

例えば，次のような評価値を持つA, B, Cが互いに競合する場合を考えてみる．

A：形状評価値 4 , 関係評価値 3
B：形状評価値 4 , 関係評価値 2
C：形状評価値 4 , 関係評価値 1

初めにBとCの比較が行なわれるとすると $B > C$ より，Cの取消しが起こり，次に，BとAの比較が行なわれる．そして，$A > B$ よりBが取消され，Aが最終的な結果として残る．これは，3つの結果の評価が $A > B > C$ という一意な順序になっているので，どの順番に比較を行なっても，Aが最終的な結果として残る．

ところが，関係評価値は他の物体との関係を評価した値であり，他に新たな物体が認識されたり，取消しが起こったりすると，変化する値である．例えば，Cの関係評価値が4に上がったとすると，$C > A$ になる．一度取消しになった認識結果も新たな関係が加わって，前回の判定が覆されるとエージェントが判断できる場合には，復活されるために再びブロードキャストを行なうことになっているので，今度は $C > A$ によって，Aが取消されて，Cが最終結果となる．さらに，Bが取消しにあった相手のAが取消されたので，Bが復活を試みるが，$C > B$ より棄却される[1]．こうして，評価値が上がった場合は復活が行なわれ，常に評価値が高いものが最終的な結果として採択される．

ところが，逆に関係評価値が下がった場合には，自分から進んで取消されるようなことはしない．過去に関係評価が高かったということは，その時点では，存在を支持する様な証拠がそれだけ存在したということであり，自分から進んで取消されるようなことまではする必要はないという考えからである．

本システムでは，基本的に一意に評価の順番が決まるようにして，取消しと復活のループの形成を回避し，関係評価値が増加した時のみ復活を行なう．基本的に対策はこれだけであるので，互い

[1] このBの復活は，$A > B$ と $C > A$ という 情報をBが知っているはずなので本来ならば回避できるが，現在の実装では，$C > A$ という取消しメッセージ中に含まれる情報を捨ててしまっているので，無駄な復活が行なわれている．

に関係のある複数の競合が起こった時などはループになってしまう可能性がある．したがって，本システムの無限ループ回避の対策は完全なものであるとはいえないが，本章や第5章，第6章で実装したシステムで行なった実験中においては，特に問題は起きていないので，経験的には十分であると考えている．

4.8.2 デッドロック回避のための終了判定法

本システムのような通信に時間遅れのある非同期マルチエージェントシステムでは，ある瞬間のシステム全体の状態を知るということは不可能である．そのために，終了判定は簡単な問題ではない [108]．本システムにおいては，終了判定せずに各エージェントが勝手に終了してしまうことはできない．なぜなら，他のエージェントが新たな物体候補を生成した場合や，一度取り消された物体候補が復活する場合は，すべてのエージェントに対してメッセージがブロードキャストされるので，そのメッセージがきっかけで，その後，各エージェントがどのような動作を始めるかは分からない．そのため，終了はすべてのエージェントが同時に行なう必要があり，すべてのエージェントが終了可能な状態になったことを検出するための工夫が必要である．もし，それが行われないと，すべてのエージェントが通信を終了しているのにも拘わらず，他のエージェントからのメッセージを待ち続け，デッドロック状態に陥ってしまうことも考えられる．

本システムにおいては，終了判定は送受信の回数と，返答が必ずあるメッセージの送信と返答の数の差をチェックすることにより実現している．具体的には，通信モジュールに対してメッセージの送信がなく，メッセージ待ち状態に入った時に，返答を必要とする5つのメッセージ,『no.4 異義メッセージ』『no.5 異義＋詳細情報』『no.6 認識済み物体の詳細要求』『no.9 初期認識要求』『no.10 条件付認識要求』の送信した回数と，それに対する返答の受信回数の差を調べる．もし，差が1以上であれば，メッセージを送信した先がメッセージの処理を現在行なっているということであり，終了することはできない．一方，差が0の時はそのようなことはないので，そのエージェント自身としては終了しても構わないということになる．そこで,『no.3 エージェントの動作終了』メッセージをブロードキャストする．その時，メッセージには『no.3 エージェントの動作終了』メッセージ以外の全メッセージの送信回数と受信回数の情報を付加する．なお，この時，1回のブロードキャストは (全エージェント－1) 回の送信とし，ブロードキャストの受信も1対1の受信も同じ1回とみなすとする．こうして，認識終了メッセージをブロードキャストした後は，再び待ち状態に入り，他からのメッセージを待つ．そして，他の全部のエージェントから認識終了メッセージが届くと，全てのエージェントのメッセージの送信数と受信数をそれぞれ合計し比較し，両者が等しい場合には，すべてのエージェントが待ち状態に入っているとみなして，終了する．終了する際には，承認されているすべての結果をホストに送り，認識モジュールに対して停止要求メッセージを送信する．もし，両者が異なる場合は現在処理を行なっているエージェントがあるということであり，新たなメッセージの受信を待つ．なお，認識終了メッセージは1度送信した後も，その後メッセージの送受信があり再び待ち状態に入った場合には，再び，認識終了メッセージと新しい送受信数の情報と共に送信することとする．また，一度認識終了メッセージを送信した後で，認識モジュールに対する返答待ちになった場合には，直ちに認識終了メッセージを取消すために，取消し情報を含んだ認

識終了メッセージをブロードキャストする.

以下に，全てのエージェントのメッセージの送信数と受信数をそれぞれ合計し比較し，両者が等しい場合には，すべてのエージェントが待ち状態に入っているとみなして終了してよい訳を説明する.

エージェントはメッセージ駆動であるので，必ず何らかのメッセージを受信してからそれに対する処理を行なう．そこで，エージェントＡが何らかのメッセージを受信し，処理を行ない，終了したとすると，エージェントＢに対してメッセージを送信し，自分は待ち状態に入る．この時，返答待ちでなければ，先ほど送信したメッセージを送信メッセージの数にカウントしてから，認識終了メッセージを送信する.

この時，第三者のエージェントＣで，エージェントＡからの認識終了メッセージを受信して，全エージェントからの情報が揃ったとした場合，エージェントＢからの最新情報を受けとらない限り，必ず送信数の方が多くなる．なぜなら，エージェントＡからの情報にはエージェントＡがＢに対して送ったメッセージがカウントされているのに対して，エージェントＣがエージェントＢについて知っている情報では，そのメッセージの受信はカウントされていないからである．また，エージェントＢがメッセージ受信に対して送信を行なわなかった場合に，逆にエージェントＢからの最新情報がエージェントＡからよりも先にＣに到着した場合，受信数と送信数が等しくなってしまう可能性があるが，そうなってしまうと困る場合，つまり，認識モジュールに条件付認識要求を出していて返答待ちをしている場合は，そうならないようにエージェントＢからの最新情報が他のエージェントにブロードキャストされる前に，取消し情報を含んだ認識終了メッセージをブロードキャストしている．ただし，ここでは，エージェントＡからの取消し情報を含んだ認識終了メッセージの方が，エージェントＢからの認識終了メッセージよりも，先に全エージェントに到着するものと仮定している.

以上のように，１つでもエージェントが処理を行なっている間は必ず送信数の方が多くなる．本システムでは，すべての処理はメッセージ送信の連鎖の中で行なわれるので，一旦すべてのエージェントが処理を停止して，メッセージ待ちに入ってしまうと，そこから再び処理が始まるということはありえない．したがって，ある１つのエージェントが他の全部のエージェントのメッセージ送信数と受信数の合計が等しいことを知ったならば，他のすべてのエージェントは待ち状態に入っているといえる.

4.9 認識モジュールの設計

認識モジュールは入力画像を直接解析し，認識対象とするある 1 つの一般名称で表現される物体のみを認識することがその役目である．つまり，ある特定の物体のみを専門に認識するモジュールである．したがって，認識のための知識や方法はそれぞれの対象に適したものを使用し，通信の方法以外は個々のモジュール毎に異なる．通信の方法は第 4.5.2節で説明したので，本節では，本章の実験システムにおいて実装した 9 個の認識モジュールをどのような方法で，どのようなアルゴリズムを用いて実装しているか簡単に説明する．

認識モジュールで行なう処理は，画像を直接扱う処理が中心になる．基本的な処理は，画像中に含まれる対象物体を認識して物体候補を生成し，その候補に形状評価値を付けて，通信モジュールに送信することである．

認識モジュールでの基本的な認識の方針としては，本システムの目的である classification 的な認識を実現するために，認識対象の形状の多様性に対応した認識手法を用いることとする．一般に物体は同一種類であっても多種多様な形状を持っており，そのすべてを列挙することは，通常形状の種類が多過ぎるので，現実的でない．そこで，認識対象の物体のモデルを厳密に定義し，多くの厳密なモデルを与えるのではなく，認識対象のモデルを柔軟に定義することにする．具体的には，物体の構造と，その物体で最も重要な部分，人工物であればその物体の機能を実現するのに最も重要な部分，に注目して認識を行なう．例えば,「椅子」なら，初めに重要要素である座面を画像中から領域分割などの手法によって探し出す．そして，さらに，付加的な要素の足や背もたれを探す．もし，ここで，足や背もたれが見付からなくて，座面のみでも，評価値に影響するだけで (表 4.3 (p.62))，とりあえず椅子と認識し，通信モジュールに結果として送信する．このように認識モジュールでは，物体の部分的要素より物体の存在を予測する．そして，他のエージェントの認識モジュールと競合が起こった場合の競合解消の処理などの，認識以外の処理に関しては，通信モジュールに任せることにする．

認識モジュールの起動は，必ず通信モジュールからの認識要求に基づいて行なわれる．前節でも述べたが，認識要求には，初期認識要求と条件付き認識要求があり，認識モジュールでの次の 2 通りの認識処理がそれぞれ行なわれる．

- bottom-up process
- top-down process

認識モジュールは，初期認識要求を受けとると，まず，bottom-up process による認識を始める (図 4.20)．この認識は，画像以外の何の付加的な情報も用いずに画像を解析して，その認識モジュールの認識の対象とする物体を探し出す．物体を認識すると，認識モジュールは 5 段階でその認識結果に対して自己評価点，つまり形状評価値を付け，物体候補情報として通信モジュールに送信する．この結果と評価値の送信は 1 つ認識する度に行なわれ，認識した数と同じだけの回数送信される．そして，解析の結果，もうこれ以上結果として通信モジュールに送信していないものはないということになったら，認識終了のメッセージを送信して，認識モジュールはメッセージ待ち状態に入る．

こうして，一度メッセージ待ちに入ると，認識要求か終了命令のメッセージを受信するまで，メッセージを待ち続ける．もし，条件付き認識要求メッセージを受信すると，今度は，top-down process による再認識を始める (図 4.20)．このメッセージには，認識の手がかりとなるような情報が含まれている．具体的には，予想される存在位置，大きさ，関係のある物体候補の物体名と物体番号などである．認識モジュールは，これらの情報を元に，bottom-up process で認識できなかった対象を探す．つまり，再認識の要求であるといえる．Top-down process では，bottom-up process と同じ方法で認識を行なっていたら意味がないので，通常は，閾値を下げたり，認識できたとみなす最低基準を引き下げたり，あるいは，関係のある既認識の物体のエージェントに対して直接詳細情報要求のメッセージを送って，関係のある物体の認識結果の情報そのものを参照することによって，認識パラメータを変更したり，場合によっては，top-down process 専用の認識方法を用いたりして，bottom-up process で認識できなかった対象を認識しようとする．

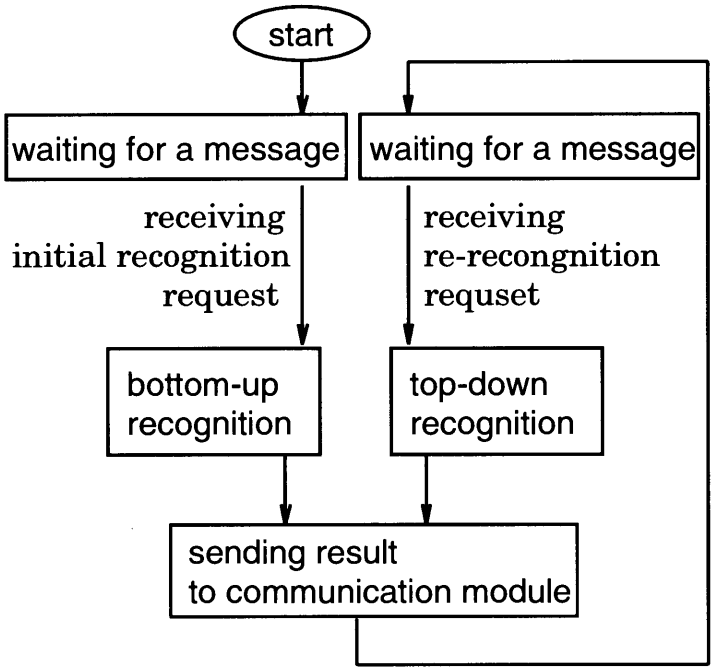

図 4.20 認識モジュールのメッセージ受信による動作

なお，すべて認識モジュールが，bottom-up process，top-down process の両方の処理を行なわなくてはいけないということはない．つまり，bottom-up process しか行なわない認識モジュール，top-down process しか行なわない認識モジュールというのも，存在しても構わないことになっている．実際に，自動車認識モジュールは，bottom-up process では，何も処理せずに，そのまま認識終了メッセージを出し，メッセージ待ちに入る．そして，道路の認識結果が出てから認識要求を通信モジュールから受けとって，top-down process を行ない，自動車の認識処理を始めるようになっている．

4.9.1 室内画像向け認識モジュール

室内画像の認識モジュールは，机 (desk)，椅子 (chair)，床 (floor)，ワークステーション (workstation)，本 (book) の 5 種類が実装されている．床以外は，基本的な構造は決まっているので，その物体の機能を最も表しているような形状，例えば，椅子なら座面，を最も重要な手がかりとして認識を行なう．一方，床のように決まった形状の存在しない物体は，領域分割によって認識を行なう．

chair

椅子認識モジュールでの椅子の定義は次の通りである．

『床から上の適当な高さに適当な大きさの水平な座面があり，その下部にそれを支えるための足が存在している．背もたれがあることがある.』

つまり，座面を重要要素とし，その他に背もたれ，足の合計 3 つの要素から，椅子を認識するという方針をとっている．Bottom-up process は，次の 2 段階で行なわれている．

1. 長さよりも間隔の方が短い垂直な平行エッジの組を画像中から抽出し (図 4.21(a))，その上にまとまった領域を探す (b)．そして，さらにその上に背もたれらしき領域があれば (c)，背もたれとみなす．

2. 次に足が検出できない場合の椅子の認識を行なう．画像全体を領域分割して，ある程度まとまった領域で，座面の楕円か菱形に近ければ，座面とみなす．背もたれも探す．

こうして，2 段階の処理を行ない，それぞれで見つかった結果を認識結果とする．

Top-down process では，探索領域が通信モジュールより指定されるので，その範囲で，1 点から広がる SNAKE [105] を用いて，座面の検出を行なう．SNAKE が座面の楕円か菱形に近い形に収束すれば，座面とみなし，認識できたことにする．

なお，評価値の付け方は，表 4.3 (p.62) に基づいて行なっている．

desk

机の定義は次のようにしている．

「床より上の適当な高さに適当な大きさの水平な机上面があり，その下部にそれを支えるための足が存在している.」

机も椅子と基本的に同じで，机上面を重要要素とし，それを足に合わせて認識している．椅子と異なるのは，机は上に他の物体が載っていることが多いために，机上面が完全に見えていることが少ないので，机上面のエッジから机上面の領域を推測してしまうことである．

Bottom-up process では，次の 2 段階で認識を行なっている．

1. 画像中から，まとまった領域を探し，平行四辺形であるものを探し，机上面であるとする．そして，その下に垂直エッジがあれば，足とみなす．これは，机上面がすべて見えていることを仮定している．

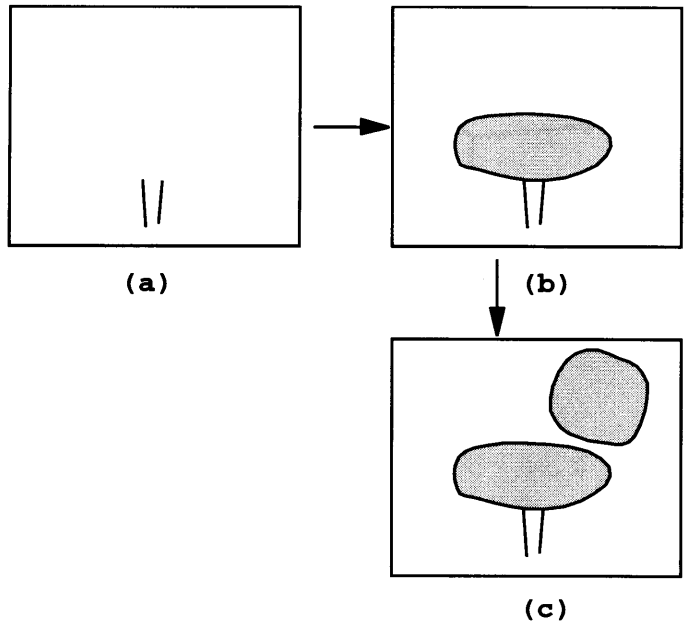

図 4.21 椅子の bottom-up process での認識方法

2. 次に机上面のエッジのみから認識を行なう．まずは，画像から図 4.22(a) の太線で書かれたような直線エッジの組合せを探す．次に，机上面が平行四辺形であるという仮定を使ってその範囲を推定し (b)，机の認識結果とする (c)．こちらの方法では，机上面にオクルージョンが発生していることを想定している．

Top-down process では，処理は bottom-up process と基本的に同じで，直線エッジ検出の閾値を下げることで再認識を行なっている．

workstation

Workstation (以下，WS と略す) の認識モジュールは，現在の実装では，ディスプレイの表示面と，キーボードのみを認識している．ディスプレイの表示面はほぼ垂直な辺の組を含む長方形，キーボードはディスプレイの表示面の横か下にある同じくらいの大きさの平行四辺形とし，両方とも認識されることが必須である．

WS は必ず机の上にあるので，bottom-up process では，認識処理だけ行なうものの結果は送らず，候補は生成しない．そして，top-down process で，与えられた範囲に含まれる認識結果のみを物体候補として通信モジュールに送る．

book

基本的に平行四辺形の認識しか行なっていない．そのため，bottom-up process では，WS と同様，認識処理だけ行なうものの結果は送らず，top-down process で，与えられた範囲に含まれる認

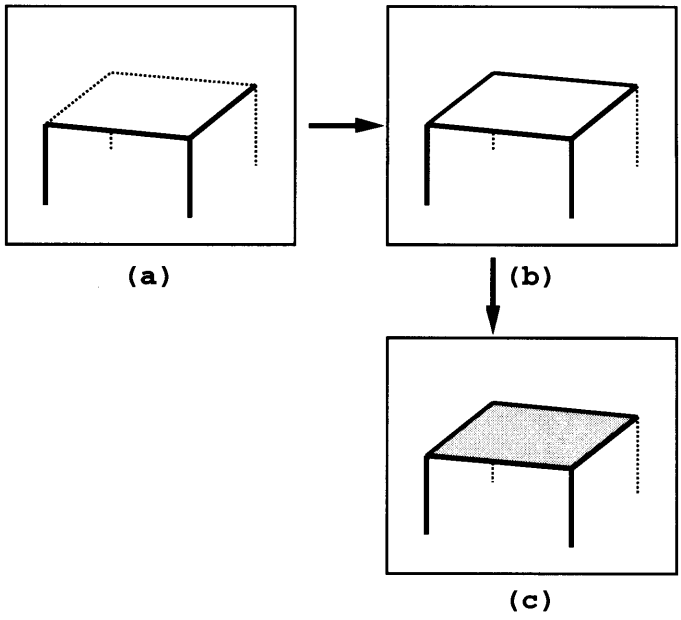

図 4.22 机の top-down process での認識方法

識結果のみを通信モジュールに送る．これは，平行四辺形の認識しか行なっていないために，多くの関係ないものまで本と認識してしまう可能性があるからである．そのため，関係知識があることによって認識要求が起こった場合のみに結果を送ることにしている．

floor

床は決まった形がないので，単純に，画像の一番下から領域を成長させて行き，得られた領域の濃淡値と複雑度から床であるかどうか判定する．簡易的な方法ではあるが，実験での認識率は 9 個の認識モジュールのうちで最も高かった．

4.9.2 屋外画像向け認識モジュール

屋外画像向けの認識モジュールは，道路 (road)，空 (sky)，木 (tree)，自動車 (car) の 4 種類が実装されている．自動車以外は，不定型な物体であるので，領域分割を主体とした方法で認識している．自動車については，2 つのタイヤに注目して認識を行なっている．

road

基本的に認識方法は，floor と同じで，単純に，画像の一番下から領域を成長させて行き，得られた領域の濃淡値と複雑度から道路であるかどうか判定する．この方法では，床と道路が重複して認識されることがあり得るが，それは，関係の評価による競合解消の比較に期待することにする．

sky

Road, floor と同様に領域成長法による領域分割を行う．画像の上方から領域を成長させ，得られた領域の濃淡値から sky であるかどうか判定している．

tree

複雑な模様がある領域を木とみなしている．関係知識として，空の下にあって道路の上にあるというのを与えて置けば，このような簡易的な認識であってもある程度の結果を期待できる．

car

自動車認識モジュールは，現在は，真横から見た場合のモデルしか実装していない．

初めにタイヤらしき 2 つの円を探す (図 4.23(a))．次に，2 つのタイヤと思われる対象の上に，同じくらいの濃淡の領域がないか調べる (b)．もしあれば，その領域の凸包を求め (c)，その中を塗りつぶす (d)．こうして求めた領域を自動車の形として適当であるか，領域の高さ，横幅と，タイヤの間隔，大きさとの関係，および，屋根と両方タイヤを結んだ線が平行かどうか調べることによって決定する．この評価から，形状評価値を 3 から 5 の間で付ける．なお，ここで凸包を求めているのは，自動車の窓ガラスの部分が領域として抽出されにくいからである．

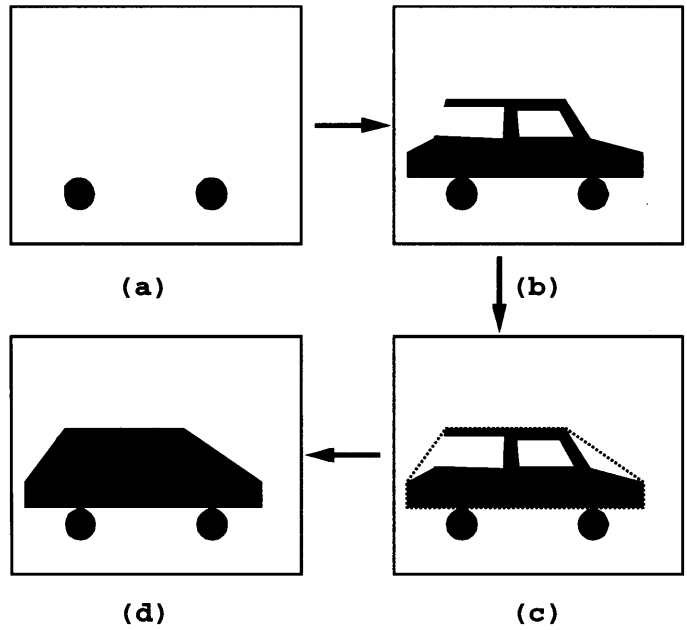

図 4.23 自動車の認識方法

4.10 実験

本節では，初めに並列計算機 AP1000+ 上での実装について簡単に説明する．そして，その後，並列計算機上で行なった 25 枚の画像に対する結果を示し，特にその中でもうまく行った 3 つのサンプル画像に対する認識実験について詳しく説明する．

4.10.1 並列計算機 AP1000+上での実装

本研究において設計したシステムは，並列計算機 AP1000+上で実装を行なった．本研究でのシステムは，通信のみでエージェント間の情報交換を行なうので，AP1000+のような分散メモリ型並列計算機には実装が容易である．実装に当たっては，1 認識モジュールと 1 通信モジュールで構成される 1 つの認識エージェントを，AP1000+の 1 つのプロセッサ・エレメントに割り当て，通信ライブラリを用いてエージェント間の通信を実現した．

本節では，初めに簡単に AP1000+のアーキテクチャについて説明し，その後で実装方法について説明する．

AP1000+のアーキテクチャ

AP1000+ は富士通が開発した分散メモリ型並列計算機で，最大 1024 台までのプロセッサ・エレメント (PE) を結合して並列処理を行なうことができる．PE には，SuperSPARC(50MHz) が使用され，それぞれの PE が 16MB または 64MB のメモリを備えている．今回実験で使用したマシンは，16PE で各 PE のメモリが 16MB の構成であった．また，AP1000+では通信コストを最小限に抑える工夫が各所になされており，分散メモリ型並列計算機の弱点であるメッセージ送受信が比較的効率よく実現されている [109].

各 PE は，図 4.24に示すように 2 次元トーラス構造の一般通信用 T-net の他に，ホストコンピュータによる放送用の B-net，ハードウェアによる同期を実現するための S-net の合計 3 種類の通信ネットワークによって結合されている．ホストコンピュータは通常のワークステーションで，プログラムファイルやデータファイルはホストコンピュータが管理している．また，プログラムの実行命令もホストコンピュータ上で行なわれ，実行命令が出されると，まずホストプログラムが起動され，その次にホストコンピュータから各 PE へ PE 用のプログラムが B-net を介して送信されて，初めてプログラムの実行が開始される．

一般の分散メモリ型並列計算機では，メッセージ送受信の際に割り込み処理による計算の停止があるため通信コストが大きくなるが，AP1000+では，受信用バッファにリングバッファを用いているため，割り込み処理が不要となり，計算と受信を並列に実行することが可能となっている．

AP1000+ への実装

実装に当たっては，1 認識モジュールと 1 通信モジュールで構成される 1 つの認識エージェントを AP1000+の 1 つの PE に割り当て，ホストモジュールをホストコンピュータ上に実装した．AP1000+では，各 PE は Solaris をベースとしたマルチタスクに対応した OS を持っており，1PE

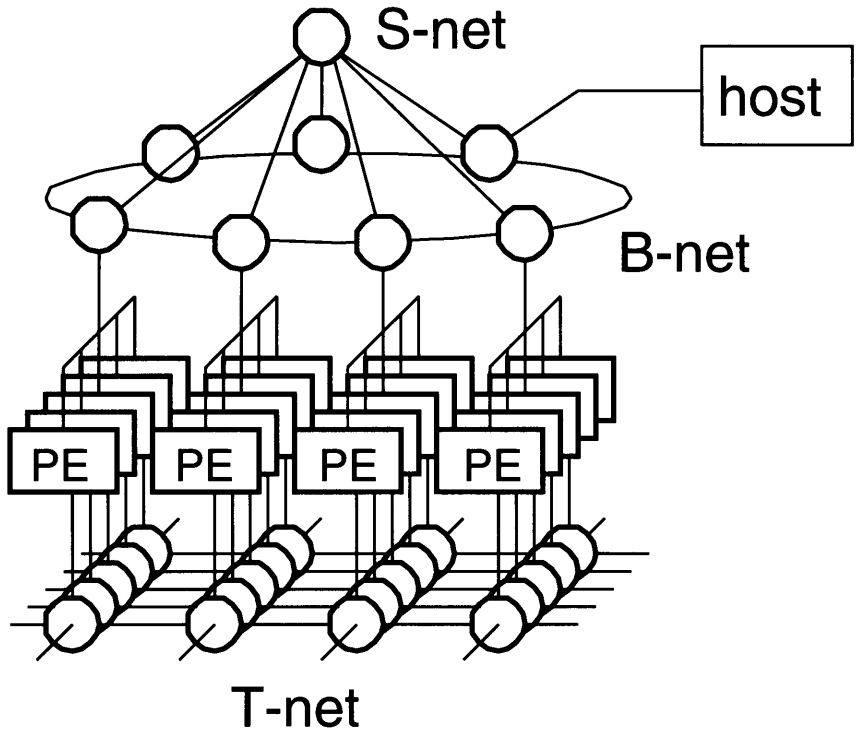

図 4.24 AP1000+のアーキテクチャ.

の中でさらに複数のモジュールを同時に走らすことができる．本システムでは，それを利用し 1PE の中で，認識モジュールと通信モジュールを同時に走らせている．

各通信モジュールの間は T-net を介したメッセージ通信が行なわれている．通信モジュールが送ったメッセージは，受信先のプロセッサの動作とは独立に受信先のリングバッファに直接書き込まれ，受信側はいつでも好きな時に読むことができる．メッセージ送受信の際に同期を取る必要がないので，本システムのような非同期通信型マルチエージェントシステムの実装には向いているといえる．

各 PE へのプログラムと画像データの送信，認識結果のホストへの返信は，B-net を使用している．本システムは C++で記述しているために，実行ファイルのサイズが大きく，1 つの認識モジュールが 1.0MB〜1.7MB の大きさで平均 1.3MB，通信モジュールも 1.1MB もある．ホストコンピュータはホストプログラムが実行されると，各 PE へ PE 用のプログラムを送信してやる必要があるが，ホストコンピュータは 1.1MB の通信モジュールをブロードキャストし，さらに，合計 11.8M の 9 個の認識モジュールを各 PE へ個別に送信してやる必要がある．そのため，ホストプログラムの実行が始まってから，各 PE のプログラムの実行が始まるまで，8.4 秒も掛かっている．実際には入力画像をホストから受けとる必要があるので，実際の認識が始まるまでには，9 秒程度も掛かることになる．

現在，本システムでは，机，椅子，床，本，ワークステーション，道路，空，木，自動車の 9 個の認識エージェントが実装されているので，16 個の PE のうち，9PE だけを使用している．使用可

能な PE の数までは，このような考え方でエージェント増やすことが可能である．

4.10.2 実験

実験は 25 枚の画像に対して行なった．画像の内訳は，サンプル画像 1 (図 4.33) や，サンプル画像 2 (図 4.39) のような研究室内を写した室内画像 22 枚と，サンプル画像 3 (図 4.44) のような大学の構内を写した屋外画像 3 枚である．すべての画像は，大きさが 320 × 240pixel の 256 階調の濃淡画像で，デジタルカメラ (CASIO QV-10) によって撮影したものである．

認識結果は，本システムで認識可能な机，椅子，床，本，ワークステーション，道路，空，木，自動車をどの程度正しく認識できたかで評価した．人が見て明らかにこの 9 種類のうちのどれかに入る物体を正しく認識できた数と，間違って他の物体であると認識したり，全く認識しなかった数とを比較し，全部正解，正しい方が多い，間違っているまたは認識されなかった方が多い，全部間違え，の 4 段階に分類した．それぞれ，画像中に含まれていて，対応するエージェントが存在する物体のうち，80%～100%，50%～80%，20%～50%，0%～20% の物体が認識された場合に分類した．それぞれ，結果は 9 枚，6 枚，5 枚と

結果を表 4.6に示す．

表 4.6 25 枚の認識結果

完全に正解	正しい方が多い	間違っている方が多い	全部間違え
5 枚	9 枚	9 枚	2 枚
(20%)	(36%)	(36%)	(8%)

4 種類の認識結果の例として，図 4.25～4.32に入力画像と，その認識結果について示す．

この実験で使用した 25 枚の画像のうち，完全に認識ができた，サンプル画像 1, 2, 3 の 3 枚について，詳しく説明する．

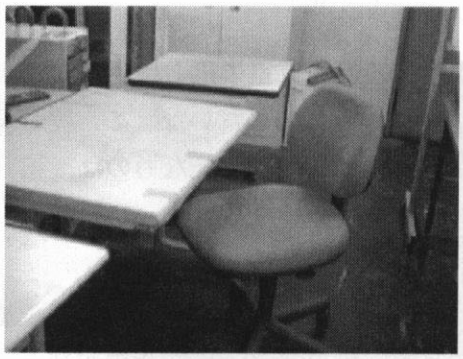

図 4.25 完全に認識できた例.

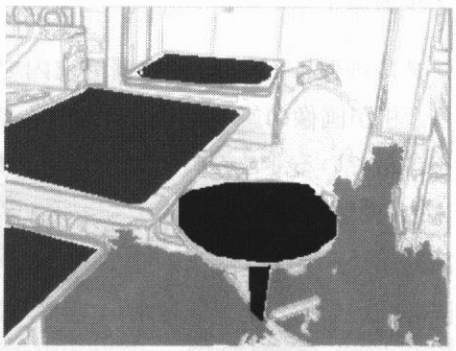

図 4.26 机, 椅子, 床が認識された.

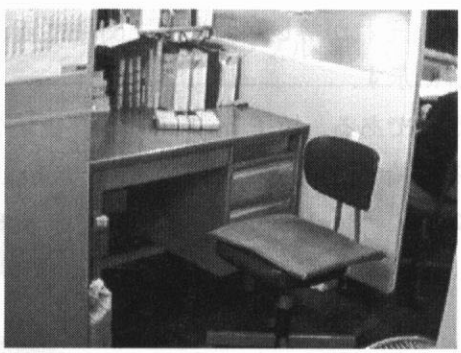

図 4.27 正しく認識できた方が多かった例.

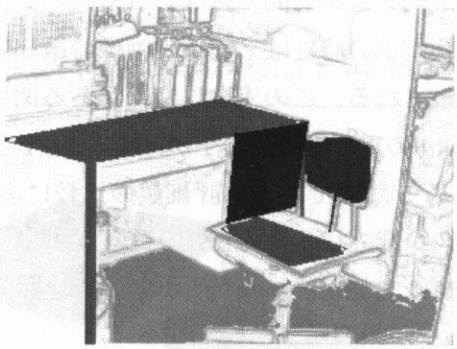

図 4.28 床, 机, 椅子が認識され，本が認識されていない

図 4.29 間違っている方が多かった例.

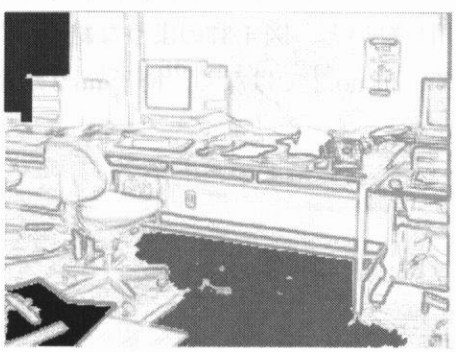

図 4.30 床だけ認識された.

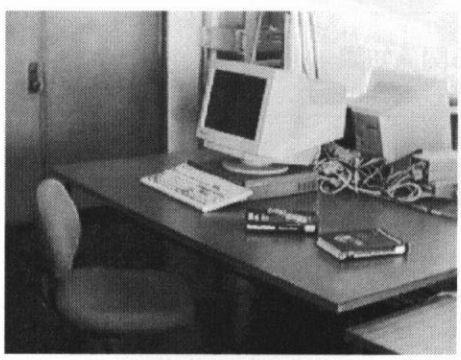

図 4.31 まったく認識できなかった例.

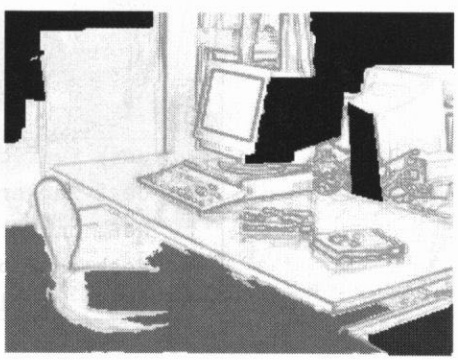

図 4.32 椅子が床の一部と認識され，机も認識できていない.

サンプル画像 1

サンプル画像 1 (図 4.33) は研究室内で撮影した画像である．研究室の机と椅子を写した実画像であり，机が画像の端で切れていたり，椅子の足がはっきりと見えていないなどの，実画像特有の問題を含んでいる．この画像に含まれる物体のうち，本システムが認識可能なものは,「机」「椅子」「床」の 3 種類であるが，その他にもホワイトボードの足やロッカーなどが写っている．

認識結果は図 4.34である．3 つの机，1 つの椅子，床が認識できている．

実際の認識の処理の過程においては，机認識エージェントは 3 つの候補 (図 4.35)，椅子認識エージェントも 3 つの候補 (図 4.36) を認識している．このうち，机の no.0 と no.2 の候補と，椅子の no.1 と no.2 の候補が重複しており，重複し合っている候補それぞれについて重複解消のための交渉がエージェント間で行なわれる．

重複解消の時には，初めに画像特徴評価値が比較され，それが等しい時には，関係評価値によって比較される．この実験に関係がある関係知識を表 4.7に示す．

表 4.8の重複解消その 1 が，椅子 no.1 と 机 no.0 の比較である．椅子 no.1 の形状評価値が 5，机 no.0 も 5 なので，関係評価値の比較によって決定する．椅子の関係は机 no.1 との "`chair near desk`" が成立し，評価値 1，机 no.0 は，椅子 no.0 との間に "`desk near chair`" と B タイプの関係 "`desk larger chair`" が成立し，関係評価値は 1.5 となる．よって，椅子 no.1 が取り消され，机 no.0 が最終結果として残る．

表 4.8の重複解消その 2 では，椅子 no.2 と 机 no.2 の比較が行なわれている．こちらは形状評価値の比較だけで，机 no.1 に決定している．

こうして，形状評価値と関係評価の比較によって，競合の解決が図られている．もし，関係知識を利用しないと，図 4.37のような結果になってしまう．これは，関係評価値が同値なので，面積によって，机 no.2 ではなく，椅子 no.2 が選ばれたためである．

表 4.7 利用している関係知識

```
desk near chair
desk larger_than chair
desk higher_than chair
desk on floor
chair on floor
```

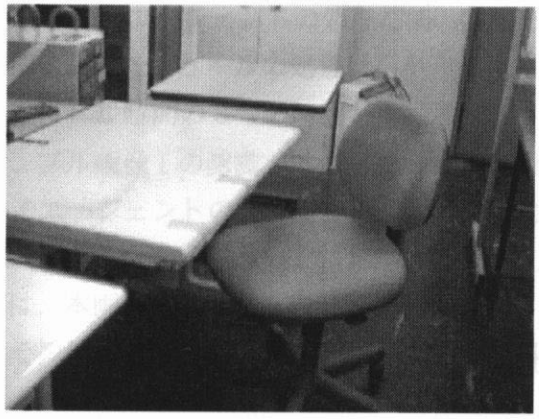

図 4.33 サンプル画像 1

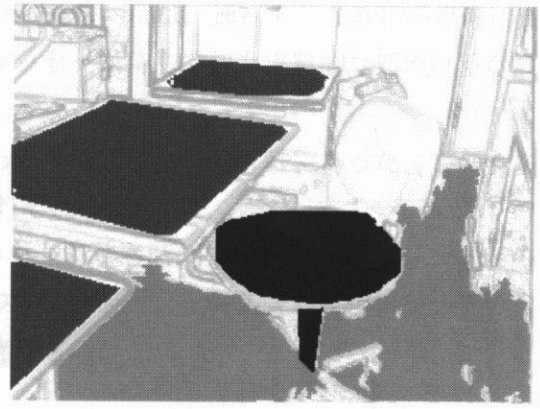

図 4.34 認識結果. (3つの机, 1つの椅子, 床)

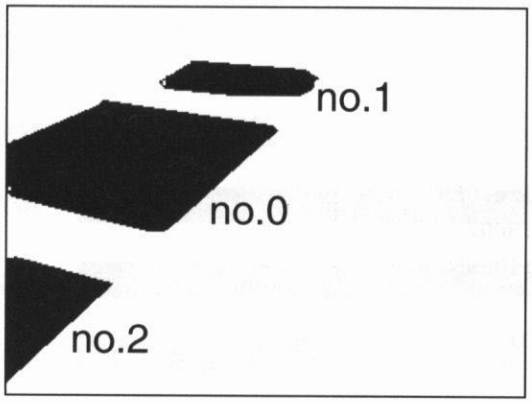

図 4.35 机の候補

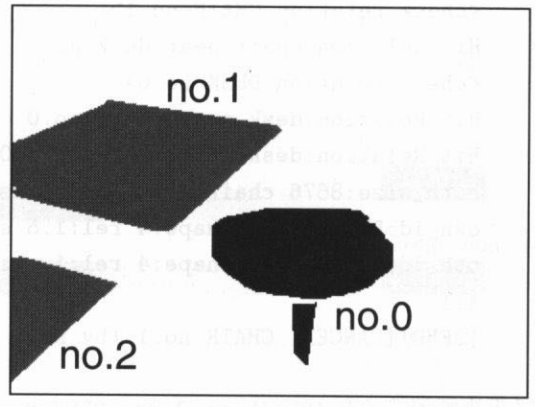

図 4.36 椅子の候補

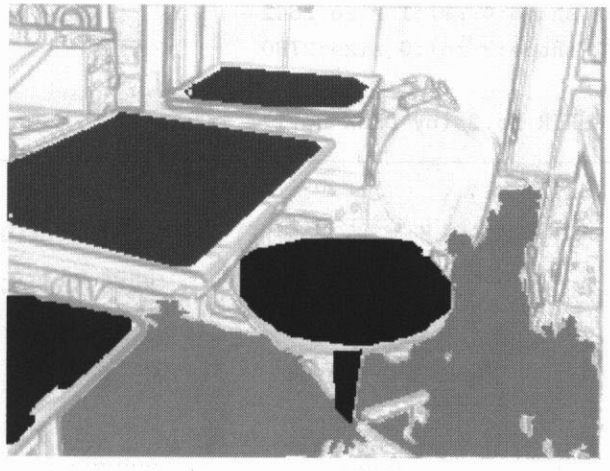

図 4.37 関係知識がない場合. 誤った結果が出た.

表 4.8 重複解消の動作

```
【重複解消その 1 (CHAIR no.1 vs. DESK no.0)】
   <check relation CHAIR no.1>
   Hit Relation:chair near desk no.1
   <check relation DESK no.0>
   Hit Relation:desk near chair no.0
   Hit Relation:desk larger chair no.0
   both_size:8676 chair_size:10295 desk_size:8697
   own_id:DESK  no.0 shape:4 rel:1.5 size:8697
   oth_id:CHAIR no.1 shape:4 rel:1   size:10295
   A:1
   [SEND][CANCEL] CHAIR no.1 (by DESK no.0)

【重複解消その 2 (CHAIR no.2 vs. DESK no.2)】
   <check relation CHAIR no.2>
   no hit.
   <check relation DESK no.2>
   Hit Relation:desk near chair no.0
   both_size:2541 chair_size:2780 desk_size:2642
   own_id:DESK  no2 shape:4 rel:1 size:2642
   oth_id:CHAIR no2 shape:3 rel:0 size:2780
   A:1
   [SEND][CANCEL] CHAIR no.2 (by DESK no.1)
```

各エージェントの処理時間，通信時間の評価

表 4.38に各エージェントの処理時間の内訳を示す．処理時間は，通信モジュールの処理に掛かった時間，認識モジュールの処理に掛かった時間，メッセージ待ちで何も実質的な処理は行なっていないアイドル時間の 3 つの時間からなる．

サンプル画像 1 の認識では，認識の処理に掛かった全時間[2]は，45.78 秒であった．このうち，机と椅子エージェントの認識モジュールの実行にそれぞれ 30 秒程度，WS と本エージェントの認識に 13 秒程度掛かっている以外は，ほとんどがメッセージ待ち時間である．通信モジュールの実行時間は，本画像に含まれている床，机，椅子の 3 つのエージェントが 2, 3 秒程度掛かっているものの，その他はほとんど無視できるくらいの時間である．

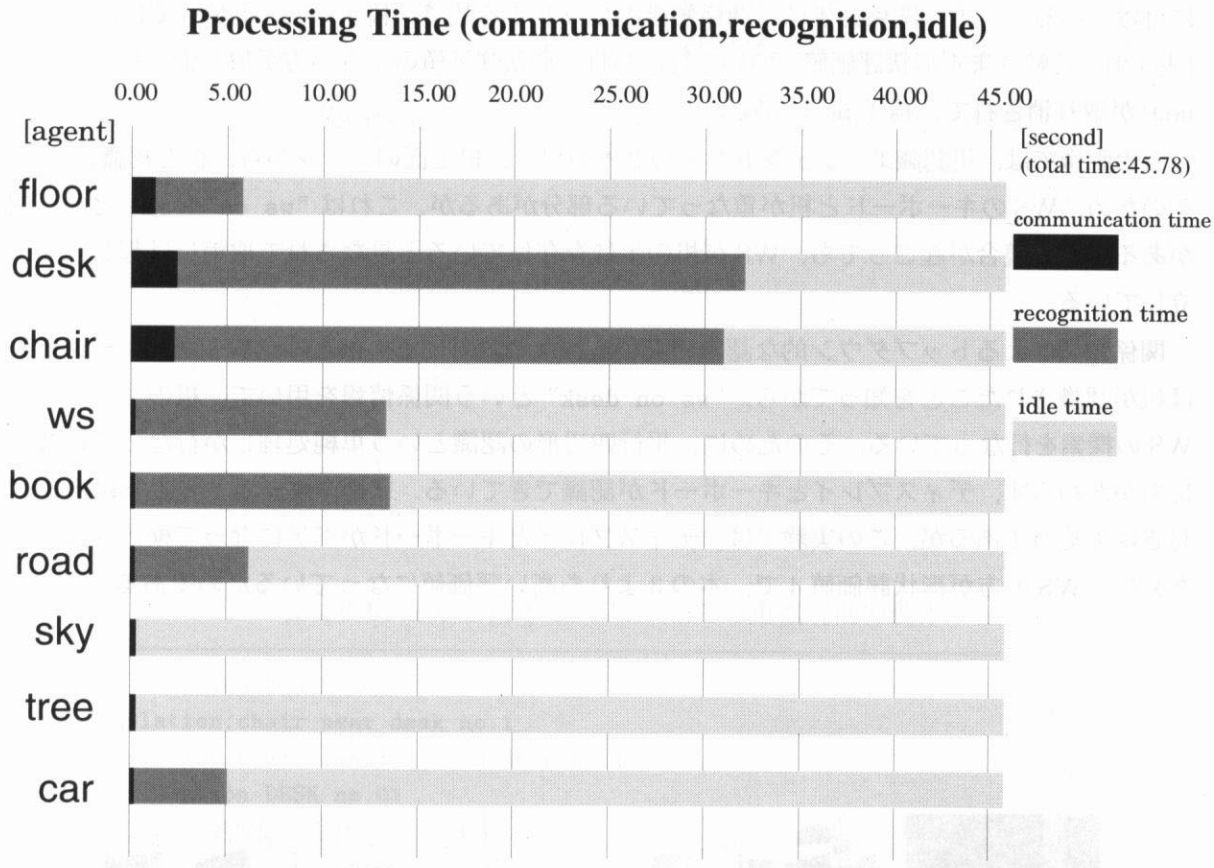

図 4.38 各エージェントの処理時間の内訳 (通信モジュールの処理時間，認識モジュールの処理時間，アイドル (メッセージ待ち) 時間)

[2]ここでは，エージェントが実行を開始してから，終了するまでの時間を認識に掛かった時間とみなすことにする．

サンプル画像 2

サンプル画像 2 (図 4.39) は，机の上にワークステーションが置いてあって机の机上面の大部分が隠されてしまっている．実験 1 の画像よりも複雑な画像である．

机認識エージェントは 2 つの候補 (図 4.41)，椅子認識エージェントも 2 つの候補 (図 4.42)，床認識エージェントは 1 つ，ワークステーション認識エージェント (以下 WS と略す) は 2 組のディスプレイとキーボードを認識した．なお，椅子の no.0 の候補は no.1 の候補の背もたれと重なっている．このうち，机の no.0 の候補と，椅子の no.1 の候補が重複しており，重複し合っている候補それぞれについて重複解消のための交渉がエージェント間で行なわれる．

椅子 no.1 が必要条件の座面に加えて，足，背もたれが見付かっているので形状の評価値 5，一方机 no.0 は机上面のみなので形状評価値 3 となっている．この評価値は各認識モジュールが主観的に付けている．一方，関係評価値は関係知識のヒットする数で，椅子は 2，机が 0 でになっている (表 4.9)．比較はまず形状評価値について行なわれ，形状評価値の小さい方が取り消されるので，机 no.0 が取り消されて，椅子 no.1 が残る．

この画像では，机認識エージェントが机の 2 本の足と，机上面のエッジから，机を認識している．そのため，WS のキーボードと机が重なっている部分があるが，これは "`ws on desk`" という関係があるので，競合が起こっても，WS が机の上に存在していると見なされて取消しは起こらず，両立している．

関係知識によるトップダウン的な認識の実現も行なわれている．例えば，WS 認識エージェントは机が認識されたことを知ってから，"`ws on desk`" という関係情報を用いて，机の上方に関して WS の探索を行なっている．そのために，平行四辺形の認識という単純処理しか行なっていないのにもかかわらず，ディスプレイとキーボードが認識できている．また，キーボードと本は見分けが付きにくそうであるが，この実験では，ディスプレイとキーボードがペアになって検出されているために，WS の方が形状評価値 4 で，本の 3 よりも高い評価値になっているためである．

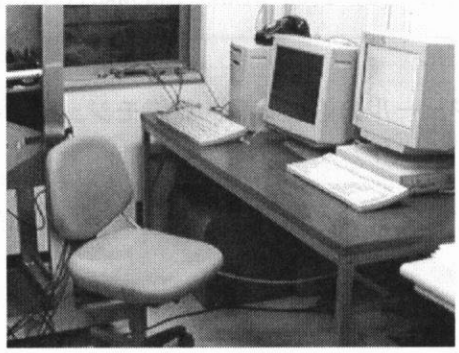

図 4.39 サンプル画像 2

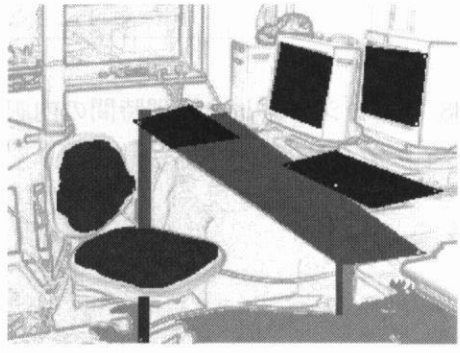

図 4.40 認識結果．(机，2 つの WS，椅子，床)

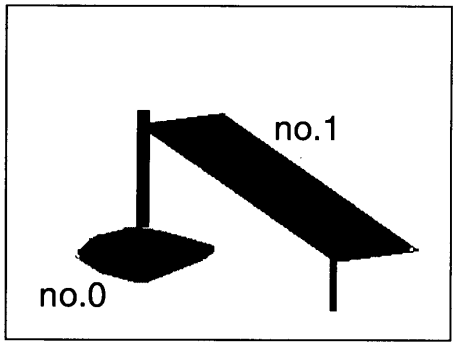

図 4.41 机の候補

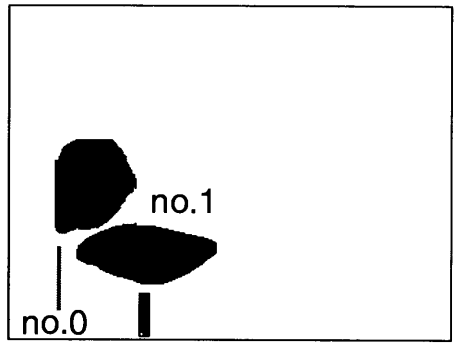

図 4.42 椅子の候補

表 4.9 机 no.0 と椅子 no.1 についての重複解消の動作

```
<check relation CHAIR no.1>
 Hit Relation:chair near desk no.1
 Hit Relation:chair smaller_than desk no.1
<check relation DESK no.0>
 no Relation.
<Compare>
 OverLap:2598 Desk:2627 Chair:5944
 Desk:  no.0 shape:3 rel:0 size:2627
 Chair: no.1 shape:5 rel:2 size:5944
 Cancel:Desk no.0 (by Chair no.1) reason:shape
```

各エージェントの処理時間，通信時間の評価

全体で 76.03 秒掛かっている．認識モジュールの実行時間は，机が 60 秒程度，椅子が 40 秒程度，WS と本が 25 秒程度づつ掛かっている．また，通信モジュールの実行時間は，画像中に含まれている床，机，椅子，WS の 4 つエージェントで 1〜4 秒程度掛かっている．

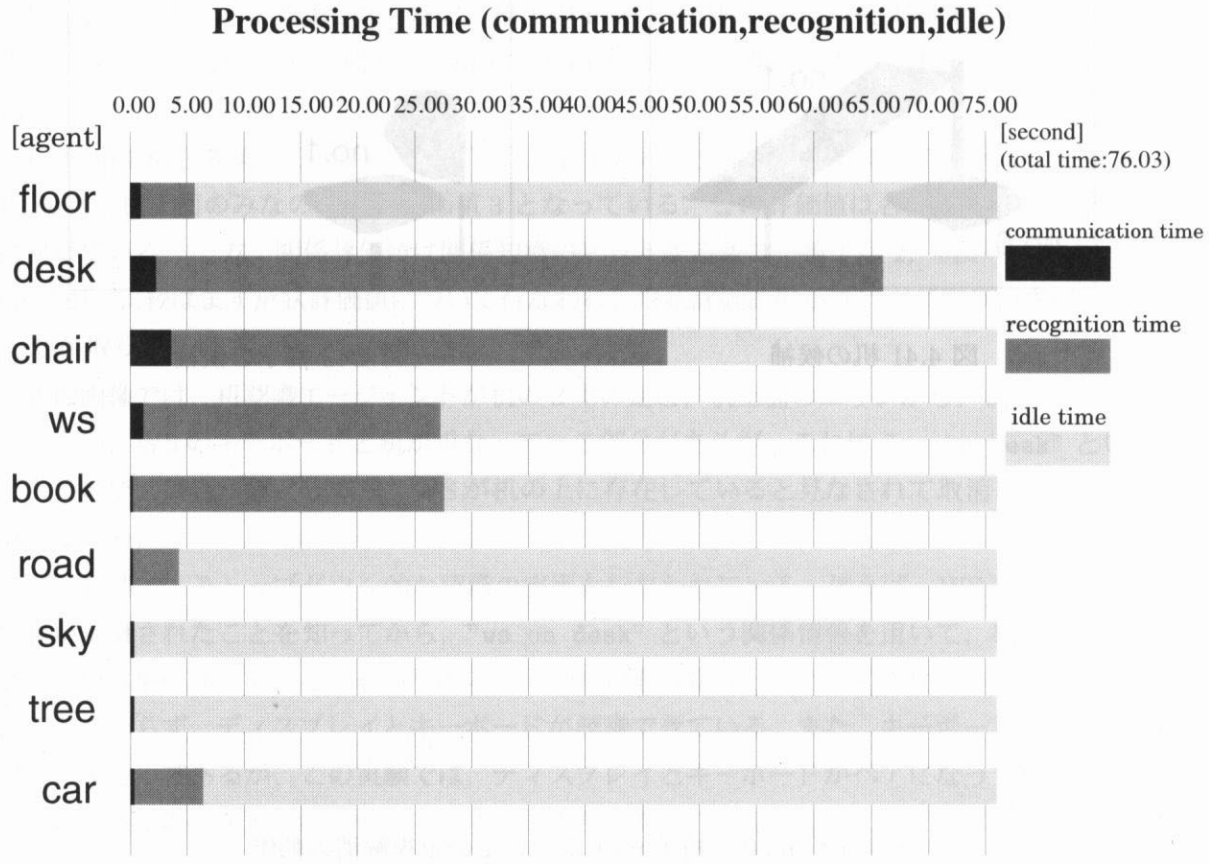

図 4.43 各エージェントの処理時間の内訳 (通信モジュールの処理時間，認識モジュールの処理時間，アイドル (メッセージ待ち) 時間)

サンプル画像 3

サンプル画像 3 (図 4.44) は，サンプル画像 1，2 とは異なり，屋外画像である．道路，空，木，自動車が写っており，それぞれのエージェントが認識を行なっている．結果 (図 4.45) では，道路，空，木，自動車がそれぞれ認識できている．

ここでの認識では，表 4.10に示すそれぞれの物体の相対的な位置関係に関する関係知識が主に使われている．これらの関係知識のうちの “car on road” によって，見分けが付きにくい道路と床を間違えることなく，正しく認識が行なわれている．

実験では,「空」「自動車」「道路」「木」「壁」「床」の候補が 1 つづつ生成された．これらの候補生成によって，競合が 1 回，取り消しが 1 回が行なわれ，最終的に図 4.40の認識結果が得られた．

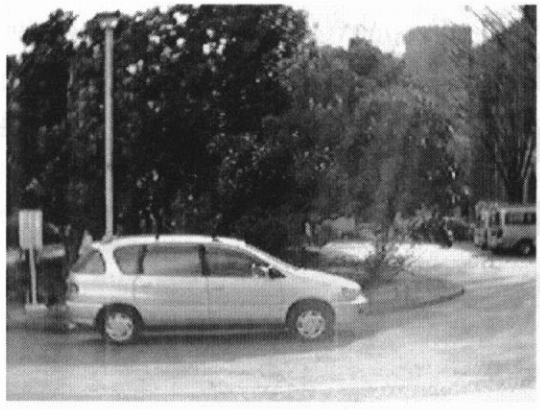

図 4.44 サンプル画像 3

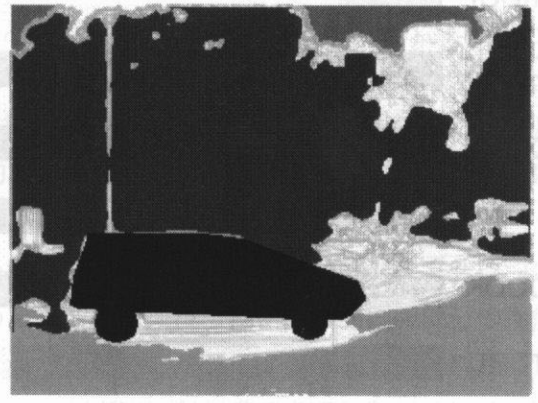

図 4.45 認識結果 (道路，空，木，自動車)

表 4.10 利用している関係知識

```
tree on road
car on road
tree higher car
sky higher tree
sky higher road
```

各エージェントの処理時間，通信時間の評価

図 4.46によると，本サンプル画像の認識時間は 66.41 秒であった．机，椅子，WS，本の 4 つのエージェントが 60 秒前後の認識モジュールの実行時間を示しており，他のエージェントの 5〜6 倍も掛っている．これは，木の部分が複雑なエッジを持っているために，エッジから対象の認識を行なっているエージェントの認識モジュールの処理に時間が掛かってしまっているからである．

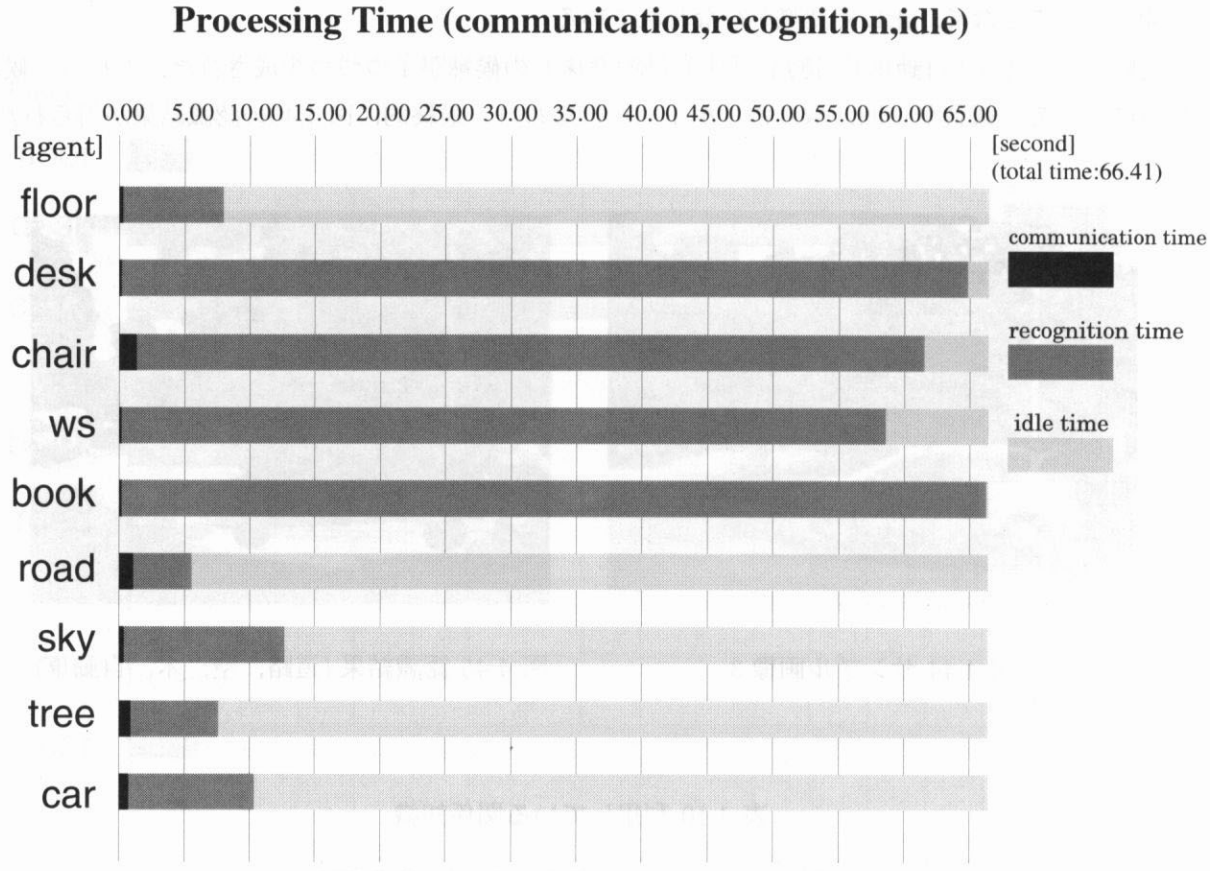

図 4.46 各エージェントの処理時間の内訳 (通信モジュールの処理時間，認識モジュールの処理時間，アイドル (メッセージ待ち) 時間)

4.11 考察

前節では，25 枚の画像に対する認識実験を行ない，成功した 3 つ例について詳しく説明した．本節では，それらの認識結果に対する考察を行なう．

4.11.1 実験結果全体に関して

25 枚の画像に対する実験では，完全に正しい結果が 20%，間違って認識した物よりも正しく認識した物の方が多い結果が 36%という結果が出た．これは，両方合わせても 56%であり，5 割を僅かに上回っているに過ぎない結果である．しかし，本実験で認識の対象とした画像は，デジタルカメラで通常の研究室内を撮影した実画像であり，実験用にセッティングした画像ではない．そのために，物体にオクルージョンが生じていたり，画像の端で切れているのは勿論のこと，照度不足のためコントラストがはっきりしなかったり，もともと色が似ているために区別が付きにくいような，実画像特有の問題を多く含んでいて，認識の対象とする画像としては難しい画像である．そのような画像に対して，5 割以上認識ができたということは，決して満足な結果とはいえないが，マルチエージェントによる認識という本システムの基本的な枠組をさらに拡張して行くことで，さらに認識率の向上を期待することができる，希望のある結果であると我々は受け止めている．

本システムは，通信部分の実装に主に重点が置かれているいるために，認識部分の実装については十分なものとなっているとはいえない．そのため，画像を直接扱う低レベルな認識部分の性能が十分とはいえず，物体候補の領域を見つけ出すことができないことがしばしば起こる．例えば，図 4.29 (p.81)，図 4.31では，机も椅子もワークステーション (WS) も認識できていない．いくら関係知識を利用した認識を実現しているといっても，初めに床しか認識されていない状況では，せっかく机の上に WS があるという関係知識があっても，それを生かすことができない．やはり，初期段階での認識によって，ある程度の物体が検出されるくらいの性能をそれぞれのエージェントの認識モジュールが持っていることが必要である．

4.11.2 成功した結果に関して

サンプル画像 1 (図 4.33 (p.83)) に対する認識は，画面の端で切れてしまった机や，足が暗くて良く見えない椅子などを検出することができた．これは，机の機能を表している形状，机の機能を表している形状のような，それぞれの物体に本質的である部分に注目した結果，認識可能となったものである．また，床に関してもおおよそうまく認識できている．

サンプル画像 2 (図 4.39 (p.86)) に対する認識では，WS が上に載っていてその大部分が隠れてしまっている机を検出することができた．これは，机を主に足と机上面のエッジから全体の形を推測することによって認識を行なった結果である．つまり，他の物体によって隠されている部分のその物体であると認識してしまっている．通常，机の面とその上に載っている物の領域に切りわけは，コントラストがはっきりしている場合を除いて，上に載っている物体の形などの特性に関する知識がないと正確に行なうことは難しい．したがって，こうすることによって，机の形状のみの知識しか持っていない机認識モジュールが，机の面とその上に載っている物の領域に切りわけを行なわな

くて済んでいる．もし，机の上の物体が認識されて，机の領域を競合を起こしても，関係知識 “on” さえあれば，机が机上の物体に下にあるとみなされて，両者は両立する．これは，従来のシステムになかった特徴であるといえる．

サンプル画像 3 (図 4.44 (p.89)) は屋外の画像であるが認識ができている．これは他の画像に対して実験を行なった時とまったく同じエージェントの構成，関係知識の下で実験を行なった．つまり，本システムは，屋内画像と屋外画像というまったく異なる種類の画像を両方，同じシステムで認識することができた．これは，従来のシステムにはない特徴である．従来のシステムでは，複数の種類の画像の認識を行なう場合は，それぞれに対応した知識ベースを用意し，それらを交換してやることが多かった．なぜなら，同時に複数の種類の画像に対する知識ベースをシステムに与えてやると，内部で矛盾を起こして正しい結果が得られなかったからである．ところが，本システムでは，関係知識を使うことによって，自然に対象画像の文脈を考慮した認識が実現できており，屋内と屋外という異なる環境に存在する物体を認識するエージェントを共存させることが可能になっている．例えば，このサンプル画像 3 においては，床エージェントも道路エージェントも同じものを認識結果としてブロードキャストしたが，自動車や木などとの関係知識によって，道路の方が最終結果として選ばれている．

4.11.3 失敗した結果に関して

図 4.27 (p.83)，図 4.29，図 4.31などは，うまく認識ができていない例である．これらに共通していえるのは，物体の境界のコントラストがはっきりしていないために，認識モジュールがうまく認識できていないということである．特に，このことは，図 4.31のようにまったく認識できていない例に顕著である．

図 4.27では，机が認識されていて，机の上に本があるという関係知識もあったために，本エージェントの認識モジュールに対して再認識要求が出されていたのにも拘らず，本が認識できていない．これは本認識モジュールの画像に対する低レベル認識部分の性能が十分でないことを示している．

図 4.29では，机も WS も認識できていない．どちらも単独では画像からはっきりとした手がかりが得られないので認識することが難しいが，人間は WS と机の関係を知っているのでどれが机で，どれが WS か認識可能である．本システムでは，どちらかが認識できて初めて関係知識が利用できるようになっているので，このように関係は存在するがその関係している物体の両方が認識できない時はうまく行かない．これに関しては，関係知識の利用についての改良が必要である．

4.12 まとめ

本章では，画像認識システムのためのマルチエージェントアーキテクチャMORE(Multi-agent architecture for Object REcognition) を提案し，それに基づくシステムを実現し，実験を行った．

実現したシステムは同一種類の物体のみを認識するエージェントの集合体として構成されるマルチエージェントシステムであり，各エージェントの認識結果はエージェント間の交渉によって常にシステム全体で整合性がとられる．システムの特徴は，次の通りである．

- 通信のみによるエージェント間での情報交換．
- エージェントを認識モジュールと通信モジュールによって構成．
- エージェント間の競合解消に関係知識を利用．

また，本システムは，非同期で通信に時間遅れのあるマルチエージェントシステムであるので，デッドロックや終了判定に関する対策が必要であり，結果の取消と復活の無限ループの防止や，エージェントの処理の終了の検出に工夫がなされている．

以上まとめると，本章における研究では，マルチエージェントシステムの構築を通じて，認識の手法とシステム構成の両方の面から，一般性と柔軟性のある物体認識の実現のための研究を行なった．その具体的な成果としては，次の２点が挙げられる．

- システム構成の面からは，異なる認識領域で認識行なうための知識と手法を統合できる構成法．
- 認識手法の面からは，画像中の領域のみに基づくのではなく，物体の本来の形状を推測することによって認識を行なう手法．

システム構成の面からの「異なる認識領域で認識行なうための知識と手法を統合できる構成法」については，実験によって，屋内画像と屋外画像というまったく種類の異なる画像を同じシステムで全く変更することなく認識することができることが示された．

認識手法の面からの「画像中の領域のみに基づくのではなく，物体の本来の形状を推測することによって認識を行なう手法」についても，実験で机の上の WS を認識することができ，有効な手法であることが示された．しかし，画像によっては机しか認識されなかったり，机すら認識されないこともあった．これは，個々の認識モジュールの認識能力の問題が大きく影響しており，今後，より多くの手法を統合して，個々の認識モジュールの性能の改善をしていく必要がある．

今後は，以下に述べるような課題を解決していくことが必要である．

- 認識モジュールの単体の認識能力の向上．
  - 同一エージェント内に，認識手法の異なる複数の認識モジュールを用意し，結果を統合して利用する．結果を統合する際，複数のモジュールから認識された対象はその形状評価値を高くすることによって，より柔軟な認識が実現できると考えられる．

- カラー画像を用いる．カラー画像には濃淡画像よりも多くの情報が含まれている．領域の抽出の精度が向上する．
- 認識モジュールが 1 つ 1 つ人手による構築 (hand-coding) になっているので，ある程度認識モジュールの構築法を体系立てる必要がある．また，多数の認識モジュールを人手ですべて構築することは困難であるので，今後は学習によって認識モジュールを構築する方法を検討することが重要である．
- 各認識モジュールの処理は現在はそれぞれが独立に行なっている．そのため，複数のエージェントが同じ画像処理を行うという，無駄が起っている可能性がある．そこで，低レベルの画像処理の結果の共有も検討する必要がある．ただし，これは通信コストとのトレードオフである．

- 関係知識の効率的な利用．
  - 通信モジュール間で行なわれた交渉の結果や，他の認識結果と関係知識を照合した結果に得られた情報を，認識モジュールにフィードバックして，再認識を行なう枠組をもっと強力なものとする．
  - 3 次元空間の構造を扱うことが必要である．現在は画像を完全に 2 次元的に扱っているが，2 次元的な扱いのみでは，物体間の空間的な関係を扱いきれない．例えば，机にオクルージョンが発生している場合，上に本が載っているのと，手前に椅子がある場合では，3 次元的な空間構造は大きく異なるのに，2 次元的にはほぼ同じようにとりあつかうことになり，載っているか，手前にあるのか区別できないことがある．

    そのためには，各認識モジュールが 2 次元的な認識のみではなく，3 次元的な認識も行う必要がある．現在は 2 次元的な形状を仮定してあてはめているが，3 次元的な形状のあてはめを行うことによって物体の 3 次元構造が推測でき，物体間の 3 次元的な位置関係の情報が抽出できるようになる．

- 結果の比較方法の改良．
  - 認識モジュールが一度間違っている結果に高い形状評価値を付けてしまうと，間違った結果が最後まで残ってしまう．つまり，関係評価値よりも形状評価を優先して比較に用いていることに関して，なんらかの改良を加えるべきである．例えば，α(形状評価値) $+ (1-\alpha)$(関係評価値) $(0 < \alpha < 1)$ を総合評価値として，この値を一律に比較に用いるなどの方法が考えられる．これは，実験によって経験的に決めるしかない．

- システム全体の挙動の解析．
  - 現在はエージェントが 9 個までしか実装されていないが，これが，さらに，16 個や 64 個にまで増えた時にシステム全体の挙動がどうなるか不明．エージェントが増えると，通信，とくにブロードキャストの回数が増大するのために，すべてのエージェントが終了するまでに時間がかなりかかることが予想される．

– 取消しと復活の無限ループが発生しないことを理論的に保証することができていないので，エージェントが増えた時に問題が起らないかどうか，実験によって確かめる必要がある．