INTERPRETATION OF 2D SCENES USING A GENERAL RELATIONAL MODEL

P McAndrew and A M Wallace
Department of Computer Science
Heriot-Watt University
79 Grassmarket
Edinburgh EH1 2HJ

ABSTRACT

A general strategy for the interpretation of two-dimensional views of manufactured components is presented. The model database contains as primitives parameterised two-dimensional shape descriptions including feature type, position, and orientation where appropriate. Depending on the extent of each feature either a point or a point vector is associated with each constituent feature within the model or scene. Attributed geometric pairwise relations are formed between relevant primitives. The complexity of the search space of possible scene-model feature correspondences is limited in two ways. Firstly, the pre-formed pairwise relations are used to indicate the probable occurrence of both specific components and specific features. Secondly, a measure of saliency is employed which directs the search to those features which most precisely determine particular components. This is an extension and generalisation of previous work on the matching of two-dimensional descriptions and complements work on low and intermediate level processing of the scene data, based on the use of the generalised Hough transformation.

I. Introduction

When presented with images of an object, or objects, viewed under widely varying lighting conditions, an ideal vision system should be able to identify each object belonging to its database quickly and accurately. Current vision systems tend to fall short of this ideal in two respects restricting both the viewing position and the lighting conditions.

In this paper we concentrate attention on the reliable interpretation of stable two dimensional views of three dimensional objects. The system is designed to be tolerant to some degree of change in geometry such as that caused by laying one object upon another. Scaling between the image and previously stored models is assumed constant (this assumption is justified as the image to camera distance usually will be fixed in industrial situations). This condition

could be relaxed quite simply but the performance of the system would be degraded. Although the current concern is with stable configurations, at a later stage we intend to proceed to the identification of three dimensional objects with minimum constraint on positioning while retaining the restriction of a single two dimensional view. A modified form of the framework which we propose is extensible to the more complex case of arbitrary 2D projection.

Poor lighting will lead generally to poor segmentation of images. A system designed to operate in such conditions must therefore use inexact matching to be tolerant of the resultant inaccuracy in the scene description. For this reason, the matching strategy described here allows for the absence of several expected features and the presence of additional ones arising from the earlier processing of the scene data. A quantised representation of geometrical relations is used, and a range querying process is employed to allow for the expected errors in the segmented description.

This work is an extension and generalisation of that described by Wallace (1987) in which geometrically based relations were used to guide the interpretation of scenes containing single and multiple components. Although that system was successful in identifying objects it had some recognised limitations, namely:-

1. Only straight lines and circular arcs were considered as features.

2. A bias existed towards line pairs in the formation of relations.

3. Acceptance of a match between model and scene data was based on fixed empirical rules.

4. No account was made of the uniqueness and relative importance of specific relations or features.

5. The final hypothesis was not directly confirmed by re-examination of the unsegmented data.

The improved system allows for any feature type to be incorporated easily into the system using a general relation, and is complementary to the work being undertaken elsewhere to extract features from the scene using the generalised Hough transformation (Kittler and Illingworth, 1987). A measure of saliency is introduced to further direct the search for matching features within the database of
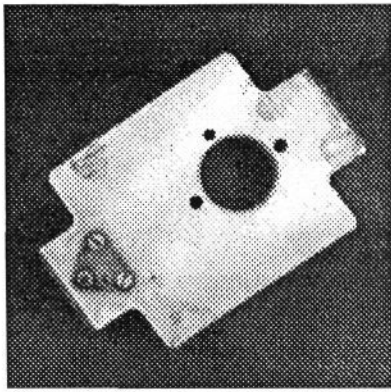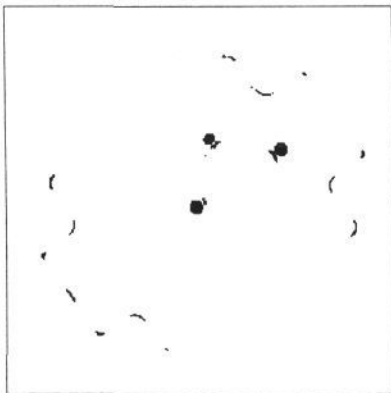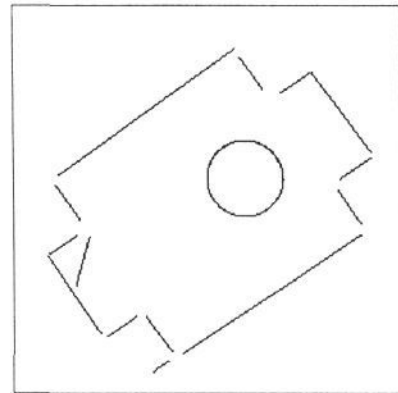
a: Original image
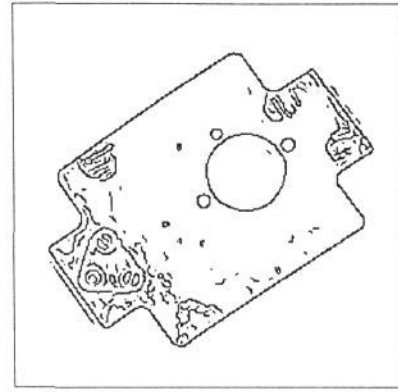
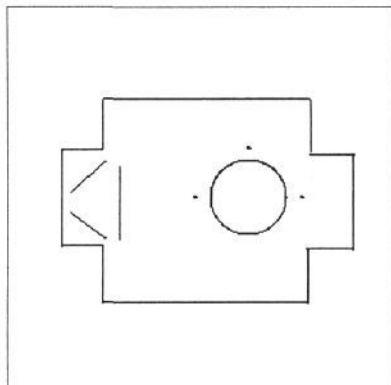b: Line detected (Canny, 1986)

c: Corner detected (Paler, 1984)

d: Line segmented

e: Stored model

Figure 1: Stages in processing an image

processing. The low and intermediate level processing is designed to locate features within the image by such means as edge detection, Hough transformation and segmentation. Figure 1 shows the effect of these operations being performed in sequence on a raw image. Increasingly static-to-static low level data transformations such as edge detection, and to a lesser extent static-to-dynamic data transformations such as Hough transformation are being performed in hardware using discrete, semi-custom or custom design techniques with the result that greater emphasis is placed on the design of software algorithms applied to segmented images in forming hypotheses.

II. Representation of Image and Model Segments

A. Representation of Single Features

The information about a feature available from segmentation may vary in complexity from coordinate and type to an arbitrarily complex shape. The full information for each feature is entered into the database but for the

models. As a further extension, we also intend to include direct confirmation of the presence of an object by correlating small, previously identified, sections of an image of the model against that section of the image where the part should lie.

In an overall system context, the processing may be considered as falling into low, intermediate and high level

NORMAL

MID-POINT

a: Straight line

MID-POINT

CENTRE

b: Circular arc

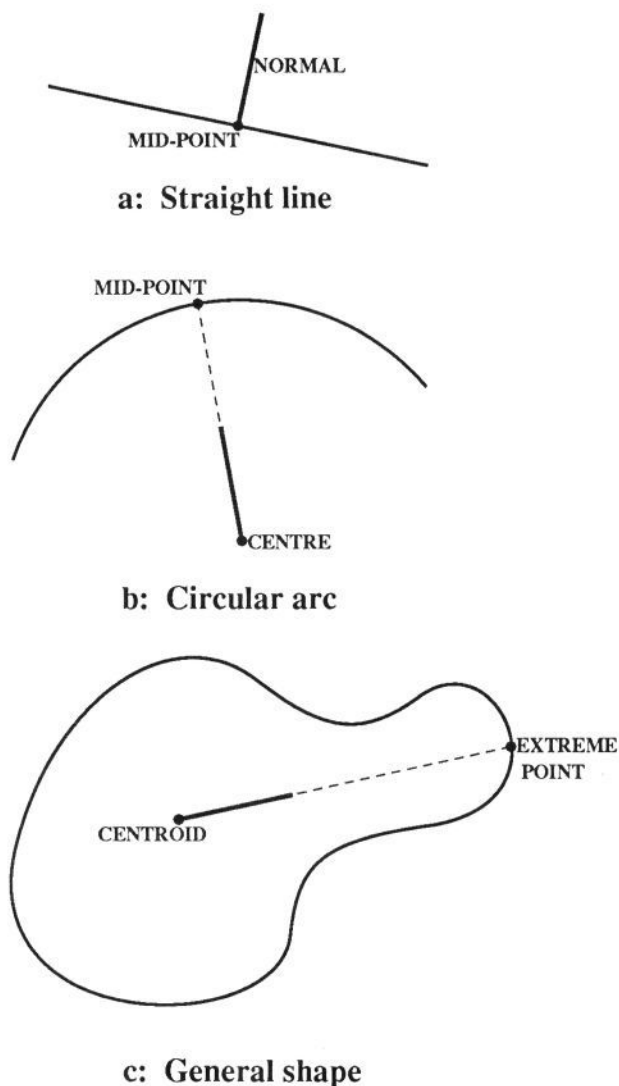EXTREME
POINT

CENTROID

c: General shape

Figure 2: Example vector representations

matching process single features extracted from the scene data are classed as one of two basic types; either an attributed position vector or an attributed point. A feature described by a position vector is defined by a central point and a direction. Examples of vector features include straight lines, described by their midpoint and normal at that point, circular arcs and higher order curvilinear features, described by the centre of curvature and direction from there to the midpoint, and closed contours, defined by a central point and an orientation vector. Illustrations of each of these exemplary categories are included in Figure 2. Examples of point features include corners, small holes and surface markings. Exceptionally, a direction vector and position can be associated with a feature of this type, for example using the bisector of a corner, to form a relation in the same manner as the midpoint and tangent direction of a linear feature.

If regular shapes such as squares, equilateral triangles and circles form part of the database, allowance should be made for the ambiguities caused by the symmetries which do not lead to a unique reduced representation. These cases may be most easily handled by allowing multiple entries in

the database — one for each symmetry. The particular case of a circle is solved by considering circles as point features with the position given by the centre point.

B. The Use of Pairwise Relationships

In general, pairwise or higher order relations between features may or may not have attributes. Examples of the latter category include such relations as "above" or "connected-to" when, for any ordered pair of features, the relation either holds or does not hold. This type of approach has been proposed by various authors (e.g. Henderson and Samal, 1986, and Shapiro and Haralick, 1985), and leads by necessity to a discrete matching strategy, which is, in our opinion, intolerant of imperfections in the segmented data. Alternatively, relations between features may be attributed, but may yet not hold between all pairs of features. An example of this was the "connect" relation which was used by Wallace (1987), which held as attributes the type of intersection and length of the intersecting features.

Although relations may be formed between any number of segmented features pairwise relations are considered exclusively in this work. These relations are formed on the reduced representation of each feature and may be of three possible types, illustrated in Figure 3.

1. Between point vectors: in this case the relation is a triple consisting of the distance between the points, the angle between the point vectors, and the normal distance from one of the points to the other vector.

2. Between a point vector and a point: two attributes are stored, the perpendicular distance from the point to the extended vector and the distance between the points.

3. Between points: only the distance between the two points is recorded.

The use of geometric pairwise relations in initial hypothesis formation is more robust when the image is degraded than the use of single feature attributes, and provides a rotation and translation independent method of indexing particular model and individual feature correspondences. By using a cellular structure to index model features on the basis of attributed relations a rapid match between scene and model segments is obtained (Wallace, 1987). The number and type of such matches located for each model can then be used to indicate the preferred model. Following choice of a model hypothesis, a similar process can be used to indicate which features belong to that model (rather than be part of some other object in the same image). Using this system to order the models and features, and so direct the search, the feature to feature matching necessary is reduced and total time spent in reaching a conclusion should be less.
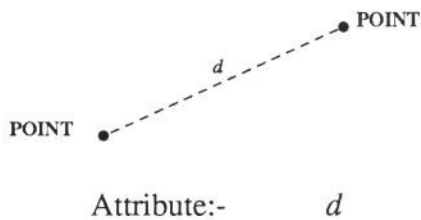
Pairwise geometric relationships have been used by a number of authors in 2D and 3D (Ayache and Faugeras, 1986, Grimson and Lozano-Perez, 1985, Dhome and Kasvand, 1986, Pollard et al, 1987, Murray, 1987) to limit
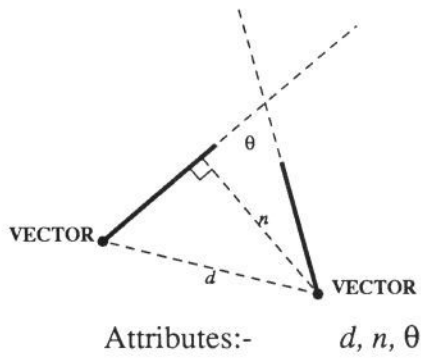
Attributes:- d, n, θ

a: VECTOR-VECTOR relation
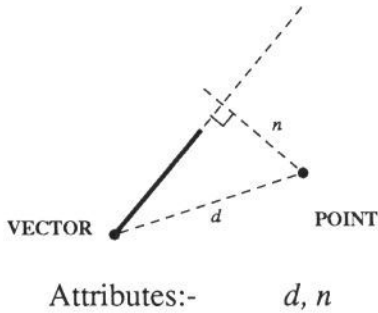
Attributes:- d, n

b: VECTOR-POINT relation

Attribute:- d

c: POINT-POINT relation

Figure 3: Three forms of the general relation

the combinatorial explosion of matching all scene to model features. Several optional attributes for pairwise relations were considered in this work, but the representation of Figure 3 possesses several advantages, summarised below. Firstly, for each pair of features extracted from the scene or model data, a single pairwise relationship exists. Secondly, as information is lost by replacing vector representations with point representations the relations lose an attribute but are otherwise unchanged. This brings practical advantages in organisation of the database and query structure and allows cases where both representations are valid, such as arcs which may form complete circles. Thirdly, the attributes are selected to provide at least one stable measure for cases involving vector representation. The formation of vectors from straight lines will, in general, lead to accurate prediction of the vector direction but, where lines are fragmentary, uncertainty in the midpoint position. Conversely, when circular arcs are being considered the position of the point corresponding to the centre of curvature should be stable but the associated direction may be unstable. The knowledge that one attribute is more stable than another can be used subsequently in deciding the range query for relation comparison — allowing more or less

leeway in matching of an attribute according to the case. Finally, the general principle is extensible to the 3 dimensional case in which case normal vectors are drawn from the centre of area of planes, or from the centre of curvature through the centre of area of a sphere, for example, and relations are formed which represent the transformation and rotation required to align these vectors.

III. The Incorporation of Saliency

When building a database of features and the relations between them it is found that, for a typical group of manufactured objects, there will be relations common to several objects, relations common to few objects, and relations unique to a particular object. This data can be used to provide a measure of saliency for the relations in the database and, by examining the features involved in those relations, a ranking on the saliency of the features.

This is essentially an offline process performed on the collection of models before the data from the test image is entered into the system. Using the scheme described in Appendix I the relations within the database are examined, and the number of models sharing relations in the neighbourhood of each relation found. The relations can then be ranked according to this number; in the current implementation this results in a partitioning into salient and non-salient relations.

Previous applications of a saliency measure are described in two papers by Turney, Mudge and Volz (both 1985). In Reference 12 shapes are divided into "subtemplates" which are particular line shapes occurring on the boundaries of the objects. Each of these is weighted by performing a cross-correlation in θ-s space with all other subtemplates. In Reference 13 unique pairs of these subtemplates are sought for each model and, if found, the model judged to be within the image.

This application depended on the subtemplates each having a distinct shape. The approach used here differs in that the features in the database are assumed to be only distinguished by class and attributes within each class. With the further assumption of the possibility of poor segmentation it would be unreliable to base a measure of saliency on those attributes. In addition, when dealing with views of more than one object, accidental alignments will be more common and the appearance of only a few salient features cannot be used as a sound basis for the decision as to an objects presence. The saliency measure can be used instead to guide the search process.

An idea of saliency has in some cases (Ayache and Faugeras, 1986, Pollard et al, 1987) been attached directly to features on the basis of length. This choice of attribute seems somewhat arbitrary, depending on a preconceived notion that the longest features will be the most important. This clearly need not be the case as it may be only in the shorter features that similar objects differ. Measurement of length may also not be stable as fragmentation of linear features may occur in the poor images considered.

Henderson and Samal (1986) also used a approach based on saliency to index preferred models, but based on point features (e.g. holes), and interfeature vectors drawn between them. Two quantities were employed, the magnitude of individual feature vectors and the angle between two interfeature vectors. These lengths and angles were formed into a histogram (using some quantisation) and those with a low number of occurrences considered most significant and thus sought out in the early stages with later stages proceeding conventionally.

The saliency ratings described in this paper are used in two ways to guide the search process. Firstly, the total number of shared salient and nonsalient relations are used to decide the most likely model to attempt to match with the image. Secondly, when attempting to match model features to image features, the features are given an order of priority, again based upon the shared relations.

Thus features which provide a large number of unique relation matches are preferred initially. This should lead to the inclusion of feature pairs unique to the model during the early stages of matching increasing the likelihood of correct identification.

Initially the salience figures were used to guide a depth first graph search, but this proved expensive in cases where the initial guiding was not accurate, although in other cases the correct solution would be found very quickly — sometimes without exploration of any incorrect paths. This algorithm was replaced with a version of the maximal clique algorithm (Reingold et al. 1977). The cliques consist of nodes containing model to image feature matches with arcs between nodes when the relation between the model features in each node are judged to be similar to the relation between the image features in each node. Here the saliency information, applied to the features from their involvement in unique relations, is used to preorder the assignments between model and image features. This approach together with limits applied to the size of maximal clique sought serves to guide the expansion. It is sometimes claimed (Bolles, 1979) that the cost involved in making particular selections of nodes in building cliques outweighes any advantage gained in reducing the expansion. This statement may be true in general but in the case considered gains can be made. By only seeking sufficiently large cliques, rather than all cliques, the presorting of the data that can sensibly be applied due to the saliency calculations will limit fruitless expansion. This limitation of expansion is important in this case as the computational cost of the adjacency calculation is relatively high. This consists of removing from the current list of possible matches those which do not share relations with the node just added, only those whose model and image features relate similarly to the latest model and image feature will be retained. The matching strategy used need not be considered fixed — it should be possible to make use of the saliency information within almost any search system. In particular this information could be used to assign initial probabilities for an application of probabilistic relaxation and, as was initially implemented, a graph search can be used. In this implementation the technique of maximal cliques does offer the advantage over the graph search of reducing duplication when significant expansion is required.

This procedure will in some cases produce a sufficient number of matches to establish the presence of an object. More commonly, several cliques of the same, largest, size will be found. These then form the the the basis for hypotheses of a match with the necessary transformation (rotation and translation) being calculated initially on the basis of the already matched features. Features can then be added, if the transformed model feature coincides with a feature in the image, and the transformation updated. At this stage allowance is made for a feature from the model to match to more than one feature in the scene data; this is necessary to allow all parts of fragmented features to be matched to the original feature stored in the model database and be pruned from the data should a successful model match occur. A final best match is selected firstly using the new number of feature matches found and secondly the standard deviation of the transformation parameters required when matching the features. A quality measure similar to that used by Ayache and Faugeras (1986) is then calculated by summing the length of model segments matched and normalising by the total possible length. If this quality measure exceeds a threshold the match will be reassessed as successful if necessary.

The introduction of this two tier check of matching was found to increase significantly the reliability of the matching particularly in allowing both simple objects and relatively complex objects to be found without changes in parameters.

IV. Local matching as confirmation

Following application of the search procedure described above, if an object from the model database has been identified and located within the image a verification process can be applied. This makes use of the local detail which otherwise would be absent from the simplified, feature based, models stored in the database.

The search process will have produced an estimate for the location of the object from which the position and orientation of the local detail within the image can be predicted. A stored template of the local detail can then be directly compared with the original image data in a region around the predicted position. The matching is determined on the basis of a normalised direct correlation between the image and template over a search area and range of orientations around the predicted values. As the search area should be small this process can be performed rapidly. This use of a direct check is similar in some respects to that reported by Gregory and Taylor (1984) where specific features were confirmed by performing comparisons between one dimensional intensity profiles of the model and scene data.

This mechanism provides support for the decision making process and is used to confirm decisions made by the experimental system. Currently, the identification and storage of the local template is manual, for a more practical

implementation this would need to be automated and integrated within the system.

V. Experimental results

As an illustration of the performance of the system the programme has been run on a database of eleven objects shown in Figure 4. These objects were initially hand segmented to provide a database for testing. It should be noted that it is equally possible to use automatic segmentation on good images of the objects to form the model database. Images of each of the objects were then presented to the camera, captured using a CRS 4000 frame-grabber at a resolution of 256x256 pixels, 256 grey-levels, and automatic segmentation performed. These segmented images were then processed by the recognition system implemented in C on a Sun 3/160C workstation. The results are summarised in Table 1. The parameters used to determine matches were a clique size of seven or a quality measure of 60%. The failure to match occurred for a curved object, the eye bracket (eye_br), which could not be accurately segmented into the line and arc segments used. The particular component possessed no sharp changes in surface curvature so that no distinct boundary could be detected in the grey level image.

Average time	4.3s
Successful on clique size	5
Successful on quality	5
Failed to recognise	1

Table 1: Results on single objects

A second, more complex, example is shown in Figure 5 which shows a bracket, identified as anadex4, and a clamp. The histogram (Figure 6) shows that on the basis of a normalised count of the saliency weighted relations the anadex4 is preferred and in fact is matched subsequently. A similar histogram built on the pruned data then selects and correctly matches the clamp. The total processing time to match both objects in this image was approximately 10 seconds after segmentation, using the same parameters as before.

VI. Conclusions

The use of geometrical relations, as opposed to topological relations, has become more common in recent research, in particular where practical analysis of degraded scenes is attempted. Compared to previously published work the approach described here offers several advantages, specifically pre-organisation of the model database, the use of generalised features to represent a wider range of possible objects, and the use of the saliency of those features to order the likelihood of firstly the presence of a particular component, and secondly to guide the matching of specific features.
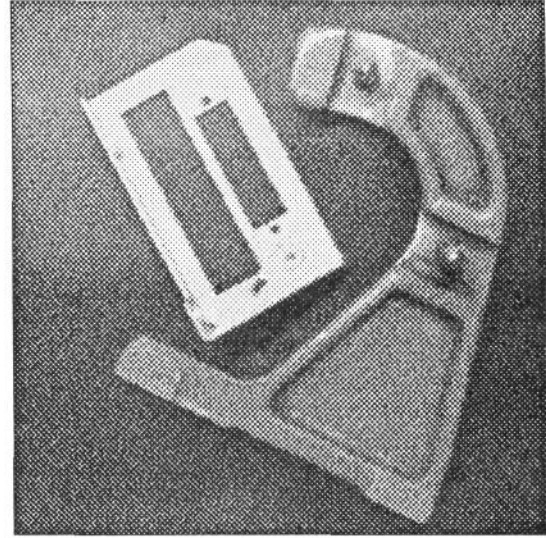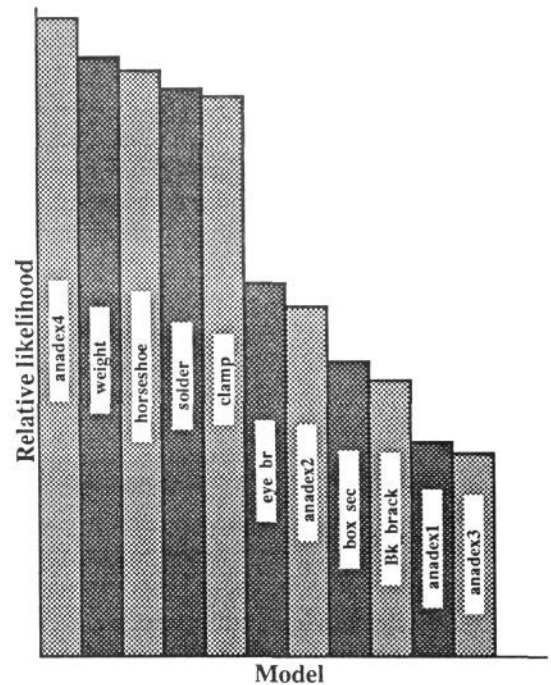

Figure 5: View of clamp and Anadex4


Figure 6: Ordering of models for first match

Assuming the existence of a database of models it seems desirable to perform as much processing as possible on the database before presenting any image data. This can be used to organise the data into the form most useful for rapid matching; calculation of relations, formation into a quantised structure for fast querying and the attachment of
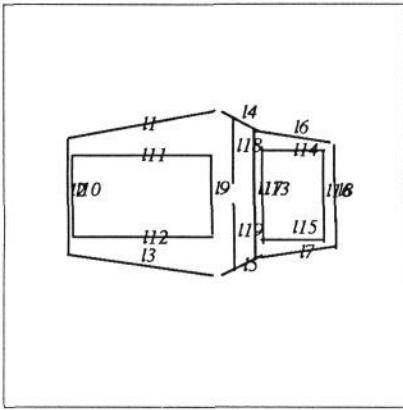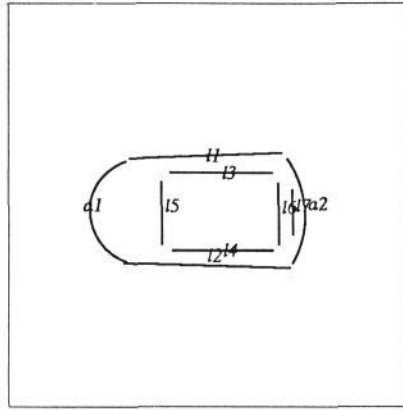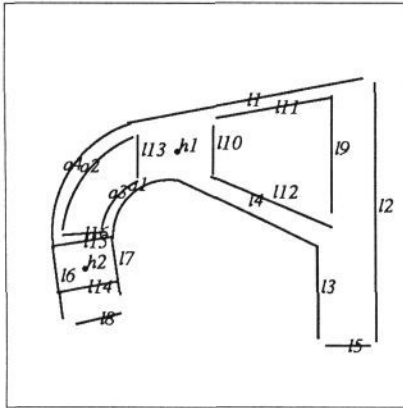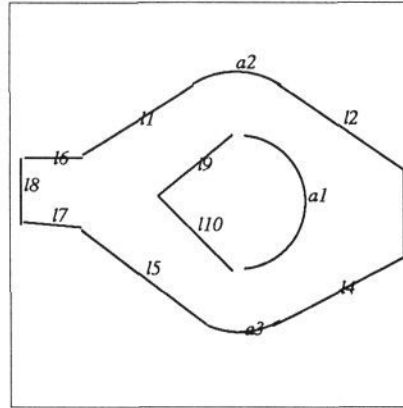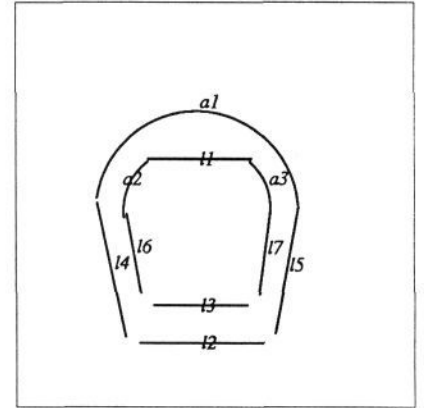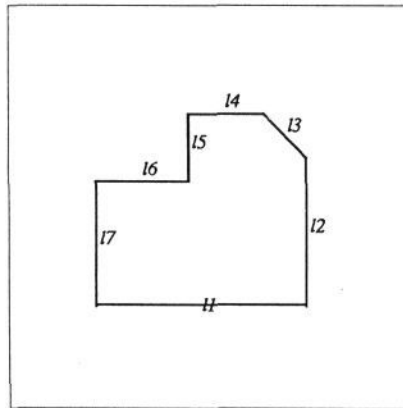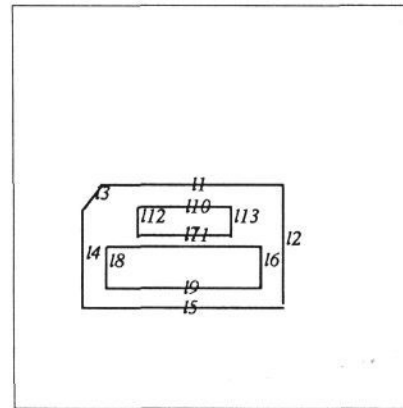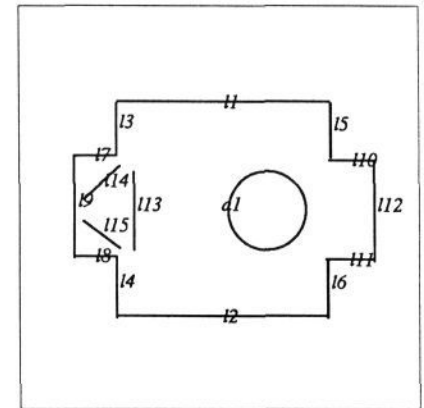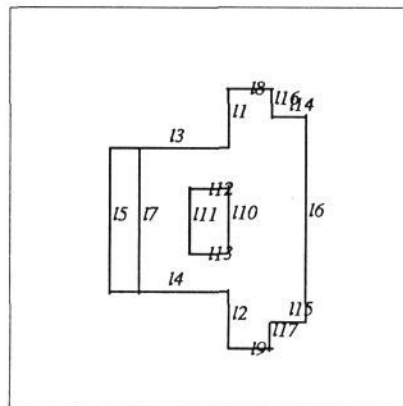
Figure 4: Models contained in the database

113

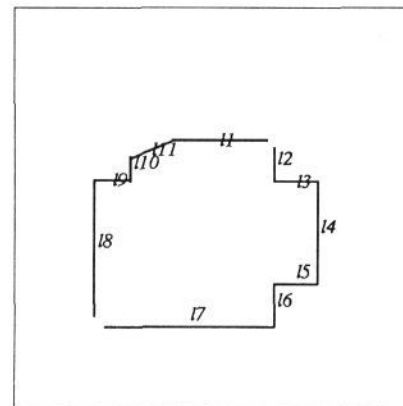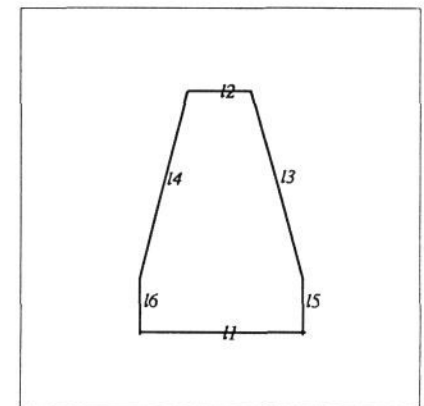saliency to relations by cross comparison within the database all fall into this category. The gains from using these relations to order models are hard to quantify but the actual cost following presentation of the data is minimal — calculation of the preference for each model being performed during formation of the relations to be found in the image.

The use of the generalised feature, whereby each feature is reduced to a simplified form by associating a point and (where possible) a direction, allows the inclusion of a wide range of segmented features within the matching process. Such features may be produced by a general or specific Hough transformation or an alternative intermediate strategy. Pairwise relations can then be formed between any two features regardless of their type. This simplification does however imply a cost as features which differ may possess the same reduced representation. This cost will increase with the complexity of the original feature, for example in the system as currently implemented both points and lines are represented completely but concentric arcs cannot be distinguished; two arcs which differ only in radius will be identical when reduced to point vector form. Although this may lead to incorrect feature matching this does not imply an incorrect model match as the reduced representations will be indistinguishable so that either the right or wrong choice of feature match support the correct conclusion equally.

The use of relations as a basis for initial searching allows interpretation of images having significant degradation caused by noise and visual clutter. These relations can also be compared between the models in the database to produce a measure of their importance or saliency. By calculating the saliencies of the relations in which a feature is involved this technique allows a salience measure to be attached additionally to features — without a requirement for complex, easily distinguishable, features. The saliency attached to the relations can be used in the selection of a model from the database for comparison to the scene data. The saliency of the features is then used to control the expansion of the search process, in this case a maximal clique algorithm, towards the most likely sets of feature matches and to limit the search process.

The presence of a particular feature in both the model and image may be confirmed by direct correlation to provide a check on correct performance. When this is integrated within the system it should reduce the inherent weakness of feature based systems to fail to consider explicit non-edge based information present in the original image data.

APPENDIX I   The matching process

The set of models is defined

$$M = \left\{ M_1, M_2, ..., M_m \right\}$$

The set of sets of features is defined

$$D = \left\{ D_1, D_2, ..., D_m \right\}$$

where model M_i has features D_i ($i = 1, ..., m$), and

$$D_i = \left\{ d_{i1}, d_{i2}, ..., d_{in_i} \right\}$$

where d_{ik} is a particular feature possessed by M_i.

Then R is an attributed pairwise relation and r is a relation mapping defined from the domain D_R, the union of all possible feature pairs, to A, the range of all possible attribute values, i.e.

$$r : D_R \to A$$

Then r_i defines a mapping within the i th model

$$r_i : D_i \times D_i \to A$$

For image I with the feature set

$$F = \left\{ f_1, ..., f_{n_m} \right\}$$

the possible matches within model M_i to a particular related pair of features (f_j , f_k) are given by:

$$r^{-1}(r(f_j , f_k)) \cap (D_i \times D_i)$$

In order to apply the principle in practice the matching has to be tolerant of inexact matches. This can be achieved by quantising the attribute space so that nearby points are considered equal and by taking the inverse of a neighbourhood of $r(f_j , f_k)$ rather than $r(f_j , f_k)$ itself. The size of this neighbourhood is controlled in two ways. Firstly, the quantisation of the system and the default size (typically half the quantisation in each dimension) are selected according to the known repeatability of the segmentation system. Secondly, if insufficient matching is found to allow a reasonable hypothesis to be formed, the image is assumed to be degraded and conditions for match further relaxed and the process repeated.

REFERENCES

1. Wallace, A.M. "Matching segmented images using pairwise relationships between features", *Image and Vision Computing*, 5:2, pp114-120, May 1987.

2. Kittler, J. and J. Illingworth "A Survey of the Hough Transform", Paper to be presented at this conference.

3. Canny, J. "A computational approach to edge detection", *IEEE Trans. PAMI*, 8:6, pp679-698, 1986.

4.   Paler, K., J. Foglein, J. Illingworth and J. Kittler "Local ordered grey levels as an aid to corner detection", *Pattern Recognition*, 17:5, pp535-543, 1984.

5.   Henderson, T. and A. Samal "Multiconstraint shape analysis", *Image and Vision Computing*, 4:2, pp84-96, May 1986.

6.   Shapiro, L.G. and R.M. Haralick "A metric for comparing relational descriptions", *IEEE Trans. PAMI*, 7:1, pp90-94, 1985.

7.   Ayache, N. and O.D. Faugeras "HYPER: a new approach for the recognition and positioning of industrial objects", *IEEE Trans. PAMI*, 8:1, pp410-421, 1986.

8.   Grimson, W.E.L. and T. Lozano-Perez "Recognition and localization of overlapping parts from sparse data in two and three dimensions", *Proc. IEE Int. Conf. Robotics and Automations*, (St. Louis, 1985), pp61-66.

9.   Dhome, M. and T. Kasvand "Hierarchical approach for polyhedra recognition by hypothesis accumulation"", *Proc. 8th Int. Conf. on Pattern Recognition*, Paris, October 1986, pp88-91.

10.   Pollard, S.B., J. Porrill, J.E.W. Mayhew and J.P. Frisby "Matching geometrical descriptions in three-space", *Image and Vision computing*, 5:2, pp73-78, May 1987.

11.   Murray, D.W. "Model-based recognition using 3D structure from motion", *Image and Vision computing*, 5:2, pp85-90, May 1987.

12.   Turney, J.L., N.M. Mudge and R.A. Volz "Recognizing partially occluded parts", *IEEE Trans. PAMI*, 7:4, pp410-421, (1985).

13.   Turney, J.L., N.M. Mudge and R.A. Volz "Recognizing partially hidden objects", *Proc. IEE Int. Conf. Robotics and Automations*, (St. Louis, 1985), pp48-54.

14.   Reingold, E.M., J. Nievergelt and N. Deo "Combinatorial algorithms: theory and practice", Prentice-Hall, 1977.

15.   Bolles, R.C. "Robust feature matching through maximal cliques", *SPIE Vol.182, Imaging Applications for Automated Industrial Inspection & Assembly*, pp140-149, 1979.

16.   Gregory, P.J. and C.J. Taylor "Knowledge based models for Computer Vision", *Proc. of 4th Int. Conf. on Robot Vision and Sensory Control*, (London,1984), pp325-330.