物体間の支持関係を利用した室内画像の認識

柳井　啓司†　　出口光一郎††

Recognition of Indoor Images Employing Supporting Relation between Objects

Keiji YANAI† and Koichiro DEGUCHI††

あらまし　本論文では，複雑なオクルージョンを含む室内シーン画像に対する画像認識システムを提案する．従来のシステムでは，物体が十分に画像中に現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現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．我々は，こうした認識を我々が従来より研究しているマルチエージェント型の画像認識システムとして実現した．本論文では，システムについての詳細と，実際にインプリメントしたプロトタイプシステムによる実験，結果について述べる．

キーワード　一般物体認識，シーン理解，画像理解システム，支持関係

1. まえがき

我々の研究では，一般の実世界シーンの単一画像に対して，そのシーン中に含まれる物体の一般名称と，物体同士の位置関係を定性的に認識する画像理解システムの実現を目的としている．本論文では，シーンを人工物で構成される室内シーンに限定して，その中で問題となるオクルージョンに対処できる認識手法について提案する．

通常，室内シーンにおいては，床の上に机があって，その上に計算機があるというように，複数の物体が積み重なって存在している（図 1）．そのため，手前にある物体が奥にある物体を隠してしまうというオクルージョンが発生しており，室内シーンを認識する場合はオクルージョンに対応することが必要である．従来のオクルージョンを含むシーンに対する認識は，あらかじめ正確な形状モデルが既知である物体を対象に行われることが多く，部分的な特徴にモデルを当てはめて，隠れている部分を推定することを行っていた．こうした方法は，工業部品などのあらかじめ形状が決まっている物体については有効であるが，一般のシーンを対象にする場合，物体の多様性の問題，つまり，同一の一般名称をもつ物体であっても形状は様々で，物体の正確な形状モデルが既知であることは通常はあり得ないために，有効でない場合がほとんどである．そこで一般のシーンに対しては，対象物体の典型的な構造特徴を記述するモデルを用いることになるが，図 1 のようなオクルージョンが多く発生している一般の複雑な

図 1　室内シーン画像の一例
Fig. 1　An example of an image of indoor scene.

† 電気通信大学情報工学科，調布市
Department of Computer Science, The Univ. of Electro-Communications, 1–5–1 Chofugaoka, Chofu-shi, 182–8585 Japan
†† 東北大学大学院情報科学研究科，仙台市
Graduate School of Information Sciences, Tohoku Univ., Aramaki-aza-Aoba 01, Aoba-ku, Sendai-shi, 980–8579 Japan

シーンの場合には，オクルージョンのない物体を除いては物体の一部分を検出することさえ難しい．

本研究では，従来のオクルージョンを含むシーンに対する認識においては物体単体を認識の対象としていたためにあまり利用されていなかった物体同士の位置関係を効率的に利用することによって，正確な物体形状が与えられていない状況での複雑なオクルージョンを含んだシーンに対する認識方法を提案する．

従来の我々のシステム [13], [14] や，過去の一般の実世界シーンを対象とした画像理解システムの一部，例えば，The Schema System [4] などは，シーン中に含まれる物体同士の位置関係の情報を認識の手掛りとして用いていた．しかし，対象が風景画像であったために室内画像ほどオクルージョンは多くなく，またオクルージョンがあっても，認識の対象が「道路」「空」「木々」などの物体の構造や形状よりも色やテクスチャを手掛りとして認識する対象であったので，オクルージョンが問題になることはなかった．また，認識対象が色やテクスチャを主な手掛りとして認識する対象であったため，対象とするシーンは 3 次元シーンではあるが，基本的には領域分割と各領域に対するラベル付けという形で認識が行われており，認識は 2 次元的に行われていた．このため，物体間の位置関係の判定は単純に画像上での上下左右で行われていたが，それである程度うまく行っていた．こうした 3 次元シーンに対する 2 次元的な認識は，屋外の遠景画像のように広範囲のシーンを写した画像の場合は，認識対象となる物体自体の奥行がシーンの奥行に対して比較的小さいために有効である．ところが，室内画像のように近景の画像の場合は認識対象となる物体の奥行が無視できないので，2 次元的な認識では不十分であり，3 次元的な位置関係を推定することが不可欠である．

それに対して，人間は実世界に存在する物体の大まかな 3 次元構造を知識としてもっているために，単一の画像からでも物体間の 3 次元的な位置関係をある程度推測することができる．また，物体の構造に関する知識に加えて，物体は支えがないと下に落ちるといった物理法則の定性的な知識をもっているので，例えば，机の足が見えていなくても，平面があってその上に計算機が見えれば，机が計算機を支えているということが推測できる．そして，更に机には適当な長さの足があって，足の下には床があって，机を支えているということも，画像中からボトムアップ的に認識することは困難であっても，知識から推測することが可能である．2 次元の画像から 3 次元世界の構造を認識するシステムが，こうした実世界の定性的な物理法則に基づく 3 次元推論の能力をもつことは，より自然な画像の解釈を実現する上で必要な能力であると考えられる．

そこで，本研究では，従来の領域分割とラベリングによる 2 次元的な認識ではなく，物体の機能を反映した構造モデルの定性的モデル当てはめによる物体の定性的 3 次元構造の推定と，その結果に基づく物体間の支持関係に関する推論を行うことによって，オクルージョンを多く含むような室内画像に対する認識システムの提案を行う．最初のモデルの当てはめは，画像から抽出したエッジ，領域をグループ化した画像特徴に対して行う．本研究では，同一種類であっても多様な形状，多様な見え方をもつ実世界の物体により広く対処できることに重点を置いているので，こうした同一物体の多様な見え方に対応できる方法を採用している．そして，物体が他の物体の上に載っているという関係を表す物体間の支持関係の推論によって，実在しない物体の候補を消去したり，部分的にしか見えていない物体の存在を推定したりすることが可能となり，モデル当てはめによる物体構造の推定の不正確さを補うことができる．このようにして，正確な物体形状が与えられていない場合でも，複雑なオクルージョンが発生している室内シーンの単一画像を認識することが可能となる．

本論文では，まず，定性的モデル当てはめによる物体個々の認識と，その結果に基づく物体間の支持関係に関する推論について述べ，続いて，物体間に通常考えられる関係をあらかじめ記述した関係知識と，それを利用した物体の候補の評価値の計算について述べる．そして，次に，我々が提案したマルチエージェントによる画像認識システム構成法 MORE (multi-agent architecture for Object REcognition) [13], [14] によるシステムの実現について述べ，最後にプロトタイプシステムによる動作例と 20 枚の画像に対する実験結果について述べる．

2. 物体個々の認識方法

「机」「椅子」などの一般名詞で表現される物体は，同一種類であっても様々な形状をもつために，正確な形状 3 次元モデルをあらかじめ用意しておくことは不可能である．そこで，本研究では，モデルは人工物であれば物体の機能など認識対象の本質を表しているような構造 [9], [10]，例えば，椅子なら座面と足，机なら

机上面と足などを表現するようにし，同一種類の物体でなるべく共通となるようなプロトタイプモデルを用意する．そして，画像から得られる物体の部分的な特徴や支持関係から予想される特徴に対してモデルを当てはめることによって，物体の存在を予想すると同時にその物体の定性的な 3 次元構造を推定する．モデルは物体によっては複数個用意して，その場合はその中で後述する画像特徴評価値が最も高いものを選択することにする．こうしたモデルの当てはめによる認識では，広い範囲の物体を認識することが可能になる代わりに，異なる種類の物体間での区別が難しくなり，物体間で認識結果の競合が起こる場合がある．そうした場合は，モデル当てはめの正確さに加えて，後述する他の物体との関係も加味した上で競合を解決して，対象の同定を実現する．

2.1 モデルの表現

モデルは，多角形，線分で表現されるモデル要素 (model element)，及びモデル要素同士の接続関係を表現するモデルグラフ (model graph) によって定義される（図 2 (a)(b)）．図 2 の机の例では，モデル要素は，四つの頂点 (fl,fr,rr,rl) をもち底辺と斜辺の長さがそれぞれ a, b である平行四辺形 (PG) と，二つの端点 (t,b) を上下にもつ長さ c の垂直な直線 (VL) で，それぞれ，実世界中では，水平な長方形の面，垂直な棒であると定義されている．そして，平行四辺形の四つの頂点それぞれに垂直な直線の上の端点 (t) が接続していることをモデルグラフが表現している．このように，「机」なら水平の机上面と垂直な足 (図 2 (e)),「椅子」なら水平な座面と垂直な足というように，多くの種類の「机」「椅子」を代表するような典型的な構造をモデルとする．

図 2 (c) では，後述する支持関係の推論のために，物体が他の物体の上に載っているときにその接面となると推定される支持必要面と，その物体が他の物体を上に載せて支持することができると推定される領域である支持可能面の情報を記述していて，システムがその物体の支持必要要素 (to-be-supported elements) と支持可能要素 (supportable elements) を推定できるようになっている．この例では，平行四辺形 (PG) の面すべてが支持可能要素，四つの垂直な直線 (VL) の下の端点 (b) が支持必要要素であることを表している（図 2 (f)）．この 2 種類の要素の情報は，次節で述べる支持関係のチェックで用いられる．また，図 2 (c) の二つの不等式は，モデル当てはめ時に使われる各モデル要素の画像中での大きさの比の範囲を表している．図 2 (d) は，後述する画像特徴評価値の計算に用いられる各モデル要素についての重み値を定義している．

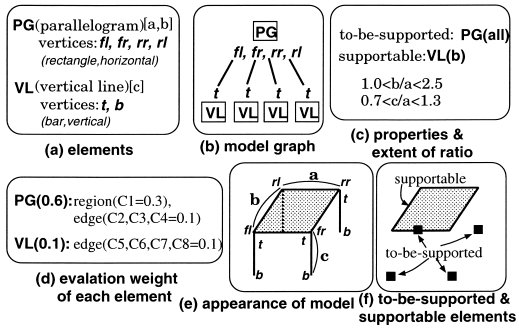

図 2 机のプロトタイプモデルの一例
Fig. 2 An example of model representation of “desk.”

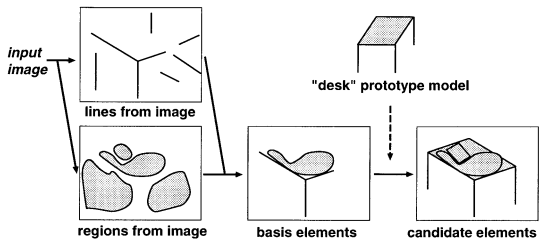

図 3 抽出された特徴に対して，3 次元構造モデルを当てはめて候補要素を推定する
Fig. 3 Flow of estimating an object candidate of “desk.”

2.2 モデルの当てはめ方法

はじめにシステムは与えられた画像から，モデル当てはめに必要な画像特徴として，エッジ，領域を抽出する．これらの領域，エッジはモデル当てはめの根拠になる画像要素なので，根拠要素 (basis element) と呼ぶことにする．基本的には，根拠要素の抽出は既存のアルゴリズムを用いることとし，Canny edge detector [2] によるエッジ検出，Hough 変換による直線抽出，領域成長法による領域分割などのアルゴリズムを組み合わせて行う．そして，更に Perceptual grouping [6] の手法を用いて，抽出した直線をグループ化して，端点が近接している直線対，平行直線対，平行直線対に 1 本の直線が加わった U-shape，平行四辺形を抽出する．

次に，根拠要素に対して各モデルのモデル要素を定性的に当てはめることによって，物体の存在を予想し，

更に単一画像に対する物体の 3 次元構造の推定を行う（図 3）．当てはめは，全部のモデルについて，モデル要素の最も大きい平面要素から行い，順次，小さい要素を当てはめていく．当てはめが成功するかどうかは，モデルにあらかじめ与えられている各要素の画像中での大きさの比の範囲を満たした上で，後述する画像特徴評価値がある一定の値以上の値になるかどうかで判断する．もし，当てはめが成功すれば，物体候補が生成できたとみなす．

また，モデル要素の水平，垂直の属性，支持必要面，支持可能面の属性を用いて，物体のどの部分が水平，垂直で，どの要素が支持可能要素，支持必要要素であるかを推定できる．あくまでも定性的当てはめなので，定量的な正確さはないが，定性的推論に必要な程度の物体のおおよその大きさや向き，位置など推測することはできる．以上のモデル当てはめに基づいて推定された物体の存在が予想される領域及びエッジをまとめて候補要素 (candidate element) と呼ぶことにする．ここでの候補要素はオクルージョンがない場合に本来見える物体全体の見え方を推定した場合の領域である．

モデル間での競合を解消する場合に用いられる評価値である画像特徴評価値 V_{im} は 0 から 1 の間の値をとる値で，候補要素の各部分と根拠要素との対応の割合に応じて計算される．根拠要素が候補要素に近いほど評価値が 1 に近くなるように，V_{im} を以下のように定義する．

$$V_{im} = \min\left(\sum_{i=1}^{n} C_i \frac{b_i}{e_i}, 1\right) \qquad (1)$$

式中の n は合計のモデル要素数である．ただし，机上面のような平面で表現されるモデル要素は，内部の領域と輪郭線を別々に考えるので，図 2 の机のモデルでは，机上面の領域，輪郭，4 本の足の合計で $n = 6$ となる．C_i は各モデル要素の重要度を表現する重みで，現在の実装では机の場合は机上面の領域の重みを 0.4，輪郭を 0.3，足をそれぞれ 0.1 としている．b_i は各要素の根拠要素の画素数，e_i はモデル当てはめによって得られた対応する候補要素の画素数をそれぞれ表している．

なお，机，椅子などの物体とは別に，特定の形状をもたない平面的な「床」「壁」などの通常，背景となる物体は，図 4 のように得られた根拠要素に対して，平面的なモデルを当てはめることにする．これら背景物体の場合は，モデル要素は平面領域のみとなり，e を候補要素の画素数，b を根拠要素の画素数とすると，画像特徴評価値は

$$V_{im} = b/e \qquad (2)$$

となる．

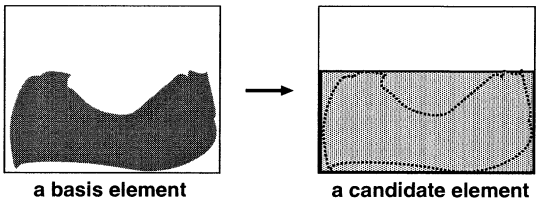

図 4　得られた根拠要素から,「床」の候補要素を推定する
Fig. 4　Estimating an object candidate of “floor.”

3. 支持関係のチェック

支持関係とは，下にある物体が上にある物体を支えている関係のことである．実世界のすべての物体には重力がかかっているので，その下にその物体を支える他の物体がなければならない．人間はこうした物理法則を経験的に知っていて，それは無意識のうちに知覚に影響を与えていると考えられる．そこで，本システムでは，この物体が物体を支えるという関係を物体の認識に利用することとする．具体的には,「床」「壁」などの背景物体を除くすべての物体について，物体候補が生成されたら，その物体を支持することのできる物体の候補が既に生成されているかどうか調べ，もしなければ支持可能な物体候補の存在を認識するように要求を出す．そして，もし最終的にその物体を支持する物体候補が見つからなければ，どの物体からも支持されていない物体は実在しないとみなして，候補を消去する．

支持関係のチェックは，支持必要要素の領域と他の物体候補の支持可能要素の領域が重複しているかどうかを調べることによって行う（図 5）．もし，支持可能要素領域が支持必要要素領域のほとんどを含んでいれば，支持可能要素をもつ物体が支持必要要素をもつ物体の下にあって支えているとみなし，その両方の物体の間には支持関係があるとする．支持関係の成立は，後述する関係知識とは無関係であり，2 種類の領域の関係のみで判断する．なお，図 5 では，床が本を直接支持しているとも判断できるが，物体候補 A が物体候補 B を支持していることを $A \Rightarrow B$ と書くとす

ると，床 ⇒ 机，机 ⇒ 本　という関係があるので，床 ⇒ 机 ⇒ 本 と判断できる．

もし，物体候補の支持必要要素に対して支持する物体候補がなければ，その支持必要要素を仮想根拠要素 (virtual basis elements) として，支持する可能性がある物体の根拠要素の一部とみなすようにする（図 6）．そして，その仮想根拠要素を含む物体候補を生成可能であるかどうか，後述する関係知識にその候補を支持する可能性があると記述されている各モデルについて調べる．

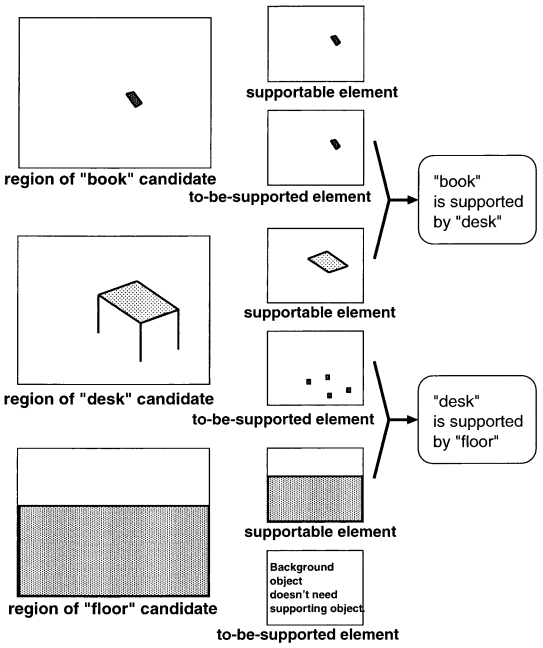

図 5 「支持関係」のチェック
Fig. 5 Checking “supporting relation.”

図 6 では，はじめにワークステーション（以下，WS と略す）の候補が生成されて，その支持物体が存在しないので，WS 候補の支持必要要素を WS 候補を支持する候補の仮想根拠要素とみなす．そして，仮想根拠要素を根拠要素であると仮定して，仮想根拠要素の周辺部に更に根拠要素となる画像特徴が存在するモデルを探す．この場合，机が WS を支持する可能性があるという関係知識が存在し，更に，仮想根拠領域の周辺に机の根拠要素を見つけることができたので，両者を合わせて机の根拠要素として，モデルを当てはめることによって，机を検出することができた．

このように仮想根拠要素の考え方を導入することによって，上に物体が載って大きなオクルージョンが発生しているために認識不可能であった物体が認識可能となる．

4. 関係知識と物体候補の評価

4.1 関係知識

システムは，自分の物体と他の物体の間の通常考えられる関係についての知識，関係知識をあらかじめもっている．これは必ずしも成り立つ必要はないが，多くの場合成り立っている関係で，物体候補の関係に

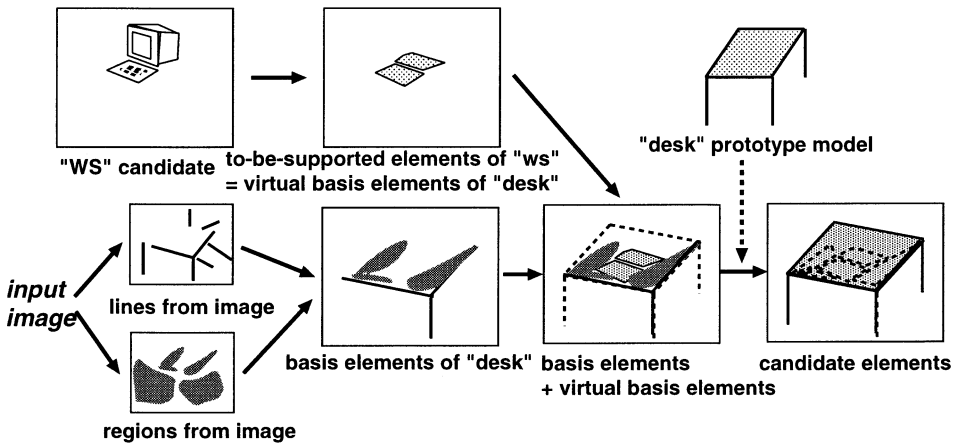

図 6 仮想根拠要素と通常の根拠要素を統合してから，モデル当てはめを行うことによって支持物体を認識
Fig. 6 Estimating a “desk” candidate that supports the “workstation” candidate by combining virtual basis elements and basis elements.

関する評価，及び支持物体の探索に利用される．関係知識に記述されている関係が周辺の物体候補との間に成り立っていると，その物体候補の関係についての評価値が高くなる．本システムで用いられている関係知識は 2 物体間の相対的な関係を記したもので，あらかじめシステムに与えておく．具体的には 2 物体間の位置関係である．これらの関係はすべて定性的なものとして表現される．

関係知識は「関係 (物体名, 物体名)」という形で表現される．例えば,「本は机の上にある」「椅子と机は同じ平面上にある」という関係は，‘‘on(book,desk)’’, ‘‘next-to(chair,desk)’’ というように表現される．現在の実装では，この 2 種類のみを利用している．

関係知識の評価は，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う．‘‘on(A,B)’’ の場合，A 物体候補を支持している B 物体候補があれば，その関係知識が成り立っているとみなす．‘‘next-to(A,B)’’ の場合は，A 物体候補と B 物体候補が同じ物体候補に支持されていれば，つまり，同じ支持平面上に載っていれば，関係知識が成り立っているとみなす．

ある一つの物体候補について，その物体と他の物体の間に成立している関係をすべて調べて，各関係の成立した数に基づいて計算したものが，0 から 1 の間の値をとる関係評価値 V_{re}(式 (3)) となる．関係評価値は，物体候補が他の物体候補との間にどの程度通常成り立っていると期待される関係が成立しているかを示した評価値であり，その物体候補のシーン中での存在の自然さを表現している値であるといえる．関係評価値の計算式（式 (3)) は，成立した関係の数が 0 個と 1 個では関係評価が大きく違うが，5 個と 6 個だとそれほど大きな違いがないということを反映した式となっている．

$$V_{re} = 1 - \exp\left(-k \sum_{i=1}^{r} c_i n_i\right) \qquad (3)$$

r は関係の種類数，c_i は関係 i についてのあらかじめ決められている重みで関係の重要度を表現している．on の場合 1.0, next-to の場合 0.5 に設定している．n_i は関係 i について成立した数をそれぞれ表す．k は定数であり，現在の実装では実験から求めた値である 0.4 に設定している．

4.2 物体候補の評価

複数の物体候補の仮想根拠要素を除く根拠要素が重複して，競合が起こった場合，画像特徴評価値 V_{im} と関係評価値 V_{re} から計算される候補評価値 V を用いて解決する．

$$V = (V_{im} \times S' + V_{re} \times w)/(S' + w) \qquad (4)$$

$$S' = \min(S, 2w) \qquad (5)$$

S は候補要素の画素数を表す．w は，画像特徴評価値と関係評価値の重みのバランスを決める定数で，候補要素の画素数が $2w$ 以上のときは V_{im} と V_{re} の重みが 2 : 1，それ以下の場合は $S : w$ になる．w は現在の実装では 2500 に設定している．

候補評価値 V が最も大きい候補を最終的な候補として採用する．一方，評価値が小さかった候補は取消しになる．

5. システムの概要

5.1 システムの基本構成

システムは，我々が提案したマルチエージェントによるシステム構成法 MORE (multi-agent architecture for Object REcognition) [13], [14] に基づいて構築する．システム構成法 MORE では，システムは単一種類の物体のみを認識する複数のエージェントのみから構成され，中央管理機構は存在しない（図 7）．そのため，エージェントの追加によりシステムを拡張することが可能であり，また，エージェントごとに異なる知識表現，認識手法を用いることができるために，大規模な画像理解システムの構築に向いている．それに加えて，処理の流れが固定されていないために，トップダウン処理とボトムアップ処理を柔軟に融合できるという特徴もある．

各エージェントは，単一クラスの物体を認識する認識モジュールと，エージェント間での協調を行う通信

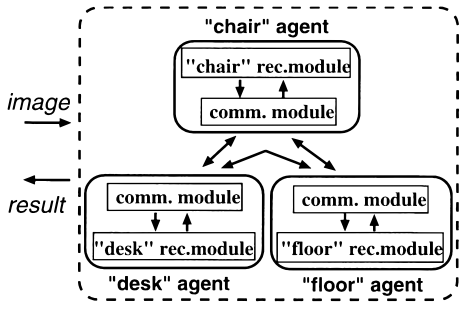

図 7 システムの基本構成
Fig. 7 Basic structure of the system.

モジュールの二つから構成される．

認識モジュール (recognition module) は，物体モデルをもっており，画像中に含まれるある単一種類の物体を認識する．認識モジュールは，新しく候補を発見するたびに，根拠要素，候補要素，支持必要要素，支持可能要素，画像特徴評価値などの情報を通信モジュールに送信する．

通信モジュール (communication module) は，認識モジュールの認識結果がシステム全体で整合性が保たれるように他のエージェントに結果の提示と競合解消のための交渉を行う．通信モジュールは，認識モジュールが生成した物体候補の情報を受け取って，システム内で矛盾した結果が存在しないように，互いに情報を交換し合い，常にシステム内のすべてのエージェントの認識結果の整合性が保たれるように処理を行う．また，通信モジュールは支持関係のチェックや，保持している関係知識を利用して関係評価値の計算を行う．

5.2 システムの動作の概要

エージェント及びそれを構成するモジュールの動作は，すべてメッセージ駆動によって行われる．

はじめに認識対象の画像が全エージェントの認識モジュールに送られ，通信モジュールが「初期認識要求」を認識モジュールに送る（図 8(1)）．認識モジュールは動作を開始し，一つ物体を認識するたびにその認識結果を物体候補として通信モジュールに送る（図 8(2)）．それを受け取った認識モジュールは，他の全エージェントに対してその物体候補の情報を送信する（図 8(3)）．更に，物体候補が背景物体でない場合に，他の物体の候補情報と照合した結果，支持物体を発見できないなら，仮想根拠要素を生成して，その情報を「支持要求メッセージ」として送信する（図 8(4)）．物体候補の情報を受け取った他のエージェントの通信モジュールは，それが既に認識されている自分の物体候補の根拠要素と重複がないかチェックする．もし重複があれば，そのエージェントは「異義メッセージ」を返信し（図 8(5)），両エージェントの間で競合解消の処理が行われる．また，「支持要求メッセージ」と仮想根拠要素の情報を受け取ったときは，その情報を認識モジュールに送って，再びモデル当てはめの処理を行う（図 8(6)）．

競合解消によって候補が消去されると，通信モジュール内にその競合候補の識別番号を記憶しておいて，競合候補が更に別の候補に取り消された場合，若しくは，関係評価値が上昇して競合解消の結果が逆転した場合に，その候補を復活させることにする．こうして，常に互いに整合性のとれた認識結果のみを残すようになっている．このように，常にシステム全体で認識結果の整合性がチェックされ，やがて，すべてのエージェントの認識が終了し，メッセージ待ち状態になると，システム全体の認識が終了する．

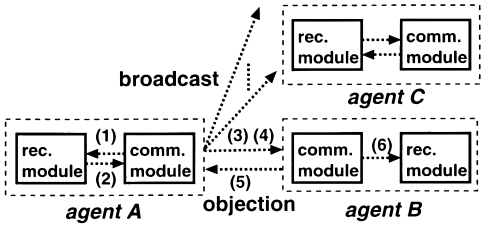

図 8 メッセージの流れ．(1) 初期認識要求．(2) 物体候補の情報．(3) 物体候補の情報のブロードキャスト．(4) 支持要求メッセージ．(5) 異義メッセージ．(6) 再認識要求

Fig. 8 Flow of messages. (1) Initial recognition request. (2) Information of a new object candidate. (3) Broadcasting information of a new object candidate. (4) To-be-supported request message. (5) Objection message. (6) Re-recognition request.

6. プロトタイプシステムによる実験

「机」「椅子」「床」「本」「ワークステーション（以下 WS と略す）」「壁」の 6 種類のエージェントを構築し，6 台の PC (Intel Celeron 450 MHz) からなる PC クラスタ上に PVM [5] を用いて，プロトタイプシステムを実装した．プロトタイプシステムの実装においては，負荷分散については重点を置いていないので，1 エージェントを 1PC として実装し，各認識モジュールは独立に根拠要素の抽出の処理を行っている．根拠要素の抽出の処理を認識モジュール同士で共有することは可能ではあるが，現在の実装では行っていない．本章では，比較的単純な室内シーンとやや複雑なシーンの 2 枚の画像に対するシステムの動作の説明と，20 枚の室内画像に対する実験結果を示す．

6.1 動 作 例

システムは画像（256 階調濃淡画像）が与えられると，画像を各エージェントの認識モジュールに画像を送信する．そして，認識モジュールがエッジや領域などの画像特徴の抽出を開始する．入力画像としてサンプル画像 1 (図 9, 480×360) が与えれたとすると，まず，各エージェントは，エッジ抽出，直線抽出，領域分割などの特徴抽出処理によって，直線エッジ（図 10）

図 9　サンプル画像 1
Fig. 9　Sample image no.1.

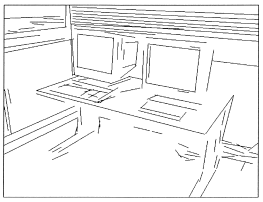

図 10　直線エッジ
Fig. 10　Straight edges.

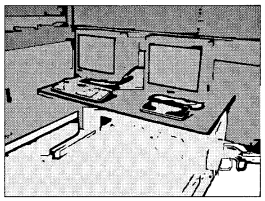

図 11　領域分割の結果
Fig. 11　Regions.

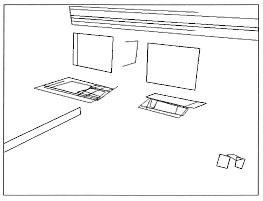

図 12　直線エッジからグループ化によって抽出した平行四辺形及び U-shape
Fig. 12　Parallelograms and U-shapes.

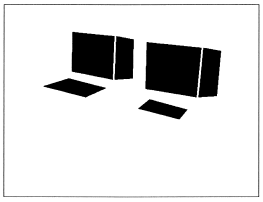

図 13　WS 候補
Fig. 13　Two “WS” candidates.

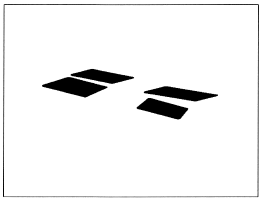

図 14　WS の支持必要要素
Fig. 14　To-be-supported elements of two “WS” candidates.

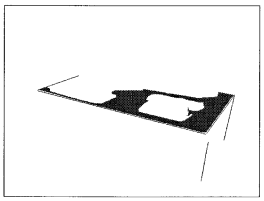

図 15　机の根拠要素
Fig. 15　Basis elements of a “desk.”

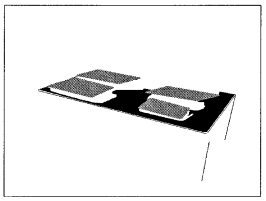

図 16　机の根拠要素と仮想根拠要素
Fig. 16　Virtual elements and basis elements of a “desk” candidate.

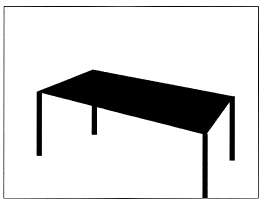

図 17　机　候　補
Fig. 17　A “desk” candidates.

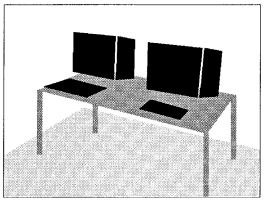

図 18　最終認識結果
Fig. 18　Recognition result.

や分割領域（図 11）が得られる．次に，直線エッジをグループ化して，端点が近接している直線対，平行直線対，平行直線対に両端点を結ぶ 1 本の直線が加わった U-shape，平行四辺形を抽出する（図 12）．すると，WS エージェントは，最も顕著な特徴として，WS のディスプレイの表示部分の平行四辺形を二つ抽出し，領域分割の結果からも同じ位置に平行四辺形領域を抽出する．そして，更にその手前のキーボードが一つは平行四辺形として，もう一つは U-shape として抽出する．次に，これらの特徴に対して，平行四辺形のキーボード，左右の辺が画像中で垂直方向である平行四辺形のディスプレイの前面を当てはめることによって，WS 候補を生成する（図 13）．WS 候補の右側面はディスプレイ前面の右奥にある垂直直線エッジから推定することができた．そして，推定したディスプレイの前面と側面の四辺形の下部の辺から，その 2 辺をもつ平行四辺形領域を推定して，その領域とキーボードの領域を，支持必要要素の領域とする（図 14）．こうして，WS エージェントは二つの WS 候補を生成するが，どちらの候補もその候補を支持する支持物体が存在していない．そこで，WS エージェントは新規に生成された WS 候補の情報を他の全エージェントに送信するときに，同時に仮想根拠要素の情報を含んだ「支持要求メッセージ」も送信する．

一方，机エージェントは，最初は直線エッジと分割領域の画像特徴だけからは，特徴が十分でなく，机を検出することができない．ところが，しばらくすると，WS エージェントから仮想根拠要素の情報を含んだ「支持要求メッセージ」が送信されてくる．机エージェントは，関係知識 `on(desk,WS)` をもっているので，WS 候補の支持必要要素を仮想根拠要素とみなして，仮想根拠要素を含む机候補を認識しようとする．すると，仮想根拠領域の周辺に，図 15 のような直線エッジと領域を見つけ，それを根拠要素として，仮想根拠領域と併合することによって，図 16 のように全体として十分な根拠要素を発見できる．そして，この根拠要素に対して，モデル当てはめを行うことによって，机候補を認識することができる（図 17）．また，更に，机候補からの「支持要求メッセージ」によって，床候補が正しく検出される．最終的には，図 18 のように二つの WS，机，床が認識された．

次に，サンプル画像 1 に比べるとやや複雑な画像の例として，サンプル画像 2 (図 19, 640×480) についての認識について述べる．この画像に対する実験で

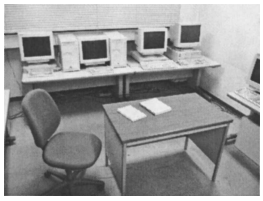

図 19　サンプル画像 2
Fig. 19　Sample image no.2.

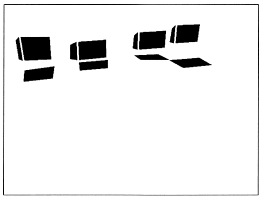

図 20　四つの WS 候補
Fig. 20　Four “WS” canditates.

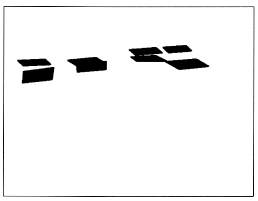

図 21　四つの WS の支持必要要素
Fig. 21　To-be-supported elements of four “WS” candidates.

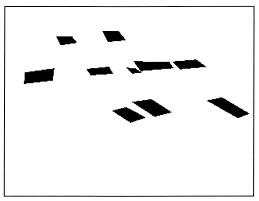

図 22　本　候　補
Fig. 22　“Book” candidates.

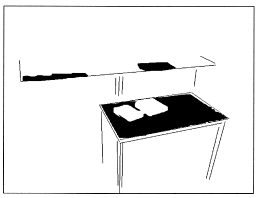

図 23　机の根拠要素
Fig. 23　Basis elements of “desk” candidates.

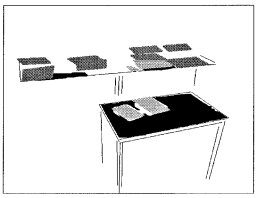

図 24　仮想根拠要素と根拠要素
Fig. 24　Virtual basis elements and basis elements of “desk” candidates.

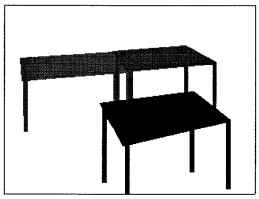

図 25　三つの机候補
Fig. 25　Three “desk” candidates.

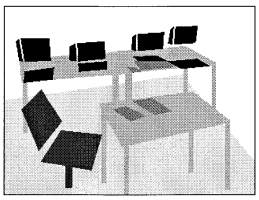

図 26　最終認識結果
Fig. 26　Recognition result.

は，初めに，机上の四つの WS(図 20) が認識され，その支持必要要素は図 21 のようになった．また，四つの WS の認識とほぼ同時に本の候補も認識され，候補が 10 個生成された（図 22）．そのため，右から 2 台目の WS を除く 3 台の WS と本の間に競合が起きているが，競合解消の処理によって，三つとも WS が残り，本が消去された．例えば，最も右の WS のキーボード部分と本との競合では，それぞれ画像特徴評価値 V_{im} が 0.78, 0.59，関係評価値 V_{re} が 0.33, 0.33 となり，候補評価値 V は 0.62, 0.43 となったために，本候補が消去された．なお，右から二つ目の WS 候補はキーボードの位置が誤って認識されたため，本来のキーボードが本として認識されてしまっている．

また，サンプル画像 1 の場合と同様に，WS 候補，競合解消で残った本候補には支持物体が存在しないので，「支持要求メッセージ」が発行されて，机エージェントは仮想根拠要素である WS の支持必要要素（図 21）を机の根拠要素（図 23）及び本の候補の支持必要要素と統合して（図 24），オクルージョンが多く，机上面があまり見えていないにもかかわらず机を認識することができる（図 25）．ただし，競合が起こらなかった七つの本候補のうち，三つについては支持物体を発見できず，最終的には支持物体なしとなって，消去された．実際，これらの本は誤って検出されたもので，支持関係のチェックによって，正しく消去された．最終的には，図 26 に示すように，後方右の机上に誤って二つの本が認識された以外は，WS，机，本，椅子，床などがほぼ認識ができた．なお，ここでは，WS エージェントの認識モジュールを暫定的に WS をディスプレイとキーボードの組で認識するように実装したため，後方左の机の上にある二つの WS 本体の箱は認識されていない．

6.2　実 験 結 果

20 枚の室内画像（画像サイズはすべて 480×360）に対して，実験を行った．実験で用いた 20 枚の画像のうちの一部を図 27 に示す．上段が単純な画像，下段は複雑な画像，中段は中程度の複雑さの画像を示しており，それぞれについて，7 枚（サンプル画像 1 を含む），7 枚，6 枚（サンプル画像 2 を含む）用意した．結果は「ほぼ認識できている (almost correct)」「半分程度認識できている (half correct)」「ほとんど認識できていない (almost incorrect)」の 3 段階に分けて評価し，それぞれ，9 枚，6 枚，5 枚となった（表 1）．

6.3　実験結果に対する考察

20 枚の室内画像に対する実験において，「ほぼ認識できている」画像としては，先に説明したサンプル画像 1 (図 9) がその一例である．支持関係と仮想根拠要素を用いた認識によって，机上面が 2 台の WS によってほとんど隠されてしまっているシーンであるが机の認識が可能となっている．

次に，「半分程度認識できている」画像の例を図 28 に示した．この画像では，机の上にあるノート PC，積み重ねられた本，本棚と並べられた本，広げられた

図 27　評価に用いた画像の一部．上段が単純な画像，下段は複雑な画像，中段は中程度の複雑の画像
Fig. 27　12 out of 20 images for experiments. Images in the upper row are quite simple, ones in the lower row are complex, and ones in the middle row have middle complexity.

表 1　20 枚の画像に対する認識結果
Table 1　Results for 20 images.

almost correct	half correct	almost incorrect
9	6	5

図 28　認識に部分的に成功した画像の例
Fig. 28　An example image whose results are half correct.

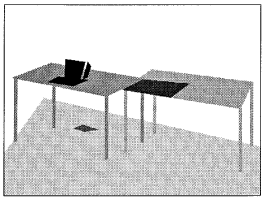

図 29　認識結果
Fig. 29　Recognition result.

図 30　認識に失敗した画像の例
Fig. 30　An example image whose results are almost incorrect.

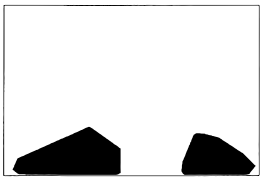

図 31　認識結果
Fig. 31　Recognition result.

ノートなどが示されているが，認識結果では図 29 に示したように，ノート PC が WS，広げられたノートが本と認識された．そのため，支持関係から，机の存在がほぼ正しく推定されている．しかし，本を平行四辺形として認識しているので，平行四辺形若しくは U-shape が検出されないと本を認識することができず，積み重なっている本や，本棚に入っている本は認識できていない．関係知識 `on(book,book)` を利用したとしても，一番上の本が認識できないと，その下の本も認識できない．このような場合に対処するには，本が数冊まとまって存在している場合のモデルを用意する必要があると思われる．

「ほとんど認識できていない」画像の例は図 30 で，認識結果は図 31 に示す．この画像は非常に複雑であり，最終結果では，床の一部しか認識できていない．WS のディスプレイは明確に写っているものの，机の上のキーボードは机の面と色が似ているために認識することが困難で，認識できていない．そして，そのため，WS の下の机も認識できていない．これは画像の解像度の限界によるところが大きく，入力画像に解像度の高い画像，例えば 2400×1800 くらいの画像を利用して，はじめは低解像度の画像を解析して，必要に応じて詳細な画像を利用するというような多重解像度解析を利用するのが望ましいと考えられる．また，中央の机の上には，モデルにない物体が置かれており，これも認識が不可能となっている．中央の机の最も手前に置かれているのは，定型がない包装材であるが，この種の物体は現在の認識方法ではモデルを用意することが難しい．動的輪郭やテクスチャ解析などの手法を取り入れることが必要であると考えられる．

7. 関連研究

本研究の題材とするようなあらかじめ物体の完全なモデルが得られない場合のシーンの認識は，古くは Tenenbaum [12] 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 Ohta [8]，The Schema System [4]，SIGMA [7] などの画像中の物体ごと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，ボトムアップだけでなく，ボトムアップとトップダウンの融合を実現した．本システムもこうしたシステムの流れを汲む知識ベース型の画像理解システムである．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．

本研究におけるシステムでは，部分的な特徴に対して定性的 3 次元構造モデルを当てはめることにより，物体の支持必要面，支持可能面を推定し，物体間の支持関係に基づく，物体候補の検証をしているのが特徴である．このように，まず個々の物体ごとに候補を生成して，その後物体間の関係で候補の検証を行うシステムとしては，Strat らによる CONDOR [11] がある．彼らはこのような認識手法を context-based recognition と名づけている．ただし，Strat らのシステムは屋外シーンを対象とした基本的に 2 次元的な認識であり，我々のような定性的 3 次元認識を行っているわけではない．

また，ある物体が他のある物体によって支えられているといったような物体間の力学的物理法則を考慮してシーンを理解する研究として Cooper らの研究 [3] や Brand による研究 [1] がある．これらの研究では，定性的な力学法則の知識に基づいて画像によるシーンを解析を行っている．ただし，これらの研究では物体の認識が目的ではなく，物体の物理的作用の画像からの理解に焦点が当てられている．そのため，対象となる物体は，ブロックなどの単純なものが多い．

本研究では，机，椅子などの一般名称で表される物体を認識する際にその物体の機能を提供するような本質的な構造に注目して認識するが，こうした認識は Stark らによって提唱されている function-based recognition [9], [10] の考え方である，人工物の本質は人間にとっての機能の提供であり，認識においても機能の提供の有無で人工物を判断するべきであるという考えを取り入れている．

8. むすび

本研究では，定性的モデル当てはめによって物体候補の定性的な 3 次元構造を推定し，更に物体間の支持関係を確かめることにより，物体の候補の検証して，全体として整合のとれた認識を実現する方法について提案した．そして，更に，こうした認識を実現するためのプロトタイプシステムを我々が従来より研究しているマルチエージェント型の画像認識システムとしてプロトタイプシステムを実装し，実験により複雑なオクルージョンを含む室内画像に対応できることを示した．

現在のシステムでは，画像中の物体の大きさがある程度より大きくない場合，認識モジュールの 3 次元構造モデルの当てはめのうまくいかないことがあるので，今後の課題として，多重解像度解析を導入し高解像度の画像を入力画像とすることが挙げられる．また，より実用的なシステムを目指すために，認識物体の種類を容易に増やすことができるように学習機構を取り入れていくことを検討している．

今後エージェントの数を増やしてシステムを大規模化する場合には，現在のシステムでは実験中に特に問題となっていない認識結果の収束性や一意性などについて，システムの挙動を解析する必要がある．

謝辞　本研究の一部は，稲盛財団研究助成金による．

文　献

[1] M. Brand, “Physica-based visual understainding,” Computer Vision and Image Understanding, vol.65, no.2, pp.192–205, 1997.

[2] J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.5, no.2, pp.140–150, 1986.

[3] P.R. Cooper, L.A. Birnbaum, and M.E. Brand, “Causal scene understanding,” Computer Vision and Image Understanding, vol.62, no.2, pp.215–231, 1995.

[4] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman, “The schema system,” Int. J. Computer Vision, vol.3, no.2, pp.209–250, 1989.

[5] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek, and V. Sunderam, PVM: Parallel Virtual Machine, The MIT Press, 1994.

[6] D.G. Lowe, Perceptual Organization and Visual Recognition, Kluwer Academic Publishers, 1985.

[7] T. Matsuyama and V.S. Hwang, SIGMA: A knowledge-based aerial image understanding system, Plenum Press, New York, 1990.

[8] Y. Ohta, Knowledge-Based Interpretation of Outdoor Natural Color Scenes, Pitman Advanced Publishing Program, Boston, 1985.

[9] L. Stark and K. Bowyer, “Achieving generalized object recognition through reasoning about association of function to structure,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.13, no.10, pp.1097–1104, 1991.

[10] L. Stark and K. Bowyer, “Function-based generic recognition for multiple object categories,” Computer Vision and Image Understanding, vol.50, no.1, pp.1–21, 1994.

[11] T.M. Strat and M.A. Fischler, “Context-based vision: Recognizing objects using information from both 2-d and 3-d imagery,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.13, no.10, pp.1050–1065, 1991.

[12] J.M. Tenenbaum and H.G. Barrow, “Experiments in interpretation guided segmentation,” Artif. Intell., vol.8, pp.241–274, 1977.

[13] K. Yanai and K. Deguchi, “An architecture of object recognition system for various images based on multi-agent,” Proc. 14th Int. Conf. Pattern Recogni-

tion, vol.1, pp.278–281, 1998.
[14] 柳井啓司，出口光一郎，“マルチエージェントによる多様な画像に対応した物体認識システムの一構成法,” 情処学論，vol.39, no.2, pp.170–177, 1998.

（平成 12 年 10 月 2 日受付，13 年 1 月 29 日再受付）

柳井　啓司

1995 東大・工・計数卒．1997 同大大学院情報工学専攻修士課程了．1997 年 10 月より電気通信大学情報工学科助手．画像理解システム，画像データベース，並列処理などに興味がある．情報処理学会，人工知能学会等各会員．

出口光一郎　（正員）

1976 東大大学院修士課程了（計数工学）．同年より東京大学工学部助手，講師を経て，1984 山形大学工学部情報工学科助教授，1988 東京大学工学部計数工学科助教授．1998 東北大学大学院情報科学研究科教授．この間，1991〜1992 米国ワシントン大学客員準教授．コンピュータビジョン，画像計測，並列コンピュータの研究に従事．情報処理学会，計測自動制御学会，形の科学会，IEEE 等各会員．