SIGMA : A FRAMEWORK FOR IMAGE UNDERSTANDING INTEGRATION OF BOTTOM-UP AND TOP-DOWN ANALYSES

Takashi MATSUYAMA
Department of Electrical Engineering
Kyoto University
Sakyo, Kyoto 606 JAPAN

Vincent HWANG
Center for Automation Research
University of Maryland
College Park, Maryland 20742 U.S.A.

ABSTRACT

The framework and control structure of an image understanding system SIGMA arc presented. SIGMA consists of three experts: Geometric Reasoning Expert (GRE) for spatial reasoning. Model Selection Expert (MSE) for appearance model selection, and Low Level Vision Expert (LIVE) for knowledge—based picture processing. Tills paper mainly describes the control mechanism for the spatial reasoning by GRE, where bottom —up and top—down analyses are integrated into a unified reasoning process.

1. INTRODUCTION

Many experimental image understanding systems have been developed to test the feasibility of image understanding! 1—7]. The followings are some problems in building an image understanding system which have not yet been treated successfully.

(1)Segmentation

There are many methods of segmenting an image to extract, objects. Each method has its advantage and disadvantage. How to select and/or combine appropriate methods is a basic problem in image understanding.

(2)Diversity in Appearance

2D appearances of a 3D object vary greatly depending on viewing angles. On the other hand, an object has many diverse appearances. For example, houses in a suburban area have many possible shapes, sizes, and colors. How to limit the number of possible appearances and intelligently select the ones to try (search) is another problem.

(3)Representation and utilization of Domain Knowledge

An image understanding system needs to have domain knowledge to construct an interpretation of the image. Usually, the sources of knowledge are diverse and redundunt. Requirements that must be satisfied by an object are specified in many different ways, and each of them gives only a weak constraint. Knowing that only some of the constraints for an object are satisfied is not enough to assign the object label to an image feature(e.g. region). On the other hand, failure to satisfy some of the constraints does not indicate that the image feature cannot be an object. How to organize and use domain knowledge is another problem.

In this paper, we describe the framework and control structure of an image understanding system SIGMA. The followings are the basic ideas incorporated into SIGMA to solve the above three problems.

(1)Knowledge—Based Segmentation

It is advantageous to use a knowledge—based segmentation system to process an image[8][9][10]. Many studies have been done on picture processing operators. Their characteristics have been well studied, such as effectiveness in extracting given types of image features(e.g. region, line) in a given environment, required cost of computation, and possible artifacts caused by the operators. A knowledge-based segmentation system uses such knowledge about the operators to realize efficient and reliable extraction of required image features.

(2)Evidence Accumulation for Spatial Reasoning

An image understanding system builds interpretations by establishing relations among objects, and searching for missing objects by analyzing the image. Objects found (object instances) can be used to predict missing objects by generating hypotheses. Hypotheses from various sources can be combined to guide the searching process[5, 6]. Such accumulation of evidence from different sources decreases the total amount of effort spent in the search and increases the reliability of the analysis.

(3)Appearance Model Selection Based on Contextual Information

As described above, an object can have many diverse appearances. When an image understanding system searches for a missing object, it should use contextual information to predict the most likely appearance(s) of the object. Given an appropriate description of the context where the target object is embedded, its possible appearance(s) in the image can be predicted.

2. OVERVIEW OF THE SYSTEM

Fig. 1 shows the organization of the entire system. It consists of the following three experts.

(1)Geometric Reasoning Expert (GRE)

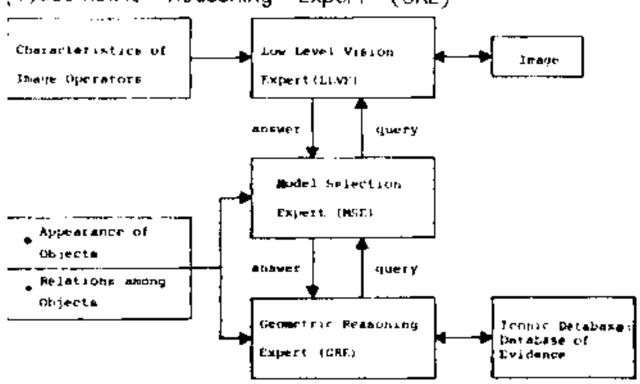


Fig. 1 Overview of the system

This export is the central reasoning module in the *system,* and utilizes a symbolic hierarchical model for the possible spatial organization of objects in the world. The geometric reasoning performed by this expert (evidence accumulation) in teg rates both bottom-up and top-down analysis processes into a unified reasoning process. All of the partial evidence obtained during the interpretation are stored in a common database (Iconic Database in Fig. 1), where consistent pieces of evidence are accumulated. GRE first establishes local environments (contexts) using the accumulated evidence. Then, either the bottom-up analysis to establish a relation between objects or the top—down analysis to find a new object are activated depending on the nature of a focused local environment. In the top-down analysis, GRE first reasons about its goal (the target object to be delected) and where to analyze the image. Then it asks Model Selection Export to perform the analysis.

(2)Model Selection Expert (MSE)

This expert reasons about the most promising appearance models to use in searching for the object in the image. This model selection is performed based on the contextual information provided by GRE. Knowledge about objects is represented at several levels of specificity. For example, an object class "house" is a generalization of many specifically shaped types of houses. GRE determines the general class of objects to search for (e.g. "house") while MSE determines which specialization (e.g. rectangular house) should be looked for. In addition to this reasoning, MSE performs geometric transformation from the scene domain to the image domain.

(3)Low Level Vision Expert (LLVE)

Tho appearance model determined by MSE is given to this expert. LLVE performs picture processing to extract the image feature corresponding to the specified appearance model. It selects appropriate picture processing operators and determines efficient and effective process sequences based on the knowledge about picture processing methodsias for details of LLVE, see[10j).

3. EVIDENCE ACCUMULATION FOR SPATIAL REASONING

It is widely accepted that image understanding systems should incorporate both bottom—up and top-down analyses. The use of geometric relations, however, is very different in the two analysis processes: consistency verification in bottom—up analysis end hypothesis generation in top-down analysis. An important characteristic of our evidence accumulation method is that it enables the system to integrate both bottom—up and top—down processes into a single flexible spatial reasoning process.

3.1 Principle of Evidence Accumulation

The spatial reasoning using the evidence accumulation method is perfomed by the Geometric Reasoning Expert. Its principle is as follows.

Let REL(01,02) denote a binary geometric relation between two classes of objects, 01 and 02. This relation can be represented using two functional expressions:

01 - f(02) and 02 = g(0l).

Given an instance of 02, soy s, function f maps it into a description of an instance of 01, f{s), which satisfies the geometric relation, REL, with s. The analogous interpretation holds for the other function g.

In SIGMA, knowledge about a class of object is represented by a frame [9], and a slot in that frame is used to represent a function such as f or g. A slot contains a group of production rules, each of which consists of a precondition and an action. A precondition represents a set of conditions specifying when the function can be activated. An action represents a computational procedure corresponding to the function, which produces the description of the related object.

Whenever an instance of an object is created and the conditions are satisfied, the function is applied to the instance to create a "hypothesis" (expectation) for another object which would, if found, satisfy the geometric relation with the original instance (Fig. 2(a)). A hypothesis is associated with (Fig. 2(b))

(i) a prediction area (locational constraint) where the target object instance may be located, and

(ii) a set of constraints on the target object instance.

All pieces of evidence (hypotheses and object instances) are stored the iconic database(Fig. 1), where accumulation of evidence (i.e. recognition of consistent hypotheses and instances) is performed. This database contains an iconic data structure (i.e. two dimensional array for 2D scene analysis) to represent locational constraints associated with the stored pieces of evidence. They are represented as regions on this array. GRE uses overlaps among the regions to index mutually consistent pieces of evidence. Note that this array represents the world under analysis and its coordinate system is defined independently of that of the image. Resides this locational information, symbolic information such as relations among and properties of object instances is also stored in this database.

Suppose object instance s creates hypothesis f(s) (based on relation KEL) for object 01, which overlaps with an instance of 01, t (Fig. 3(a)). If the set of constraints associated with f(s) is satisfied by t, these two pieces of evidence are combined to form what we call a "situation". That is, a situation is defined by a set of mutually consistent pieces of evidence. GRE unifies f(s) and t, and establishes the relation REL from s to t as the result of resolving the situation. This is the bottom—up process to establish a geometric relation between a pair of object instances.

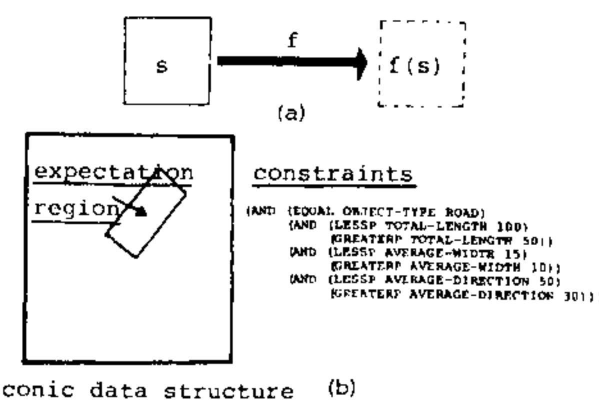


Fig. 2 (a)Hypothesis generation
(b)Description of a hypothesis

On the other hand, a situation may consist ot overlapping hypotheses alone(Fig. 3(b)). Then their unification leads *CUE* to search for an instance of the required object in the image. The expert asks MSE to detect the instance, which in turn activates LlVE. If the instance is detected, it is inserted into the iconic database, and the relations between the new instance and the "source" instances, which generated the hypotheses, are established. This process is the top-down analysis to find a missing object. Fig. 4 shows goal specifications to MSE and LLVE in the top-down analysis.

3.2 Handling Part-Whole Relations

Two types of geometric relations are used in our system: "spatial relation" (SP) and "part-whole relation" (PW). These two types of relations are used differently ill the system. PW relations specify hierarchies which represent objects with complex internal structures, while SP relations represent geometric relations between different classes of objects. While hypothesis generation by an SP relation is done as explained above, the use of PW relations is different.

Suppose the PW hierarchy illustrated in Fig, 6(a) is given. The system uses PW relations both to group parts into a whole and to predict missing parts. In general, the objects corresponding to leaf nodes in the hierarchy are instantiated first, because their appearances are simple and correspond directly to primitive image features. The presence of a higher level object instance is represented symbolically by an instantiated hierarchy. This implies that no iconic description (i.e. region} representing higher level object instances is stored in the iconic database. (Note that hypotheses for higher level objects have iconic representations and as a result, can interact with other pieces of evidence.)

Let a denote an instance of object class 01 (Fig-Ma)). Then, it can directly instantiate its parent object through the PW relation instead of generating a hypothesis as in case of SP relations(Fig. 5(b)).

This bottom-up instantiation through a PW hierarchy is controlled by a "kernel list" associated with each object class. An object instance in our system is in one of two instantiation states: fully-instantiated and partially—instantiated. The kernel list is used to discriminate these two states. The list consists of a set of sublists. Suppose object 0 is composed of part objects P1, P2,...,Pn. Each sublist in the kernel list of O specifies a subset of }P1, P2,...,Pn{. Object O is fully instantiated if all parts objects in at least one sublist are fully instantiated. Otherwise it is partially instantiated. Only a fully instantiated object can instantiate its parent object via a PW relation.

When a parent object is instantiated, it may then generate hypotheses for its missing part objects (Fig. 5(c)). That is, no instantiated part object generates hypotheses for the other part objects at the same level of the PW hierarchy.

PW relations are represented in the same way as SP relations. That is, a PW relation is represented by

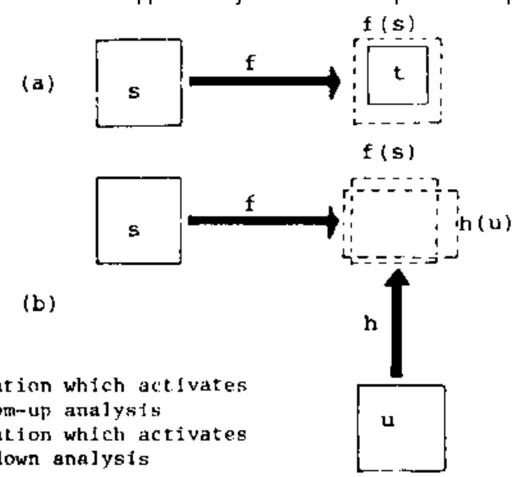


Fig. 3
(a)Situation which activates bottom-up analysis
(b)Situation which activates top-down analysis

```
((GOAL  (AND (EQUAL OBJECT-TYPE HOUSE)
             (AND (LESSP AREA 475)
                  (GREATERP AREA 250)))
  (LOCATION (AND (LESSP X 1000)
                 (GREATERP X 100)
                 (LESSP Y 7000)
                 (GREATERP Y 300)))
  (CONTEXT (HOUSE-GROUP002, HOUSE-GROUP004, ROAD0005)))
```

(a) Goal specification to MSE
(unit is (square) feet and (X,Y) is in the world coordinate)

```
((GOAL  (AND (EQUAL OBJECT-TYPE RECTANGLE-HOUSE)
             (AND (LESSP AREA 400)
                  (GREATERP AREA 200)))
  (LOCATION (AND (LESSP X 1000)
                 (GREATERP X 100)
                 (LESSP Y 2000)
                 (GREATERP Y 300))))
```

(b) Goal specification after selecting a specific object model

```
((GOAL  (EQUAL IMAGE-FEATURE-TYPE RECTANGLE)
        (AND (LESSP AREA 230)
             (GREATERP AREA 125)))
  (LOCATION (AND (EQUAL START-I 24)
                 (EQUAL START-J 40)
                 (EQUAL END-I 120)
                 (EQUAL END-J 230))))
```

(c) Goal specification to LLVE
(unit is number of pixels and (I,J) is in the picture coordinate)

Fig.4 Goal specification in a top-down analysis

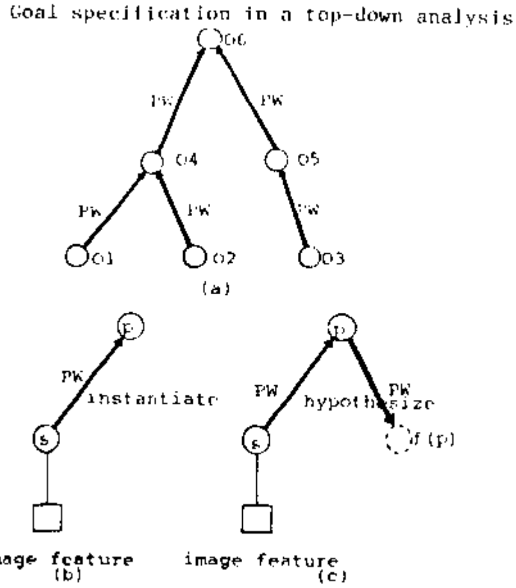


Fig. 5 (a)PW hierarchy (b)Bottom-up Instantiation (c)Top-down hypothesis generation a .slot in a frame, where a set of production rules are stored. Conditions for both the bottom-up instantiation and top—down hypothesis generation through a PW hierarchy are represented by preconditions of the production rules*. Computational procedures to generate parent instances and hypotheses for part objects are represented as actions of the rules. (The interpretation process to construct instantiated PW hierarchies (i.e. recognition of complex objects) will be described in Section 4.5.)

1. AERIAL IMAGE UNDERSTANDING BY A PROTOTYPE SYSTEM

This section describes the knowledge organization and analysis process of a prototype system for aerial image understanding. The system is implemented on VAX 780 and is written in FLAVOR. LISP, and C. The model selection and low level vision experts in this system are realized by simple functions written in LISP and *C.*

4.1 Knowledge Representation

Fig. 6 illustrates the knowledge organization used in the prototype system. As described above, an object class is represented by a frame, which consists of slots. Information stored in the slots includes attributes of the object and its relations to other objects. Besides these slots, a set of constraints among object attributes are stored in a frame to represent their allowable value ranges[7]. These constraints are basic- requirements to be satisfied in object recognition.

The relations used to associate frames are:

(1)PW represent geometric structures of objects with complex internal structures
(Z)SP :geometric relations between objects
(3)AKO specialization/generalization relations among objects
{4)10 represent instances of a class of object
(5)ICW :Some pairs of objects cannot occupy the same location in an image. For instance, a region cannot be interpreted as both house and road at the same time. Pairs of frames representing object classes which cannot occupy the same location are linked with an in-conflict-with (ICW) relation.
(6)APO :represent appearance of object

4.2 Initial Segmentation

The first analysis of an image is initiated by MSE. At the very beginning of the analysis, there is no object instance in the system. MSE examines the knowledge stored in the system and selects objects with simple appearances. Then, it asks LLVE to extract image features which match the selected appearances. The basic constraints on object attributes are associated with the goal specification to LIVE. All image features found by LLVE are returned to MSE. which then instantiates corresponding object instances and inserts them into the iconic database. These instances are seeds for reasoning by GRE.

Fig. 7(a) shows an aerial photograph (black and white) used in the experiment (250 X 140 and six bits for each pixel). Figs. 7(b)(c) illustrate the instances of house and road-piece extracted by the initial segmentation. Note that the segmentation in our system is dynamically performed on request and that no fixed set of image features (e.g. regions) to be interpreted are formed by the initial segmentation.

4.3 Interpretation Cycle of GRE

GRE iterates the following steps until no change is done in the iconic database.

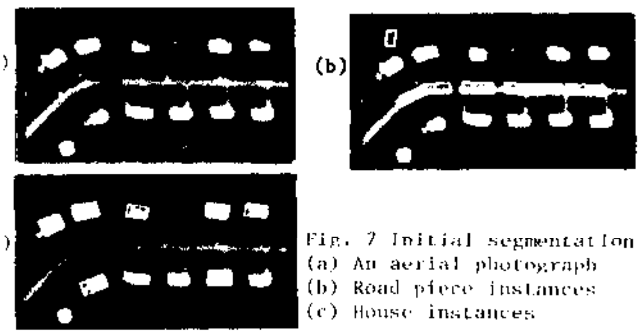


Fig. 7 Initial segmentation
(a) An aerial photograph
(b) Road piece instances
(c) House instances

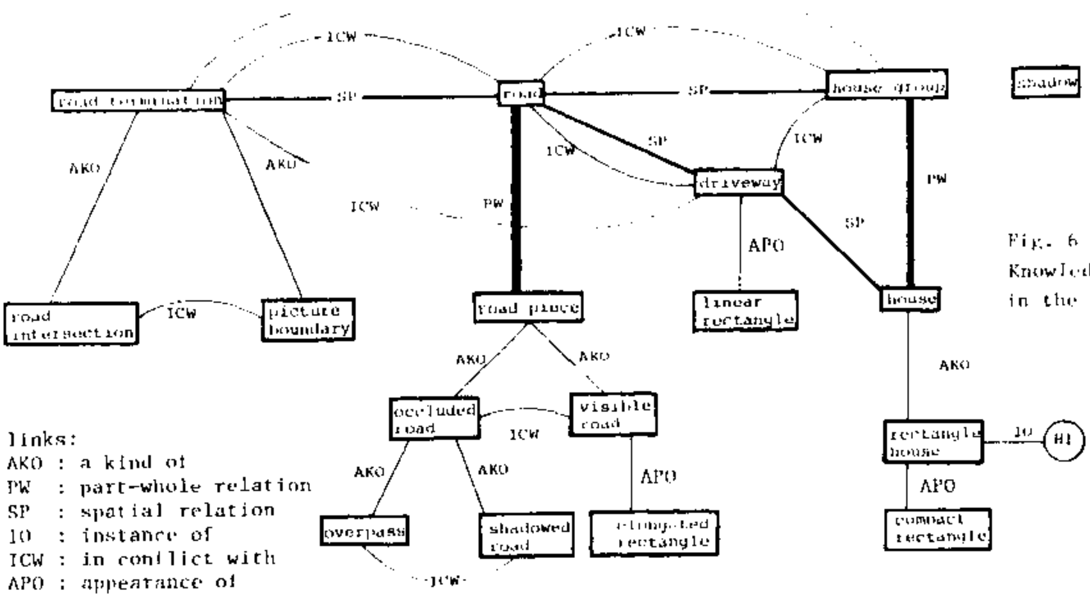


Fig. 6 Knowledge structure in the prototype system

(1) Each instance of an object generates hypotheses about related objects using functions stored in the object model. Object instances in PW hierarchies instantiate! their parent objects, which then generate hypotheses for missing parts.
(2) All pieces of evidence (both hypotheses and instances) are stored in the iconic database.
(3) Consistent pieces of evidence are combined to establish "situations",
(4) Focus of attention : since there are many situations, the most reliable situation is selected. Each piece of evidence has a reliability value, and the reliability of a situation is computed from those of its constituent pieces of evidence.
(C>) The selected situation is resolved, which results either in verification of predictions on the basis of previously detected/constructed object instances or in top-down image processing to detect missing objects.

The system has one additional post-processing: During the analysis by GRE, conflicting pieces of evidence may be generated. Comparing Figs. 7 (b) and (c), for example, two road-piece instances overlap with house instances. These interpretations are considered as conflicting. GRE maintains all possible interpretations throughout the analysis. The final interpretation process then selects the maximal consistent interpretation. At this stage, all partially instantiated objects and their parts are removed, because enough evidence to support their existence has not been obtained from the analysis.

4.4 Consistency Examination among Evidence

The consistency among pieces of evidence is examined based on:
(1) prediction areas of hypotheses and locations of instances
(2) object categories of evidence
(3) constraints imposed on properties of hypotheses and instances
(4) relations among sources of evidence.

•1.4.1 Intersection of Prediction Areas

Fig. 8(a) shows all intersections formed from four pieces of evidence El, E2, E3, and E4 in the iconic database. A partial ordering on intersections can be constructed on the basis of region containment. Intersection 0P1 is less than 0P2 if region 0P1 is contained in region 0P2. Fig. 8(b) shows the lattice representing the partial ordering among the intersections in Fig. 8(a). Each intersection consists of some set of hypotheses and instances. Situations are only formed among intersecting pieces of evidence (i.e. satisfying locational constraints). In other words, this lattice is an index to search for consistent pieces of evidence. To examine the consistency among pieces of evidence, if is sufficient to examine all intersections containing only a pair of pieces of evidence and then to propagate the results through the lattice.

4.4.2 Conflicting Evidence?

Let OP be the intersection arising from evidence {E1, E2} and let OBJ 1 and 0BJ2 denote the object categories of El and E2. respectively. If 0BJ1 and 0BJ2 are linked by an CW relation, then El and E2 are said to be conflicting, and OP is removed from the lattice. The removal of OP is propagated through the lattice, and any intersections contained in OP are also removed.

In the above case, if both E1 and E2 are instances, GRE records them as conflicting interpretations and performs independent analyses based on them. (See Section 4.4.4.)

4.4.3 Constraint Consistency

After eliminating all conflicting intersections from the lattice, the remaining intersections are checked to determine if their associated sets of constraints are consistent. Let E1 and E2 denote the non-conflicting evidence under consideration. One of the following conditions must hold:
(a) The object categories of E1 and E2 are the same,
(b) there is a path between the two categories consisting of PW relations,
or
(c) one of the two categories is a subcategory of the other, according to the AKO hierarchy.

As shown in Fig. 9, suppose a grandparent object in a PW hierarchy has been instantiated by an instance of a leaf node object s. Let p and q denote instances of the parent and grandparent objects. q as well as p generates hypotheses for its missing parts, say g(q). Suppose that g(q) itself has parts and one of them has already been instantiated. Let t denote that instance.

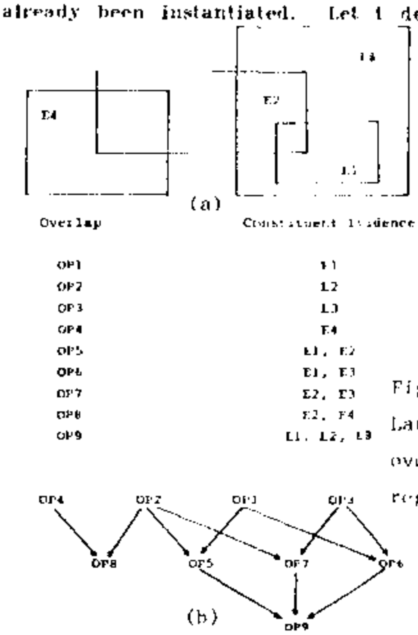


Fig. 8
Lattice to represent overlaps among regions

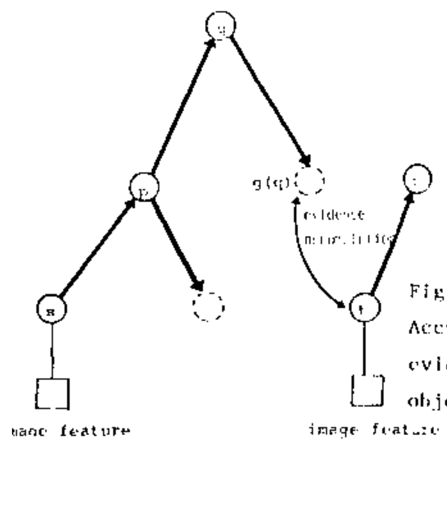


Fig. 9
Accumulation of evidence of different object classes

Then, if instances s and t are really parts of the same object, regions of g(q) and t will overlap with each other and will be consistent. (Note that instance u cannot intersect with g(q) directly since no iconic description is associated with u in the database.) In this case, although the object classes of g(q) and t are different, they can be consistent since their object categories are linked by a sequence of PW relations.

In such a case, since the names of the attributes used in the constraints associated with El (g(q) in Fig. ft) and E2 (t in Fig. 9) are different, they cannot, in general, be directly compared. In this ease, the constraints associated with the lower level object (i.e. t) ore translated into those for the higher level object (i e g(q))) by using PW relations. Currently, this translation is done simply by rewriting the attributes (slot names) of the lower level object into appropriate attributes of the higher level object using a "attribute translation tablc" for the PW relations *(Fig. 10)*). The similar attribute translation is used between object categories linked by AKO relations.

The properties of and/or constraints associated with both pieces of evidence must be consistent. Both constraints associated with a hypothesis and properties associated with an instance represented by a set of linear inequalities in one variable. A simple constraint manipulation system(7| is used to check the consistency between the sets of inequalities.

4.4.4 Relations between Sources of Evidence

Sources of accumulated evidence involved in a situation must not be conflicting. Let SI and S2 denote the source evidence of El and E2 respectively. If a piece of evidence is a hypothesis, its source evidence is the instance which generated the hypothesis. An instance is the source evidence for itself. It is possible that SI and S2 are mutually conflicting (belonging to conflicting interpretations), but that El and E2 themselves are consistent. *In* such a ease, we do not combine El and E2 into a situation; analysis based on such conflicting interpretations should be performed independently.

4.5 Resolving a Situation

As described in Section 3.1, one of two actions is taken in order to resolve a situation: confirm relation between instances or activate top-down analysis. After the action is taken, GRE provides a description of its proposed solution to the situation to all instances involved in that situation. Each instance then evaluates the proposed solution according to its specific expectations.

4.5.1 Resolution Process

In what follows, the process of resolving a situation is described by using the example shown in Fig. 11. Suppose GRE selected the overlapping region between two hypotheses generated from two road instances RD1 and RD2 (Fig. 11(a)). In the symbolic data structure, RD1 and RD2 are linked to their part road-piece) instances RP1 and RP2 by PW relations, respectively.

Since this situation consists only of hypotheses, the system activates top-down analysis to find a road-piece in the overlapping region. This request is issued to MSE together with the supporting evidence (i.e. RD1 and

```
(AND (EQUAL OBJECT-TYPE ROAD-PIECE)
     (AND (LESSP LENGTH 19)
          (GREATERP LENGTH 14))
     (AND (LESSP DIRECTION 60)
          (GREATERP DIRECTION 45)))
```

Slot name of road-piece frame	Slot name of road frame
Length	Total-length
Width	Average-width
Direction	Average-direction

```
(AND (EQUAL OBJECT-TYPE ROAD)
     (AND (LESSP AVERAGE-LENGTH 19)
          (GREATERP AVERAGE-LENGTH 14))
     (AND (LESSP AVERAGE-DIRECTION 60)
          (GREATERP AVERAGE-DIRECTION 45)))
```

Fig. 10 Attribute translation

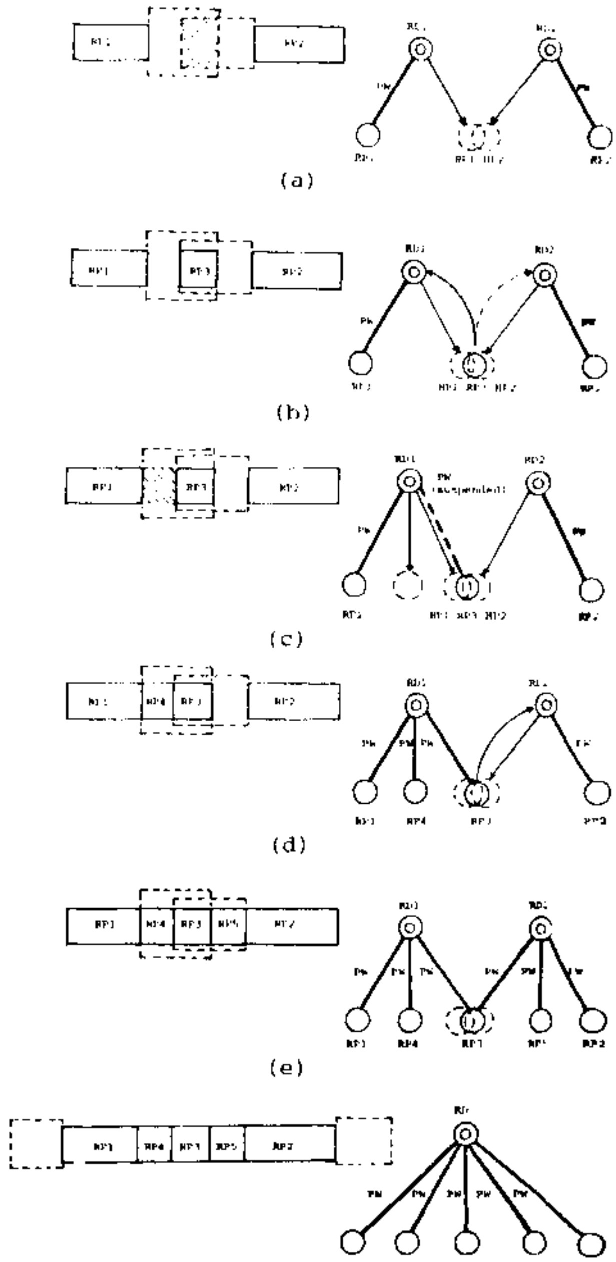


Fig. 11 (see text) (f)

RD2). so that the expert can use any available; contextual information of such supporting evidence.

ASSUME' that a new road-piece instance, RP3, is created (Fig. 1Kb)). Then, GRE provides this result to the instances involved in the situation. RD1 and RDL2

Suppose RD1 is the first to be informed of the proposed solution. RDI examines -whether or not RP3 satisfies all constraints required to establish the PW relation with itself. In this case, however, RP3 fails, because RP3 is not adjacent to RP1. (The constraints associated with the hypothesis do not include this type of relational count Taints.) This failure activate an exception handler, which is also a production rule stored in the corresponding slot in the road frame. Then it issues a top-down request to find a road-piece between RP1 and RP3 (see Fig. 11(c)).

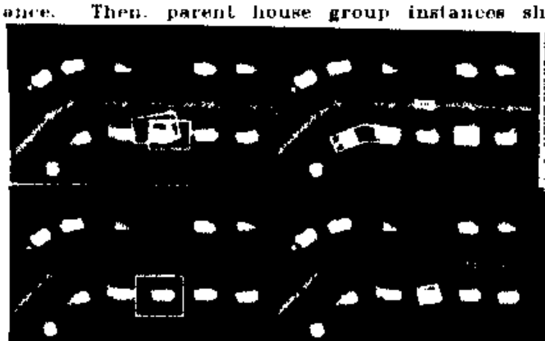
Assume that another new road-piece instance, RPl, is detected (Fig. 11(d)). Since RP4 is adjacent to RPl. RDI establishes a PW relation to RP4, and then to RP3.

Fig. 11(e) shows the interpretation after the same analysis is performed by RD2. In this case, however, when RD2 establishes a PW relation to RP3, an exception handler in RP3 is triggered, because RP3 has two different parents. More specifically, after RD2 establishes a PW relation to RP3, RD2 asks RP3 to ckeck its reverse relation from RP3. An exception handler is activated as a result of this checking process. This handler issues a request to GRE to examine the consistency between two parents. If they are consistent, GRE merges the two PW hierarchies below them into one (Pig. 11(f)).

The hypotheses generated by RDI and RDL' are removed from the iconic- database. The resultant new road instance in Fig. 1 1(f) generates new hypotheses for its adjacent road—pieces at the beginning of the next interpretation, cycle.

Figs. 12 and 13 illustrate an example of this process. First top-down analysis is performed to extract a house instance. Then. parent house group instances sharing a common house instance (i.e. a new detected house instance) are merged into one instance.

(a)



(b)

(c)

(d)

Fig. 12 Top-down detection of a house
(a)Selected situation (b)Source instances
(c)Composite hypothesis (d)Proposed solution

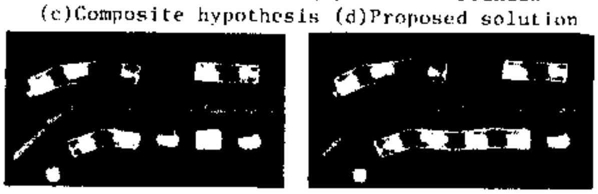


Fig. 13 Unification of house group instances
left:House group instances before the analysis
right:House group instances after the analysis

4.5.2 Error Analysis

There are several stages in the above example where the top-down request might have failed.

In general, MSE has the ability to deal with such failures. For example, MSE analyzes the request to find RP3 (Fig. 11(a)) by first assuming the road-piece to be detected is a visible road(Fig. 13), and issues a request to FIVE. If this request fails, MSE switches to the other appearance of a road-piece, i.e. an occluded—road (Fig. 6). The selection between overpays and shadowed road is done based on the cause of the failure returned from LIVE.

If all efforts by MSE fail, this is reported to GRE. Then. GRE reports this to RDI and RDL', which trigger their relevant exception handlers (if any). Since different new hypotheses may be generated by such exception handlers, no immediate further analysis is activated.

Fig. 14 illustrates an example of this, where a road instance is reported that its hypothesis for an adjacent road—piece cannot be verified. Then it removes that hypothesis and newly generates a hypothesis for a road terminator (Fig. ft). assuming that it comes. to an termination.

If a top—down request issued by an instance fails, the instance reports this to GRE. Then GRE activates another instance involved in the focused situation. In the prototype system, failures of this type are not taken into accout in any way.

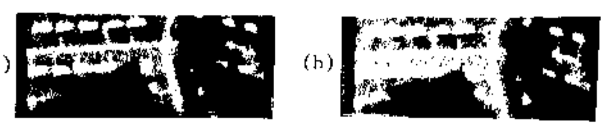
4.5.3 Merging a Pair of Partial PW Hierarchies

If a part instance is shared by two parent instances, the part issues a request, to check the similarity between the parents. If they are similar, the system merges them into one.

Similarity examination involves checking whether or not the two parents instances denote (perhaps different pieces of) the same object. For example, RDI and RD2 in Fig. 11(e) should be merged into one, although they do not denote the same (portion of) road.

In practice, according to the request from the system, the more reliable of the two parent instances to be merged, ckecks whether or not the part instances of the other instance[1] are consistent with that more reliable parent. The more reliable parent may decide to merge with the other parent, that such a merge is not (and will never be) possible, i.e. both parents are mutually conflicting, or that sufficient information is not available to make a decision.

(a)



(b)

(c)

Fig. 14 Changing hypothesis
(a)Road instance
(b)Road segment hypothesis
(c)Road terminator hypothesis

Fig. *15* illustrates an example of the third case. Suppose that definition of a house group is a group of regularly arranged bouses which face the same side of the same road. As shown in Fig. l5, if two house group instances share a house instance, the similarity examination is performed. If both house group instances face the same side of the same road instance, then they are similar and merged into one. On the other hand, if one {or both) of them has not established such a "faring" relation, then it is not possible to verify the similarity between them. Moreover, even if the two house group instances have established "facing" relations to different road instances, it is still possible for them to he similar, because those road instances may be merged later.

If the result of the similarity examination is "inconclusive", the system records its causes and suspends the action of establishing a new PW relation from a parent instance to the shared part instance. In the case shown in Fig. 15, the relation between HGI and H3 is suspended. The system records all suspended actions together with their causes. The suspended action can be reactivated if its cause is resolved by analysing other situations.

At the final stage of the analysis, the system makes copies of shared part instances involved in the suspended actions, and separates overlapping interpretations. The system does not regard these interpretations as conflicting. but considers them as possible interpretations.

5. CONCLUDING REMARKS

Fig. 16 shows the final results of analyzing the aerial photograph shown in Fig. 7. Although there are several mis-interpretations, they can easily be removed since such interpretations are isolated and/or conflicting with the correct (maximally consistent) interpretation.

SIGMA is not a completed system and has several problems to be solved First, no negative sources of evidence are considered in assessing the reliability of a situation. Introduction of negative evidence requires a more general method of combining evidence. The most, difficult problem would be how to coordinate interpretations excecuted in parallel in local areas In our system, the interpretation of an object with many parts can be initiated from any part in parallel, and partial interpretations are merged into one global interpretation if they are consistent. This merging process is triggered by a local interpretation shared by multiple partial interpretations. The complication explained in Section 4.5.3 in merging partially instantiated PW hierarchies arises from the intrinsic parallelism involved in our reasoning process. Since the current global control is sequencial, the suspension of the reasoning process was introduced to cope with the parallelism. More general and powerful control scheme is necessary to manage such parallel reasoning.

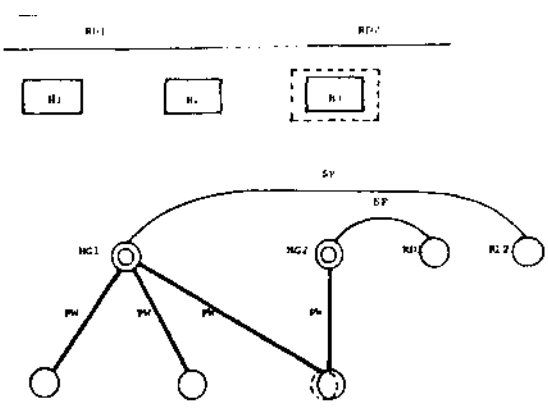


Fig. 15 Suspending unification process

ACKNOWLEDGEMENT

The authors sincerely appreciate helphul comments and constructive discussions by Profs. A. Rosenfeld and L.S. Davis of University of Maryland. The first author thanks Prof. M. Nagao of Kyoto University for giving him this research opportunity. The support of the U.S. Air Force Office of Science Research under Contract F-49620-83-C-0082 is gratefully acknowledged.

REFERENCES

[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] Nagao, M. and Matsuyama, T., "A structural analysis of complex aerial photographs," Plenum, 1980.

[3] Tsotsos, J.K., "Temporal event recognition: an application to left venticular performance," Proc. of 7th IJCAI, pp.900-907, 1983.

[4] Havens, W. and Mackworth, A., "Representing knowledge of the visual world," IEEE Computer, Vol. 16, No. 10, pp.90-98, 1983.

[5] 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.

[6] Selfridge, P.G., "Reasoning about success and failure in aerial image understanding," Ph.D. Thesis, TR 103, Univ. of Rochester, 1982.

[7] Brooks, R.A., "Symbolic reasoning among 3D models and 2D images," Artificial Intelligence, Vol. 17, pp.285-348, 1981.

[8] Nozif, A.M. and Levine, M.D., "Low level image segmentation: an expert system," IEEE Trans. Vol. PAMI-6, No. 6, pp.555-577, 1984.

[9] Minsky, M.L., "A framework for representing knowledge," in The Psychology of Computer Vision (P.H. Winston ed.), McGraw-Hill, 1975.

[10] Matsuyama, T. et al, "An expert system for top-down image segmentation," (in Japanese) Technical Report of Working Group on Computer Vision of IPS Japan, WGCV-36-3,1985.

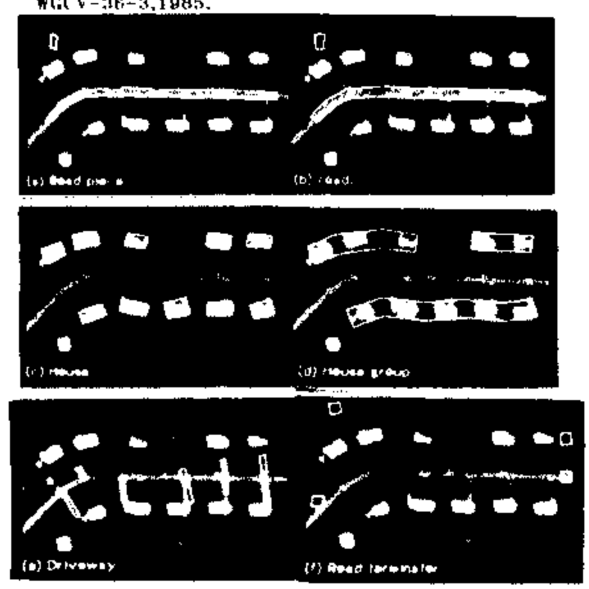


Fig. 16 Final object instances