IAPR Workshop on CV - Special Hardware and Industrial Applications OCT.12-14, 1988, Tokyo

HYPOTHESES OF GEOMETRIC STRUCTURE GENERATED IN AN ACTIVE COMPUTER VISION SYSTEM

Brunnström, K., Eklundh, J-O., Olofsson, G.
Computer Science and Numerical Analysis
Royal Institute of Technology

S-100 44 Stockholm
Sweden

ABSTRACT

We will discuss how hypotheses of geometric structure can be generated in an active computer vision system. These hypotheses are built on sparse but reliable observations and can communicate with higher level hypotheses as well as the basic classified data hypotheses. We present experiments with such hypotheses, where we have chosen to form hypotheses about the existence of planar surfaces in a scene, given classified edges and junctions. The results of the experiments illustrate clearly the benefit of the approach. A goal directed active vision system could easily use this idea to make both powerful verifications of generated predictions as well as new hypotheses and predictions at different level of abstraction.

INTRODUCTION

A robot vision system in a static environment should be able to resolve the 3-D structure of its immediate surroundings. The human visual system is able to produce a stable idea about the world using mainly 2-D information, i.e. images, and a priori structural knowledge in the given context. We think of it as a process for hypothesis verification using a continuous flow of perceived information. Some similar process ought to be appropriate for an automatic vision system as well.

In a robot system high level ideas cannot emerge out of thin air, and therefore image data are used for generating some initial hypotheses. A verification process can then be used at various levels of complexity simultaneously to get a more complete and stable set of structural hypotheses. This kind of intermediate clues should also trigger more conceptual hypotheses about objects and their interrelations in the scene.

In this process, new information will be received that can be used for improving the completeness of the scene interpretation. We will let the system set up goals to resolve ambiguities and incompatible interpretations of parts of the scene by looking closer or getting a new projection of the scene.

We can now talk about the system as a continuous dialogue between the bottom-up hypothesis generation on one hand and the top-down hypothesis verification and view planning on the other.

A recurrent problem in 3-D vision systems is the explosive growth of the internal models. We want to address this problem by letting the system generate simple, conservative hypotheses from sparse but reliable data.

Geometric 2-D structure is very useful in the interpretation of 3-D scenes [1][2][3][4]. We have chosen to begin with a test of our ideas by constructing a system to generate intermediate hypotheses about the existence of planar surfaces and their interrelations in the scene. Extracted edges, lines and junctions are classified and used in order of validity for generation of hypotheses about instances of planar surfaces.

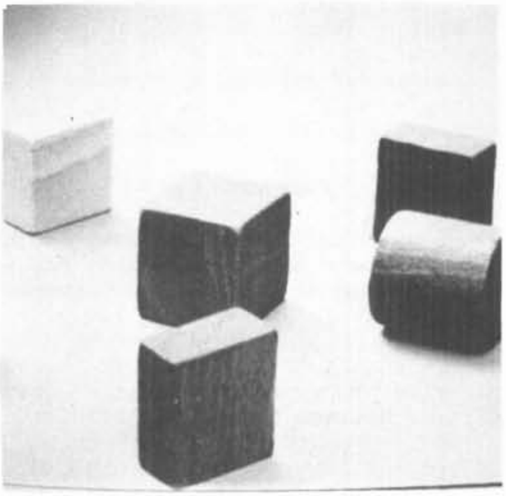

Figure 1. The test image.

THE ACTIVE HYPOTHESIS

In general

The general idea is to use sparse sets of reliable image data and hypotheses in an active reasoning system to achieve stable interpretations of a static 3-D scene. We mentioned above the continuous prediction–verification process on and between all levels of abstraction. This process is the "reasoning machine" we are aiming at. The goals of the reasoning could be of several types, e.g. specifically towards determining the existence of a particular object. Then we would start with predictions of the existence of different intermediate image cues from at least one projection of the scene. The verification would be to try to get correspondence between such predictions and sparse sets of conservative low level hypotheses about the interpretations of image properties. These low level hypotheses have been extracted independently and are verified by prediction of other low level properties or intermediate hypotheses. If these predictions are verified, then the predicting object also gets at least partly verified. An example is when junctions and edges mutually verify each other as being junctions and edges respectively. This can be done e.g. by matching edge ends to the junction, perhaps even in directions predicted by the junction.

Other subjects for reasoning could be to verify the existence of a particular texture, geometric structure or object motion, or to verify some magnitude of scale, distance or motion.

The Experiment

As mentioned in the introduction, we have chosen to test the prediction–verification idea on hypotheses about the existence of planar surfaces in a static scene. The particular choices of hypotheses and data extractions are unimportant. The essential is the function of the reasoning process which we think we can illustrate and evaluate with these experiments.

One very strong indicator of the existence of planar surfaces is two connected non-collinear lines. We use a line to trigger the hypothesis about the existence of a planar surface on either side of it. If the line also has a line as neighbour, then we try to verify the planar surface hypothesis. It is supposed to be verified if a closed contour can be formed by predicted and verified edges around it. We do not say that there is only one plane, but that there is at least one. We select each connected edge according to a somewhat lengthier rule that favors convexity of the plane on a given side of the contour. This means that contours that have an edge pointing towards its middle are not verified as planes at this stage of the process even if they should have been closed if it was not for this edge. Such contours and non-closed contours can be used to predict more edges or reclassifications of edges and junctions. These edges and junctions, that we use to predict and verify the plane hypothesis, are themselves first extracted and classified by some methods from the image data and then verified as mentioned above by simple combination of edge ends and junctions etc. This method also reduces the amount of reliable primitives.

We use the most reliable hypotheses only. This means that edges classified as lines with both ends matched to a junction are used before other edges in the prediction of sectors. Less reliable hypotheses on all levels could either be used in further predictions of lower or higher level instances of "knowledge" or could themselves be involved in another hypothesis' verification process as being predicted by another hypothesis. In the latter case the two hypotheses might verify each other.

The existence of a planar surface should, of course, also generate higher level hypotheses on, for example, their real contours, orientations, interrelations or even what objects they might belong to. We have not yet implemented such methods.

GENERATING PLANAR SURFACE HYPOTHESES

To illustrate the strength and weaknesses of the approach we will use a picture of a simple scene, as shown in Figure 1, and generate hypotheses of planar surfaces from it. The output will only point out the possible existence of such surfaces and the final verification is left to higher level reasoning.

As input to the hypothesis generator we use a set of junction points, and a set of lines and curves. The junction points are obtained by processing the image with a technique for finding interest points developed by Kakimoto [5], based on Moravec's approach [6]. The number of points have been reduced by thresholding. The threshold has been set quite low which means that we will use quite a few points, about 500 of 1500, as shown in Figure 2.

The lines and curves have been generated with the *edge focusing* method developed by Bergholm [7] combined with the strategy of classifying edges as distinct and diffuse [8], as shown in Figure 3. When the diffuse edges (grey in Fig. 3) have been filtered away and the edges have been linked, the edge segments are hand classified as line and curve segments. This could also have been done with a curve approximation method, such as the one by Bengtsson et. al [9]. In the experiment only the line segments have been used, but there is nothing in the method that require this.

The generation of hypotheses of planar surfaces will be done in three different steps. A block diagram can be seen in Figure 4. First, all the junctions will be traversed to find all those that have at least one line or curve connected to it, see Figure 5. At this stage we only use closeness as a condition, but it is possible to judge the directions of the edges around a junction and this information could also be used as additional conditions and at predictions.

The next step is to take all those junctions which have a matching edge and try to find a junction on the other end of the edge; these edges are shown in Figure 6. The criteria for a match is that the junction point is in a small rectangular area around the end point of the edge. Those edge elements that have matched one junction point, but failed to find the other, they still have some support for being important; therefore, we think it is motivated to perform a predictive search for other junction points or end points of edge segments. First, the rectangular search area will be made bigger, then we search in a sector with small opening angle in the end direction of the edge segment some pixels out from the end point.The size of the sector is in this experiment set totally arbitrarily to an opening angle of fifteen degrees and a length of thirty-five pixels. If both an end point and a junction point are found the junction point will be selected. We consider those edges which meet the condition of having one matched junction on each end of it as a very strong cue of something important in the scene.

As a trigger condition for generating a planar surface hypothesis, we demand that one of these *strong-cue-lines* has a non-collinear connected line on the other side of one of its junction points. Those lines that meet this condition can be seen in Figure 7. If this condition is met a search will be started for finding a closed boundary. The search is performed first in one direction and if it is not possible to come back to the start point again, the search will continue in the other direction. Primarily the search for the next boundary segment goes via a matched junction point, but if this fails, the condition is relaxed to find an edge segment with an end point in a small neighbourhood around the end point of this segment. The planar surface hypotheses with closed boundary can be seen in Figure 8.

At this level no assumptions of the shapes of the planar surfaces are made, so curved and concave parts will also be possible. Hence it is probable that some hypotheses may be incorrect, for instance if an edge is missing on a polyhedral object two different planes may be hypothesized as one plane. This goes also for boundaries belonging to different objects, but any found T-junction will be registered, though, for later use. These false hypotheses will have to be resolved on a different level, where interaction between close lying planar surfaces and geometric reasoning will reject the false hypotheses and verify the correct ones. This part has not been a goal for this experiment and therefore has not been incorporated here.

DISCUSSION OF THE EXPERIMENTAL RESULTS

Starting at a low level; the interest points signals for possible junctions points, but there are only a few of them that actually describe something important in the scene. It tends to be a correspondence between interest points with strong values and significant features in the image; however, it is hard to find a good threshold value and thresholding will of course also remove important points which have low values. By letting the interest points and the edge segments verify each other, an intelligent choice will be made among the two sets of data, as shown by Figure 2, 3 and 6. There is very low reduction of edge elements in this experiment, however. The reason for this is that the edge data is very good and has been filtered of diffuse edges. It can be seen here and have been observed when using other edge detection schemes, that the number of edges due to

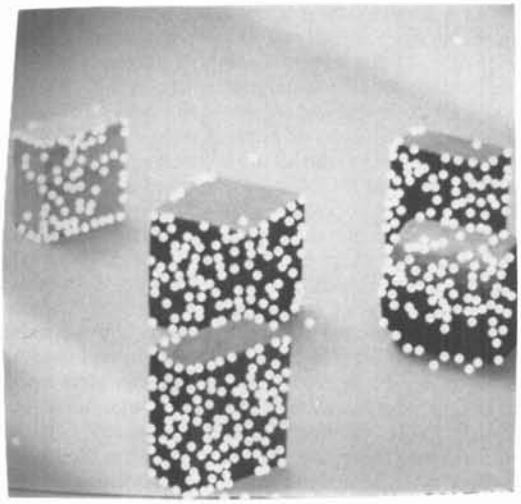

Figure 2. The original set of interest points

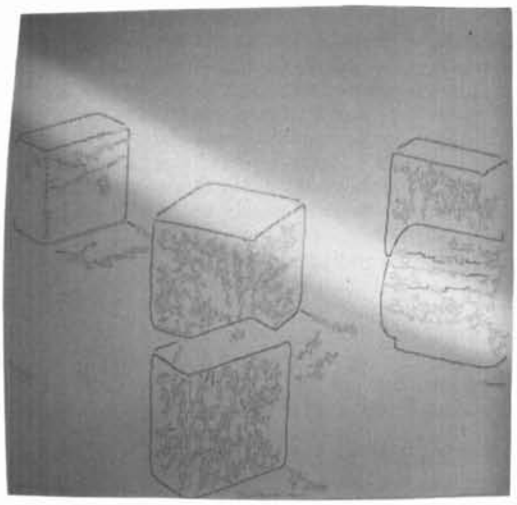

Figure 3. All edges; grey means classified as diffuse.

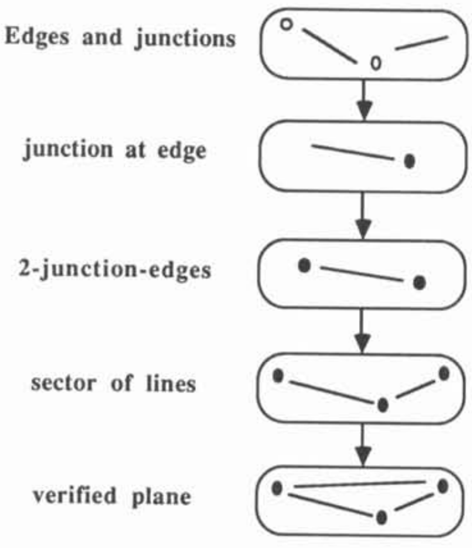

Figure 4. Scheme of the geometric grouping.

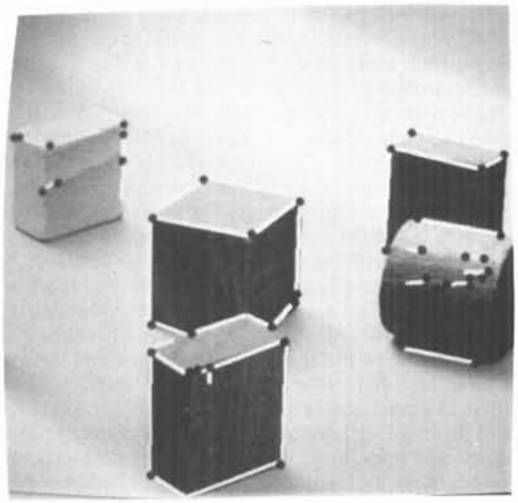

Figure 5. Junctions with at least one connected edge.

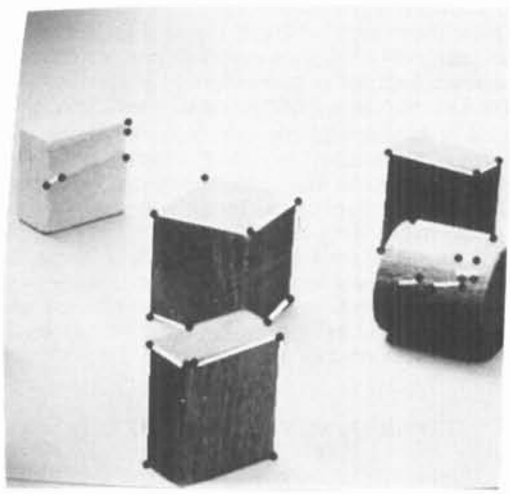

Figure 6. Edges matched to 2 junctions, including predicted d:o.

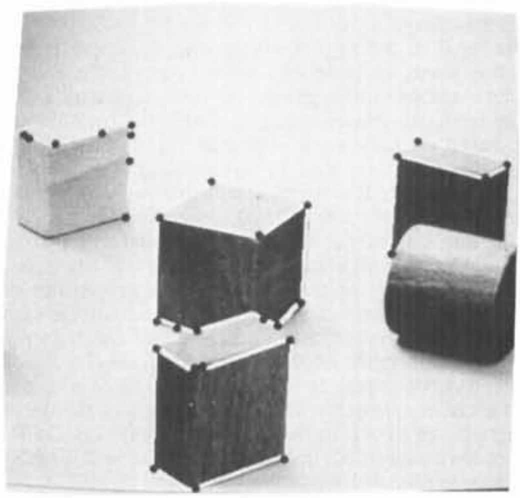

Figure 7. All lines that form planar surface hypotheses.

noise have been brought down. It could be noted that there are important edges that are filtered away as diffuse edges. We do not think that this matter so much, since this comply with our notion that the initial hypotheses should be made on a sparse set of reliable data.

At higher level, very strong hypotheses are found, as those shown by Figures 7 and 8, with a simple and straightforward search strategy using no special knowledge, but some basic facts about planar surfaces. This is part of our strategy, that at an early state form high level hypotheses and based on these make top-down predictions. The two planes found with closed contours in Figure 8, could be used for generating hypotheses for object recognition. These could be matched against other edge elements and planar surface hypotheses in its immediate surroundings and then make predictions about missing and incomplete boundary elements.

The process has not succeeded to find closed boundaries at a couple of places, in spite the fact that all necessary edge segments and junction points seems to be there. For instance; at the rectangular prism closest to the camera, there is at one place two junction points at close distance, which have given rice to a mismatch; and at the plane, on the cube partly occluded by the front toy block, the predictive search, described in last section, has failed to find the proper junction point due to the order in which the search is performed. There are some other places also, where a better performance would have been expected. These are flaws that every algorithm at this level will have, but the important point is that the planar surface hypotheses have been generated in an intuitive manner and could therefore be used with few rejections by higher level reasoning.

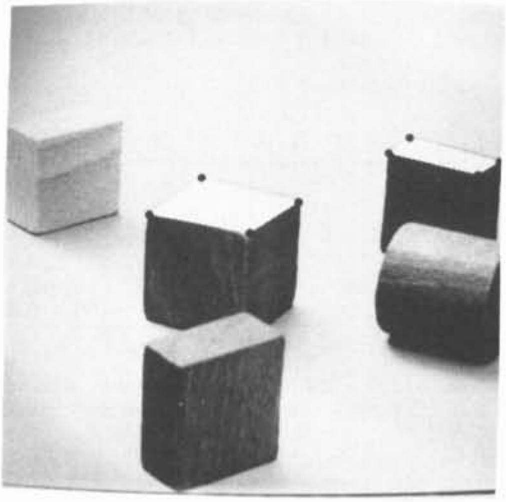

Figure 8. Verified plane hypotheses.

CONCLUSIONS

The aim of our work has been to demonstrate the power of modelling visual scene interpretation as a reasoning process. Hypotheses triggered on different levels of abstraction have been verified by the use of very simple geometrical grouping techniques over sparse but reliable sets of image data and hypotheses. The sparseness of data and hypotheses together with simple geometrical prediction have been proved in our experiments to be highly effective. The predictions are verified by the use of simple techniques for geometric grouping in a manner of least commitment; the strongest clues for verification are used first and then we still don't over-interpret the results. We call this conservative hypotheses.

By the use of these principles in our experiments we have been able to trigger few but correct intermediate hypotheses about the geometrical structure of the scene. From the combination of a small set of edges and junctions we have predicted and verified the existence of planar surfaces.

The implications for future work include the incorporation of active visual search for predicted geometrical scene properties and the use of other hypotheses about geometrical structure on low as well as higher level of abstraction. For example hypothesized curves, hypotheses about the existence of non-planar surfaces and their relations to each other and the planar surface hypotheses. We think we can resolve much of the geometrical 3-D structure of a static but complex scene by the use of such kinds of hypotheses and their interaction with a goal directed active visual search process. Predictions of scene structure can for example be verified by getting more data from a new view point with the camera.

ACKNOWLEDGEMENT

We gratefully acknowledge The National Swedish Board for Technical Development. This work was done under the basic research program for image analysis

REFERENCES

1. Brooks, R. A., *Symbolic Reasoning Among 3-D Models and 2-D Images,* Artificial Intelligence, vol 17, Aug. -81.
2. Malik, J., *Interpreting Line Drawings of Curved Objects,* Int. Journal of Comupter Vision, 1, pp 73-103, -87.
3. Olofsson, G., *Experiments with an Algorithm for Line Extraction,* TRITA-NA-8682, Computer Science and Numerical Analysis, Royal Institute of Technology, Stockholm, Sweden.
4. Brunnström, K., *Target-directed Understanding of 3-D Objects in a Knowledge-based Vision System,* Proc. 5th Scandinavian Conference on Image Analysis, Stockholm, June -87.
5. Kakimoto, A., *An Algorithm for Finding Interest Points,* TRITA-NA-8807, Computer Science and Numerical Analysis, Royal Institute of Technology, Stockholm, Sweden.
6. Moravec, H. P., *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover,* Stanford Artificial Intelligence Laboratory Memo AIM-340.
7. Bergholm, F., *Edge Focusing,* IEEE PAMI vol. 9, No 6, Nov. -87.
8. Bergholm, F., Sjöberg, F., *Extraction of Diffuse Edges by Edge Focusing,* Pattern Recognition Letters, 7:3, pp 181-190, (March 1988).
9. Bengtson, A., Eklundh, J. O., Howako, J., *Shape Representation by Multiscale Contour Approximation,* TRITA-NA-8607, Computer Science and Numerical Analysis, Royal Institute of Technology, Stockholm, Sweden. To appear in IEEE PAMI