Hypothesis and Verification in 3-D Model Matching.

Francis P. Sykes, Stephen B. Pollard and John E.W Mayhew
Artificial Intelligence Vision Research Unit,
University of Sheffield,
Sheffield, S10 2TN.

This paper addresses 3-D hypothesis and mixed 3-D and 2-D verification as a means of obtaining high performance closed loop model-matching. Performance includes the following criteria: low computational expence; low failure rate; and good localisation.
Hypotheses are based upon congruencies identified between 3-D scene and model descriptions. Verification is provided by quantitative evaluation of the 3-D match and the identification of inconsistency (relient upon properties of opaque objects).
Although the results are dependent upon the model representation and the quality of the input (raw images), the ability of a verification system to discriminate between a correct and a wrong match has been demonstrated in a number of experiments.
Furthermore, symmetries are identified by the system and exploited to guide the search for correct matches.

The architecture of the matcher/verification algorithm

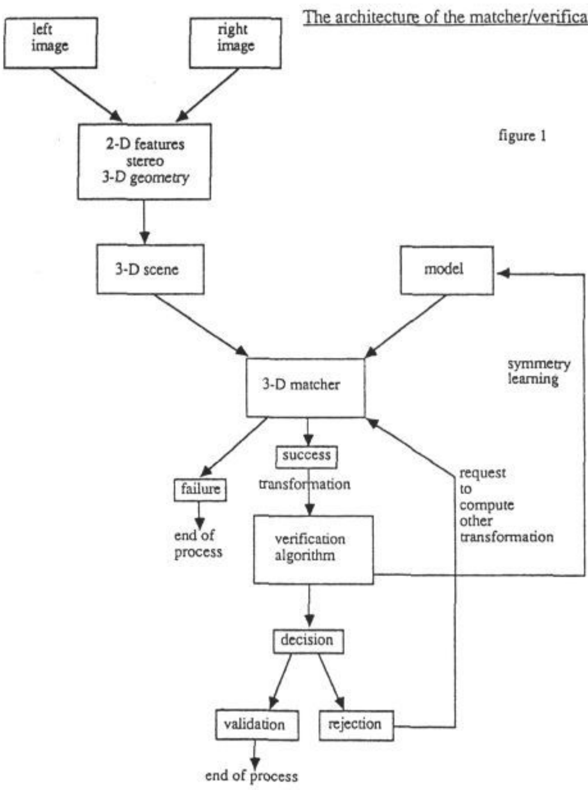


figure 1

F.P. Sykes is a visiting research student from ENST, 46 rue Barrault, 75013 Paris, and gratefully acknowledges SAGEM and the British Council for their support.

This paper describes the 3-D model matching algorithm used within the TINA system at AIVRU Sheffield (an early version of the TINA system is presented in Porrill et al 1987). The adopted strategy (see figure 1) is to base initial matching hypotheses on congruencies identified between 3-D scene and model descriptions, and then to employ a model-based verification strategy exploiting both 2-D image and 3-D scene descriptions to determine the correctness of the hypotheses. Approaching the problem of model matching in this fashion allows us to combine all the good properties of 3-D scene and model matching (primarily computational tractability and geometrical accuracy), with the robustness and completeness of methods based upon the back-projection of the model.

The 3-D scene descriptions encountered in the current TINA system are obtained through the processes of edge and line-based binocular stereo and are presently restricted to linear segments. The object models may also include surface information (in the form of a surface tessellation). However, matching hypotheses are restricted to those linear features corresponding to discontinuities in the model's surface.

1. Match Hypothesis

Matches are hypothesised using the SMM model matcher, (an early version is presented in Pollard 1987). The SMM algorithm exploits ideas from several sources: the use of a partial pairwise geometrical relationships table to represent object model and scene description from Grimson and Lozano-Perez (1984), the least squares computation of transformations by exploiting the quaternion representation for rotations from Faugeras et al (1985), and the use of focus features from Bolles et al (1983).
Given a model $M = \{m_1....m_i....m_n\}$ and a scene $S = \{s_1....s_j....s_m\}$ the matching hypothesis strategy proceeds as follows:

(1) A number of focus features are chosen from the model $F = \{f_1....f_i....f_n\}$ where $F \subset M$.

(2) A number of groups $G_i \subset M$ of features from the model of cardinality at least C are chosen; these have two properties:

- Each group has a principle focus feature $pf(G_i)$ (though it is allowed to include additional focus features).
- Each focus feature is the principle focus feature for at least one group.

These groups form the basis for hypothesis.

AVC 1989 doi:10.5244/C.3.2

(3) Potential matches for each of the principle focus features of each group are selected and represented as the pair $(pf(G_i),s_j)$.

(4) For members of each group, consistent matches (in terms of a number of pairwise geometrical relationships) are sought in the context of each principle focus feature match, to form a set of matched groups. For example the matched focus feature $(pf(G_i),s_j)$ gives rise to the matched group of pairs (m_l,s_k) where $m_l \in G_i$ and $pwgr(pf(G_i),m_l) \equiv pwgr(s_j,s_k)$.

(5) Each matched group is searched for maximal mutually consistent (in terms of the pairwise geometrical relationships) cliques of cardinality at least C, each of which can be thought of as an implicit transformation.

(6) Each mutually consistent clique is further constrained by insisting that the explicit transformation they represent (recovered using the method described by Faugeras et al (1985)) includes each of the constituent matches.

(7) Hypothetical matched groups are ranked on the basis of their cardinality and total match length. The matched group with highest rank provides the best hypothetical match.

The relationships between pairs of lines from scene or model used here are:

- orientation differences.
- minimum separations between (extended) lines.
- distance to the beginning and end of each physical line with respect to the point of minimum separation and in the direction of the line (only applicable for non-parallel lines).

1.1. Lazy Evaluation

The advantage of the above matching strategy, in addition to the fact that exhaustive global search is avoided, is that it can be subject to lazy evaluation at stages 4, 5 and 6. Not all matched groups require examination at each of these stages. Consider the use of a conservative evaluation function to provide an upper bound upon the match quality at each stage. If at stage 5 (the identification of maximal cliques within each matched group of line segments), for example, we rank matched groups according to this bound, then it is possible to halt the max-clique search if a mutually consistent clique is found with a value greater than the bound of the remaining match groups. If subsequently the best hypothesis's evaluation reduces below this bound (or is rejected by the verification process) then the max-clique search will recommence.

2. Model Driven Verification

As a precursor to the verification process, the clique of consistent matches of the current hypothesis is extended. Given our initial estimate of the transformation it is possible to examine the set of potential matches to identify those that are consistent with it. The best new transformation, in the least squares sense, can be computed from the current set of candidate matches. The process of estimation and consistency checking continues until the set of consistent matches achieves saturation.

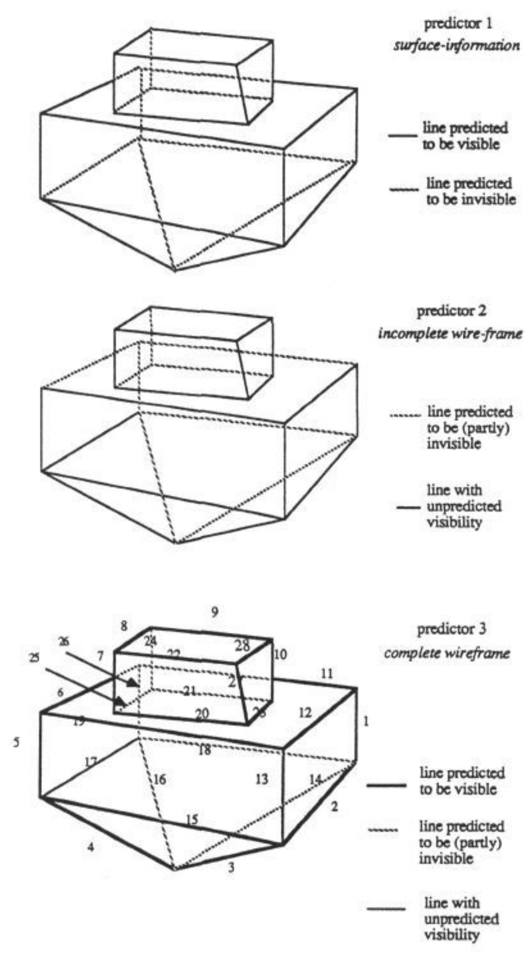
2.1. Determining Visibility

The first step in the verification procedure is visibility prediction, to derive for every model line its a priori visibility. Each line in the model is segmented into sublines (ordered along their length) which are appended with predicted visibility (often a line includes only a single subline with a single visibility label). The level of sophistication and completeness of this scheme depends upon the level of surface information, and topological completion of the model representation. However even in situations where incomplete models with little or no surface information are used some visibility prediction is still possible. Three levels of surface description are considered here (illustrated in figure 2):

- full explicit surface descriptions
- implicit surface description from full wire-frame description
- no surface description

Comparing labeling in the three visibility predictors

figure 2



The process of prediction is straightforward, if laborious, when explicit surface representations is available. If this is not the case, implicit surface knowledge can be used; for example, if model and/or scene lines intersect in their projection on an image plane, either both share a common depth or one is situated behind the other in which case we can make certain assumptions concerning visibility.

Given complete wire-frame descriptions the following strategy can be adopted.

(1) identify external boundaries of the object with respect to current viewpoint (these must be visible).

(2) the point which is closest to the camera's optic centre is also visible.

(3) when a line that has been labeled partly-visible is occluded by an intersecting line, the visibility of the former is propagated (locally) to the latter.

(4) when a line that has been labeled visible meets a vertex, the purely local examination of the hypothesised surfaces at the vertex enables visibility to be propagated to some incident lines. This process identifies the set of lines that are visible whatever the surface distribution of an object, given its wire-frame representation.

2.2. Line Based Verification

Following their projection into the left and right image planes a series of line searching process attempts to locate the predicted model subline segments. The various search process place initial emphasis on the left image.

According to the results of visibility prediction and line search, sublines from the model are classified according to one of the following six categories:

(1) lines that were expected to be visible and have indeed been matched.

(2) lines that were expected to be visible but have not been matched.

(3) lines that were expected to be invisible and have not been matched.

(4) lines that were expected to be invisible but have been matched.

(5) lines whose visibility has not been predicted and have been matched.

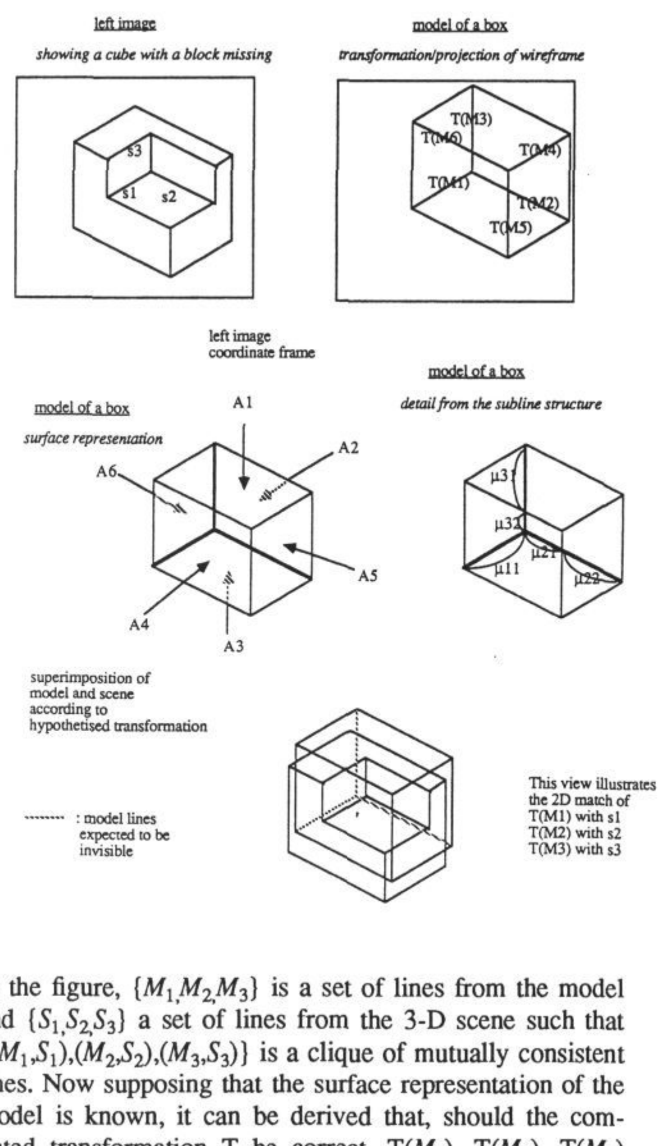
(6) lines whose visibility has not been predicted and have not been matched.

In the cases where lines that are not predicted as invisible have been matched (ie cases 1 and 5) the length of the projection of the matched line in the left image is added to the partial sum of the length of visible lines.

If the measured length of the scene line is considerably greater than that of the model line, (and if the informations on length, in the model representation are reliable) then this is considered a geometrical inconsistency.

In the case of lines that were expected to be invisible but have been matched (ie 4), false alarms are avoided by also searching for the line in the right image. If this search is successful, then there is strong evidence that the current transformation was computed on the basis of an implausible matching between model and scene lines. Such a situation is illustrated in figure 3:

figure 3 The detection of inconsistencies



In the figure, $\{M_1, M_2, M_3\}$ is a set of lines from the model and $\{S_1, S_2, S_3\}$ a set of lines from the 3-D scene such that $\{(M_1,S_1),(M_2,S_2),(M_3,S_3)\}$ is a clique of mutually consistent lines. Now supposing that the surface representation of the model is known, it can be derived that, should the computed transformation T be correct, $T(M_1)$, $T(M_2)$, $T(M_3)$ should be occluded. So, the 3-D match is inconsistent with the knowledge of surfaces because the transformation it yields matches M_1 with S_1, M_2 with S_2, M_3 with S_3, yet from the position of the camera's optic centre relative to the model, M_1, M_2, M_3 should be occluded by A_1, A_4, A_5.

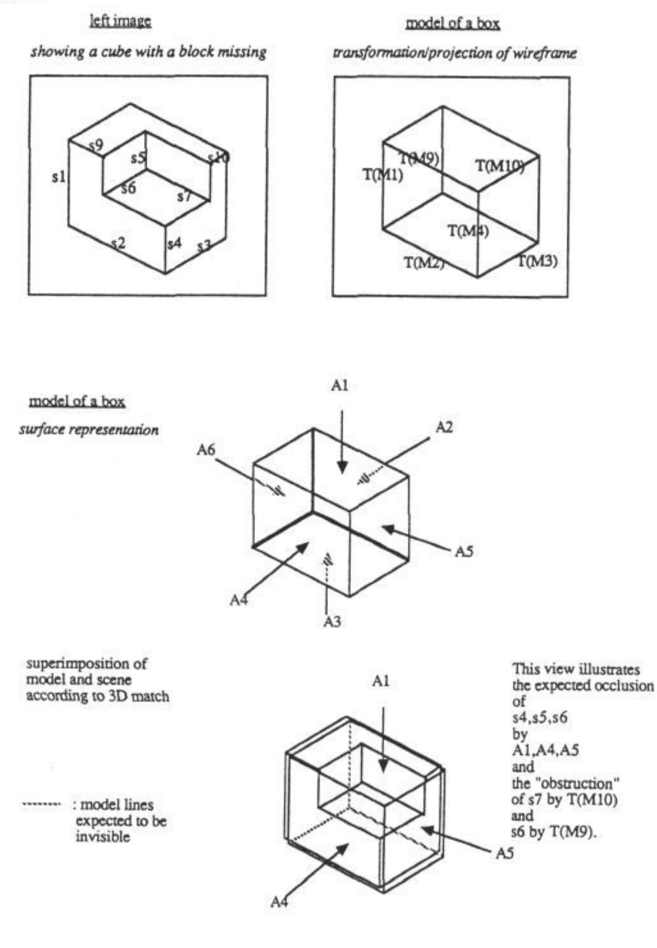
If explicit surface information is not available, the same idea can be exploited: In figure 3, the subline structure of lines M_1, M_2 and M_3 is (μ_{11}), (μ_{21}, μ_{22}), (μ_{31}, μ_{32}) respectively. Visibility prediction using wire-frame representation states that two neighbouring sublines cannot be matched simultaneously. An inconsistency is revealed by the fact that μ_{21}, μ_{22} have both been matched successfully.

If however the match in the right image is unsuccessful, then it is probable that the matching of the projected model line in the left image is coincidental and therefore the match is not doubted by this information. It is possible of course that this approach is over conservative for this particular line and indeed the 2-D match in the left image reflected the matching of 3-D features but that owing to poor quality in the right image the projection of the model line on the right image is unsuccessful. In such cases it is likely that other inappropriate matches also exist and at least one of these will be able to doubt the currently hypothesised transformation.

In the case of (2)or(6), a more thorough line search is undertaken, which if successful will result in the relabelling of the line. If this process also proves to be unsuccessful the projection of the line into the right image is also considered.

(v) If sufficient information has not already been obtained to falsify the currently hypothesised matched, The verification process continues by searching for potential inconsistencies with scene data. The inspection of those scene lines that project in 3-D within the hypothesised model can be used to reveal wrong 3-D match hypotheses. This point is illustrated in figure 4.

figure 4 Exploiting object opacity



High confidence scene descriptions should lie in front of the hypothesised model and not behind it. In a similar fashion the absence of match for a line predicted visible can be justified in this way (that is; if a high confidence scene description does lie in front of it then there is a good chance of an occluding object). If no surface information is available, a similar reasoning can be held. In figure 4, $s_1,..,s_{10}$ be a set of 2-D scene lines corresponding to 3-D lines $S_1,..,S_{10}$. The matcher has matched $S_1,..,S_4$ to $M_1,..,M_4$. Let T be the transformation associated with the 3-D match, $T(M_1),..,T(M_4)$ are matched in 2-D to $s_1,..,s_4$. The study of the intersection of s_6 and $T(M_9)$, S_7 and $T(M_{10})$ contravenes the 3-D match.

2.3. Quantitative evaluation of the 3-D match

So far, we have been investigating ways by which we can derive inconsistencies when the 3-D match is wrong. However, the absence of such information does not imply that the 3-D match is correct, so we need to take into account the quantitative results of the 2-D matches.

To this end, we can compute various criteria that give some idea of the quality of the match. We will express the criteria in terms of the ratio between a measured quantity (number of lines matched in 2-D, length of these lines in 2-D or 3-D) and the corresponding predicted quantity (derived from visibility prediction). We have investigated the following criteria

$$\text{criterion 1} = \frac{Nmatched}{Ntot}$$

$$\text{criterion 2} = \frac{matched\_vis\_length}{max\_vis\_length}$$

$$\text{criterion 3} = \frac{matched\_vis\_length}{max\_vis\_unobstructed\_length}$$

$$\text{criterion 4} = \frac{matched\_vis\_length}{max\_not\_invis\_length}$$

where

Nmatched is the number of 2-D matched lines.

Ntot is the total number of lines expected to be visible.

matched_vis_length is the sum of the lengths of all lines in the image matched in 2-D to the projection of a line from the model.

max_vis_length is the sum of the length of all projected model lines expected to be matched in 2-D.

max_vis_unobstructed_length is the sum of the lengths of all projected model lines expected to be matched in 2-D accounting for occlusions.

max_not_invis_length is an upper bound of the optimal length of model lines matched in the image.

The strategy preferred is to insist both criterion 2 or 4 (depending upon the availability of surface information) to be above a lower threshold and criterion 3 to be above a higher threshold.

The 3-D match is validated if the above requirement is met and if no geometrical inconsistency has been identified.

If the 3-D match is rejected by the verification process, the latter invokes the matcher which attempts to compute another transformation. The process terminates when the verification process validates a match or when the matcher is unable to compute a (further) transformation.

3. Learning about Partial Symmetries

When a hypothesis generated by the 3-D matcher has been rejected by the verifier, the list of lines from which the assumed transformation is computed can provide useful information.

M defines the relationship "are consistent matches within the transformation" so that M(scene_line, model_line) can be read: in the transformation, scene_line is matched to model_line. Let scene and model lines be indexed as scene(k) and model(l) for $k \in S$ and $l \in M$.

The transformation can be described by the following relationship:

There is a function T defined from $V \rightarrow M$ and a set V, ($V \subset S$) such that for all $k \in V$:

$$M(scene(k),model(T(k)))$$

Let T', V', M' be parameters associated with another transformation. For all k in V', M'(scene(k),model'(T'(k))). Hence we can define a relationship R by

R(T(k),T'(k)) <-> M(scene(k),model(T(k))) and M'(scene(k),model'(T'(k))).

This means that the one set of scene lines {scene(k)}, for $k \in V \cap V'$, embodies the same configuration as {model(T(k))} or {model'(T'(k))}. If model = model', (the model used is the same in both transformations), then R unveils a symmetry within the model. If model≠model', then R unveils a similarity between the two models. This property is only interesting if the cardinality of $V \cap V'$ is not to small (what would be the point in using the fact that model and model' both have a right angle).

This can be used in two modes:

- When in the process of building a clique of mutually consistent lines the result is a subset of {model(T(k))} for $k \in V \cap V'$, the model matcher could be asked to extend the clique so as to disambiguate the configuration, in other words, to find a line model(T(x)) with x not in $V \cap V'$ to complete the clique.
- An other way of looking at it would be to say: if the

3-D match of the scene-based on the matching of {scene(k)} to {model(T(k))} is rejected by the verifier then try the 3-D matching of {scene(k)} to {model(T(k))}.

4. Experiments

Experiments were carried out using an "L" shaped object as a model, and stereoscopic images containing a "widget", Lego house or the "L". These images represented the object seen from various viewpoints and lit in various conditions. Some images included obstructing objects. The results of experiments were displayed by superimposing on the raw images the projected model lines, the colour of which indicated their category (1 to 6). Scene lines that were interpreted as occlusions or as geometrical inconsistencies were also outlined. In the figures presented below, dotted lines represent scene lines and plain lines represent model lines.

Image showing a Lego house:

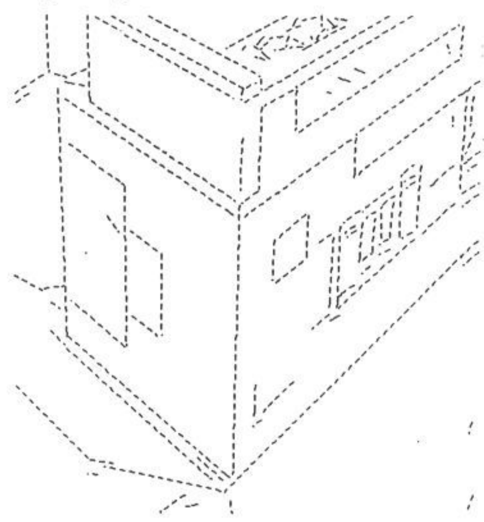
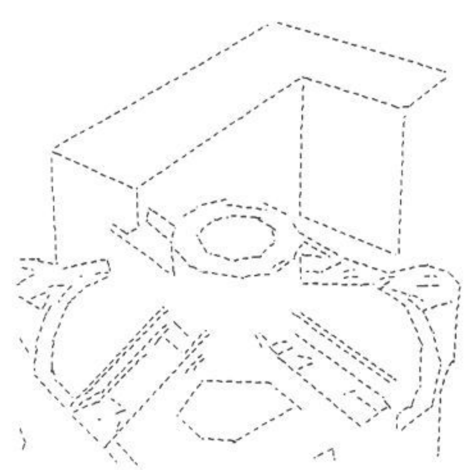


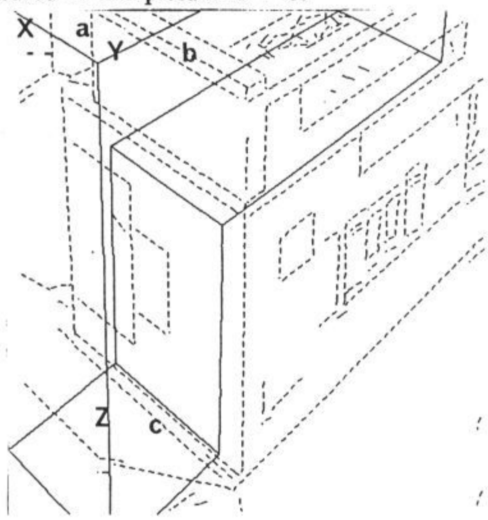
Image showing the "L" partly occluded by an object:



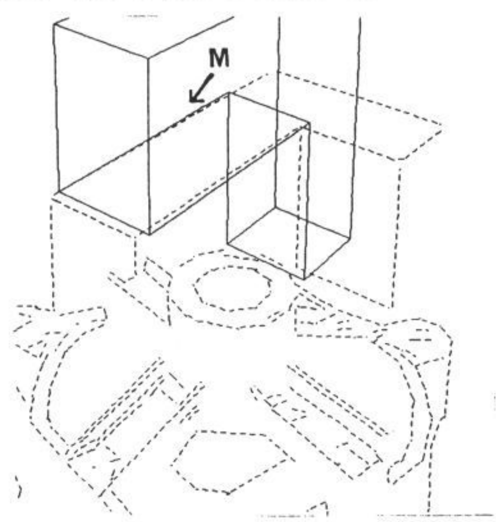
The 3-D matcher never succeeded in matching the widget to the "L".

Some settings of the matcher's parameters allowed the matcher to match the house to the "L". All matches were later discarded by the verification process, as the value of criterion 2 was never more than 10 %. In most cases the rejection was corroborated by the detection of geometrical inconsistencies.

In the figure below, scene lines a, b and c were detected behind model lines X,Y and Z respectively, which discards the hypothesised 3-D match. The value of criterion 2 measured in this picture is 4 %.

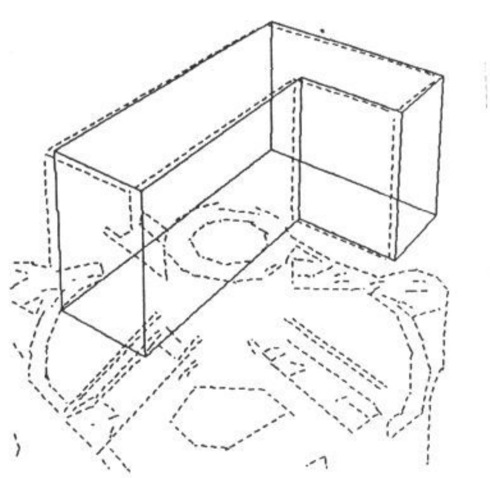


The "L" being a highly symmetrical object, in images where it appeared the matcher often yielded matches that corresponded to location errors. Such an error can be seen in the following picture. Here the 3-D match is discarded as model line M, which was predicted invisible, has been matched successfully in the left and right images. The value of criterion 2 is 24 %.



All correct matches were validated by the verification system.

The following figure illustrates a correct match on image where the "L" was partly obstructed by an object. The value of criterion 2 was 65 % and criterion 3 was 80 %.



In 5 % of the cases it also validated wrong matches (a wrong transformation was computed on the right object) as the quantitative assessment was considered high enough and no geometrical inconsistencies were identified.

In many cases, matches were rejected by the verification system on grounds of geometrical inconsistency.

References

Bolles RC, P Horaud and MJ Hannah (1983) 3DPO: A three dimensional part orientation system, Proc. IJCAI 8, Karlshrue, West Germany, 116-120.

Faugeras OD and M Hebert (1985) A 3D recognition and positioning algorithm using geometric matching between primitive surfaces, Int. J. Robotics Research.

Grimson WEL and T Lozano-Perez (1983) Model based recognition from sparse range or tactile data, AI Memo 738, Dept of AI, MIT, Cambridge Mas.

Pollard S B, Porrill J, Mayhew J E W and Frisby J P, *Matching geometrical descriptions in three space,* Image and Vision Computing, Vol5, No 2, 73-78, 1987.

Porrill J, Pollard S B, Pridmore T P, Bowen J B, Mayhew J E W & Frisby, J P (1987), *TINA: The Sheffield Vision System,* Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Milan, 1138-1144.