

画像理解システム SIGMA†

—ボトムアップ、トップダウン解析の統合—

松山 隆 司** ビンセント ハング***

本論文では画像理解システム SIGMA の構成と幾何学的推論における制御構造について述べる。SIGMA は次の三つのエキスパートから構成されている。(1)対象物間の幾何学的関係や幾何学的構造に関する知識に基づいた推論を行う Geometric Reasoning Expert (GRE), (2)対象物の具体的なモデルや画像上での見え方を推論する Model Selection Expert (MSE), (3)画像処理に関する知識を利用して柔軟で効率的な画像データの解析を行う Low Level Vision Expert (LLVE)。GRE における幾何学的推論では、ボトムアップ、トップダウンの二つの解析プロセスが一つの推論過程に統合され、状況に応じていずれかのプロセスが起動される。また、トップダウン解析においては GRE, MSE, LLVE の三つのエキスパートが互いに役割を分担し、未知の対象物を検出する。本論文では、GRE における幾何学的推論のメカニズムを中心に SIGMA の機能を説明し、プロトタイプシステムによる航空写真の解析結果を示す。

1. はじめに

1970年代の後半において、多くの実験的な画像理解システムが作られ、種々の応用分野でその有効性が示された¹⁾。こうした一連の実験の結果、次のような問題点が明らかになった。

(1) 不十分なセグメンテーション能力：セグメンテーションに関しては、画像処理研究の当初から多くの手法が提案されてきたが、いずれの手法も完全ではない。

(2) 対象物の多様な見え方に対する対応策：3次元物体は、見る角度によって見え方が大きく変化する。また、一つの対象物のクラスには多種多様な特徴をしたものが含まれる。こうした多様な見え方をする対象物をうまく認識するには、認識対象のモデルの柔軟な表現法とその効率的な利用法が不可欠となる。

(3) 知識の明示的な表現：従来の多くの画像理解システムでは、知識が各プロセスに分散されていたり、プログラムの形で表現されていたりした。そのため、どの知識がどのように用いられるのかが明確でなく、機能の拡張や異なった分野への適用が困難である。

これらの問題のうち、3次元物体の認識に関しては

コンピュータビジョン研究として現在活発に研究が進められている^{2)~4)}。一方、人工知能研究の考え方を画像理解に応用し、知識に基づいた画像の解析を行うとする研究も各所で進められている^{5)~9)}。

本論文で述べる画像理解システム SIGMA は、いわゆるフレームによる知識の表現を利用し、(1)~(3)の問題に対処しようとするものである。ここでは、SIGMA の全体的構成と幾何学的推論における制御方式について述べ、航空写真を対象とした幾つかの実験結果を示す。

2. システムの構成

SIGMA は次の三つのエキスパートから構成されている(図1)。

(1) Geometric Reasoning Expert (GRE): このエキスパートは、システムにおける中心的な役割を果たすエキスパートで、対象物の幾何学的構造および対象物間の幾何学的関係に基づいた推論を行う。GRE における幾何学的推論では、トップダウン、ボトムアップの両解析プロセスが一つの推論過程として統合され、状況に応じていずれかのプロセスが起動される。

トップダウン解析においては、このエキスパートで得られた情報が次のエキスパートに渡され、そのエキスパートが実際の解析を行う。

(2) Model Selection Expert (MSE): このエキスパートは、GRE からの要求によって起動され、いま注目している対象物の具体的なモデルとその画像上での見え方を推論する。すなわち、情景における対象物のモデルと画像におけるその見え方との間の情報変

† SIGMA: A Framework for Image Understanding—Integration of Bottom-up and Top-down Analyses—by TAKASHI MATSUYAMA (Department of Electrical Engineering, Faculty of Engineering, Kyoto University) and VINCENT HWANG (Center for Automation Research, University of Maryland).

** 京都大学工学部電気工学第二学科

*** メリーランド大学

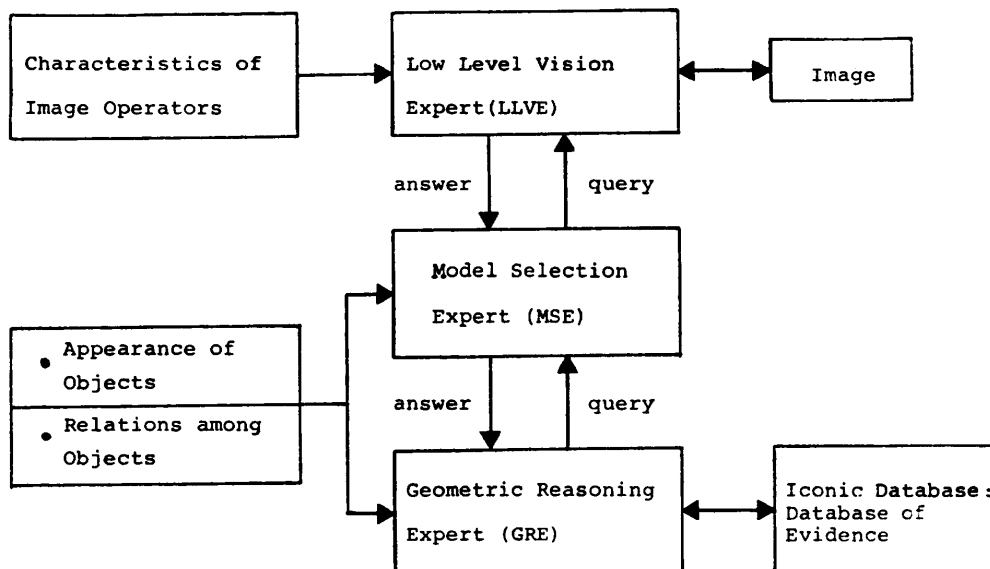


図1 システムの構成

Fig. 1 Overview of the system.

換を行う。

(3) Low Level Vision Expert (LLVE): ここでは MSE で決められた画像上での対象物の見え方に基づき、実際に画像を処理し、指定された画像特徴を抽出する。その際、このエキスパートは各種の画像処理に関する知識を利用し、効率的かつ効果的に画像処理を実行する。

3. トップダウン、ボトムアップ解析の統合

3.1 トップダウン、ボトムアップ解析における幾何学的関係の利用法

REL (01, 02)を二つの対象物のクラス間の幾何学的関係とする。通常のボトムアップ解析では、(1)まず、認識対象である 01, 02 自身の性質 (色, 形など) を利用し画像から対象物の候補となる画像特徴を抽出する。(2)次に、関係 REL (01, 02) がこれらの画像特徴の間で成り立つかどうかを調べ、この関係を満たさない画像特徴は対象物ではないとして候補から取り除く。Constraint Filtering, 弛緩法はこの解析法の典型的な例である。

しかし、複雑な情景の画像理解においては、処理の第一段階で対象物に対応するすべての画像特徴を完全に抽出することは非常に困難である。そのため、(1)の処理で見逃された画像特徴を発見するためトップダウン解析が必要となる。そのようなトップダウン解析では、REL (01, 02) が未検出の画像特徴の位置、形、

大きさなどを予測 (仮定) するのに使われる^{7), 8)}。

このように、複雑な情景の画像理解においてはボトムアップ、トップダウンの両解析法の組合せが不可欠であるが、これらの解析プロセスにおける幾何学的関係の利用法はまったく異なっている (ボトムアップ解析: 制約条件の検証, トップダウン解析: 仮定の生成)。このため、従来の画像理解システムでは、この両者をどのように組合せればよいのか、あるいはいつ一方から他方へ解析を切り換えればよいのか、といったことがあいまいであった。

3.2 SIGMA における幾何学的推論の基本的考え方

ここでは GRE における幾何学的推論の基本的考え方および MSE, LLVE の役割について述べる。

一般に、二つの対象物のクラス 01, 02 の間の関係 REL (01, 02) は、 $01 = f(02)$, $02 = g(01)$ のように二つの関数を用いて表現することができる。02 のクラスのインスタンス (具体的実体) を s とすると、関数 f は、 s をそれと関係 REL を満たすクラス 01 のインスタンス $f(s)$ に写像する。同様のことが関数 g についても成り立つ。

SIGMA では、ある対象物のインスタンスが見つかり、それとある幾何学的関係を満たす対象物に対する仮定 (上の $f(s)$ に当たる) が生成される。仮定は、

- (1) 目的とする対象物が存在すべき位置を表す領域

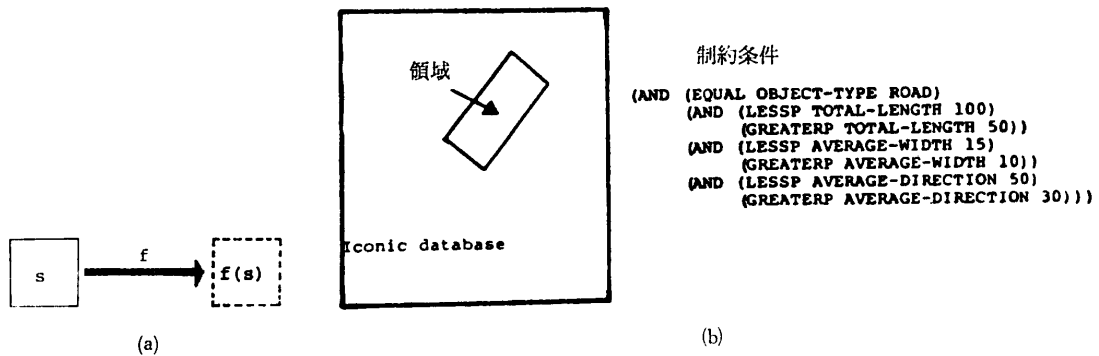


図 2 (a) 仮定の生成, (b) 仮定の記述
 Fig. 2 (a) Hypothesis generation, (b) Description of a hypothesis.

(2) 目的とする対象物の特徴に関する制約条件の集合として表現される (図 2)。

GRE 内のすべての部分情報 (evidence)—対象物のインスタンスおよびそれらによって生成された仮定—は、共通のデータベース (図 1 の Iconic Database) に蓄えられ、それらの間の整合性が調べられる。このデータベースには、対象世界における対象物の位置情報を表すための幾何学的空間 (具体的には多次元配列) が用意されており、すべての部分情報はこの配列上の領域として表現される。GRE は、この配列上での領域間の位置関係 (重なり) を利用し、関係のある部分情報を検索する。この配列は対象世界を表し、その座標系は画像の座標系とはまったく異なる (後述するプロトタイプシステムでは、航空写真を対象としているため、2次元配列を用いているが、対象世界が3次元の場合は3次元配列となる)。こうした位置情報のほかに対象物のインスタンスや仮定に関するシンボリックな情報 (属性や関係) もこのデータベースに蓄えられる。

いま、02 のクラスのインスタンス s が、関係 REL に基づいて 01 のクラスの対象物に対する仮定 $f(s)$ を生成したとし、Iconic Database 中で $f(s)$ とクラス 01 のインスタンス t とが互いに重なったとする (図 3 (a))。このことは、 $f(s)$ に課せられた位置的制約条件が t によって満たされたことを意味する。このとき GRE は、 $f(s)$ に付けられた他の制約条件と t のもつ属性とを比較し、それらの間の整合性を調べる。もし $f(s)$ と t との間に整合性があると、それらは一つの状況 (situation) を形成すると言う。すなわち、状況とは、互いに整合性のある部分情報の集合として定義される。図 3 (a) の場合、GRE は整合性のある二つの

部分情報 $f(s)$ と t を統合 (同一視) することによって、その状況の解を求める。この統合の結果、対象物のインスタンス s と t の間に関係 REL が設定される。このプロセスが SIGMA におけるボトムアップ解析による対象物の関係付けである。

一方、図 3 (b) に示すように、同じクラスの対象物に関する二つの仮定 $f(s)$ と $h(u)$ とが、互いに相手の制約条件を満たし一つの状況を形成している場合、GRE は次のような推論を行う。 $f(s)$ と $h(u)$ との間に整合性があるので、前のプロセスと同じく両者を統合する。しかし、今の場合、状況を構成しているすべての部分情報が仮定であるので、前のようにインスタンスの間に関係を設定することによってその解を求めることができない。そこで GRE は、統合された仮定 (各仮定に付けられた制約条件の積) を満たす対象物のインスタンスの検出を MSE に依頼する。この要求は MSE からさらに LLVE へと引き継がれ、実際の

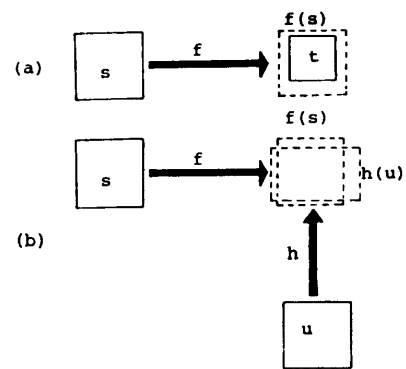


図 3 (a) ボトムアップ解析を起動する状況, (b) トップダウン解析を起動する状況

Fig. 3 (a) Situation which activates the bottom-up analysis, (b) Situation which activates the top-down analysis.

画像の中から目的とする対象物に対応する画像特徴が探索される。要求された対象物のインスタンスが検出されると、その結果が GRE に返される。GRE は、MSE より返されたインスタンスを Iconic Database に登録し、その新しいインスタンスと、仮定を生成していた元のインスタンス (図 3 (b) では s と u) との間にそれぞれ対応する関係を設定する。こうした解析プロセスが SIGMA におけるトップダウン解析である。

以上述べたように SIGMA では、部分情報の蓄積、統合を通じてボトムアップ、トップダウンの二つの解析が一つの推論過程として統合され、状況に応じて一状況の中にインスタンスが含まれるかどうかによっていずれかの解析プロセスが起動される (この推論過程の論理的意味は文献 10) 参照)。

トップダウン解析において GRE から MSE へ出される要求は、目的とする対象物の存在範囲、属性に関する制約条件、およびその周囲の環境条件 (すなわち、その状況を支持しているインスタンス、図 3 (b) では s と u) から成り立っている。MSE は、こうした情報に基づき、目的とする対象物の具体的な見え方を決定する。また、対象世界の座標系から画像の座標系への変換を行い、要求された対象物を画像上での特徴へ変換する。その後、MSE は、推論の結果得られた画像特徴が実際に画像上に存在するかどうかを LLVE に確かめるよう要求する。

LLVE は、要求された画像特徴を画像データから抽出するための最も良い方法を、画像処理に関する知識を用いて決定し、実際に画像を処理する。LLVE では、画像特徴のタイプ、ある画像特徴から別の画像特徴への変換を行う画像処理操作、単純な画像処理操作を効果的につなぎ合わせるためのヒューリスティックが画像処理に関する知識として蓄えられ、柔軟な処理が可能となっている (LLVE の詳細は文献 13) 参照)。

3.3 部分全体関係

SIGMA では 2 種類の幾何学的関係を用いている (図 5 参照)。一つは SP 関係で、二つの異なった対象物間の幾何学的関係を表す。もう一つは PW 関係で、部分全体関係を表す。SP 関係が対等な対象物の間に定義されるのに対し、PW 関係は複雑な内部構造をもつ対象物とその部分との間に定義される。したがって、PW 関係の場合、その関係で結ばれる二つの対象物の間には上下関係が存在し、それを用いた幾何学的推論は、以下に示すようにして行われる (SP 関係

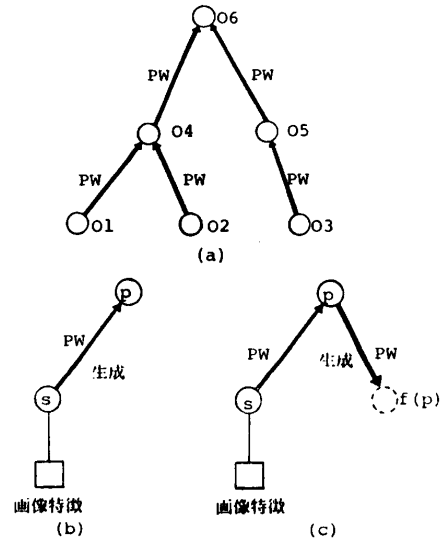


図 4 (a) PW 関係, (b) ボトムアップ的な上位インスタンスの生成, (c) トップダウン的な仮定の生成

Fig. 4 (a) PW relation, (b) Bottom-up instantiation, (c) Top-down hypothesis generation.

による推論は上で述べたものと同じ)。

図 4 (a) に示すような PW 関係によって作られる木構造を考える。一般に、これらの対象物のクラスのうち、最初に具体的なインスタンスが生成されるのは、木構造の最下位に位置するクラスである。なぜなら、最下位のクラスの対象物の見え方は単純な画像特徴に直接対応し、画像から容易に抽出できるからである。

クラス O1 のインスタンスを s とすると、 s は PW 関係を用いてその上位のクラス O4 のインスタンス p を生成する (図 4 (b))。すなわち、PW 関係を下から上へ使うときは、仮定ではなく上位のクラスの具体的なインスタンスが生成される。これは、SP 関係では二つのクラスが対等であるのに対し、PW 関係では上下関係があり、下位のクラスのインスタンスが存在すれば上位のクラスのインスタンスが存在すると考えられるからである。しかし、生成された上位インスタンスは純粋にシンボリックに記述されるだけで、Iconic Database 上の領域としては表現されない。

こうした PW 関係によるボトムアップ的なインスタンスの生成は、無条件に行われるわけではない。PW 関係による木構造に属する各対象物のクラスには、その一つ下のクラスの対象物のうちいずれのものがインスタンスをもてば、その一つ上のクラスのインスタンスを生成すべきかを示す情報が次のような 2 重リストの形で蓄えられている。

((部分 a 部分 b) (部分 c 部分 d) (部分 a 部分 d))

このリストは、部分 a と部分 b、部分 c と部分 d、あるいは部分 a と部分 d のいずれかの組合せに対応する部分のインスタンスが存在するときのみ、このクラスのインスタンスは完全なものとなり、その上位のクラスのインスタンスを生成することを表す。

図 4 (b) において s によってその上位のインスタン

ス p が生成されると、 p はその下位のクラスの対象物でまだ検出されていないものに対する仮定を生成する (図 4 (c))。このトップダウン的な仮定の生成においても、上位のクラスのインスタンスがその下位のすべてのクラスの対象物に対する仮定を生成するとは限らず、ある条件が満たされたときに仮定が生成され

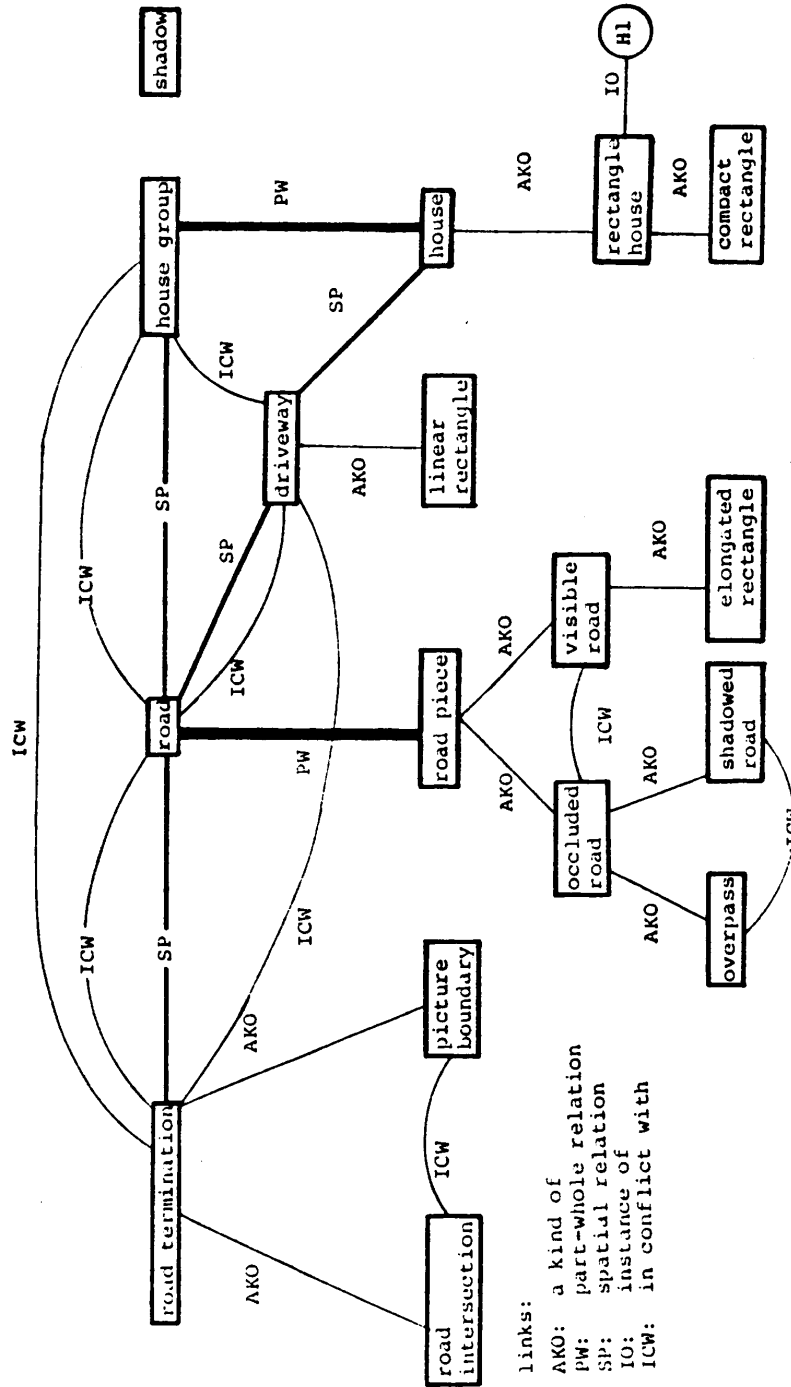


図 5 プロトタイプシステムにおける知識構造
 Fig. 5 Knowledge structure in the prototype system.

る。また生成された仮定は、SP 関係によって生成されたものと同じく Iconic Database 上で領域として表現される。

幾何学的推論過程では上記のような PW 関係に基づくボトムアップ、トップダウン的推論が組合わされ、複雑な構造をもつ対象物が認識される（詳細は 4.6.1 項で述べる）。

4. SIGMA による航空写真の解析

ここでは、SIGMA のプロトタイプシステムによる航空写真の解析について、知識の表現および具体的な解析プロセスを実験結果とともに述べる。このプロトタイプシステムは、VAX 780 上でインプリメントされており、言語としては FLAVOR, LISP, C を用いた。また、このシステムでは MSE, LLVE は簡単な LISP や C のプログラムとして実現されており、あまり柔軟な解析はできない。

4.1 知識の表現

対象世界に関する知識は、フレームを用いて表現される。図 5 はプロトタイプシステムで用いた知識を表したもので、各対象物（のクラス）がそれぞれ一つのフレームによって表現される。フレーム間の関係としては、

- 1) PW: 部分全体関係
- 2) SP: 対象物間の幾何学的関係
- 3) AKO: 対象物のクラスの上位下位関係
- 4) IO: あるクラスに属する対象物のインスタンス

```

FRAME_NAME : Road
SLOT_NAME  : PROPERTY : Total_length
                          Average_direction
                          Average_width
      SP    : Left_connecting_road_terminator
                          Right_connecting_road_terminator
                          Left_neighboring_house_group
                          Right_neighboring_house_group
      PW    : Left_adjacent_road_piece
                          Right_adjacent_road_piece
                          Component_road-piece

FRAME_NAME : Road_piece
SLOT_NAME  : PROPERTY : Length
                          Width
                          Direction
                          Local_coordinate_system
      PW    : Road
  
```

図 6 フレームによる知識表現

Fig. 6 Knowledge representation by frame.

を表す

- 5) ICW: 対象世界において同じ位置に存在できない対象物間の関係を表す

がある。

図 6 は、Road と Road-Piece フレームの記述を表し、それらがもつ属性や関係を示すスロットから構成されている。長さや幅、方向といった属性はすべて対象世界における単位や座標系を基準としており、画像データとは直接関係ない。また、Left-adjacent-road-piece, Right-neighboring-house-group などの幾何学的関係を表すスロットには、いわゆるプロダクションルールが書かれており、指定された条件が満たされると 3 章で述べた f や g に対応する関数が起動され仮定が生成される。3.3 節で述べた PW 関係による Δ 構造に基づいたボトムアップ、トップダウン解析における上位インスタンス、下位の対象物に対する仮定の生成過程で現れた各種の条件は、こうしたプロダクションルールの条件部として表される。

4.2 初期セグメンテーション

解析の当初は対象物のインスタンスがまったくなく、純粋なボトムアップ解析が MSE によって起動される。MSE は、図 5 に示した知識を調べ、各対象物の最も簡単な見え方に対応する画像特徴の抽出を LLVE に要求する。この際、MSE は、フレームのスロットに書かれているデフォルトを利用し目的とする画像特徴の属性に対する制約条件を決定する。図 5 の場合、コンパクトな長方形と細長い長方形が MSE, LLVE によって抽出され、それぞれ家、道路セグメントのインスタンスとして Iconic Database に登録される。図 7 は実験に用いた航空写真の一つで、大きさは 250×140 、濃度は 6 ビットの白黒の画像である。図 8 に初期の解析で得られた家と道路セグメントのインスタンスを示す（白枠で囲まれたもの）。

4.3 GRE における認識サイクル

GRE は以下に示す認識過程を新たな変化が生じなくなるまで反復することにより、画像の解釈を進める。

- (1) 新たに生成された対象物の各インスタンスが、その対象物のクラスに対応するフレームに蓄えられている知識に基づき、それと関係のある他の対象物に対する仮定を生成する。

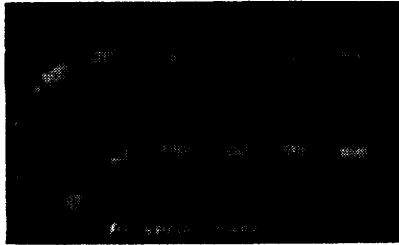


図 7 航空写真

Fig. 7 An aerial photograph.

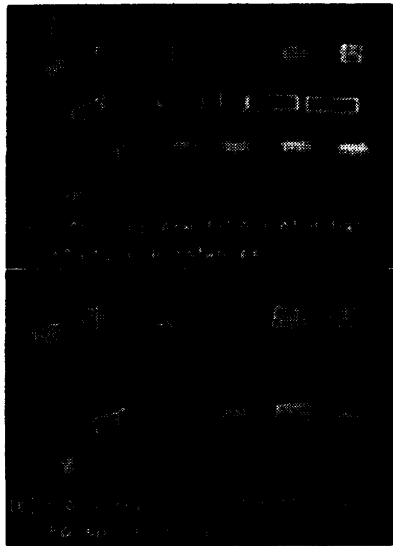


図 8 初期インスタンス

Fig. 8 Initial object instances.

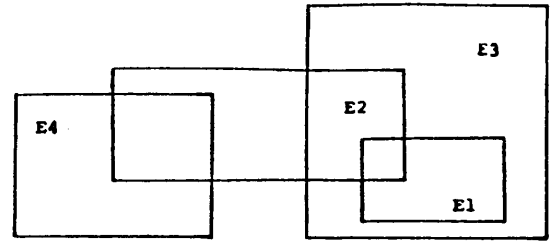
(2) すべての部分情報 (インスタンスと仮定) は Iconic Database に蓄えられる。

(3) 相互に整合性のある部分情報が蓄積され、状況を形成する。

(4) 注意の集中：一般にある時点での状況は一つとは限らず、多くの状況が同時に存在する。GRE は、信頼度の最も高い状況の一つを選び、その解析に焦点を絞る。

(5) 選ばれた状況の解を求める。その結果、ボトムアップ的なインスタンス間の関係の設定、あるいはトップダウン的なインスタンスの検出が行われる。

後で述べる例からもわかるように、SIGMA ではすべての可能な解釈を最後まで保持する。これは、解析途中ではまだ発見されていない情報が存在する可能性がたえずあり、互いに矛盾する解釈のうち、いずれが正しいのかを決定する判断をできるだけ避けているためである。こうした考え方はいわゆる least commitment principle と呼ばれる原理¹¹⁾に基づいている。



Overlap

Constituent Evidence

OP1	E1
OP2	E2
OP3	E3
OP4	E4
OP5	E1, E2
OP6	E1, E3
OP7	E2, E3
OP8	E2, E4
OP9	E1, E2, E3

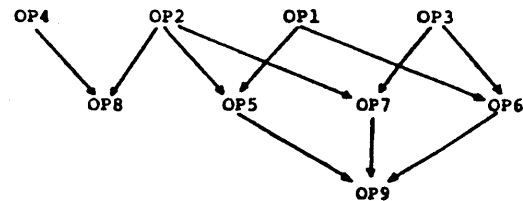


図 9 部分情報の重なりを表すラティス構造

Fig. 9 Lattice structure to represent overlaps among regions.

したがって、上の認識サイクルが終了した後、GRE は得られた多くの解釈のうち最も信頼性のあるものを選び、それを最終結果とする。

4.4 部分情報の蓄積 (状況の形成)

解析過程の任意の時点において、Iconic Database 中には多数の部分情報が蓄えられており、GRE は次の四つの観点から部分情報間の整合性を調べる。

(1) 領域間の重なり：先に述べたように、対象物のインスタンスや仮定の位置的制約条件は、Iconic Database 中の 2 次元配列上の領域として表現される。これらの領域の重なりは図 9 に示すようなラティス構造によって表され、GRE は互いに重なった領域に対しそれらが表す部分情報の間の整合性を調べる。

(2) 対象物のクラス：OP を二つの部分情報 (領域) E1, E2 の重なり; OBJ1, OBJ2 を E1, E2 の表す対象物のクラスとする。このとき、OBJ1 と OBJ2 とが ICW の関係で結ばれていると (図 5)、OP は不整合とみなされ、ラティス上で OP および OP の

下位にあるすべての重なりが除去される。さらに、このとき E1 と E2 がともにインスタンスであると、GRE はそれらが互いに矛盾する解釈であるとみなし、以後それぞれの解釈に基づく解析を独立に行う（下記の(4)参照）。

次に GRE はラティス上に残された領域の重なりを調べ、次のいずれかの条件を満たすもののみを整合性があるとみなす。

- a) OBJ1 と OBJ2 が同じ対象物のクラスである。
- b) OBJ1 と OBJ2 の間に AKO, PW による上位下位関係が存在する。

(3) 制約条件の整合性：SIGMA における制約条件には、対象物のクラスの定義に付属したもの、仮定に付けられたもの、および対象物のインスタンスに付けられたものがある。これらの制約条件は、すべて属性を表す変数に対する不等式として表現される⁹⁾。

上の例で E1, E2 に付属する制約条件の集合を C1, C2, それらが属す対象物のクラス自身に付けられた制約条件を C とする。このとき GRE は、C1, C2, C の共通集合が空でなければ（すべての不等式を満たす解が存在すれば）、E1, E2 の間に整合性があるとみなす。ここで C を用いる理由は、E1, E2 が他のクラスの対象物のインスタンスによって生成された仮定である場合、E1, E2 がその対象物のクラスが満たすべき制約条件 C を必ずしも満たすとは限らず、C と整合が取れるかどうかを調べる必要があるからである。

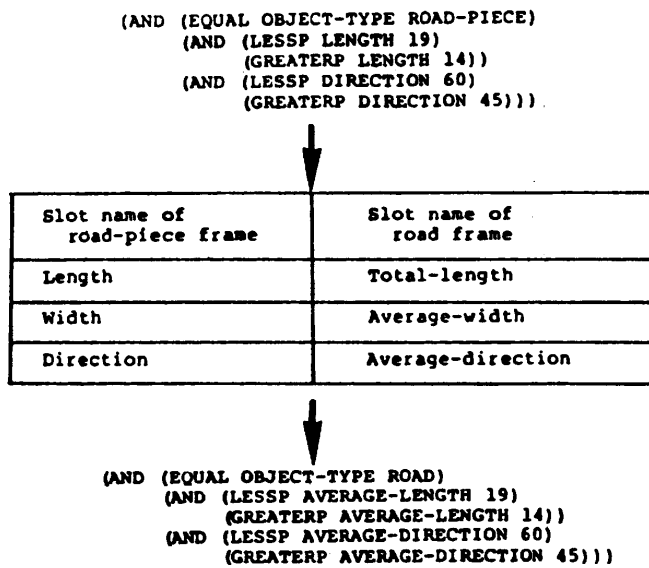


図 10 属性の変換
Fig. 10 Attribute translation.

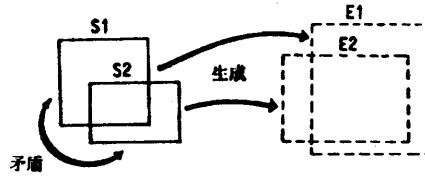


図 11 矛盾する解釈から生成された部分情報の分離
Fig. 11 Hypotheses generated by conflicting evidence.

上の b) で述べたように、OBJ1, OBJ2 が同じ対象物のクラスでなくても E1, E2 の間に整合性があることがある。この場合、E1 と E2 に付けられた制約条件はそれぞれクラス OBJ1, OBJ2 の属性に対するもので、不等式に現れる変数の名前が異なる。したがって、それらの間の共通集合が直接求められない。こうした問題に対処するために各 PW (AKO も) 関係には一方のクラスの対象物のもつ属性を他方のクラスの対象物の属性に翻訳するための規則が蓄えられている。現在のシステムでは図 10 に示すような簡単な属性変換テーブルを用いて、下位のクラスの属性を上位のクラスの属性に置き換えている。

(4) 矛盾する解釈：(2) で述べたように ICW で結ばれた二つの対象物のクラスのインスタンスが同じ位置に存在すると、それらは矛盾する解釈とみなされる。E1, E2 を生成した元の部分情報（仮定の場合はそれを作ったインスタンス、インスタンスのときはそれ自身）を S1, S2 とする。このとき、S1 と S2 が互いに矛盾する解釈に属している場合（図 11）、E1, E2 自身の間にも整合性があっても、これらは蓄積されずその重なり OP は状況とはみなされない。

以上述べた四つの観点からの整合性の検証を満たしたもののみが状況として登録される。GRE は領域間の重なりを表すラティス構造を利用し、まず二つの領域の間の整合性をすべて調べ、その結果を伝播させることにより、多くの領域の重なりに対する整合性を検証する。

4.5 注意の集中

4.4 節で述べたプロセスによって互いに整合性のある部分情報から状況が形成される。各部分情報にはその信頼度を表す数値が付けられており、GRE はそうした信頼度を用いて各状況の信頼度を計算する。プロトタイプシステムではすべての部分情報に同じ信頼度を与えている。GRE は最も信頼度の高い状況の一つを選び、その解を求める。

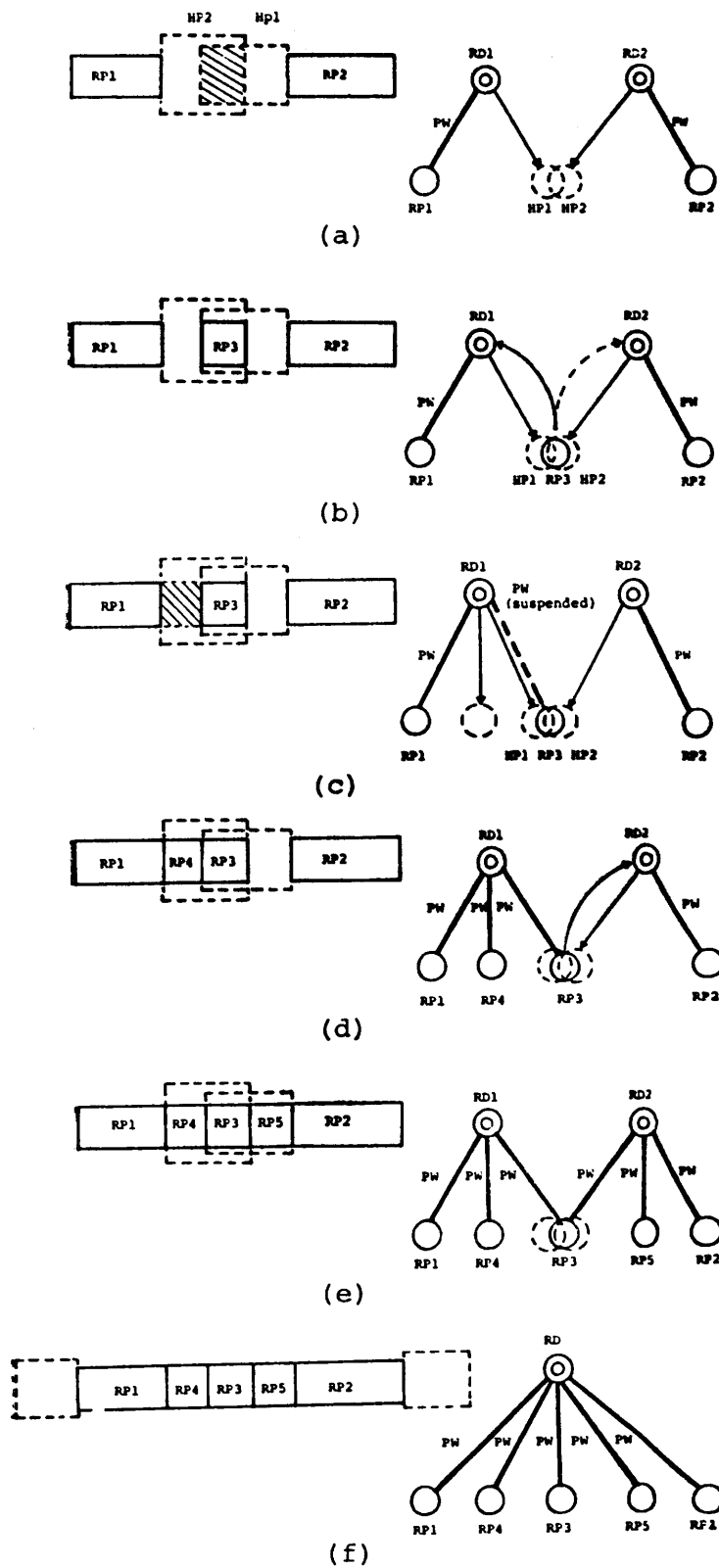


図 12 PW 関係による部分木の統合
 Fig. 12 Unification of partially instantiated PW trees.

4.4 節で述べた部分情報の整合性検証過程からわかるように、SIGMA では否定に関する情報は取り扱っていない。そのため、整合性のある部分情報は互いに強め合い、さらにそれらと整合性のある新しい部分情報をトップダウン解析を通じて作り出す。その結果、一種の positive feedback がかかり、初めある程度の量の部分情報の間に整合性があると、それを種として急速に整合のとれた解釈が広げられる。しかし、一方では解析の途中で生じた誤った解釈は否定されることなく最後まで残されることになる。さらに、医療診断などでは、ある情報によって特定の病気の可能性が否定されることがあり、負の情報が積極的な意味をもつ。こうした正負 2 種類の情報をうまく利用することは今後の課題の一つで、そのための方法としては Dempster & Shafer 理論¹²⁾の導入が考えられる。

4.6 状況の解析

4.6.1 PW 関係による部分木の統合

ここでは、図 12 を用いて GRE による状況の解析（部分全体関係に基づく推論）過程を説明する。

いま、図 12(a) に示した状況が選ばれたとする。これは、道路セグメント（図 5 では road piece）に対する二つの仮定 HP1, HP2 から構成されており、それらは道路のインスタンス RD1, RD2 によって生成されたものである。また、道路のインスタンス RD1, RD2 は、それぞれ道路セグメントのインスタンス RP1, RP2 と PW 関係で結ばれている。この状況は、仮定のみから構成されているので、GRE はトップダウン解析を起動し、MSE, LIVE に仮定された道路セグメントの検出を要求する。

このトップダウン解析によって新たな道路セグメントのインスタンス RP3 が求めたとする（図 12(b)）。GRE

は、この結果を仮定を生成した元のインスタンス RD1 と RD2 に伝える。

いま RD1 が初めに自分が生成した仮定に対する結果を得たとする。RD1 は RP3 が必要なすべての制約条件を満足するかどうかを調べる。RP3 の属性に関する制約条件はすでに仮定に付けられた制約条件として調べられており、ここでは RD1 と RP3 との間のより詳細な関係が調べられる。具体的には、RD1 は RP3 との接続関係を検査する。

上の例において、RP3 は RP1 に連結しておらず、RP3 は RD1 の期待する条件を満たさない。その結果、RD1 は RP3 との間の PW 関係の設定を保留し、エラー処理ルーチンを起動する。このルーチンもまた、道路フレームの中に蓄えられており、条件チェックの失敗によって起動がかけられる。現在のシステムでは、このエラー処理ルーチンが再び RP1 と RP3 の間の道路セグメントを探すようトップダウン要求を出す (図 12(c))。これによって RP4 が抽出されたとする、それが RD1 に返される。この場合は、接続性に関する条件が満たされ、RP4 が RD1 と PW 関係で結ばれ、その後 RP3 も RD1 の部分として PW 関係で結ばれる (図 12(d))。

同様の解析が RD2 側に対して行われ、RP5 に続き RP3 が RD2 と PW 関係で結ばれる (図 12(e))。このとき、RP3 は二つの異なる道路インスタンス RD1, RD2 を親としてもつことになり、“エラー”が検出される。このエラーは GRE に報告され、GRE は RD1 と RD2 との整合性を調べ、整合性があればそれらを一つに統合する。その結果、二つの木構造が一つにまとめられる (図 12(f))。その後、RD1, RD2 によって生成されたすべての仮定が消去され、RD1, RD2 の統合によってできた道路インスタンスによって新たな仮定が生成される。

図 13 は家のトップダウン解析による検出、図 14 は

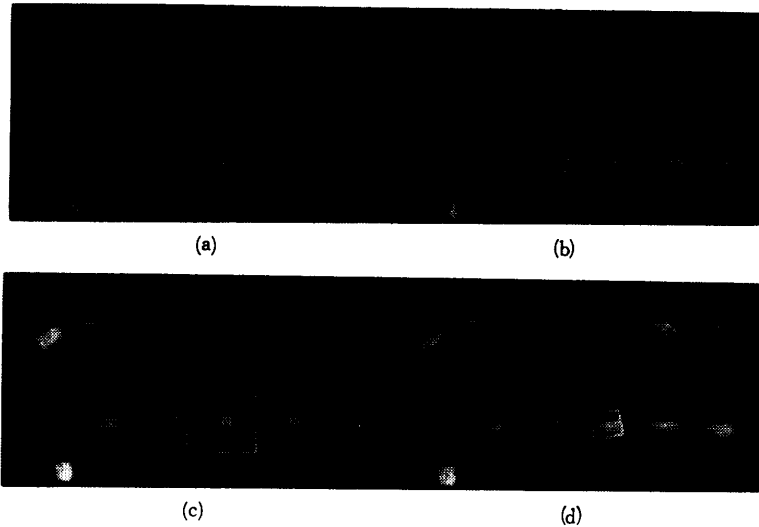


図 13 家の検出

- (a) 選ばれた situation, (b) 仮定を生成した元のインスタンス,
(c) 統合された仮定, (d) 検出された家のインスタンス

Fig. 13 Top-down detection of a house.

- (a) Selected situation, (b) Source instances,
(c) Composite hypothesis, (d) Proposed solution.



図 14 家並のインスタンスの統合

- (a) 解析前の家並のインスタンス, (b) 解析後の家並のインスタンス

Fig. 14 Unification of house group instances.

- (a) House group instances before the analysis,
(b) House group instances after the analysis.

その結果生じた家並 (図 5 では house group) のインスタンスの統合過程を表している。

4.6.2 エラーの解析

上で述べた状況の解析過程の途中には様々な失敗が生じる可能性がある。まず MSE が道路セグメントの検出を行う際、その道路セグメントは visible road (図 5) と仮定して LLVE に抽出を要求する。それが失敗した場合、MSE は道路セグメントの別の見え方の可能性を調べる。図 5 に示した知識では道路セグメント (road piece) には、visible road のほかに occluded road が AKO 関係で結ばれており、それからさらに overpass と shadowed road がつながっている。そこで MSE はこれらの可能性を逐次試みる。

MSE によるすべての解析が失敗したときは、それ

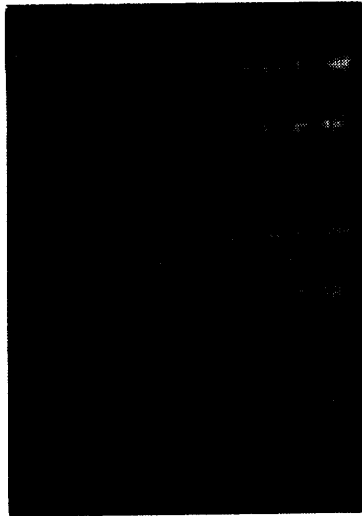


図 15 仮定の切り換え

下: 道路インスタンス, 中: 道路セグメントの仮定,
上: 交差点の仮定

Fig. 15 Changing hypothesis.

bottom: road instance, middle: road segment,
hypothesis, top: road terminator hypothesis.

が GRE に報告され、GRE はいま注目している状況に含まれるすべての仮定を生成した元のインスタンスに失敗を知らせ、その状況の解析を終える。一方報告を受けたインスタンスはそれが生成した仮定が誤っていたと考え、別の可能性を試すことができる。たとえば図 5 において、通常、道路は道路セグメントの連続として定義されており、道路セグメントに対する仮定を生成する。しかし、画像の端や交差点では道路セグメントが見つからずトップダウン解析が失敗する。こうした失敗が GRE から報告されると、道路のインスタンスは SP 関係による仮定の生成を起動し、道路端 (road terminator) の解析に進む。すなわち、道路フレームにおいて、道路端との SP 関係を表すスロットに蓄えられる仮定生成用のプロダクションルールが、

IF (道路セグメントの検出が失敗) THEN……

となっている。図 15 は、道路セグメントから道路端への仮定の切り換えの例である。

図 12 の例では、道路のインスタンス RD1, RD2 が直接トップダウン解析を要求した (図 12(c))。この場合、RD1 による RP4 の検出が失敗すると、RD1 は GRE にそれを知らせるが、現在の GRE ではそれに対しては特別な処理は行わず、RD2 側の解析を続ける。

4.6.3 解析の中断・保留

あるインスタンスが PW 関係を通じて同時に二つ

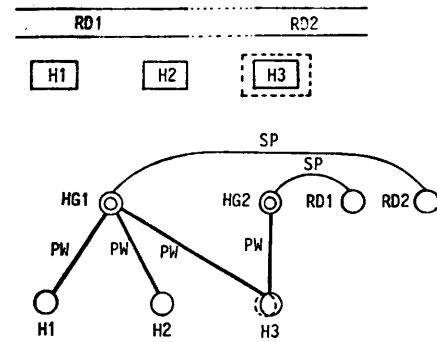


図 16 解析の中断, 保留

Fig. 16 Suspending unification process.

以上の上位のインスタンスに関係付けられた場合 (図 12(e)), そのインスタンスは GRE に上位のインスタンスの間の整合性を調べるように要求する。このとき GRE はそれぞれの上位インスタンスの属性を調べ、それらの値に整合性があれば一つのインスタンスに統合する。こうした PW 関係による部分木の統合には難しい問題がある。

いま、家並が“共通の道路の同じ側に規則的に並んだ家の集合”として定義されるとし、図 16 に示す例を考える。この場合、家のインスタンス H3 が二つの家並のインスタンス HG1 と HG2 の両方に PW 関係で結ばれ、H3 が GRE にこの両者の整合性の検証を要求している。GRE は HG1 と HG2 の属性を調べるが、その際、SP 関係で結ばれる家並に面する道路インスタンスが、HG1 では RD2, HG2 では RD1 となっており、その値が異なっている。

このような場合 (HG1, HG2 の一方あるいは両方が、それに面する道路インスタンスをまだ検出していない場合も含む)、GRE は HG1, HG2 の統合過程を凍結する。これは図 16 において、RD1 と RD2 とは異なるインスタンスとなっているが、後の解析において両者が一つのインスタンスに統合される可能性があり、この図の示す時点においてただちに判断を下すことができないことによる。

GRE はすべての凍結された統合過程を記憶しておき、新たな状況が解析されるごとにそれらを調べ、凍結を解除できるものがあればその統合過程を再開する。

こうした複雑な処理が必要な理由は、GRE における推論が本質的に並列になっているためである。すなわち、図 12, 16 の状況は、互いに独立な (並列な) 推論過程によって作られた部分的な解釈 (RD1, RD2

や HG1, HG2) が共通する局所的な解釈 (RP3 や H3) を通じて一つに統合されることを表している。こうした推論方式は本システムの大きな特徴であるが、並列推論という難しい問題を含んでいる。

すべての解析 (4.3 の認識サイクル) が終了した後、GRE は凍結されているそれぞれの統合過程において、二つの上位インスタンスに共有されている部分インスタンス (図 16 では H3) のコピーを作り、二つの PW 関係による木構造を分離する。現在のシステムでは分離された木構造をそれぞれ可能な解釈と考え、矛盾とはみなさない。このためある一つの対象物のインスタンス (H3) が異なった複数の上位対象物のインスタンスに含まれることになる。

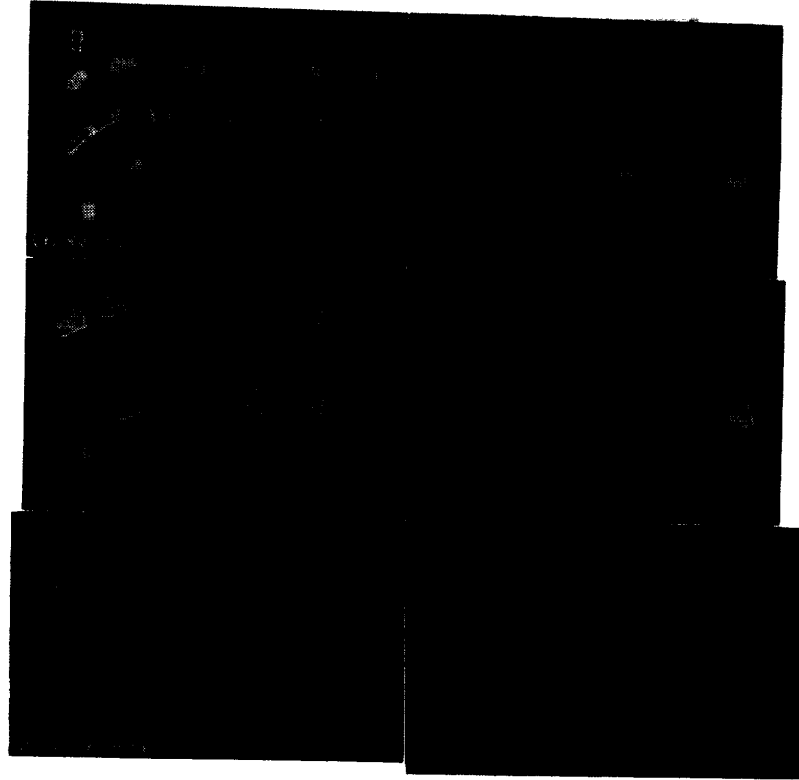


図 17 最終結果

Fig. 17 Final result.

5. おわりに

図 17 に図 7 の航空写真の最終的な解析結果を示す。ここではすべての解釈をそのまま出力しており、三つの誤った道路インスタンスがある。これらはすべて初期セグメンテーションにおけるエラーのためである。しかし、図より明らかなように、これらは正しい解釈の家のインスタンスと重なっていたり孤立した解釈となっていたりし、容易に誤りであることがわかる。

以上述べたように、SIGMA では部分情報の蓄積・統合によってボトムアップ、トップダウンの二つの解析過程が一つの推論過程に統合され、柔軟な制御方式が実現される。また、トップダウン解析では、目的とする対象物の画像上での見え方を推論し、それに対応する画像特徴を画像処理に関する知識を利用して抽出する。こうした各種の機能によって、多様な対象物からなる複雑な情景を柔軟かつ効率的に解析できる。

SIGMA はまだ完成されたシステムではなく、幾つかの問題が残されている。まず、プロトタイプシステムは、GRE を中心にインプリメントされ、MSE, LLVE は GRE から起動できる簡単な LISP 関数として実現されているにすぎない。LLVE は現在京都大学で開発中であるが、MSE に関してはその推論方式

の詳細な検討が必要である。とくに、現在のところは応用として航空写真の解析を考えているため対象世界を 2次元にしているが、3次元世界の情景を解析する場合には、MSE が 3次元情景と 2次元画像との間の変換を受け持つことになり、かなりの推論機能が要求される。

GRE における問題の一つは、部分情報や推論過程の依存関係の取り扱いにある。すなわち、図 16 で示した例では、ある関係の設定が他の関係の存在に依存して決められる。また、互いに矛盾する解釈間の関係やそれらに基づいた推論方式、負の部分情報の取扱いなどについてもより深い考察が必要である。さらに、GRE の推論過程に含まれる並列性の解析とその意味付けも今後の課題である。より柔軟な推論を実現するためには、こうした各種の複雑な依存関係の統一的な記述、操作法の開発が不可欠である。

謝辞 最後に、この研究を行うにあたり有意義な示唆と建設的な議論をしていただいた米国メリーランド大学の Rosenfeld 教授、Davis 助教授、また米国での研究の機会を与えていただいた京都大学の長尾教授に心から感謝いたします。

参 考 文 献

- 1) Binford, T.O.: Survey of Model-Based Image Analysis Systems, *The International Journal of Robotics Research*, Vol. 1, No. 1, pp. 18-64 (1982).
- 2) Brady, M.: Computational Approaches to Image Understanding, *ACM Comput. Surv.*, Vol. 14, No. 1, pp. 3-71 (1982).
- 3) Barnard, S. T. and Fischler, M. A.: Computational Stereo, *ACM Comput. Surv.*, Vol. 14, No. 4, pp. 553-572 (1982).
- 4) 情報処理, コンピュータビジョン特集, Vol. 24, No. 12 (1983).
- 5) Tsotsos, J. K.: Temporal Event Recognition: an Application to Left Ventricular Performance, Proc. of 7th IJCAI, pp. 900-907 (1983).
- 6) Havens, W. and Mackworth, A.: Representing Knowledge of the Visual World, *IEEE Comput.*, Vol. 16, No. 10, pp. 90-96 (1983).
- 7) Russell, D.M.: Where Do I Look Now?: Modeling and Inferring Object Location by Constraints, Proc. of Pattern Recognition and Image Processing, pp. 175-183 (1979).
- 8) Selfridge, P.G.: Reasoning about Success and Failure in Aerial Image Understanding, Ph. D. Thesis, TR 103, Univ. of Rochester (1982).
- 9) Brooks, R. A.: Symbolic Reasoning Among 3D Models and 2D Images, *Artif. Intell.*, Vol. 17, pp. 285-348 (1981).
- 10) 松山, Hwang, V.: 画像理解におけるトップダウン, ボトムアップ解析の統合, 信学会, パターン認識と学習研資 (1984. 12).
- 11) Marr, D.: Early Processing of Visual Information, A. I. Memo 340, MIT (1975).
- 12) Barnett, J. A.: Computational Methods for a Mathematical Theory of Evidence, Proc. of 7th IJCAI, pp. 868-875 (1981).
- 13) 松山, 尾崎: トップダウン・セグメンテーションのための画像処理エキスパートシステムの開発, 情報処理学会研資, コンピュータビジョン, 36-3 (1985).

(昭和 59 年 11 月 30 日受付)

(昭和 60 年 2 月 21 日採録)