# OBJECT RECOGNITION FOR A FLEXIBLE MANUFACTURING SYSTEM

Y. Huang & J. C. Trinder & B. E. Donnelly

School of Geomatic Engineering
The University of New South Wales
Sydney NSW 2052, Australia

ISPRS Commission V, Working Group 3

**KEY WORDS :** Edge, Extraction, Object, Reconstruction, CAD, Model and Identification.

## ABSTRACT

3D object recognition is a difficult and yet important problem in computer vision. It is a necessary step in many industrial applications, such as the identification of industrial parts, the automation of the manufacturing process, and is essential for intelligent robots equipped with powerful visual feedback systems. In this paper, a complete procedure is described to recognise 3D objects, using model-based recognition techniques. Objects in the scene are reconstructed by digital photogrammetry, while models in the database are generated by CAD system. A detailed comparison between the potential matching graphs of an object and a model determines the identification of the sensed object, its position and orientation.

## 1. INTRODUCTION

Digital photogrammetric procedures of machine vision are being investigated for their application in a flexible manufacturing system (FMS). Flexible manufacturing enables multiple products to be fabricated on a single assembly line under computer program control. The system is managed by work transfer robots which are required to recognise objects, as they pass along the assembly line, and to determine the next appropriate action that should be taken on them. For the recognition of objects, it is necessary to extract visible features on multiple digital images of the object by image analysis procedures. These features form the basis of the reconstruction of the objects in terms of 3-dimensional geometric primitives. This representation of the object is then compared against entities in a model database, which contains a description of each object the system is required to recognise.

The development of such model-based recognition techniques has occupied the attention of many researchers in the computer vision community for years (Besl and Jain, 1985; Chin and Dyer, 1986; Brady et al., 1988; Fan, 1990; Flynn and Jain, 1991). Many machine vision systems developed so far have been mostly based on range images which contain direct 3D properties of objects. Using range images, the ambiguities of the feature interpretation which usually occur in an intensity image, such as shadows, surface markings or illumination, are eliminated. However, an intensity-based vision system is still acceptable not only because of its relevance to biological vision but also because of the robustness of passive sensing for industrial and other applications. There are a number of advantages in the use of intensity imaging system, including: the intensity data is viewable by an operator and can reveal more than geometric information, eg. colour, texture, blemishes; features such as edges and faces can be extracted from the object by image processing, provided that these features are apparent in the image; lighting can be varied to accentuate various elements in the object.

One problem of object recognition is related to the representation of models in a database and objects in scenes. The representation of models should be compatible with the description of the sensed object, so that the matching of elements from models and objects can be identical. One can match objects with models at many different levels or descriptions with some tradeoffs: the lower the level of a description, the easier it is to compute them. However, such a description is not invariant to viewing directions, which makes it difficult to find correspondence between objects and models. The higher level descriptions, on the other hand, maintain their invariance but the known algorithms to compute them are often weak and error prone (Fun, 1990). The appropriate level of description to be used for matching, thus depends on the expected variations in the scenes and on the state-of-art in computing descriptions of models.

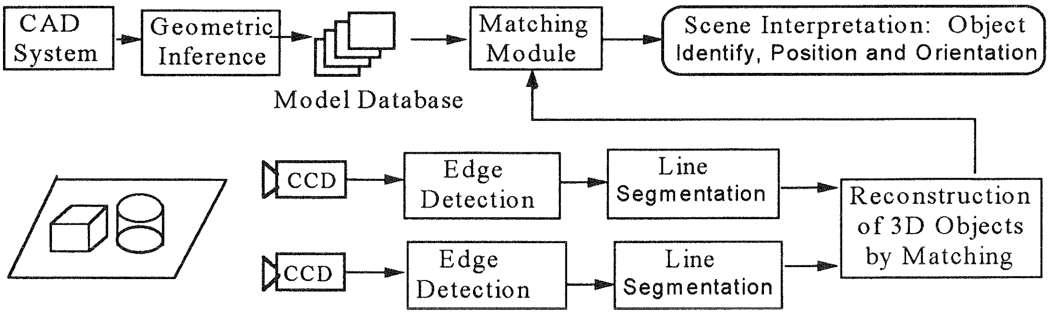

Figure 1 : Components of an object recognition system

In this paper, a relational graph representation, in terms of object boundaries, is developed. The graphic representation is described by planar surfaces which are bounded by straight lines or regular curves. These surfaces are grouped in terms of their normal directions and are stored together with their areas and perimeters as matching elements. The topological relations between surfaces are constructed in terms of the centre of each surface and the common edge of two surfaces. The automatic procedure for a machine vision system for FMS includes three steps: image processing techniques for the extraction of features on the industrial components and hence the 3-dimensional measurements of these components; representation of the 3-dimensional objects in an efficient and convenient data structure; artificial intelligence procedures for the recognition of the objects by comparing with models stored in the design database. Figure 1 illustrates a schema of the procedure by means of which objects in the scene can be reconstructed and recognised from digital stereo images.

## 2. RECONSTRUCTION OF OBJECTS IN THE SCENE

Images of industrial components are generally characterised by sharp discontinuities which represent features on the objects. In order to describe features by their boundaries and to use feature-based image matching in photogrammetry, it is necessary to extract the complete details of linear features of the object from its edges, and to represent them in a suitable data structure such as straight lines and smooth curves. By matching these geometric features in stereo images, the 3 dimensional geometry of objects can be reconstructed.

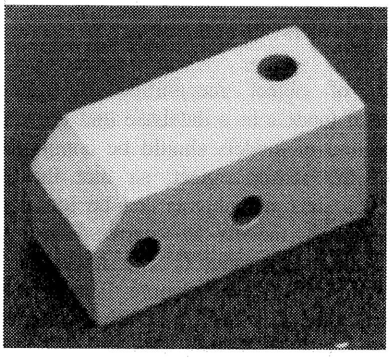

Figure 2 : An image of a block

### 2.1 Edge Detection

Edge detection is low level image processing, which serves to simplify the analysis of images by drastically reducing the amount of data to be processed, while at the same time preserving useful structural information in images. The edge detection method (Trinder & Huang, 1993) used in the research is based on a linear model which locates an edge point in the operator window with subpixel accuracy. The linear model comprises two aspects: one is to determine the peak of the intensity change in gradient direction, which is performed by the Förstner Operator (Förstner, 1987); the other is to limit the unstable edge location in the edge direction by the introduction of a linear constraint which passes through the window centre and meets the edge at right angles.

To improve the accuracy of the edge location, an edge point is determined by a weighted average of two points derived from both sides of the edge. Another implementation is to use a round operator window instead of a square window, so that the result will not be influenced by the difference of edge orientations. The scattered edge points are then chained by following neighbouring pixels, based on the minimum local distance. The direction of the edge chaining corresponds to the local direction of the edge, which is attached to each edge point as a primitive for the succeeding process. Figure 3 displays the edges detected from the image of an industrial component shown in figure 2. The edges can be located with a precision of 0.1 pixel, depending on the contrast of edges. Generally, high contrast, which reduces the influence of noise, results in high precision of edge location.

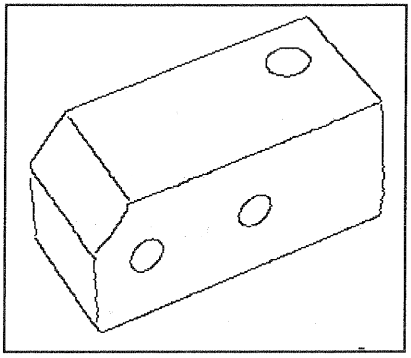

Figure 3 : Edges on the block

### 2.2 Line Segmentation

In industrial environments, regular shapes such as ellipses and straight lines, often occur as elements of object boundaries. In order to interpret objects in the scene, it is generally more relevant to present the boundaries of objects revealed in the images in geometric form, because simple geometric functions provide more reliable information and are easier to calculate.

The suitability of the approach developed in the research for line segmentation is demonstrated in (Huang & Trinder, 1994), based on the analysis of the local directions of edges. The local direction of an edge point is a 1-D value which clearly reflects the trend of the edge at the edge point with respect to the next point. If a group of edge points contain the same trend in edge direction, it means that these edge points are of similar geometry. The method attempts not to find corners, but directly to find the basic components of straight lines or regular curves by the assessment of the local direction at the edge points along a complete edge. The principles of the line segmentation are: all edge points on a straight line should indicate approximately the same edge direction; and all edge points on a regular curve should indicate the same sign of the difference in direction.

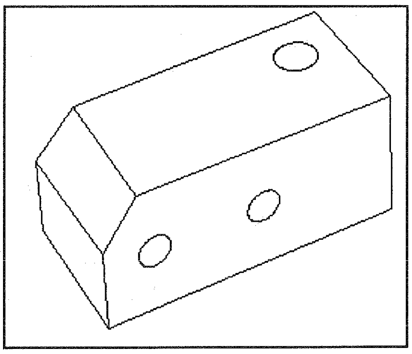

Figure 4 : Results of line segmentation

The method starts with finding basic components of straight lines or regular curves by the assessment of the edge direction at the edge points along a whole edge. Straight lines or regular curves are then extended to their end points using geometric parameters determined by their components. Those lines with the same properties extracted from different edge sections can be merged in terms of their geometric similarity. In order to close boundaries of objects in image space, straight lines and open regular curves are linked at their terminal points. Finally, surface patches are generated by constrained chaining the geometrical lines. Figure 4 illustrates the results of line segmentation.

## 2.3 Stereo Matching

Using two or more CCD cameras, 3D surfaces of an object can be constructed by matching straight lines and ellipses. The reliability of the matching is confirmed by epipolar geometry, based on the camera orientation. The correspondence of a pair of straight lines in stereo images is determined in terms of their terminals, which should satisfy the epipolar constraint. A 3D straight line is simply presented by 3D coordinates of its two terminals, using ray intersection.

To find the correspondence of regular curves, it is necessary to consider their size and location. An epipolar line, which is tangential to a given ellipse in one image, must be tangential to a matching ellipse in the other. This condition includes the basic requirement for the size and location of matching ellipses. The matched elements of an ellipse can be presented as a straight line with two tangent points at its ends as shown in figure 5. In a similar manner to matching straight lines, the ellipses being matched should satisfy the condition that their corresponding tangent points lie on the same epipolar line.

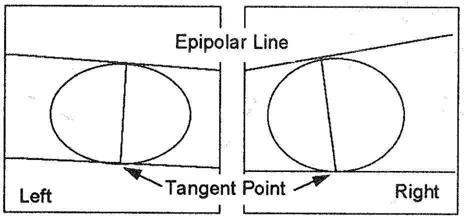

Figure 5 : Matching of an ellipse

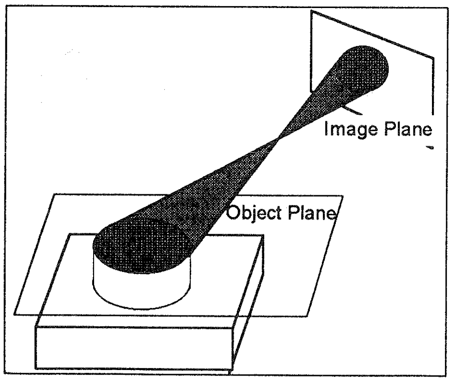

Figure 6 : Intersection of a conic surface with planes

The calculation of 3D ellipses is based on the assumption that a special ellipse is an intersection of an object plane with a conic surface, whose apexes are at the projection centre. An ellipse in an image can also be referred to as the intersection of an image plane with the conic surface, as shown in figure 6. The parameters of an ellipse on two different planes can be transformed, if the relations of the two planes are known. The determination of an ellipse in object space, therefore, contains two aspects: one is to determine an object plane; the other is to establish the relation between image and object planes.

An object plane is determined by a few intersection points on the corresponding ellipses in the stereo images. The relation between image and object planes is established in terms of the camera orientation. The ellipse on the image plane is then transformed onto the object plane. A 3D ellipse can be described by a plane in object space and the 2D curve parameters on the plane.

## 3. GENERATION OF MODELS IN DATABASE

In an industrial environment, CAD systems are usually used to design objects for the manufacturing task. Automatic generation of the same recognition code for object information used for design and manufacture would be an efficient, cost-effective approach. AutoCAD is a general purpose Computer Aided Design program for preparing two dimensional drawings and three dimensional models. The speed and ease with which a drawing can be prepared and modified using a computer, offer a significant advantage over hand preparation. Using the AutoCAD system, models can be created and output in a DXF file, a drawing interchanging file. CAD models serve as a basic description of object geometry. Inference procedures of various sorts are then applied to the CAD models to produce a graphic presentation in the database for object recognition.

## 3.1 CAD Output: The DXF Format

DXF files are standard ASCII text files, which can easily be submitted to other programs for specialised analysis. Since a DXF file is a complete representation of the drawing database, for the presentation of matching features, it is not necessary to use all information in the file. In the research, the attention is concentrated on that portion of the DXF standard devoted to the description of 3D geometry.

A DXF file is subdivided into four editable section, plus the END OF FILE marker. The HEADER section contains settings of variables associated with the drawing. The TABLES section contains several tables, each of which contains a variable number of table entries. The BLOCKS section contains the entities that make up the blocks used in the drawing, including anonymous blocks generated by associative dimensioning. All basic geometric elements in the sequence design stage of a model are stored in this section. The ENTITIES section contains entity items, which can also appear in the block sections. The appearance of entities in the two sections is identical, but this section provides the final drawing of the CAD design.

A DXF file is composed of many groups, each of which occupies two lines in the DXF file. The first line of a group is a group code, used to indicate both the type of the value of the group and the general use of the group. The second line is the group value, in a format that depends on the type of group specified by the group code. For example, a line is presented by two points. The codes for the coordinates of a start point are (10, 20, 30), and for the coordinates of a end point are (11, 21, 31). Each code is followed by a coordinate value. A program can easily read the value following a group code without knowing the particular use of this group in an item in the file. A DXF file can often specify object geometry in terms of group entities such as: lines, circles, arcs and polylines. The basic geometry of models can be used to construct graphic presentations for object matching.

### 3.2 CAD Models to Graphic Presentation

An ideal 3D representation is unique and unambiguous, and has a rich set of representable parameters. The graphic presentation of models used in this research is constructed by deriving a subset of the basic geometric entities from DXF files. The computational burden of graphic presentation is not incurred at object recognition time, since the transformation of CAD models to graphic presentations need only be applied when a new model definition is created and the corresponding vision object is needed. Each model is handled separately, so that the addition of a model to the database does not change the representations of existing models.

#### 3.2.1 Attributes of Geometric Primitives

The basic geometric elements of object boundaries are stored explicitly in the analytic format in DXF file, in terms of lines, circles and arcs.

1) Line Segments: In the DXF specification, a line segment is characterised by a starting point and an ending point. The coordinates of points are stored in the list of vertices. A line is presented as two numbers of vertex and length.

2) Circular arc: In the DXF file, a circular arc is specified in its own (arc-centred) coordinate system $(x_a, y_a, z_a)$, in which the plane of the arc is parallel to the $x_a y_a$ -plane, and displaced from it along the $z_a$ axis. The direction of the $z_a$ axis is given and related to the world coordinate system. The primitives of an arc contain its central coordinates, z axis direction, radius, start angle and end angle. A circle is presented similarly to an arc without start angle and end angle, while an ellipse is presented by 12 arcs which link smoothly at their ends. This system computes major axis and minor axis of an ellipse from these symmetrical arcs. The attributes for an ellipse are radius (an average of major axis and minor axis), and ratio (major axis divided by minor axis). These attributes are identical with those of circles, where the ratio is 1.0.

#### 3.2.2 Planar Surface and Their Topologic

The inference system does not attempt to present objects in a complete way, but rather dominant features are used for model matching. Planar surfaces are chosen as the main features which are related to each other. In the inference system, planar surfaces are generated from basic geometric elements of object boundaries, which are classificated into two kinds: regular curves (circles and ellipses) and polygons. Each planar surface is presented by the normal direction of the plane ($\alpha$, $\beta$, $\gamma$), its central coordinates and bounded edges. A 3D regular curve for a planar patch is presented by 2D parameters projected on the plane, while a polygon is simply a group of straight lines. Additional primitives of a planar surface are radius and ratio for an ellipse or a circle, and perimeters and area for a polygon. To establish the topological relation among planar surfaces, the surfaces are grouped in terms of their normal directions. The planar surfaces are also related by their common edges and the distances between their central coordinates.

#### 3.2.3 Graphical Presentation of a Model

The graphical presentation of a model can be created from the DXF file. The presentation of each model includes: model name, range of element sizes, and orientation, ellipse, polygon, line and vertical point sections. Model name stands for a model listed in a database, which is followed by the four values for maximum and minimum lengths of lines, and maximum and minimum values of circle or ellipse radius. Since objects are constructed to high accuracy, by comparing the dimension between sensed objects and models, most models whose dimensions are beyond the range of object dimension can be ignored. An orientation section contains the main orientations of planar surfaces, each of which includes the list of ellipses and polygons. Figure 7 displays an industrial model, whose surfaces are grouped in terms of their directions. An ellipse section lists all ellipses whose elements are central coordinates, radius and ratio. The direction of each ellipse is derived from the orientation section. A polygon section contains all polygons whose elements are central coordinates, perimeter, area and the list of bounded lines. A line section lists all lines, where the first two numbers are vertices, and the next numbers are polygons between which the line lies. The final element is the length of a line. A vertical point section lists the coordinates of all vertices.

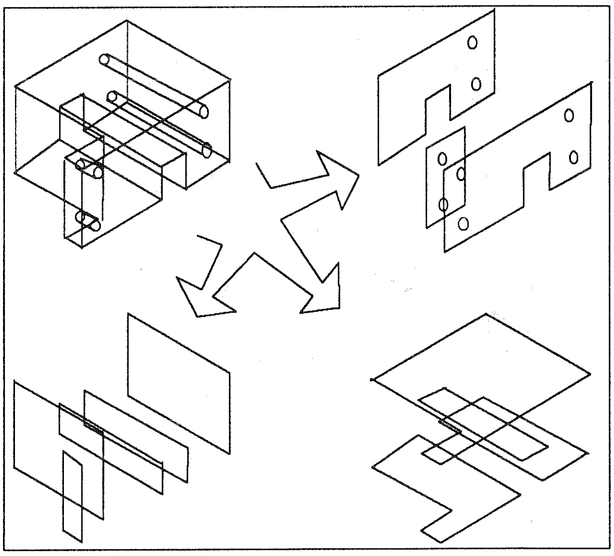

Figure 7 : Surfaces are grouped in terms of their directions

## 4. MODEL MATCHING

Matching between an objects in the scene and the models in the database is performed by a detailed comparison between their graphic presentations. A sensed object is presented in the same way as the models in the database. Since the object in the images are only partly visible, its description will not be complete. Therefore, the number of detected surfaces in an object will always be less than the number of surfaces in its corresponding model. The matching process contains two steps: the screener, in which most models unmatched to the given object are excluded; and the graph matcher, which performs a detailed comparison between the potential matching graphs and computes the 3D transformation between them.

### 4.1 Screener

In principle, the number of models in the database may be large, and evaluating each pair to find possible correspondence would be prohibitively expensive. Instead, a simple comparison between the dimensions of a scene object and models is used to ignore most models whose size is different from the size of the object. There are two elements used for dimension comparison: length of lines and average radius of ellipses. Each model has its maximum and minimum lengths of straight lines and size of ellipse, if such features exist on the model. The size range of an sensed object should be within the range of a matched model, since the number of geometric features of an object is less than that of a correspondent model. Considering the errors in image processing, the maximum value of a model increases by 10%, and the minimum value also decreases by 10%. This process limits the candidates of matched model to a very small number which are then performed in the next process.

### 4.2 Graph Matcher

The graph matching procedure consists of finding the pairs (in the model surface and object surface) forming the largest set consistent with a single rigid 3D transform. The process begins by finding all the possible pairs <m, o> where m and o are the model and object surfaces, respectively. The geometric comparison of surfaces is depended on the perimeter, the area and the number of lines which bound a polygon, or radius and ratio when the surface is an ellipse. In measuring the similarity between m of the model and o of the scene object, the normalised measure of the difference is computed for each of the following properties:

- $d_{m,o}(1) = d(A_m, A_o)$, where $A_m$ and $A_o$ represent the surface area of a polygon in a model and an object, respectively.
- $d_{m,o}(2) = d(P_m, P_o)$, where P represents the perimeter of a polygon.
- $d_{m,o}(3) = d(R_m, R_o)$ where R represents the average radius of an ellipse.
- $d_{m,o}(3) = d(Rt_m, Rt_o)$ where Rt represents the ratio of major axis and minor axis of an ellipse.

Thresholds are set for each of the differences to determine whether to accept or reject the match. One surface of an object may correspond to more than one surface of a model, as shown in figure 7, where the size of all circles are the same, so that one circle in an object may be found to correspond to eight circles in the model. If one surface of an object does not match with any surface in a model, however, it will indicate that the model does not match with the object and is rejected. The process results in each surface of sensed object having multiple corresponding candidates in a model. It is obvious that only one matching candidate is possible, if the object is corresponds to the model. Therefore multiple candidates must be reduced to a single candidate for each surface of an object. This process involves a compatibility constraint using topologic relations.

Topologic relationships exist among planar surfaces of a model or an object. If two pairs $< m_i, o_i >$ and $< m_j, o_j >$ satisfy similarity measures respectively, the relation between $m_i$ and $m_j$ should be the same as that between $o_i$ and $o_j$. Everytime a pair of nodes $< m_j, o_j >$ is selected, it is compared to all of the already matched pairs $< m_i, o_i >$ using a compatibility constraint. If this constraint is not satisfied, the chosen pair $< m_j, o_j >$ is discarded. The constraint contains the following relation checks.

- Orientation Relation ( $\xi 1$ ): Planar surfaces of a model or an object are grouped in terms of their normal directions. The angle between the orientations of two surfaces reflects the relations between their orientations. Let $\theta_m$ and $\theta_o$ denote the angles between the orientation of $< m_i, m_j >$ and $< o_i, o_j >$, and let $\Delta\theta = |\theta_m - \theta_o|$, then the pairs $< m_i, o_i >$ and $< m_j, o_j >$ are said to be $\xi 1$ compatible if and only if $\Delta\theta$ is less than a certain threshold.

- Proximity Relation ( $\xi 2$ ): Proximity relations summarise the distance between surface centres. Let $L_m$ and $L_o$ denote the distance between centroids of the surfaces $m_i$ and $m_j$, and $o_i$ and $o_j$, respectively, and let $\Delta L = |L_m - L_o|$, then the pairs $< m_i, o_i >$ and $< m_j, o_j >$ are said to be $\xi 2$ compatible if and only if $\Delta L$ is less than a certain threshold. If two surfaces are polygons and adjacent (ie. they share a common edge), the adjacency relation is checked between the two surfaces.

After all surface nodes of an object match with the model nodes and satisfy compatibility constraints, a geometric transformation is calculated between the coordinate systems of a model and object. Computing the geometric transformation between matched objects not only indicates how to bring the matched objects into correspondence, but also helps to verify the matching process. The estimate of the actual transformation between model coordinate system and object coordinate system can be given by a set of vertices of the polygons. If all elements of the object after transformation are matched with the elements of a model, the object is considered to correspond to the model, and its position and orientation are determined.

## 5. EXPERIMENTS

The system has been tested on several industrial components. The CAD models were constructed from physical prototypes whose dimensions were measured by hand. AutoCAD system designs of each model in terms of the data dimension were generated, and output in a DXF file. The geometric inferencing is then performed on models to create graphic representations which are stored in a model database. Figure 8 displays the models listed in the database.

A sensed object is captured using two CCD cameras as shown in figure 9. The image processes are then applied on the images to construct 3D objects (figure 10). The representation of the object is described in the same way as the models. The maximum and minimum of the line lengths in the object are 103 mm and 10.8 mm, and the maximum and minimum of the circle radius are the same as 7 mm. The four values are used to screen the models in the database and delete those whose ranges of dimension cannot cover the range of the object dimensions. In the database, only models 4 and 6 cover the dimension of the object, while the sizes of the other models are either too small or too large and therefore they are ignored.

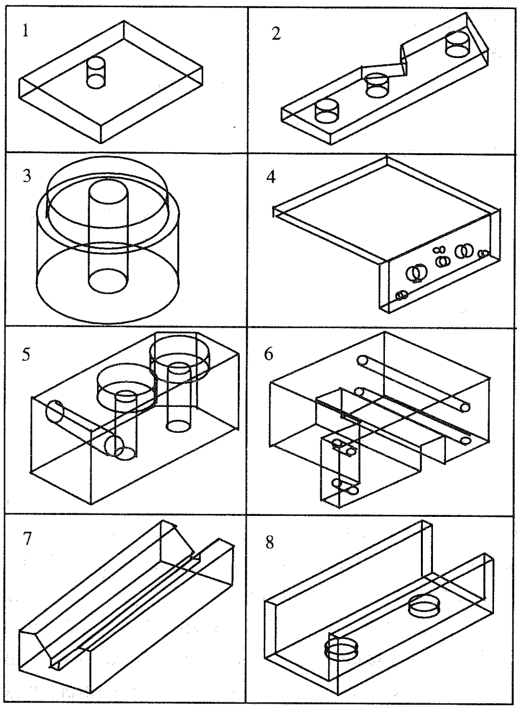

Figure 8 : CAD models in the database

The graph searching process is then used to find the correspondence between the nodes in the object and the nodes in the model. Graph matching starts with one planar surface. Once one match is established, more matches can be added if the resulting match meets the constraints of the node similarity and topologic relations. After all elements of the object are matched with those of a model, the object is transformed into the coordinate system of the model. Finally, not only the object is recognised as model 6, but also its position and orientation are determined.

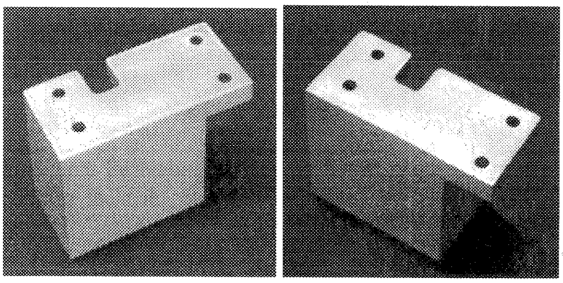

Figure 9 : Stereo images of an object

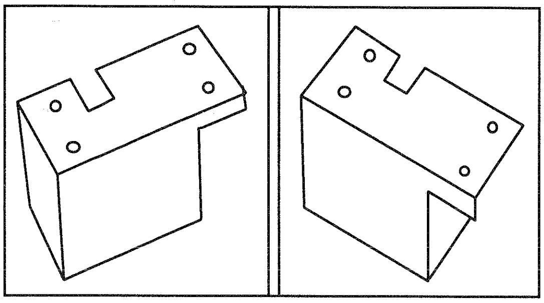

Figure 10 : Line segmentation of edges

## 6. CONCLUSION

This paper describes the elements of an automatic procedure for object recognition. The digital photogrammetry system is data-driven in that no a priori scene knowledge is required. The descriptions of the objects are computed without any knowledge about existing models, which is important when the environment is unknown. The process of object reconstruction reduces the image data to a few parameters of geometric functions, which are more meaningful and reliable. However, the system for the application of object recognition is limited to industrial components with simple regular shapes. It is not efficient for complicated objects or objects with occlusion. One reason is that the processes of edge detection and line segmentation cannot extract small detail features correctly, since there is not enough edge information.

## 7. REFERENCES

Besl P. J. and Jain R. C., "Three-Dimensional Object Recognition," ACM Comput. Surveys, Vol. 17, No. 1, pp. 75-145, 1985.

Brady J., Nandhakumar N. and Aggarwal J., "Recent Progress in the Recognition of Objects from Range Data," in Proc. 9th Int. Conf. Pattern Recognition, pp. 85-92, 1988.

Chin R. T. and Dyer C. R., "Model-Based Recognition in Robot Vision," ACM Comput. Surveys, Vol. 18, No. 1, pp. 67-108, 1986.

Fan T. J., "Describing and Recognizing 3-D Objects Using Surface Properties," Springer-Verlag New York Inc. 1990.

Flynn P.J. and Jain A. K., "CAD-Based Computer Vision: From CAD Models to Relational Graphs," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 2, pp. 114-132, 1991.

Förstner W., "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features," ISPRS Intercommision Workshop on "Fast Processing of Photogrammetric Data," Interlaken, pp. 281-305, 1987.

Huang Y. and Trinder J. C., "A Feature-Based Approach to Reconstruction of 3D Objects from Digital Images," International Archives of Photogrammetry and Remote Sensing, Vol. 30-3, pp. 391-398, 1994.

Trinder J. C. and Huang Y., "Edge Detection with Sub-Pixel Accuracy for a Flexible Manufacturing System," SPIE, Vol. 2067, Videometrics II, pp. 151-161, 1993.