

PVV - A GOAL-ORIENTED SYSTEM FOR INDUSTRIAL VISION

Augustin Lux and Viviane Souvignier

Laboratoire IMA6
B.P.62
38042 St.Martin d'Hères Cédex
Grenoble - France

ABSTRACT

The PVV system applies the principle of goal directed vision to the localisation of parts in 2-D scenes. It contains a description module working in unconstrained bottom-up fashion, or constrained by goal parameters. The goals for description are defined by the interpretation module as a function of the most plausible current hypothesis. Hypotheses are generated according to the results of description. The system has been entirely implemented on a microcomputer and is being used as vision sensor for the laboratory's robots.

I INTRODUCTION

A major problem in computer vision comes from the data explosion facing the description module. Goal directed processing is an important strategy for the solution of this problem; it introduces a top-down component to limit the processing to the data necessary for the system to succeed.

PVV - Prediction and Verification in Vision - is a goal-directed system in which interpretation and description are implemented as coroutines. It localizes and identifies one or more partially visible objects in a grey level image on the basis of two-dimensional object models. The system has the modular structure shown in figure 1.

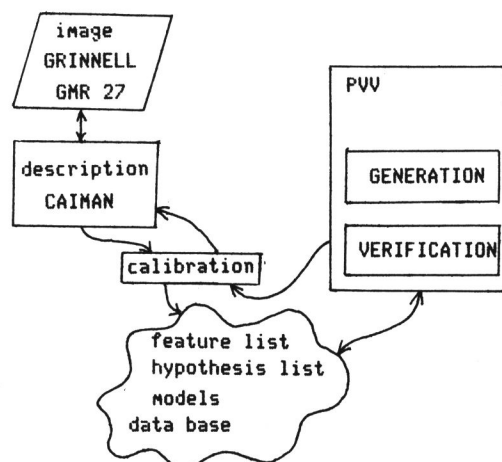


figure 1

* A "description" module based on an edge-tracking program is capable of finding features in the image (straight line segments). It can work in a bottom-up fashion, finding features in a given area of the image; and in a top-down fashion, verifying the presence of predicted image-features. Its performance depends on a large set of parameters of thresholds, tolerances, windows, and goal conditions. When no goal condition is set, the module works in a bottom-up fashion. The more precisely a goal is defined, the more the analysis is selective.

* The PVV module interprets the image features found in terms of object models in the object plane. PVV generates geometrical hypotheses combining a set of model features with a set of image features. Image features lacking in the most plausible hypothesis are used to form new goals for the description module. When a hypothesis is sufficiently plausible the system reports success.

* A calibration module converts image-plane coordinates to/from object plane coordinates. The calibration module works with an orthographic projection.

This structure gives a number of important advantages.

1/ Image description is completely controlled by interpretation. PVV uses unguided image description only to establish initial hypotheses, using conservative parameter values to assure description speed and reliability. During the prediction and verification cycles description is applied with a well-defined goal to limited parts of the image. In this case description parameters are chosen to allow for fine details.

2/ The overall system is viewpoint independent. Viewpoint information is contained in the calibration module only, it does not affect the interpretation.

3/ The choice of extended features as the basis for interpretation limits the combinatorics of the feature matching.

In this paper we describe PVV's global structure and the interpretation module. The description module is presented in a forthcoming report (Lux 1983).

Several comparable systems have been described in the literature; in particular (Perkins 1977), (Yachida

* This work has been supported by the DRET under research contract n.81/665

and Tsuji 1977), (Bolles and Cain 1982). In spite of the very similar overall specifications, the internal structure of these systems is quite different.

Perkins' system shares with PVV the use of extended features, the main differences are its use of concaves as the basis for model-to-image matching, and the multi-pass structure.

The program by Yachida and Tsuji makes intensive use of the prediction-verification paradigm but is limited to completely visible objects. It uses a richer set of feature types.

The local-feature-focus (LFF) method proposed by (Bolles and Cain) identifies and localizes partially visible objects on the basis of clusters of local features (small holes and corners). The LFF method and PVV can be considered as two complementary subsets of a more general vision system, each of which already is sufficient to solve a number of interesting vision tasks. LFF features sophisticated model analysis and the use of maximal cliques to construct initial hypotheses; PVV features global construction of hypotheses, optimal interaction between image interpretation and description, and general camera position.

IV TASK BEHAVIOR

Figures 2 to 4 illustrate the main features of the system. In the example of fig.2 the system is asked to find a lightswitch cover in a heap of small parts. The model is shown in fig.2a. Figures 2b to 2f illustrate the major phases of the analysis. The system first activates the description module to search for a line with sufficient contrast in a predefined way (dotted line on the bottom of figure 2b). While tracking this line (turning around the electric plug), three straight line segments are found (figure 2b), until the line is lost due to vanishing contrast in the white region in the upper part of the image. These segments give rise to four initial hypotheses, two of which are shown in figures 2c and 2d.

Notice that image lines and model lines do not need to coincide in a hypothesis - they only have to overlap. But allowing for partial visibility and accidental alignment increases the number of possible hypotheses. Constraining image feature length during matching can reduce by a factor of three the number of hypotheses generated in this example; only one hypothesis is generated if image feature length is limited to 1.5 times the candidate model feature length.

The most plausible hypothesis then is verified by asking the description module to find the expected missing features. Figure 2e shows the area where one of these is looked for. Figure 2f shows the line segments effectively detected at the end. A least squares fit is used to compute the best transformation superimposing the model on these segments in the object plane. Typical precision is about 1 degree in rotation, and better than one pixel in translation, which is sufficient for our robots.

Figures 3a (image) and 3b (model superimposed on found image features) show how the system can handle a problem where the detection of a low contrast line is necessary to disambiguate an object location. It would be unreasonable to search for this line without having a well justified hypothesis, since line tracking with a very low contrast threshold is bound to be time consuming and to find spurious lines.

Figures 4a (image, found segments, initial hypothesis) and 4b (final hypothesis) illustrate system performance with very noisy data. Noise in this case is produced by highlights and shadows on the object. The initial hypothesis is strongly affected by this noise. Some predicted segments are found, however, enabling the system to compute a satisfying final solution.

V THE PVV INTERPRETATION MODULE

Interpretation starts with an initial call to the description module, then control oscillates between hypothesis generation and hypothesis verification until the goal is reached or all possibilities are exhausted.

A MODELS AND HYPOTHESES

A model is a feature list describing the visual appearance of an object in its reference position in the object plane. It may contain features representing parts of the outer contour, inner contours or lines visible on the modelised surfaces. The features are not necessarily contiguous, and do not have to describe closed lines.

A hypothesis represents an object as a model instance with a given position in the object plane. A hypothesis is composed of: 1) A model feature to image feature association list. 2) A planar transformation, superimposing the model features - at least partially and within a certain tolerance - on the associated image features. 3) A plausibility and a validity value describing the current status of the hypothesis. The plausibility is a numerical function of the image features confirming the hypothesis. The validity is a number giving an account of the unsuccessful feature predictions.

B HYPOTHESIS GENERATION

The initial hypothesis generation constructs all possible hypotheses according to the initial image features extracted from different regions of the image. Several parameters allow to adapt the system to varying situations:

- the number of initial image features
- the threshold of minimal plausibility
- the tolerance for accidental alignments

On subsequent calls to this phase, which take place when new hypotheses are needed to achieve the goal, they are generated in a more selective manner. Existing hypotheses are updated and/or a limited number of new hypotheses are generated. Image features

necessary for these operations are taken by order of decreasing reliability. If these operations fail, the description module is called to search for more features.

C HYPOTHESIS VERIFICATION

The hypothesis having the highest plausibility value is submitted to a sequence of prediction-verification cycles. The system determines the most important non instantiated model feature. A window in the image is computed where this feature should be found. This prediction results in a call to the description module with a well circumscribed goal. The description module either finds one or several segments which are part of the searched feature, in which case the hypothesis' plausibility is increased. Or the description module does not find the predicted feature, and the hypothesis' validity is decreased. Predictions and verifications are cycled through until one of the following three conditions is satisfied:

- plausibility > confirmation-threshold
- validity < refute-threshold
- no unused feature left

A hypothesis being confirmed, PVV has localized and identified an object in the scene. If a hypothesis is rejected, or can be neither confirmed nor rejected once all features have been searched, the most plausible hypothesis is selected between the remaining ones. If it is not valid enough control passes to the hypothesis generation module.

VI CONCLUSION AND FUTURE WORK

The system described in this paper has been developed in FORTRAN on a large time-sharing system (MULTICS on a HB-68) and entirely implemented on a LSI-11/23 microcomputer. It is connected with the LM manipulator language interpreter (Miribel and Mazer 1982). Programs occupy about 55K words, minimum heap size is 4K words. Execution times are not yet significant, since the system spends its time swapping the program on floppy disk. The interpretation phase of the first example (fig.2) takes about 0.8sec. on the HB-68.

Work is continuing on the description module to admit more varied feature types. The interpretation module is being completed to handle multiple-object hypotheses with occlusion analysis.

ACKNOWLEDGEMENTS

- The PVV interpretation module owes much to a program written by R.Popplestone in PDP-2 during his stay at the IMAG Laboratory winter 1981.

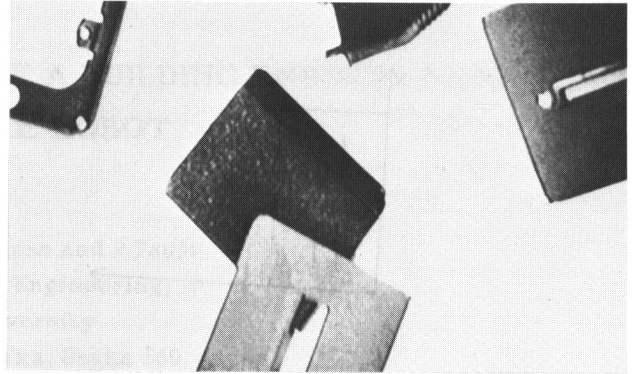


fig.3a

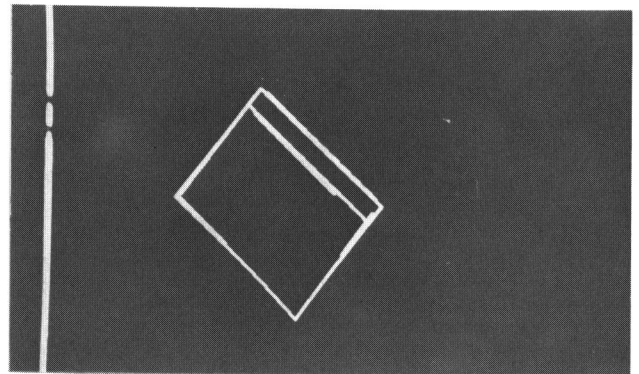


fig.3b

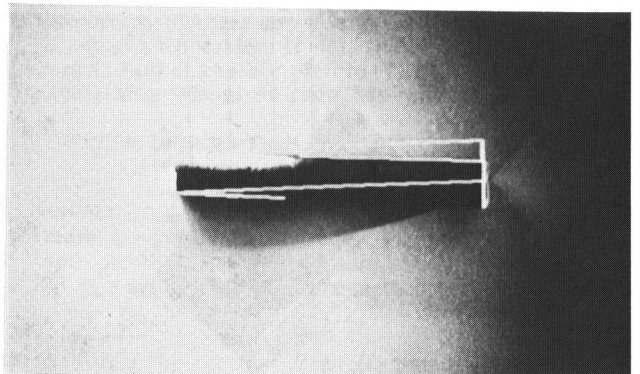


fig.4a

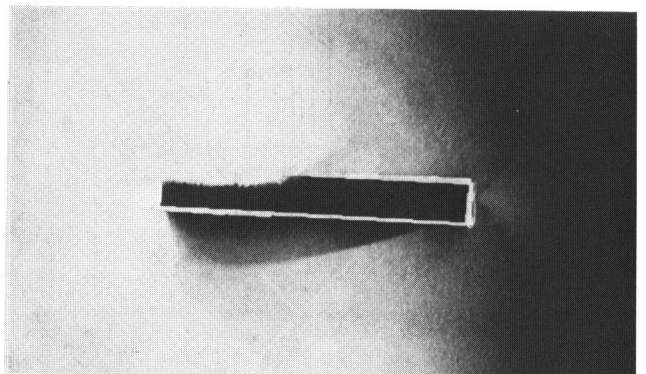


fig.4b

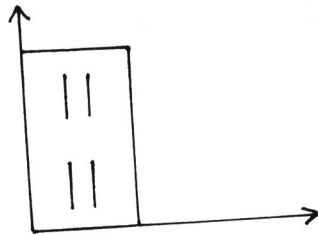


fig.2a

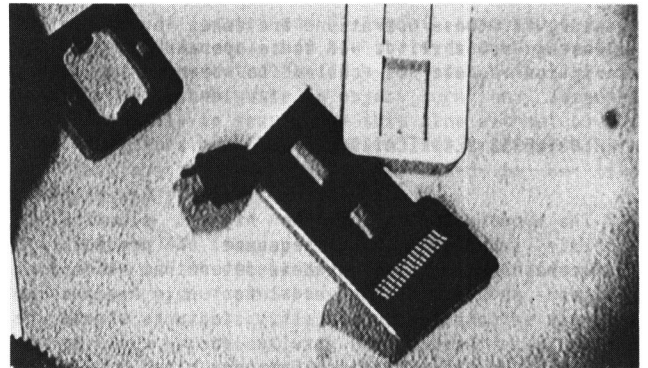


fig.2e

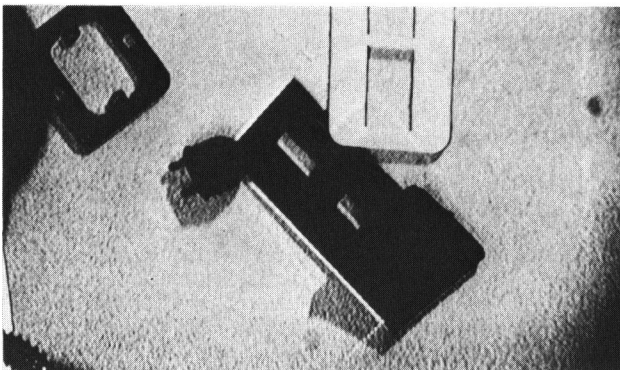


fig.2b

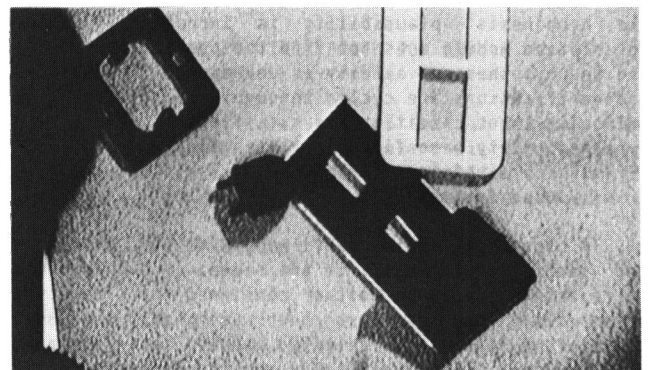


fig.2f

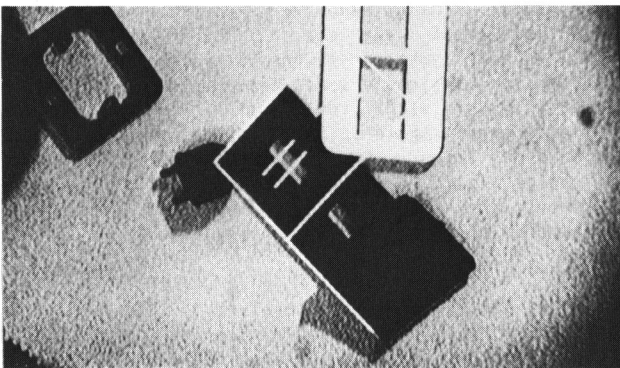


fig.2c

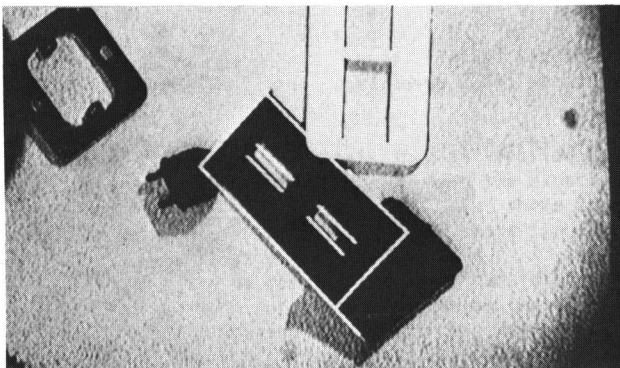


fig.2d

REFERENCES

- [1] Bolles, R.C., and Cain, R.A.(1982) "Recognizing and locating partially visible objects: The Local-Feature-Focus method", IJRR 1,3, p.57-82
- [2] Lux, A.(1983) "CAIMAN - Manuel d'utilisation", Research Report, Lab.IMAG, Grenoble - to appear
- [3] Miribel, J.F., and Mazer, E.(1982) "Manuel de référence LM", IMAG Laboratory, Grenoble
- [4] Perkins, W.A.(1977) "Model based vision system for scenes containing multiple parts", IJCAI-77, p.678-684
- [5] Yachida, M., and Tsuji, S.(1977) "A versatile machine vision system for complex industrial parts", IEEE Trans.Comput. C-26, p.882-894