METHODS FOR ANALYZING THREE DIMENSIONAL SCENES

F. Rocker and A. Kiessling
Institut for information processing and pattern
recognition, Breslauerstr. 48
D-75 Karlsruhe, West Germany

Abstract

Several methods for analyzing three dimensional scenes are discussed with respect to their advantages and difficulties in getting real three dimensional coordinates. Stereo picture processing and light intersecting methods are described along with the result of a new hardware system. On the example of a software system some special problems with the abstraction of scenes with primitive objects are mentioned. A short look ahead to the future closes this paper.

1- Introduction

When designing a robot system, we have to deal with the problem of how the robot can get information about its environment. To get full information it seems to be necessary to design a visual input facility. Therefore there is the problem of how the automaton can get a three dimensional image of the world from visual information.
This paper describes some systems, designed in the past, and tries to point out their problems and difficulties. Starting from these considerations, a new system is presented with which some of the well known difficulties may be overcome.

2. Stereo pictures

For interpretation of stereo pictures the system needs two pictures of the same scene, taken from different viewing angles (fig. 1). Provided that the two camera positions P1 and P2 and the angles a and B are known, the three dimensional coordinates of the point B of the scene can be calculated. The main problem in stereo picture processing is to define two conjugate points in the stereo pair. After the definition of one point in the first picture, the corresponding point in the second picture must be located. This can only be done if the picture point contains significant information, e.g. a corner point. This means that we have to deal with two different problems: first, we have to define a significant point in a first picture, and second, we have to recognize the corresponding point in the second picture. In order to solve the first problem we have to apply timeconsuming preprocessing methods like edge enhancement, etc. The second problem can be solved by two dimensional local correlation methods fV. Because of these difficulties, stereo methods are not used in fully automatic systems since now.

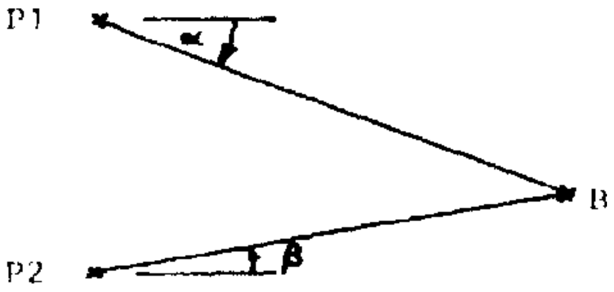

Figure 1

3. Light intersecting methods

In all methods described below, illumination of the scene is regarded as a part of the recognizing system. Then all visible points of a scene fit with a straight line of light or with an illuminating plane of light. With this information the three dimensional coordinates of all visible points can be calculated from their two dimensional coordinates in the picture. Using this principal several systems were designed.

3.1 Laser range finder

The laser range finding system uses a plane of light with known orientation in space for illumination of the scene. The plane of light is generated with a laser and a cylindrical lens. In such an illuminated scene the camera is able to see only the intersecting line of the light plane and the surface of the scene. The camera takes a complete picture of this image and transmits it to the computer for further processing and extraction of three dimensional coordinates of the intersecting line. By moving the light source step by step across th whole scene, the process of transmitting pictures and calculating spatial coordinates is repeated. It is evident that there is a large amount of data being transmitted and processed, one picture per line, containing very little information. Therefore the system is intolerably timeconsuming, Agin and Binford /5/ need about 5 to 10 minutes per scene.

3.2 Parallel grid illumination

In order to get quick results from the system, one must try to take the complete information of the scene with only one picture, i.e. the scene has to be illuminated with all planes of light at the same time. The simplest way to do this is to use an optical grid, being projected cnto the scene. Processing of the picture could be done by tracing the images of the intersecting lines in the picture. Computation of three dimensional coordinates can be done correctly only, if it is possible to identify the images of the lines of light in the picture by their mathematical equations. This problem causes some restrictions to this method:

a) The camera has to be adjusted in such a way that the pictures contain parts of the ground surface.
b) Shadows, interrupting the image of a line, cause the system to stop tracing the line.
c) The camera position must be chosen in such a way that lines, running on the top surface of an object, can be distinguished from those running on the ground surface behind the obj

d) Scenes with more than one object must not have hidden planes of objects.

In order to be able to identify the light lines by their mathematical equations it is necessary that the lines start from the ground surface with known position.
Other difficulties arise in the evaluation of inner and outer parameters of the camera (position, elevation, rotation and focus), especially when more than one parameter is variable. I.E. Sobel suggested a method to solve this problem by imaging a scene with an object of known size and orientation /2/. If the identity of one line is determined, all points on this line can be aomputed with their three dimensional coordinates. But errors may occur if a line is interrupted by shadows or other objects or if there exists a transition from one line to another, generated by the projection of the camera (fig. 2).

If all these restrictions are considered and found tolerable to the concrete application, this method seems to be a guick possibility to *analyze* three *dimensional scenes. For processing* a picture with *256 x* 256 *sanples, containing* 25 lines *of light, our* computer *CDC 3300* requires about *30* seconds to compute the complete three *dimensional coordinates of the* lines.

3.3 Grid and point illumination

If the restrictions remarked in 3.2 are intolerable better results can be achieved by combining two different pictures of one scene, taken by the same camera in the same position: one picture contains the image of the grid illuminated scene and the other contains the same scene, illuminated by a point light source. In many cases edges, missing in a grid illuminated scene, are precisely imaged in a corresponding point illuminated scene and vice versa as shown in fig.3 . *The* combination process could be done by three consecutive steps: First we process the grid illuminated scene as far as possible and compute the three dimensional scene coordinates. Then we do some contour finding process in the point illuminated image and get a list of contours. Then in a third step we generate a two dimensional image out of the computed three dimensional coordinates, containing those contours found in the first step. Then we are able to combine both two dimensional contour pictures getting a picture containing all contours from both images.

3.4 Hardware extraction of three dimensional coordinates

In all light intersecting methods for analyzing three dimensional scenes described above, one of the three coordinates is computed from the data of illumination and two coordinates from the data of the picture. This can be changed so, that two coordinates are derived from the illumination and only one coordinate from the camera. *Then* we get subsystems which can be built easily in hardware, being not too expensive.
The scene is illuminated by a single beam of light, generated by a locally fixed laser and deflected under computer control by an electrooptical deflection system. The scene is scanned systematically and in each position of the laser beam, the two equations describing the straight line of the light beam in space are well known. Projection of the scene is made by a cylindrical lens onto a one dimensional camera, designed as a line of photo diodes of OCD-technology. Thus each sensor of the camera represents a plane in space as shown in fig.4

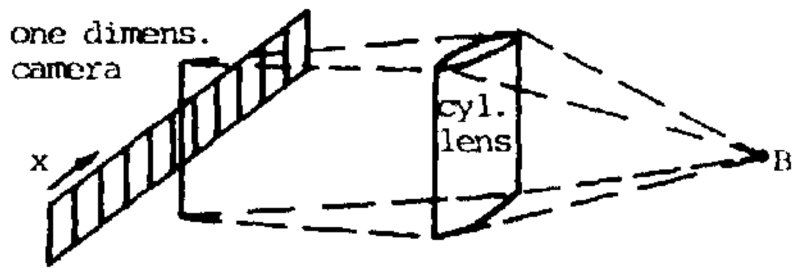

Figure 4

Only the intersecting point of this plane and the straight line, formed by the deflected laser beam, is visible to the system, so three dimensional coordinates can be computed directly without any picture processing. As this system measures three dimensional coordinates *of* cnly one point of the scene at one time, there are *no* restrictions to the scene as mentioned *in 3.2,* which derived from the parallel grid illumination.
The time required for processing a whole scene depends on the distance between object and camera because of the quadratic dependance of the sensitivity of the camera. Scanning a scene with a resolution of 256x256 points is expected to require 3 to 7 seconds /3/. Fig. 5 shows the complete system.

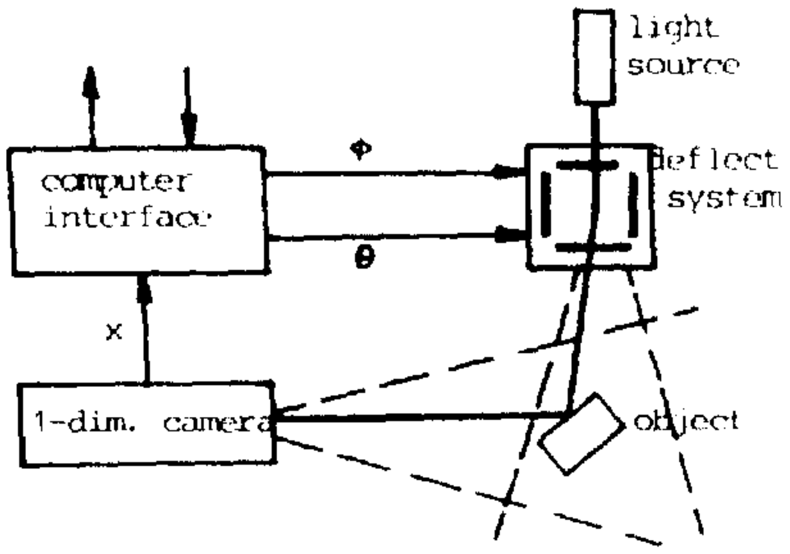

Figure 5

4. Experimental system

In order to study the above mentioned problems an experimental system was built, containing ail parts from scanning the scene until classification of objects (fig.6). As an input device a TV camera with a zoom lens is used (fig.7). Maption of data rates is made by an analog scan conversion memory. The camera writes with european IV rate (19.7 Mbit per second) into the memory and the computer reads out with its own rate (about 2 Mbit/sec) . All other parts of the experimental system are built in software running on a CDC 3300 with 32K core.

4.1 Picture processing

The scene is illuminated with parallel grid light, generated by a slide projector. Iherefore picture preprocessing such as contrast enhancement is not necessary. With a tracing algorithm, following the local maximum of grey level, the computer generates chain encoded data strings, which represent the illuminated lines in the picture. Fig. 8 shows the grid illuminated (a) and the traced (b) scene.

4.2 Abstraction of scene

The process of abstraction, described below, works with all surfaces, which could be described analytical, but the current software system is restricted to simple objects with plane surfaces like cubes, wedges, pyramides etc. In order to define the contours of these objects it is necessary to extract those points of the data strings, were a significant change of direction occurs (fig.9) in the chain encoded data. After that, three dimensional coordinates of all points on edges are computed, but that can only be done correctly if identification of lines is correct. This sometimes fails, as shown in fig. 10. The failure is caused by the narrow lines on the left vertical plane of the cube so that on top the last line is connected with a wrong horizontal line on the ground.
Because of systematic errors, the extraction of three dimensional coordinates is always within some failure range. Beyond this, by using grid illumination, in some cases corners of objects may only be extracted by interpolation between points of different edges. Both problems can be solved by using the fact that a scene is restricted to simple objects with plane surfaces. That means that some of the extracted points always form a straight line in space and that some straight lines must form a plane. The next step is to check which points form a straight line and which edges form a plane of the object. In both cases, the same decision rule *in* used. In the following, the algorithm is explained for the extraction of plane surfaces.
The straight lines are defined by their starting and ending points P . Fixed by a minimum data set (3 points for a plane) the equation for the first plane f- is computed (fig.11). Then the distance d_r between this first plane and a next point (IV) is computed and checked against a threshold £ . If the distance d_5 is less than £ , then the point P, belongs to the same plane (fig.11 b) and the equation of the plane is recomputed by regression analysis using the points P-, P-, P. and P.. This algorithm is repeated for all neighboring points. Fig.11 c-e shows the completion of a cube.

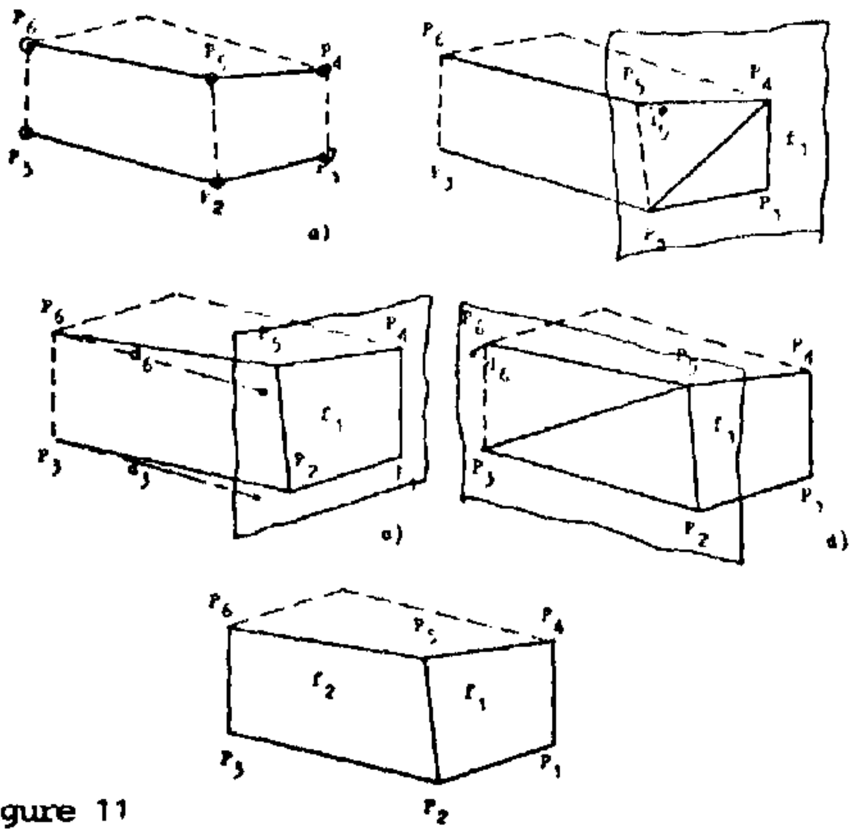

Figure 11

The intersection of the two planes f- and $f_?$ forms the final vertical edge of the cube (fig. 12). After that, all coordinates are changed so that they fit with the equations of the planes.

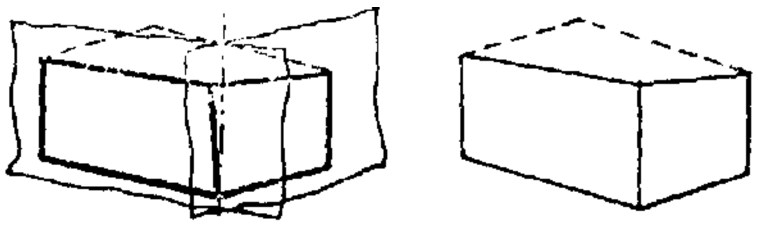

Figure 12

4.3 Data structure

With the computed points and planes a three dimensional, hierarchical, ring oriented data structure is generated. The complexity of objects is restricted to a maximum of 200 points, 200 edges and 60 planes, where up to 10 edges are allowed to touch the same point.These restrictions are given by the limited working storage capacity of the computer. In addition, the experimental system contains parts for mainpulating data structures (rotation, translation, scaling) and presenting them on a display.

4.4 Classification of three dimensional objects

Tne last block in the experimental system is formed by a classificator which compares data structures of known objects against the description of an unknown input object. The features for classification are:

a) the number of planes
b) the type of planes (e.g. number of edges)
c) the orientation of planes
d) the relative length of edges
e) the three dimensional coordinates of the corner points.

In order to be able to classify three dimensional objects it is ultimately necessary to be independant of the given orientation of the object in the scene. Therefore classification must be done invariant to rotation and translation of objects. By control of parameters the user is able to specify size invariant or size sensitive classification. Furthermore he can get results dealing with scaling and distortion of objects. Fig.13 shows an example of data structures presented on a display together with the result of a classification: DREIBEIN was identified to be the same object as DREIB10 and DREIB11.
The classification process itself is done by rotating, translating and scaling the unknown object until it fits with one or more of the known objects. The classification is very quick: To identify the object Dreibein in fig. 13, the CDC 3300 requires about 0.8 seconds.

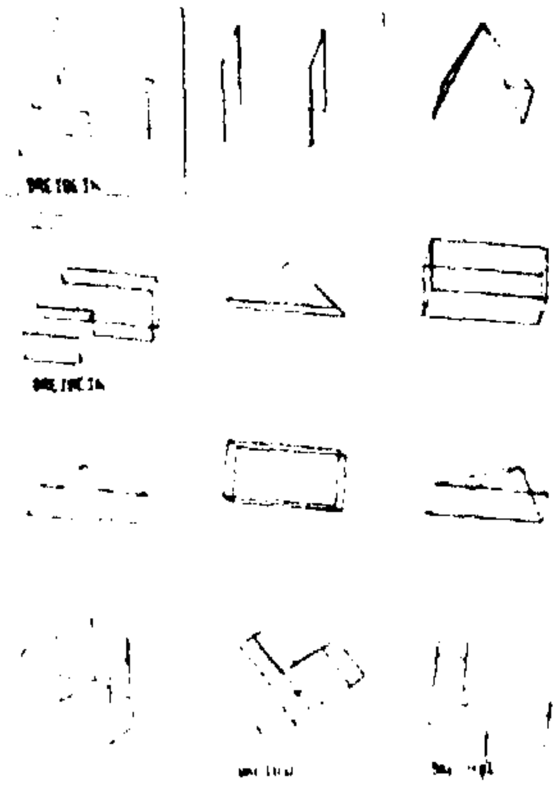

Figure 13

5. To the future

The described experimental system is able to analyze scenes consisting of one object with plane surfaces. Investigations on more complex scenes present a lot of difficulties as shown in 3.2. Therefore the system mentioned in 3.3 is proposed to avoid these difficulties.
Furthermore there are. no criteria for seperation of complex objects. Several criteria axe being considered, such as "as large as possible", "number of planes as few as possible", "as compact as posiible" etc.
First efforts with hidden lines have been published for the two dimensional case /4/. This method with a successive eliminations of objects and hypothetic completions of hidden lines seems to be well suited also for the three dimensional case.
Research in the past has shown that the light intersecting methods are practical for analyzing three dimensional scenes, especially in cases with a fixed working space. One can think of applications like, nuclear research centers in so called hot cells, for mounting on assembly belts and also for intelligent robots in space flight missions or remote sensed identification of vehicles and aircraft.

Examination of objects with curved surfaces and their description in a computer is a hard problem with no practical solution at hand. Curved surfaces can be described by approximation with many small planes /6/. Approximation by generalized cylinders reduces the amount for description, but there are no satisfactory results /5/. It seems that scanning and picture processing in robotics is a solvable problem, but description of objects requires more research in that field.

6. Literature

/1/ K. Wolferts, Special Problems in Interactive Image Processing for Traffic Analysis, Proc. 2.nd International Joint Conference (IJCPR2) on Pattern Recognition, pp 1-2, August 1974 hyngby, Denmark

/2/ I.E. Sobel , On Calibrating Controlled Cameras for Perceiving 3-D Scenes, Proc. IJCAI3, pp 648-657, August 1973, Stanford, California.

/3/ A. Kiessling, Theoretical Investigations of for Scanning 3-D Objects, internal report Nr. 20, institut for information processing and pattern recognition (FIM), Karlsruhe, West Germany

/4/ G. Grape, Computer Vision Through Sequential Abstractions, PhD. thesis, Stanford University, Stanford, California

/5/ G.J. Agin and T.O. Binford, Computer Description of Curved Objects, Proc. IJCAI3, pp 629-640, August 1973, Stanford, California

/6/ J. Encarnacao, Untersuchungen zum Problem der rechnergesteuerten räumlichen Darstellung auf ebenen Bildschirmen, Dr. thesis, Technical University, West Berlin, July 1971

/7/ K. Bers, Classification of Three Dimensional Objects, internal report Nr. 10, December 1973, institut for information processing and pattern recognition (FIM), Karlsruhe, West Germany

/8/ F. Röcker, Localization and Classification of Three Dimensional Objects,Proc. IJCPR2, pp 527-528, August 1974, Lyngby, Denmark

/9/ F. Röcker, Designing a Robot System, internal report Nr. 8, August 1973, institut for information processing and pattern recognition (FIM), Karlsruhe, West Germany

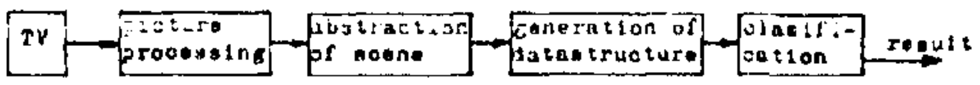

Fig. 6 Experimental system

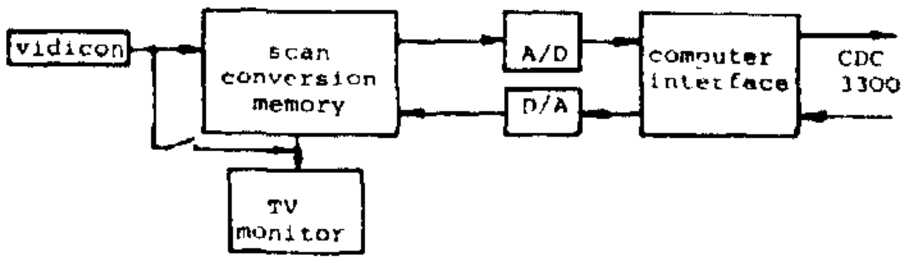

Fig. 7 Computer controlled TV camera

Figure 2 Grid illuminated scenes

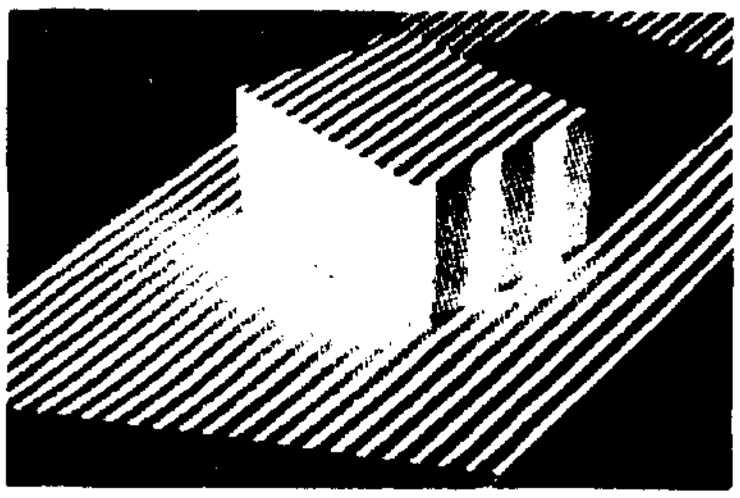

Figure 3 a) Grid illuminated cube

Figure 3 b) Point illuminated cube

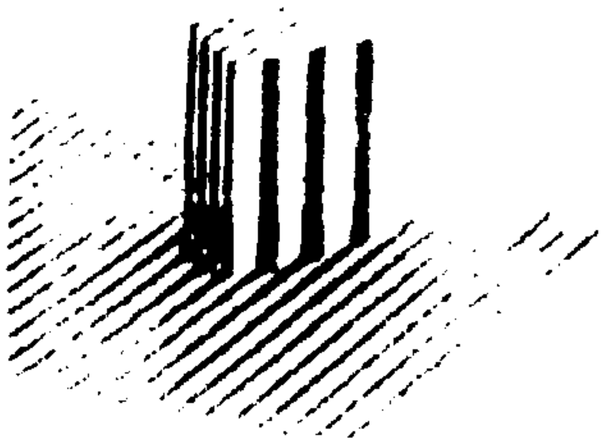

Figure 8 a) Scanned picture

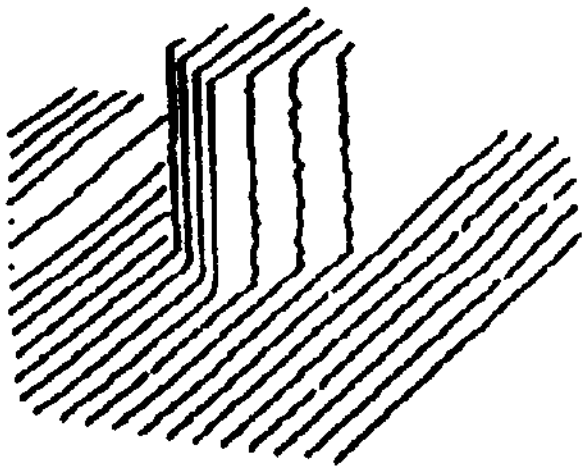

Figure 8 b) Traced picture

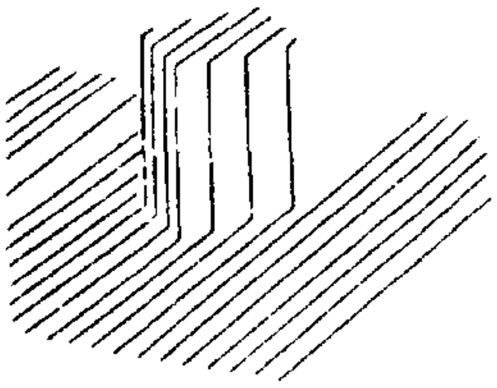

Figure 9 Computed lines

Figure 10 a)
Scanned cube

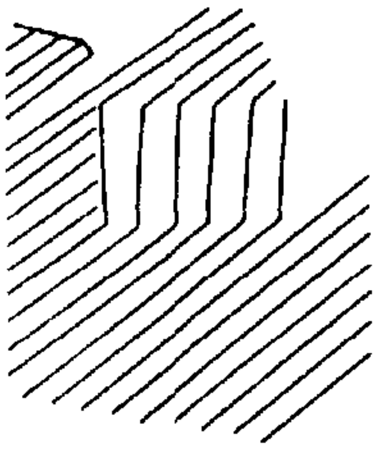

Figure 10 b)
Traced cube with failure