FOR REFERENCE ONLY

Pk'D sec

q 2>(SAC tdjl

COSMETIC QUALITY OF SURFACES:

A COMPUTATIONAL APPROACH

VELUPILLAI BALENDRAN

A thesis submitted in partial fulfilment of the
requirements of the Nottingham Trent University
for the degree of Doctor of Philosophy

November 1993

Acknowledgements

I offer thanks for the constructive suggestions, constant encouragement and unstinted assistance/inspiration provided throughout this project, to my supervisors Dr. K. Sivayoganathan, Dr. David Al-Dabass and Lindsey Evett. I am grateful for the support and advice provided by Prof. George Thompson and my head of Department Allen Hague. Thanks are also due to Jaguar Cars, Castle Bromwich, who helped not only with funds, but also with manpower and other resources for the project. In particular, Sivayogaiswaran, Don Perry, Kes Lodge, William Harris, and Nigel Simms deserve special thanks for providing insight into cosmetic quality and guiding of the project. I am also indebted to several inspectors from the body panel quality control department of Jaguar Cars, for tirelessly trying to impart their inspection skills to me. The project could not have progressed smoothly without the support of some of the amiable and very helpful technicians from the Department of Manufacturing Engineering. The electronics wizardry of Jes Keeling which brought the steel skeleton - manufactured with precision by Tony Maides, Keith Adams and Les Flear for which I am thank them - into life as an inspection robot deserves praise. I also thank Gary Griffiths for checking out the accuracy of the linear table. The research lab which was set up for this project where I worked as the only researcher, now accommodates several researchers. I have come to appreciate, learn and enjoy many things from the cosmopolitan, multi-cultuaral environment in the lab propagated by Alex Seiler, John Keat, Anne Leibar and Andrew Czerwinsky.

Summary

The use of qualitative, subjective, personal experiences for decision making purposes in manufacturing processes is examined. The dissertation offers directions for mechanising such experiences and decision making processes by taking an example from the automotive industry - the assessment of cosmetic quality of vehicle body panels - which has defied automation and still remains as an enclave of human activity within the industry. Cosmetic quality is essentially the assessment of the degree of eye appeal of body panels, taking into account the various defects on its surface, which are visible to the trained eye of inspectors under special lighting conditions. The difficulty in dealing with the problem arises due to the non-verbalizability of both the sensory content of the inspection experience, and the decision making process that follows. This is identified as due to limitations of language as well as the impossibility of gaining awareness of some of the thought processes that are involved.

The general solution offered requires the mapping of the human experience - both the sensory and the decision making - to mathematical domain with an appropriate basis. Specifically it requires;
1. a new tool and/or a methodology for mapping of the sensory content of the experience,
2. a new mathematical basis for evaluation of the experience that enables hitherto unknown forms of knowledge to be synthesised and acquired, and
3. the employment of conventional techniques in Artificial Intelligence (especially Expert systems, Machine Learning and Neural Networks), with the new knowledge as data, for the mapping of the decision making process.

With regard to cosmetic quality a novel computer vision based mechanical tool has been designed and implemented, which maps the visual experience of inspection as a mathematical "cosmetic map". A new mathematical basis is proposed for the evaluation of the cosmetic map, which results in the assignment of a descriptive classification code for any panel inspected by the tool. When this machine classification and actual human classification are taken as a pair, it embodies new inspection knowledge, and is acquired on-line in a database. While the potential of this evolving database to function as an "expert assistant" is exploited, its inability to reach generalisations is also indicated as a disadvantage. After sufficient inspection data has been acquired in such a database (over a period of time), employing machine learning techniques for unravelling the rules involved with decision making is highlighted as a way of overcoming this disadvantage. However, in the absence of such representative data, as is the case here, how synthetic data can be composed and used to determine experimentally a suitable neural network for generalising its learning and thereby functioning as an expert assistant is also indicated. The internal representation learnt by the selected network seemed reasonable, if not human-like, and strengthens the view that networks have the power to capture salient features or concepts, not explicitly stated in the data.

Contents

Chapter 1

Introduction

Cosmetic quality is a term used in industry to describe a qualitative aspect of surfaces. It is often judged, rather than measured in quantitative terms, and indicates the "goodness" of a surface. Automobile and glass industries are familiar with this term. The problem of cosmetic quality arising in the inspection of vehicle body panels in the automobile industry is addressed here with a view to automation.

In cosmetic quality, the word quality pertains to surfaces; and is assessed by human visual inspection. The word cosmetic is meant to connote a functionally non-essential, but visually essential aesthetic requirement. Briefly, cosmetic quality is a judgement of eye appeal based on the smoothness of surface form. When body panel surfaces are formed by pressing, a variety of defects arise which mar the smoothness of the surface and hence reduce the eye appeal of the automobile in its final painted and polished form. As eye appeal is a subjective concept, created and used by humans, naturally, it is best judged by them. The defects in the panel which contribute to the assessment of cosmetic quality, are visible under special lighting conditions to the trained eye of inspectors. However, as individual observers may assess surfaces differently, the subjectivity involved in the assessment of cosmetic quality is evident. (Beauty is in the eye of the beholder!). Industry would prefer to have this assessment made by machines. But given the severe limitations of today's technology in dealing with matters that are subjective, it is not surprising that cosmetic quality is still assessed manually.

The generic nature of the problem addressed in this dissertation is how a regularly used, but subjective concept based on personal experience can be accommodated by industrial automation.

An overview of inspection in industry, the methods used, and the subject of surface inspection may be in order prior to a description of the problems with cosmetic quality.

1.1 Inspection in Industry

In the broadest sense inspection is a process of comparison of the state of an item of interest with achievable standards. It involves search, fault recognition and decision making. The vital role of inspection in manufacturing processes is well recognized. In today's competitive industrial environment, the success of an industry is determined to a large extent by the quality of its products. Consumer demand for high quality products has led to the implementation of sophisticated quality control methods in manufacturing. Quality control relies heavily on inspection to differentiate good products from bad. Hence, the importance of inspection in manufacturing is clear. In order to reduce waste, cost, customer complaints as well as intangible costs such as loss of customers and erosion of reputation, inspection may be necessary not only after manufacturing, but before and (often at several stages) during manufacturing.

Traditionally, inspection has been performed manually. With appropriate tools and training, humans can perform highly complex inspection tasks. However, "human error" has its place in manual inspection, especially when the inspection task is monotonous or is to be performed at a forced pace. In the early 70s when highly mechanised and automated manufacturing processes were replacing older technologies, automated inspection was in its infancy and labour intensive methods of inspection (causing bottlenecks in production) dominated the industrial scene [Flanagan 69]. Even in the late 80s the inadequacies in the methods of inspection and hence the inability to keep pace with industrial requirements has been acknowledged [Batchelor 85][German 85].

In recent years, significant inroads have been made to overcome technical limitations in automated inspection, in order to meet the demands of modern industry for on-line, unsupervised, inspection for quality and process control. The introduction of the sense of vision for inspection in the form of "computer vision" or "machine vision" during the past decade has been a major influence in automated inspection. Though inspection systems make use of contact as well as non-contact methods for sensing, non-contact methods are more attractive since they afford remote sensing and freedom in positioning sensors (often a requirement in hazardous environments), in addition to overcoming effects of disturbance, interaction and even possible damage to either target or sensor. Some of these methods make use of techniques in optics, infra-red imaging, microwaves, ultrasonics etc.

However, the cost of implementing fully automated inspection systems has often been high and unacceptable. This is due to the high cost of customization involving design of optical and lighting systems for vision systems, complex configuration and development of image processing software [Chin 82]. As a result the number of installed systems is small. This high cost and sometimes the technical limitations in implementing fully automatic inspection systems has led industry to view partially automated implementations favourably. Still, a majority of inspection tasks are performed manually.

According to Scott [Scott 82], at least one of the following conditions must be fulfilled to justify automated inspection:-

1. inspection time and number of inspectors is very large;
2. inspection is part of the automated assembly process;
3. high cost penalties apply if product deviates from standard;
4. quality standards are objective criteria and are readily definable;
5. calculations are needed to determine the good or bad status; and
6. products are produced in large numbers, and 100 percent inspection is necessary but cannot be performed manually.

The work reported in this dissertation for automating automotive body panel inspection was mainly motivated by 3 above; the implementation of which also promised three other benefits - reduction in inspection time, 100 percent inspection and reduction if not total elimination of the subjective nature of manual inspection.

Kopardekar *et al.* [Kopard 93] categorize industrial inspection as:

1. monitoring: especially in continuous processes where an inspector monitors process parameters for deviations in order to control the process,
2. examining: where an inspector searches for defects,
3. measurement: where an inspector uses instruments and tools to assess conformity, and
4. patrolling: where an inspector checks others' work.

According to them, machines are more suitable for monitoring and measurement tasks than humans who are more adept at patrolling tasks, while examining tasks can be performed by both machines and humans satisfactorily.

1.2 Methods in Automated Visual Inspection(AVI)

Slow, subjective and error prone human performance in manual inspection are the primary reasons for shifting to computer vision based inspection [Kopard 93]. Batchelor *et al.* [Batchelor 85] provide a good introduction to several methods for automated visual inspection. They broaden the term automatic to include hybrid methods that have one or more fully automatic sub-systems that assist, rather than replace, a human inspector. Similarly, they use the term visual to encompass several other non-visual methods such as thermographic, infrared, ultraviolet, ultrasonic, microwave, nuclear magnetic resonance(NMR), neutron-beam, magnetic-field or tactile data; the argument being that these forms of data may be represented in the form of an image. They highlight the multi-disciplinary nature of AVI and the interplay of a diverse range of technologies/subsystems including but not restricted to:

- mechanical handling,
- illumination,
- optics/fibre optics,
- image sensor technology,
- electronics,
- computer architecture,
- software,
- algorithms.

Because of this multi-disciplinary nature of AVI, available literature tend to be problem specific. A broad range of problems have been solved by incorporating AVI techniques. Successful implementations can be found in the printed circuit board, food processing, manufacturing, automotive, textile, shoe making and other industries. In all these applications three common elements

- search,

- fault recognition, and
- decision making

can be clearly seen.

AVI applications invariably incorporate sophisticated optical and other methods/techniques (such as holography, interferometry, moire fringe or its variants, time of flight, triangulation etc.) and appropriate image analysis algorithms suited to the specific problem they address. They could range from simple accept/reject decision making algorithms based purely on an acquired image to complex algorithms involving artificial intelligence (AI) and/or neural network techniques. The sophisticated algorithms are a reflection of the complex nature of problems handled by present AVI applications. For example, in the automation of the inspection of wave-soldered joints in PCB manufacture, Jagannathan *et al.* [Jagan 92] use neural networks to first select features and then use an inference engine (an AI technique) to identify and classify defective solder joints.

In the area of surface inspection, references are mostly for inspection of flat or cylindrical surfaces such as paper, metal strip, glass/plastic sheets, cylinder bores etc. [Anon 74][Brook 76][Norton 77][West 77][West 82]. In these cases inspection is done primarily for detecting surface defects such as dents, bruises, scratches, stains etc., though in some cases dimensional accuracy such as thickness or width of strips of sheets are also inspected.

1.3 Surface Inspection

Inspection of surfaces are necessary to assess some property or criterion associated with the surface. For example inspection could be done for assessing:

- deviations in surface form,
- roughness,
- hardness,
- colour etc.

to name a few properties of interest. This dissertation addresses the issue of deviations in surface form of commonly encountered complex surfaces in industry.

Farin [Farin 90] defines a surface as the locus of a curve that is moving through space and thereby changing its shape. Traditionally, manufactured surfaces especially sculptured or complex surfaces, have been ill-defined in a mathematical sense. In spite of the availability of powerful mathematical techniques for design/representation of surfaces pioneered by Paul de Casteljau for Citroen, Bezier for Renault, James Ferguson for Boeing, Coons for Ford and others in the early sixties [Farin 90], (which are now standard texts in surface modelling), clay models and digitised 3D surface data from co-ordinate measuring machines or cross-sectional data are still used for surface definition. For instance most of the body panels in the automobile industry belong to this category of ill-defined surfaces. Critical dimensions in the drawing are mainly for ensuring straight forward assembly with other components (such as hinges, locks, apertures etc). As Bezier [Bezier 90] indicates, details of surfaces in the form of cross-sections (generally at 100mm intervals) given by the stylist who conceives the outline, which are sometimes modified by the plasterer who prepares the clay master model as a reference, are mainly designed for cosmetic reasons. These cross-sections and master models form the basis for preparing the stamping dies which are then used in the manufacture of body panels. The panels produced by this stamping process are usually within the specified dimensional tolerances. However, assessment of surface form is a more complex issue as explained later.

1.3.1 Cosmetic Quality of surfaces

Where surfaces are designed mathematically or otherwise for aesthetic appeal (in addition to other functional requirements), integrity of surface form is more critical than dimensional conformity with its designed or master surface. Some examples are automobile skin panels, general purpose casings, and most types of plastic objects. In these situations industry uses the term "cosmetic quality" to describe the quality of surfaces. Perry of British Glass Research Association [Perry 69] highlights this aspect when he states that "One of the values of the glass package is that its appearance can be very attractive and therefore, even though a container might be able to perform its job 100% efficiently in terms of being filled and holding the contents, it could be rejected because of its appearance. The cosmetic container is a typical example of one which has an important part to play in the selling operation because of its eye appeal".

As a bounded surface has an infinite number of points, it can only be completely defined mathematically. Such definitions are not always possible in practice as may be evidenced in sculptured surfaces. Even if a mathematical definition is given, it is of no direct use in inspection, as only a finite number of "points" of the surface can ever be verified. The only way to overcome this situation is to make use of both points and continuity properties of surfaces, such as smoothness and discontinuity. In some situations where aesthetic considerations weigh heavily only the smoothness properties may need to be inspected. Where mechanical functionality is important, often, inspection of critical dimensions of a finite number of points may be all that is required. However, there are situations where both form and dimensions are critical, as in the manufacture of lenses or turbine blades. Thus, the distinction between surface form and component dimensions is to be recognized in inspection. Dimensions deal with absolute or relative co-ordinates of (a finite number of) points (on the surface) with respect to some datum. Surface form is a concept based on the entirety of a surface (taking every point of the infinite point set of the surface into consideration). Thus we express forms of surfaces as planer, cylindrical, spherical, bi-quadric, bi-cubic etc., or a combination of these, often associating a mathematical property with a surface. It is not always necessary or even practical to categorise or analyze a surface as rigorously as described above when inspecting surface form. Given sufficiently dense points for physical inspection, we rely on the human ability to "fill in the gaps" of uninspected points by assessing the surface form. Consider for example the surface of the bottle in fig. 1.1 whose surface is made up of several clearly recognizable sub-surfaces (such as those forming the neck, body, base etc.) smoothly merging at their boundaries with one another. Apart from simple inspection for dimensional accuracy (for functional requirements such as mating of the filling head with the neck of the bottle) careful inspection may be required for aesthetic reasons. The bottle in fig. 1.2, which is out of shape will certainly be rejected by skilled inspectors on the grounds that most of the consumers may spot the defect. Explicit surface definition of the ideal or the inspected bottle are unnecessary here; eye appeal being the criterion for inspection.

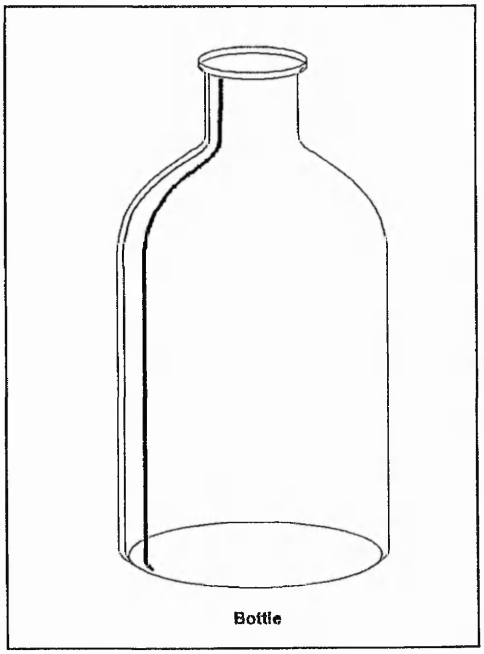


Fig 1.1 Acceptable bottle

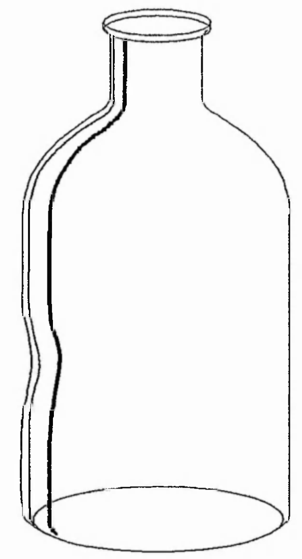


Fig 1.2 Defective bottle

Surfaces of machined components are mostly planer or surfaces of revolution and well established control of cutting conditions can usually overcome problems of surface form. However, moulded or pressed parts tend to show deviation in surface form compared to the ideal form. This is due to various process parameters and conditions that interact with the component during the formation of the surface. It is known that with pressed components, press settings, condition of dies, variation in metallurgical properties of the component etc., have an effect on the form of the surface produced; so much so that stamping is still considered an art. The deviations in surface form that are of interest, occur in localised regions of the surface which are then considered to be defective or flawed. Classification of these defects varies from industry to industry as mentioned below:

Automobile industry:

buckle, dish, pimple, ding, scratch,line, distortion, damage, recoil, split, return, lamination, roll mark, mislocation, chatter, air hole, slip line, swamp etc.

Steel industry:

pits, blisters, scabs, scratches, lines etc.

Tin plate industry:

spots, scratches, anode streaks, indentations etc.

Glass industry:

BS 3447:1962 about 50 possible faults.

Optical industry:

form error, scratches, digs, pits, spots, edge chips, greyness etc. MIL-O-13830A and DIN 3140, sheet 7 are widely accepted conventions for specifying accuracy and cosmetic quality of optical surfaces.

Paper industry:

holes, tear-outs, lumps, joins, creases, dirt spots, marks etc.

Textile industry:

floats, cracks, reediness, end-breaks, snarls, lashing-in, diamond-barring, colour, listing, pinholes, mildew etc.

Surfaces of large, thin, stamping easily deform under light load. Additional reinforcement or fabrication is necessary to make such surfaces sufficiently rigid and conform to their master model. This is the case with automobile skin panels which need to be spot welded to interior panels and fittings to obtain the desired shape. However, financial considerations demand that such panels be inspected for acceptability of surface form prior to reinforcement. This is feasible if smoothness of form is the criterion for inspection, since, the assumption that small but relatively uniform deformation of thin surfaces (as is the case with un-reinforced skin panels due to their own weight) do not change the smoothness characteristics of form of the surface significantly, is valid in practice. In other words, the surface defects from stamping are preserved before and after reinforcement. Absolute surface data from CAD models are irrelevant for comparison in these situations.

Methods that can be used for inspection of cosmetic quality are described in chapter 2. Human visual methods (such as the "green room" method in the automobile industry) are still widely employed by modern industry for the assessment of cosmetic quality. This is not surprising since eye appeal - a concept that is difficult to bring within the realms of computing, let alone automation - is the main criterion in cosmetic quality. Apart from the well known disadvantages of labour intensive methods on the shop floor, a major concern is the subjectivity in this form of inspection, due to the human involvement. If the effects of boredom/ drudgery of routine inspection are discounted, it is believed that identification of defects poses little difficulty to skilled inspectors. This is because when inspecting for cosmetic quality, a defect is one which can be seen by the naked eye; not one which exists but can not be seen. However, when assessing the cosmetic quality subjectivity is involved in two stages; firstly when assessing the severity of an individual defect and secondly when assessing overall quality based on a multiplicity of defects - especially since some defects are deemed as more critical than others. (For example, a large defect in the underside of the bottle mentioned earlier, which is usually not seen because of its location, may be regarded as less critical than a much smaller defect on the side wall of the bottle). Both these aspects of subjectivity are addressed in this dissertation.

1.4 Towards automation of cosmetic quality

It is clear that the term cosmetic quality is used in a highly subjective manner - which could mean different things to different people. It is based on personal experience. That there is no mathematical basis to it is plain to see. Hence it defies scientific study. This state of affairs in modern industry is most unwelcome. One of the main problems that is identified is the lack of a mechanical tool of some sort, even for the evaluation of a single defect, let alone one for assessing overall cosmetic quality. Thus, the first problem is to invent a tool capable not only of detecting defects, but which can also serve as a yardstick for the quantification of defects. Towards this end relevant techniques and methods that are available and used in industry are explored in chapter 2.

In chapter 3, a tool for assessment of defects is proposed. This tool is capable of:

- identifying visually perceivable defects;
- presenting the severity of defects in a manner representative of the visually perceived severity;
- quantifying severity of defects.

In effect, the tool maps the visual experience of inspection into the mathematical domain as a "cosmetic map".

The solution to the problem of overall assessment of cosmetic quality of panels (with one or more defects) is important if automation is contemplated. For this purpose an understanding of how inspectors decide to accept, reject, or rework panels has to be gained. In chapter 4, the issues related to the acquisition of inspection domain knowledge is discussed.

The lack of a mathematical basis in cosmetic quality was mentioned above. To overcome this situation, an algebra is proposed in chapter 5, for the evaluation of a mathematical entity - the cosmetic vector - for the cosmetic quality of a panel. This is based on the ability of the above tool to quantify cosmetic data.

In chapter 6, an on-line method is proposed for knowledge acquisition in the panel inspection domain. This possibility arises as a direct consequence of the new mathematical basis adopted, and is based on the system generated quality codes derived from the cosmetic vector mentioned above. A spin-off of this method is the ability gained to function as an "expert assistant". Some aspects of machine learning are also discussed with a view to finding generalisations in the newly acquired data.

In chapter 7, experiments with neural networks for generalisation of data are reported. A neural network capable of being trained initially using newly acquired inspection knowledge, and subsequently functioning satisfactorily as an "expert assistant" based on system generated quality codes is also proposed.

In chapter 8 some of the conclusions arrived at are presented and discussed along with possible direction of future work.

Chapter 2

Methods of surface inspection

2.1 Manual methods of surface inspection

The automobile industry has played a key role in initiating several methods for the manual inspection of relatively large surfaces. Manual methods are mostly (if not exclusively) used when surface inspection is to be done on ill-defined surfaces where the criterion of inspection is in the assessment of integrity of surface form i.e cosmetic quality.

2.1.1 The Green Room method

This is by far the most popular method for inspecting large components such as automobile body panels. Automobile industry throughout the world use this method.

The material of the surface to be inspected must be such that it can be coated with a thin layer of highlighting oil, which is a special grade low viscosity oil, in order to improve the reflectivity of the surface. Metallic surfaces such as those fabricated with sheet metal stamping are well suited for this as the thin layer of oil is retained as a film for a considerable period of time of the order of 5 minutes. Plastic components - especially those with matt finish - where a consistent thin film cannot be sustained are unsuitable for this method. The component is then sent to a room flood lit with green fluorescent lights - hence the name Green Room method - where inspectors evaluate the cosmetic quality by observing reflections and shadings of the lights from different points of view. The method requires a high degree of skill and is very subjective. To the novice, even detection of a flaw can be difficult; let alone the assessment of severity of a flaw.

Two aspects of this inspection have to be mentioned here. One is the ranking of individual flaws based on severity; the other the assessment of overall quality level of a surface with multiple flaws. Inspection knowledge gathered from previous experience serves as the only yardstick in these evaluations.

To set some modicum of standard to this process of inspection a panel known as the "critique panel" is used in the automobile industry. This is a panel at the lower level of acceptability and is usually changed for every batch of pressed panels, since panels of the same batch tend to have the same nature and distribution of defects. The critique panel serves more as a guide than an absolute reference to inspectors due to the subjectivity involved in perception at all stages.

2.1.2 Hand Feeling method

A very high level of skill is required for this method which is used only when green room facilities are not available. A soft glove is worn by the inspector who assesses defects from the feeling s/he gets when passing the hand gently and slowly over the panel. It is believed that a highly sensitive touch sensory perception faculty is required for this form of inspection. Inspectors claim to be able to accurately detect flaws with this method. However, due to the nature of inspection, subjectivity cannot be eliminated.

2.1.3 Grit Stone method

This method is often used when poor quality surfaces are to be reworked in order to bring them to the required cosmetic quality level. A fine-grit flat stone (or sometimes a fine flat file) is used to lightly rub or hone the surface to leave marks on the surface which may indicate high spots or valleys.

2.2 The D Sight method

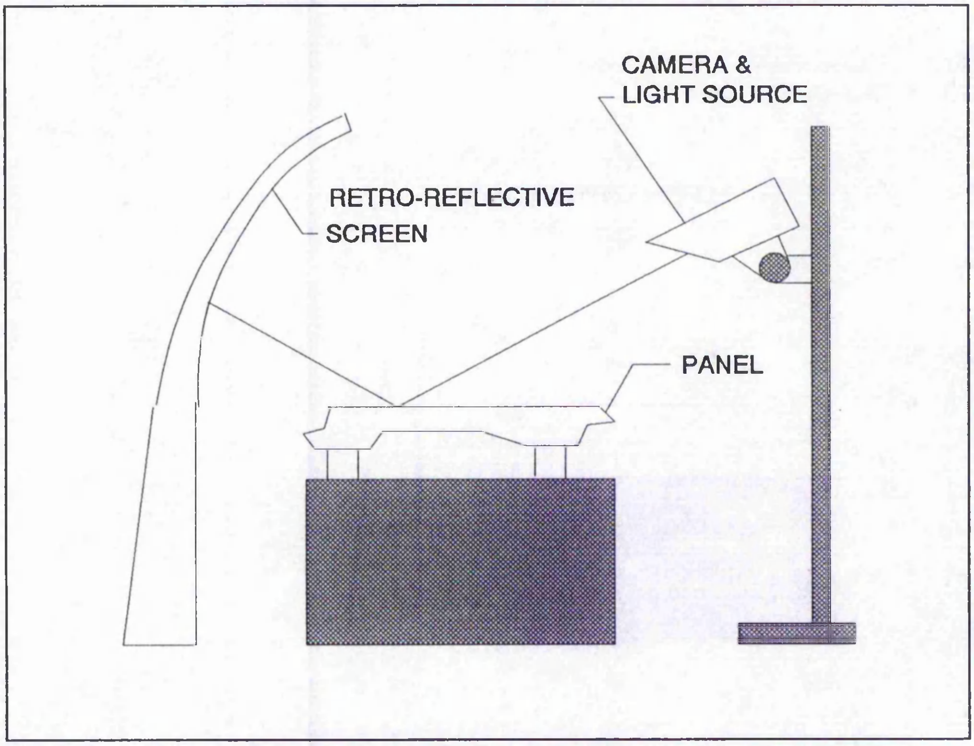


Fig 2.1 D Sight inspection system

In late 80s General Motors engineers and scientists of their Advanced Engineering Staff (AES) division were alleged to have developed along with Diffracto Inc., an associate vision company, a machine vision based surface inspection package for possible applications in automotive, plastics, aerospace, glass, tool & die, and other industries where surface smoothness is a critical factor for cosmetic or functional requirements. The arrangement of this patented system is now known to be as shown in fig 2.1 and is commercially available. The panel to be inspected is placed on a holding fixture. Incandescent bulbs attached to the camera are focused on the panel. The light beams travelling from the camera and reflected by the panels bounce on the special retro-reflective curved screen and are again reflected by the panel and finally reach the camera; the output of which can be seen in real-time on a monitor. The effect is to produce an enhanced view of the defects on the panel as shown in fig. 2.2.

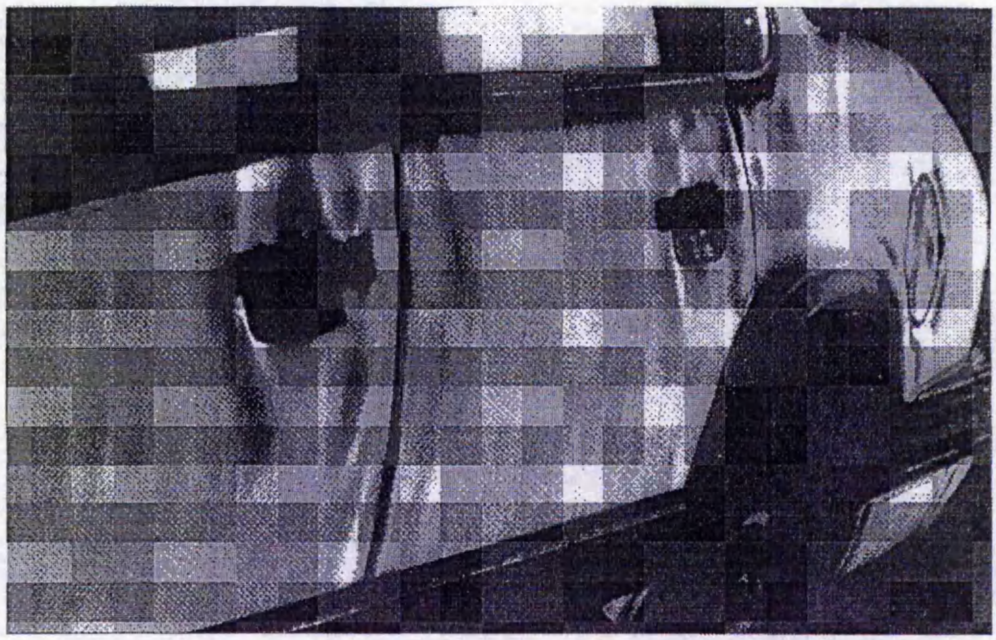


Fig 2.2 An image produced by the D Sight system

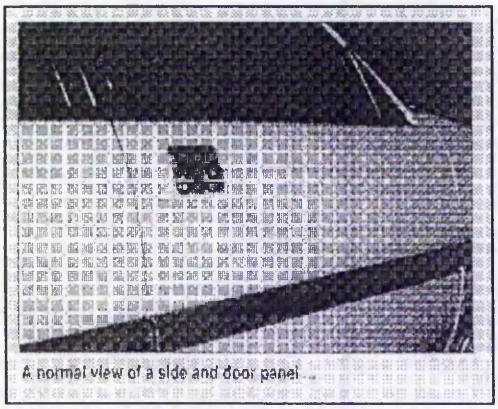


Fig 2.3 A normal view

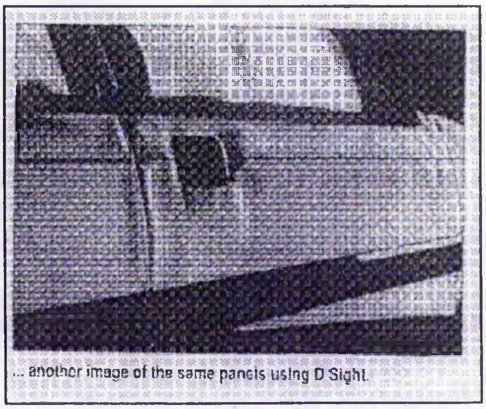


Fig 2.4 View using D Sight

Fig 2.3 and 2.4 offer a comparison of normal versus a D Sight based views of a car door panel. Since the optical phenomenon associated with this technique was not clearly understood and as explanations provided were based on intuition, Reynolds *et al.* [Reynolds 88] of Diffracto Ltd., first provided what they called the "most current and complete theory along with pertinent experimental evidence". The difficulty in explaining the phenomenon arises from the nature of the retro-reflective screen, which consists of millions of tightly packed, small (25-76 micron dia.), half-silvered glass beads, that cause a combination of reflection, refraction and diffraction of light to occur. Unsatisfied with the explanation given, they then provided in 1990 [Reynolds 90], a complete theory based on graphical ray-tracing. They mention only that the slope variations are converted by the phenomenon to grey scale information. It is not clear whether depth variations too are accounted for by this phenomenon.

2.3 Optical metrology based methods

Sophisticated optical inspection methods usually make use of holographic, moire, speckle effect or Fourier transform techniques which often involve interference or diffraction phenomena. These techniques are used for measuring the static or dynamic displacement of solid surfaces. Introduction to these techniques may be found in [Gasvik 87], [Hariharan 92]. Comparison of some of the techniques is discussed in [Creath 88].

Holographic methods used in metrology such as double-exposure interferometry or real-time interferometry compare an object in its natural state with its deformed state by way of the interference pattern produced by (original and/or reconstructed) wave fronts of the two instances. They also generally require holographic plates and are unsuitable for continuous on-line inspection processes. Further, the very stringent stability requirements necessitate vibration-isolated optical set-up, which is clearly impractical in shop-floor environments.

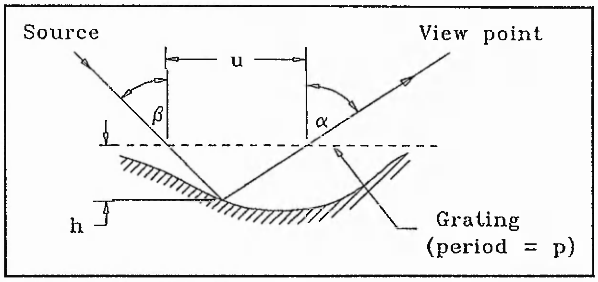


Fig 2.5 Shadow-moire method

In this respect moire techniques are more promising. The moire effect results from the superposition of gratings. The mathematical description of the moire patterns is the same as for interference patterns formed by electromagnetic waves. The shadow-moire method is particularly suitable for measurements on large non-moving surfaces that do not require high accuracy. The method requires a grating to lie over the object to be measured and be illuminated as shown in fig 2.5, causing its shadow to superpose on the object. When this shadow is viewed through the grating at a different angle, fringes can be observed as contours which can be imaged by conventional vision systems and analyzed by computers to yield 3D surface data. With a grating of period p, bright fringes are obtained for height contours given by [Gasvik 87]

$$h = np/(\tan\alpha + \tan\beta)$$

and dark fringes for

$$h = (n+½)p/(\tan\alpha + \tan\beta)$$

where n=0, 1, 2,...

Computer vision techniques are particularly attractive with projected fringe method where the fringe contours can be formed by image buffer manipulation. This method requires two different (or states of) surfaces - a reference and a test surface - though the resolution achievable by this method is better than that of the shadow-moire method. The requirement of two surfaces may be overcome by using two gratings; one for projection and one for viewing [Reid 84], [Reid 86]. The speckle effect based methods (Fourier fringe, Young fringe, speckle interferometry etc.) as well Fourier transform profilometry also require two surfaces [Su 88].

These optical metrology methods have not gained widespread acceptance in industry due to the large amount of fringe data that has to be analyzed. For this purpose microcomputer systems with general purpose processors take unacceptably high computational time. Nevertheless, some of the computational aspects are suited for implementation in parallel computing machines [Osten 87].

2.4 Co-ordinate measuring method

The Co-ordinate measuring machine (CMM) can be used for obtaining 3D data from discrete points on surfaces. Alternatively, it can be used in a scan mode to obtain a series of surface points on an incremental basis along a specific co-ordinate direction. These machines often make use of a sensitive touch trigger probe (of a given diameter) to supply 3D data related to the surface point the probe touches during the process of measurement. The accuracy achievable by most machines is in the range ± 25μm. Menq *et al.* [Menq 92] discuss some aspects of automated dimensional inspection using CMMs. In recent years non-contact probes (making use of laser based triangulation) too have been used with CMMs for measurement.

The discrete point data collection method is well suited for some inspection purposes. It is most suited for inspection of machined engineering components having simple geometric surfaces or features that can be assessed by a few measurements. For complex, free form or sculptured surfaces (for example a turbine blade with aerofoil cross-sections) the inspection requires substantial analysis of the collected surface data not least because of the offset curves or offset surfaces (see [Pham 92] for a review) that have to be calculated to determine the point of contact of the probe for proper compensation of the finite probe diameter.

Both discrete point and scan data are used for mathematical modelling of complex or sculptured surfaces of real objects that have not been designed by computer aided design (CAD) techniques thereby paving the way for what is called "reverse engineering" [Mason 90]. Though these methods are computationally expensive, the possibilities they offer for 3D surface inspection are evident.

2.5 Other possible methods

Several other methods are available for "indirect" inspection of surfaces. These methods essentially provide 3D surface data. The analysis of the data for inspection purposes is problem specific and is clearly separate from the method of data collection. (The optical metrology and CMM methods described above may also fall into this category if they are not used directly for inspection purposes but for obtaining contour or surface data for further analysis). These 3D surface data collection methods may be broadly classified as;

(a) range finding and

(b) computer vision based methods.

These methods have been of much interest in the field of robotics.

2.5.1 Range finding techniques

Range finding is a technique for obtaining 3D data of surfaces. This may make use of one of the following methods:

Time-of-flight

triangulation

range from focus

The time-of-flight method uses a collinear source signal (usually produced by laser or acoustic techniques) and an appropriate detector in order to measure the time it takes for the signal to propagate from source to target and back. Distance is then determined from this elapsed time and the transmission speed of the signal. Here the source signal used is impulsive, such as a laser source in pulsed mode. On the other hand, the method can also be employed in a continuous-wave mode where the phase difference in the wave forms of the source signal and the reflected signal can be used to determine the distance. As can be seen, the time-of-flight method does not require image processing techniques (but image analysis methods are frequently employed with range data gathered by other methods for scene analysis or object recognition purposes).

Active triangulation makes use of what may be called one-spot at-a-time method and does not require image processing techniques. In this method the image of a small spot of light from a source falling on a target is focused onto a light detector. At the instant when the detector senses the spot, knowing the directions associated with the source and the detector and also the fixed distance between the source and detector, range data can be recovered. This method is time consuming for collection of "dense" surface data and is more appropriate for obtaining "sparse" surface data.

Range from focus relies on the analysis of image phase shift that occurs when an image is out of focus to determine range information.

2.5.2 Vision based methods

Some of the methods in computer vision such as shape from shading, shape from texture are not suitable for obtaining accurate 3D information of surfaces. These methods are mainly suitable for assessing surfaces shape from 2D images, and other types of analysis and procedures must be applied to derive absolute size, orientation and location [Zuech 88]. Shape from shading make use of the observation that physical shape changes are usually accompanied by changes in luminance of the object. Shape from texture is based on the idea that cues about the surface slant can be obtained from the changes in texture coarseness. These methods make use of several other assumptions and practical applications require more research.

The triangulation technique discussed in earlier may be extended for use with computer vision systems by employing a video camera in place of the detector for obtaining range data. Data can be gathered on a point-by-point or plane-by-plane basis. In the plane-by-plane method a plane of light may be projected on to an object which is viewed from an oblique angle by a video camera (whose image containing the illuminated pixels is recorded). The location of any illuminated point of the object is determined by the intersection of the known plane of light and the ray to the camera corresponding to the illuminated pixel. (Planes of light are often projected by means of a scanner or a cylindrical lens). Structured light, light stripe, light sectioning are synonyms for this method. It may be noted that 3D information is available only for the regions of the object that are both illuminated and visible to the camera.

Bastuscheck *et al.* [Bastusch 86] describe a triangulation technique for obtaining 3D information from a single 2D view at once. Their method makes use of two images captured by a stationary camera, of an object illuminated by a slide projector; the first using a graded neutral density filter (to obtain a beam with monotonically varying intensity of light) and the second with a beam of uniform intensity. The two intensity images are divided pixel-by-pixel, which they state cancels out all factors which affect the intensity of the reflected light except the filter transmissivity. The resulting quotient or "ratio image" contains only 3D information about the surfaces of the scene.

Following the observation that human depth perception is predominantly based on binocular vision, computer vision systems have used two cameras separated by a given distance to image a common scene in order to obtain depth information of the scene (i.e 3D data). This approach is known as stereo machine vision and was pioneered by Marr and Poggio [Marr 76][Marr 82]. Since the two cameras are located at different positions, the images will not be identical. The recovery of depth relies on the solution to this disparity by way of achieving stereo correspondence on a point-to-point basis.

2.6 Evaluation

Surface inspection methods were reviewed above with a view to designing a tool for detection of surface defects on panels. More specifically, the aim was to find a suitable front-end for gathering data of panels (either as 3D surface data or preferably as cosmetic effect based data), with one or a combination of available methods.

Of the manual methods considered, the green room method which is the most popular method of panel inspection, offers the possibility of optical simulation. The other manual methods - hand feeling and grit stone - are clearly difficult to simulate. Consideration is not given to the D sight method as it is a patented method. The shadow-moire method which is the most promising of optical metrology based methods is not favoured on account of the large amounts of data generated and the accompanying penalty on processing time. The same argument could be levelled against the CMM based measurement with the additional criticism of slow rates of data collection and inaccurate contact based measurement (due to deformation of panel, errors in probe compensation etc). Range finding methods are suitable for obtaining nominal measurements. For applications requiring greater accuracy, such as the one under consideration, this method is felt to be either inadequate or costly. Shape from shading or texture do not offer proven approaches, and require further research. Also, the dependence of these techniques on illumination cannot be viewed favourably for a shop-floor based application. The vision based triangulation methods, especially the structured light method, appear to be suitable, offering adequate resolution with low cost, general purpose equipment (cameras, lasers, lenses etc).

The structured light technique was chosen as the basis for the front-end of the tool. However, it was adapted to incorporate the light reflected from panels, which is one of the main optical phenomena involved in green room inspection. The technique used is described in detail in chapter 3.

Chapter 3

A tool for surface inspection

3.1 Preliminary evaluations

This study was undertaken essentially to deal with the problems in cosmetic quality in the automobile industry, but the techniques developed can be applied directly to other areas of surface inspection. At the time when this study was initiated, the automobile industry throughout the world relied on human visual inspection to ascertain cosmetic quality. The D Sight system described in 2.2 was probably the only machine vision based system available for computerised inspection, though an operator was required for making decisions on quality of panels, thus involving subjectivity. The subjectivity of human assessment is an undesirable element in the system. The inconsistent assessment of the same surface, by different inspectors at different stages of manufacture of the car body (such as on receipt, after fabrication, after sub-assembly, after full assembly as a body etc.), often necessitates rework and hence additional production costs, which could have been minimised if the subjective nature of the inspection was eliminated.

Automation is a solution to this problem. But, a prerequisite is a tool for "measurement" of quality in the panel inspection domain, as was mentioned in chapter 1. Such a tool was developed as part of the research and is now covered by an European patent. As a minimum requirement, the tool to be used for inspection should have the following capabilities:

1. Ability to *identify* or detect visually perceivable defects at a (fixed or configurable) defect detection threshold suitable for inspection. (i.e at least capable of detecting all defects detected by inspectors if not more).

2. Ability to *depict* severity of defects in a manner representative of the visually perceived severity; i.e. in a cosmetic context, matching the visual experience of inspectors.

3. Ability to *quantify* (i.e measure) the severity of defects (according to some suitable method).

4. *Consistency* in operation.

These requirements ensure (a) the ability to detect and measure, (b) the relevance of measurement, and (c) the consistency of the process of inspection of defects. Computer vision based solutions for meeting these requirements were investigated and are dealt with below.

3.1.1 Defect detection

Several feasibility studies carried out by the author (excerpts from the first two reports of the studies given appendix C), at the behest of the collaborating automobile manufacturer proved that machine vision could compete with skilled inspectors in identifying (i.e. only detecting the presence or absence of) defects in panels [Balendran 89].

3.1.2 Cosmetic equivalence

The achievement of equivalence with human visual cosmetic inspection criteria, is an important aspect when considering automation. The popular "green room" method of inspection, which is a more critical method of assessing a surface than the other manual methods of surface inspection is chosen here for automation. The "green room" method solely relies on reflected light from the surface under inspection. Defect assessment is based on relative strengths of shading in the neighbourhood of a region. Thus, not only surface depth patterns, but more importantly the variations in surface normals play a part in the assessment. The method adopted to meet this requirement made use of light directly reflected from the surface in the assessment of quality. This was considered an appropriate mapping of the actual inspection process, since it took into account variation in both depth and surface normal (as will be shown in 3.5). Thus, rather than mapping surface height values which would only represent surface form, the effect produced by reflection of light on panels was mapped. (In the ensuing this map and the above mentioned effect produced by the reflection of light shall be referred to as the "cosmetic map" and "cosmetic effect" respectively).

3.1.3 Quantification

The cosmetic map which is a map of the cosmetic effect can be imagined very much like any 3D topological surface. The following assumptions are made in the assessment of defects:

- that smoothness of form of the cosmetic map is indicative of cosmetic quality,

- that a cosmetic map covers a substantial portion of the whole surface under inspection and hence enables the evaluation of the underlying smooth (mean) form.

- that deviations of cosmetic effect from this evaluated mean either above or below some preset threshold, signifies defects.

- that a contiguous set of defective points fully bounded by the defect free points alone, or in association with the boundaries of the cosmetic map can be classed as forming a single defect.

- that the severity of a defect is directly proportional to the sum of the deviations of cosmetic effect from the mean form within a defect boundary (disregarding sign).

3.1.4 Consistency

A controlled optical environment (a prerequisite for most if not all computer vision based methods) together with repeatable machine movements, preset machine settings and appropriate algorithms for assessment guarantee consistency of results.

3.2 Prototype model

Figure 3.1 illustrates the arrangement of the rig for automated inspection. A table **T** moves on precision linear bearings **L1** and **L2**. Four adjustment screws (not shown) at each junction of the frame **F** enable precision levelling of the linear bearings. This practically guarantees straight line motion of the table. A four axis manipulator (designed and manufactured in-house) consisting of arms **A1**, **A2** and laser head **H**, carries a 5mW Helium-Neon fan laser **L**. The axes of rotation of the arms **A1** and **A2**, and the laser head **H** are such that they are all parallel to the direction of motion of the table **T**. The mechanical attachment of the cameras **P** and **S**, and the laser **L** are such that their axes lie on what we shall refer to as the central plane. In any configuration of the manipulator, the direction of motion of the table is parallel to this central plane. The fan laser located in the laser head **H** and can swivel about a pivot axis that is perpendicular to the central plane. The axis of the panel camera **P** is inclined

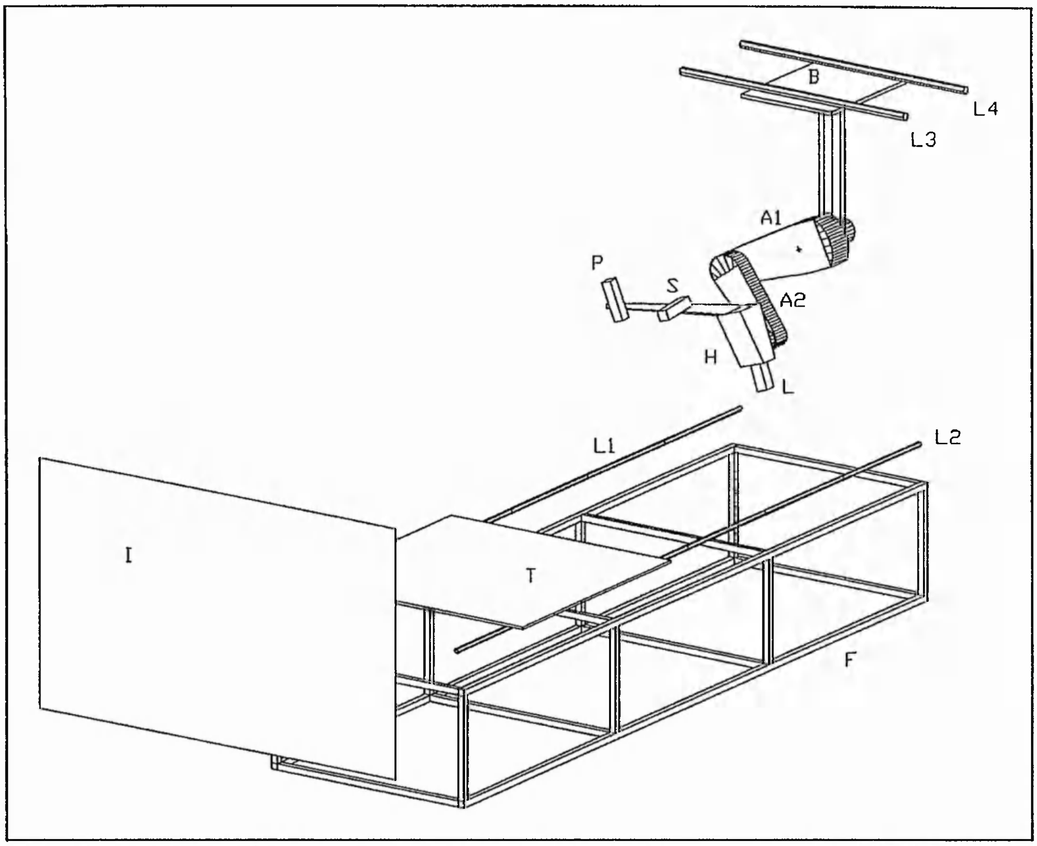


Fig 3.1 The rig

at a fixed angle τ with the plane perpendicular to the direction of motion of the table. This tilt angle is necessary for obtaining sharper images of laser traces (described below) on the panel. The controller **C1** (not shown) drives the table and the manipulator under computer control. The manipulator is suspended from a base **B** which moves on linear bearings **L3** and **L4**. The base is driven by a computer via the controller **C2** (not shown). The component to be inspected is placed on a former which is located on the table. Cameras **P** and **S** fitted with bandpass interference filters (centre wavelength 632.8nm matching the He-Ne laser wavelength, 11nm bandwidth, 80% transmission) mounted on the laser head **H** provide image information from the surface of the component and screen **I** respectively. The screen is positioned to be perpendicular to the motion of the table.

3.2.1 Overview of the method of inspection

It may not be possible to inspect large components in one operation or pass. In such situations the manipulator may have to be positioned at different locations in order to cover a region at a time for the complete inspection of the component. Facilities are provided in the controlling software for teaching these different locations for a particular type of component and storing them in a configuration files. (In addition to the series of manipulator locations, various other parameters that are relevant to the inspection are also stored in the configuration files). Loading of the appropriate configuration file for the type of component that is to be inspected ensures consistency of set up. Briefly, inspection procedure is as follows. With the laser **L** switched on, the manipulator is moved to the first location. This produces a trace (or contour) of laser light on the panel and another trace on the screen **I** due to the reflection of the laser light off the panel as shown in figure 3.2. These traces are captured by cameras **P** and **S** and the data is stored for later use. The table is then indexed by a predetermined distance and the data from the new traces are stored as before. The procedure is repeated for a preset number of indexes. The saved trace data is then used by various algorithms to evaluate defects on the region of the inspected surface (which is graphically displayed and also saved). If the component requires multiple inspection passes, the manipulator moves to the next location and the above procedure is repeated. This procedure is continued for the remaining locations in the configuration file. At this stage a graphical representation of the defects in the regions that were inspected is presented.

3.3 Laser trace data

The capture of the laser traces mentioned above is an important aspect in the inspection process. In fact this is the only purpose of using the vision system.

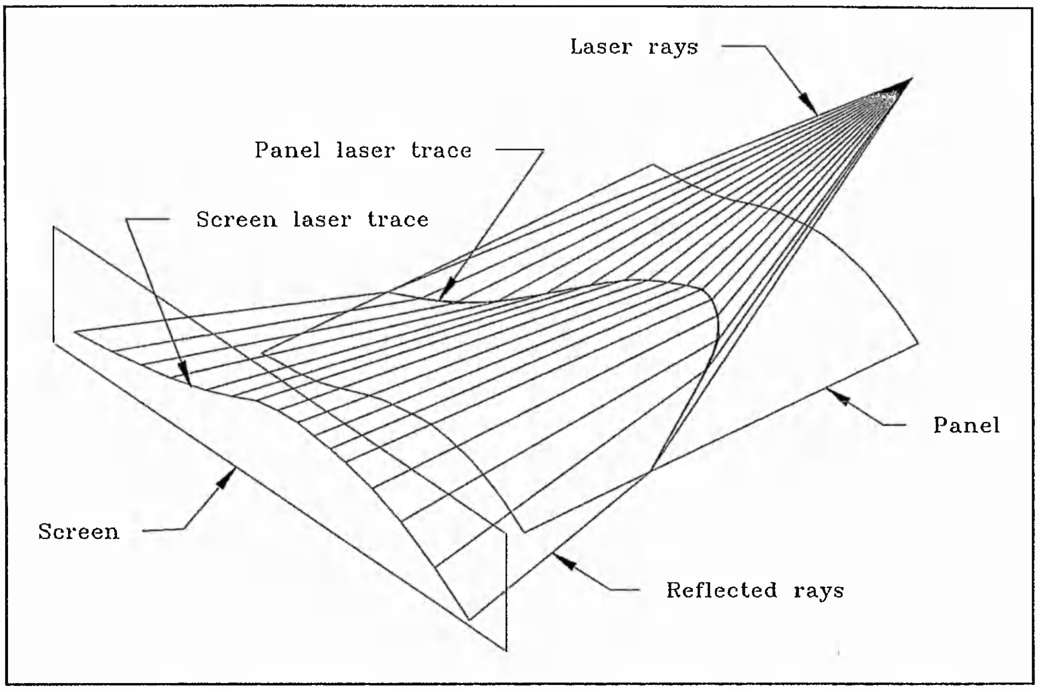


Fig 3.2 The laser traces

What is meant here by capture of a trace is the collection of pixel co-ordinates along the trace to serve as a representation of the trace. The problem faced here is similar to the problem of edge detection encountered in many computer vision applications where a physical edge (which is assumed to be characterised by sharp changes of intensity in the image) is to be determined. Here, however, it is not a physical edge that is to be detected but the boundary along which the laser plane illuminates the panel. Clearly, there is a sharp change in intensity along this boundary, and edge detection techniques are relevant. The only additional requirement imposed here is a much faster edge detection capability.

An appreciation of typical vision hardware and the methods used in edge detection should enable identification of suitable approaches and simplifications that can be employed in obtaining laser trace data.

3.3.1 Vision hardware

Typical vision hardware consists of at least one video camera and an electronic digital storage device known as a frame buffer or frame grabber which a host computer can access. The video camera outputs scenes presented to it continuously at regular intervals. Each scene is known as a frame consisting of two fields completed by raster scanning, and the frequency at which these fields are output is one of the standard frequencies of 50Hz or 60Hz, depending very much on the video standards - whether PAL, SECAM or NTSC - used in the country in which the camera was targeted to be used. Assuming a field scanning rate of 50Hz, it is clear that in an interlaced mode the interval between two frames is 40ms. If camera information pertaining to a frame can be stored or "grabbed" in the frame buffer within the 40ms, then, the frame buffer is known as a "real-time" frame buffer. Cheaper versions require several frame cycles and are thus not real-time buffers.

Assuming a monochrome (i.e. black-and-white) PAL type camera, the output of the camera during raster scanning is an analog d.c. signal representative of the light intensity along scan lines of the frame in addition to synchronization cues for detecting scan start, scan end, horizontal and vertical retrace etc. The frame buffer incorporates a fast analog to digital converter to convert the analog intensity signal from the camera on each scan line into digital information called pixel grey scale values and stores them in the frame buffer. This storage is done in such a manner that there is a direct correspondence between the digital intensity information stored in the frame buffer and the actual light intensity at the various parts of the scene. Typically only 512 line scans are used from the available 625 raster lines and the analog signal of each scan line is sampled into 512 pixels. The mean intensity is represented by a 6 or 8 bit value resulting in 64 or 256 grey scales. Thus a frame buffer can be thought of as a two dimensional array (512x512) storing pixel grey scales.

Vendors of hardware often provide facilities for accessing the pixel grey scale values in the frame buffer from the host computer. They also provide several library functions for many popular image processing tasks. The vision system used for this research is a Matrox MVP-AT real-time vision system which can be configured to accept inputs from four monochrome cameras into four separate 512x512x8 bit frame buffers or alternatively from a single colour camera into a 512x512x24 bit frame buffer. The library functions supplied have been compiled with Microsoft C and can be readily linked with user software written in Microsoft C. The functions provided (over 125) include low level hardware control and display commands, several frame grabbing commands, point-to-point pixel manipulation commands using one or two images, statistical commands such as those related to generation of histograms and thresholding, lookup table commands that enable modification of input and output lookup tables, neighbourhood commands such as frame averaging, frame comparison, convolution, erosion, dilation, thinning, edge detection (using Sobel, Prewitt, Kirsch, Laplace, horizontal or vertical edge operators), graphics command and input output commands.

3.3.2 Edge detection

Traditional computer vision based edge detection methods assume that the task is one of measuring and localising changes of light intensity in a given image. Poggio *et al.* [Poggio 88] agree with this assumption in a narrower sense - as a first step towards physical edge detection - and argue that local intensity changes alone are insufficient to determine physical edges correctly and that other higher level vision processes are required. Notwithstanding such arguments, primitive edge detection techniques have been put to practical use. An introduction to these techniques can be found in most introductory books on computer vision. eg [Rosenfeld 82], [Boyle 88], [Ballard 82]. In addition to the methods mentioned in section 3.3.1, some of the other methods are Roberts operator, Canny operator, Gaussian operators, Laplasian of Gaussian (LoG) operators, local adaptive threshold (LAT) based operators, entropy based operators etc. All these methods operate on the complete image, i.e on every pixel of the 512x512 array and are therefore computationally expensive. Their use can only be justified if the nature of the image requires analysis of the complete image. This is not the case with the laser trace images encountered here. These images contain only a single laser illuminated curve with a dark background due to the narrow band filters used with the cameras. Thus, a simple, fast and practical method which acquires the laser trace in an efficient manner was implemented as described below.

3.3.3 Acquisition of trace data

Figure 3.3 shows a typical pixel intensity profile along a single raster scan line of a laser trace image (though in practice the maximum grey level is much lower than that shown). The horizontal axis (not shown here) represents the digitised pixel number in the range 0 to 511. Pixel intensities above a threshold value as indicated here, but of a complete laser trace image in the frame buffer (i.e of all the raster scan lines) are shown graphically in figure 3.4. This threshold is a parameter in the control software which can be preset by the user. Two thresholds - one for the panel camera images and the other for the screen camera images - are available.

It is clear that only a relatively small portion of the image, approximately 5%, contains information useful for the detection of the laser trace. This fact coupled with the desire for fast edge detection (in order to maintain a low total cycle time during inspection) promotes the search for a custom method for the

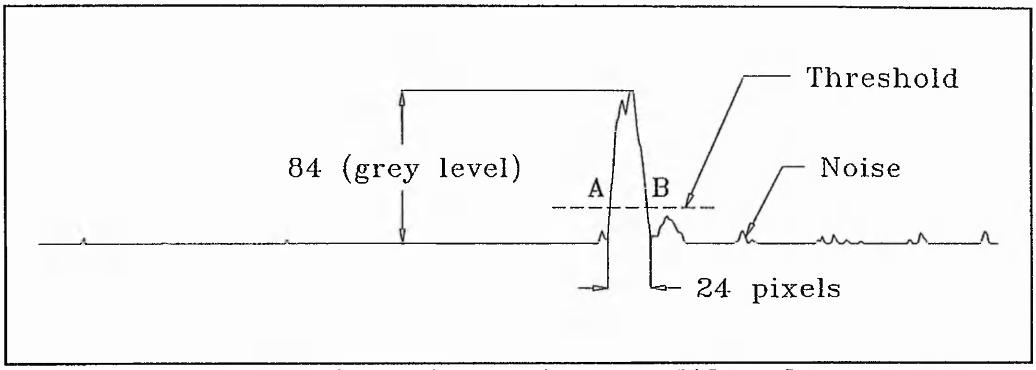


Fig 3.3 Typical pixel intensity profile along a raster line

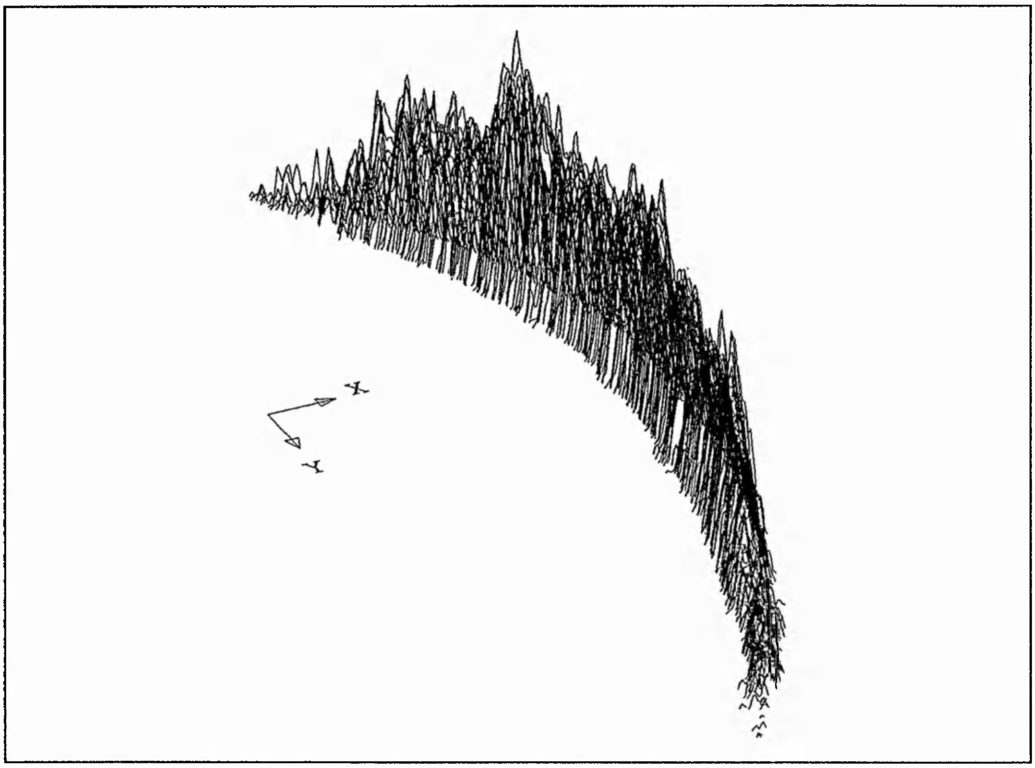


Fig 3.4 Frame buffer map above threshold plane

acquisition of trace data from the laser trace images.

The method used searches an image in a raster-by-raster manner. In any raster, it searches sequentially, for a pixel with intensity above the preset threshold value. When such a pixel has been found in a raster, as at A in figure 3.3, the pixels that follow along the raster are monitored until the one below the threshold, as at B, is reached. If the number of pixels between A and B (i.e. the pixel distance AB) exceeds a user defined, preset value (usually between 2 and 10), then it is assumed that the points A and B are the front and rear edges of the laser trace (since a laser beam has a finite width). If not, the search continues; first along the current raster line, and then along the next raster line, and so on until the image is fully searched. However, if the laser edges A and B are detected in a given raster, then, further search is abandoned on that raster and the next raster is considered. In this new raster, the search is not started from the beginning as described above. Instead, search is performed in a small range which is an extension of the range A to B of the previous raster on both sides by a (user defined) preset value. Such a method provided satisfactory performance in rapid collection of laser trace data of both screen and panel camera images.

Given the finite width of the laser beam, the question arises as to what value - whether the pixel co-ordinate of point A or of point B - is to be used for the description of the laser trace. After experimentation, it was decided to use the mean value. Use of the pixel co-ordinate of the centre of gravity of the intensity profile between A and B was also considered, but was not used on account of the time penalty involved.

3.4 Three dimensional panel surface data

Figure 3.5. is a schematic which identifies the laser, panel camera and the scene

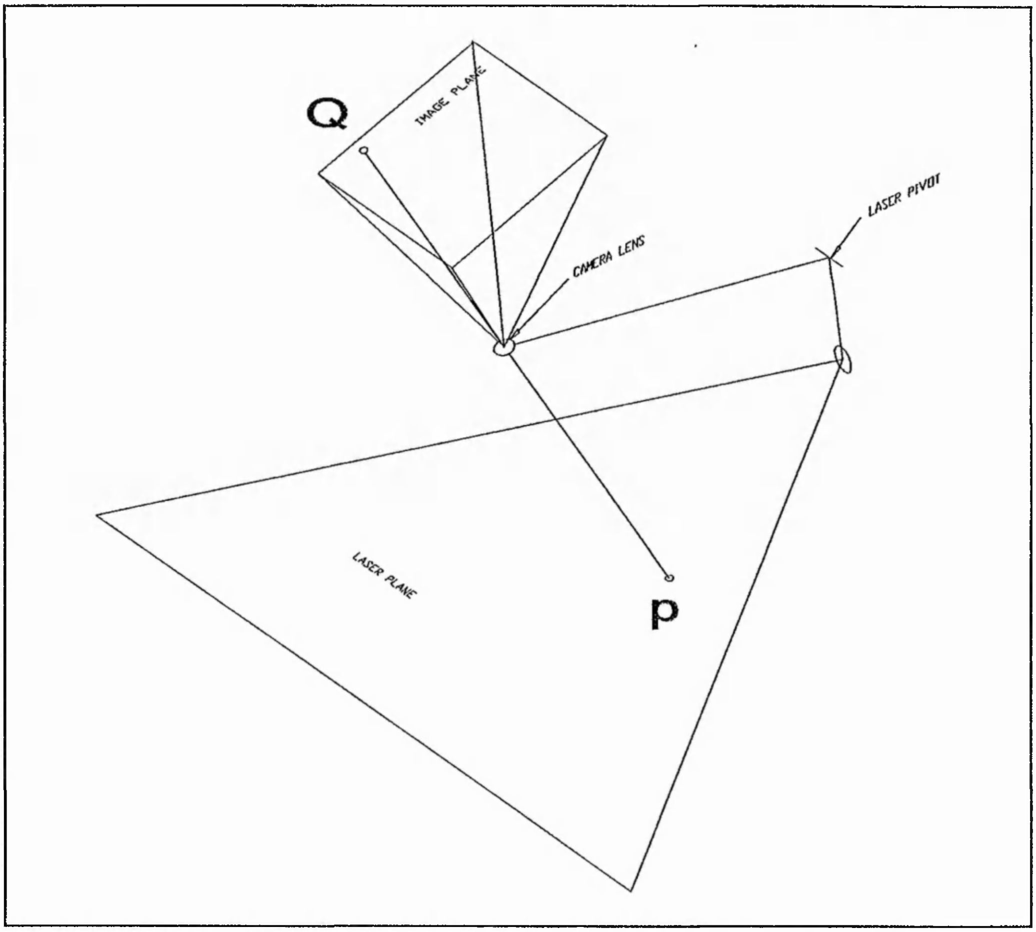


Fig 3.5 Laser and panel camera configuration

viewed when they are set up in the normal configuration. The CCD image plane of the camera is magnified for reasons of clarity. It also shows a typical point **P** in the scene presented to the camera, and its image point **Q** on the image plane of the camera. It may be worth noting here that the co-ordinates of the points **Q** can be obtained in terms of orthogonal pixel offsets by processing the frames grabbed from the camera output. The exercise we undertake here is to derive physical co-ordinates of point **P** in real space from the pixel based two dimensional camera co-ordinates of point **Q**. Note also that the laser itself can be swung about the laser pivot to obtain the best set up for inspection. The angle of the laser (measured between the laser plane and direction of motion of the table) is the only parameter that varies between different inspection set ups.

Figure 3.6 shows the trace **R** produced on a typical panel by the fan laser when the laser plane impinges on the panel surface and its camera image **S**. Since the panels inspected in practice are not fully reflective, the trace **R** is visible to the naked eye and hence to the camera due to the scattering of light that takes place when it impinges the panel. In order to obtain the best image **S** taking into account the physical constraint of camera location (imposed by the coverage area of the scene), it was found that a slight tilt away from the overhead position was necessary for the camera. The panel camera was then fixed permanently in this tilted position. It may be noted (since we propose to find 3D representation of the trace **R**), that the points of the trace **R** lie not only on the panel but also in the plane of the laser.

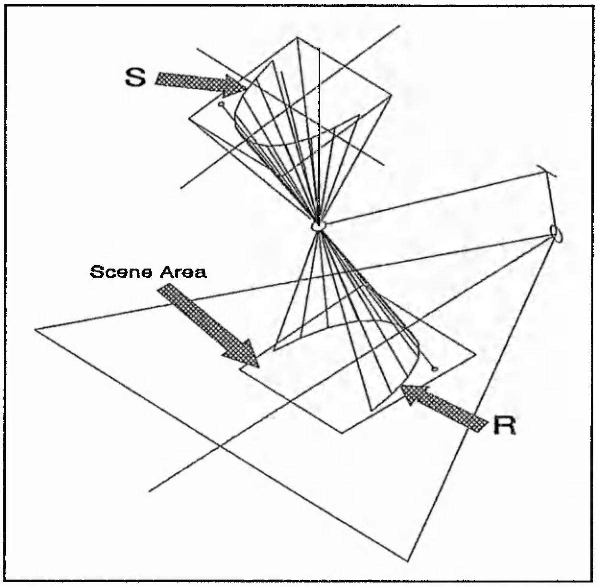


Fig 3.6 Laser trace and its camera image

Figures 3.7, and 3.8 show the co-ordinate systems used in the ensuing discussion. It is convenient to imagine the image plane of the camera as the frame buffer of the vision system itself. The brief explanation for this is that the frame buffer is a one to one map of the physical scene viewed by camera in

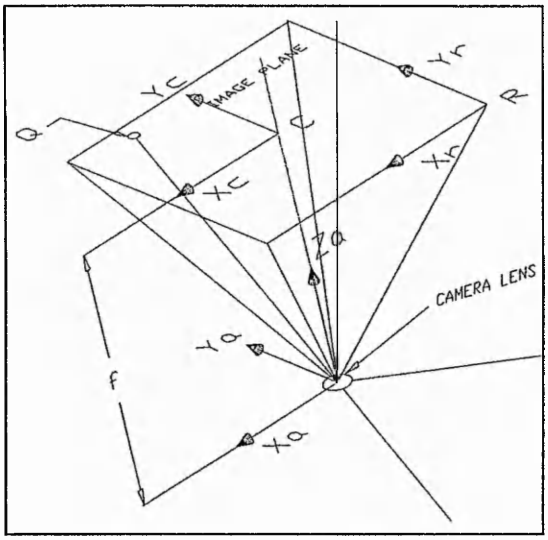


Fig 3.7 Internal Camera co-ordinate systems

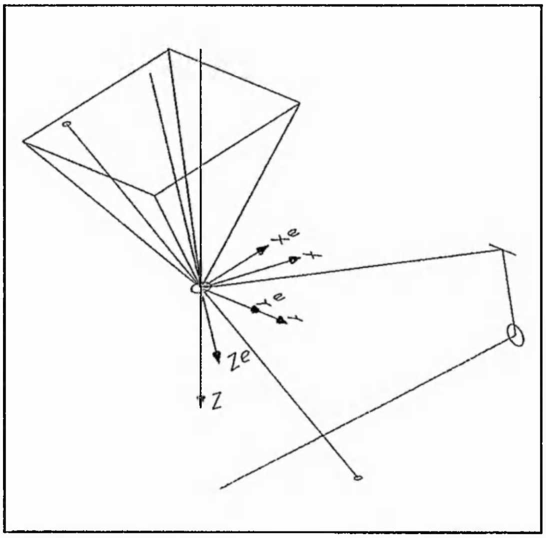


Fig 3.8 External camera and world co-ordinates

digital format, built from the analog scene content data from the camera. Henceforth we shall consider the image plane to consist of a 512x512 matrix of pixels (picture elements), same as the frame buffer. Thus any pixel may be represented in terms of its column and row positions. The two dimensional co-ordinate systems {Xr, Yr} at a corner R (column=0 and row=0) of the image plane and {Xc, Yc} at the centre C (column=256 and row=256) of the image plane, and the three dimensional co-ordinate system {Xa, Ya, Za} at the optical centre of the camera lens (with axes Xa and Za in the central plane) shown in figure 3.7 are internal to the camera and are all pixel based. That is the co-ordinates of a typical image point such as **Q** are represented in terms of its pixel column, row position and shortest distance **f** between the lens and the image plane expressed in terms of pixels. Assuming a "pinhole" model for the camera optics, this later distance **f** can be computed knowing the angle ϕ subtended by a scene view to produce a 512x512 pixel image.

$$f = \frac{(512/2)}{\tan(\phi/2)}$$

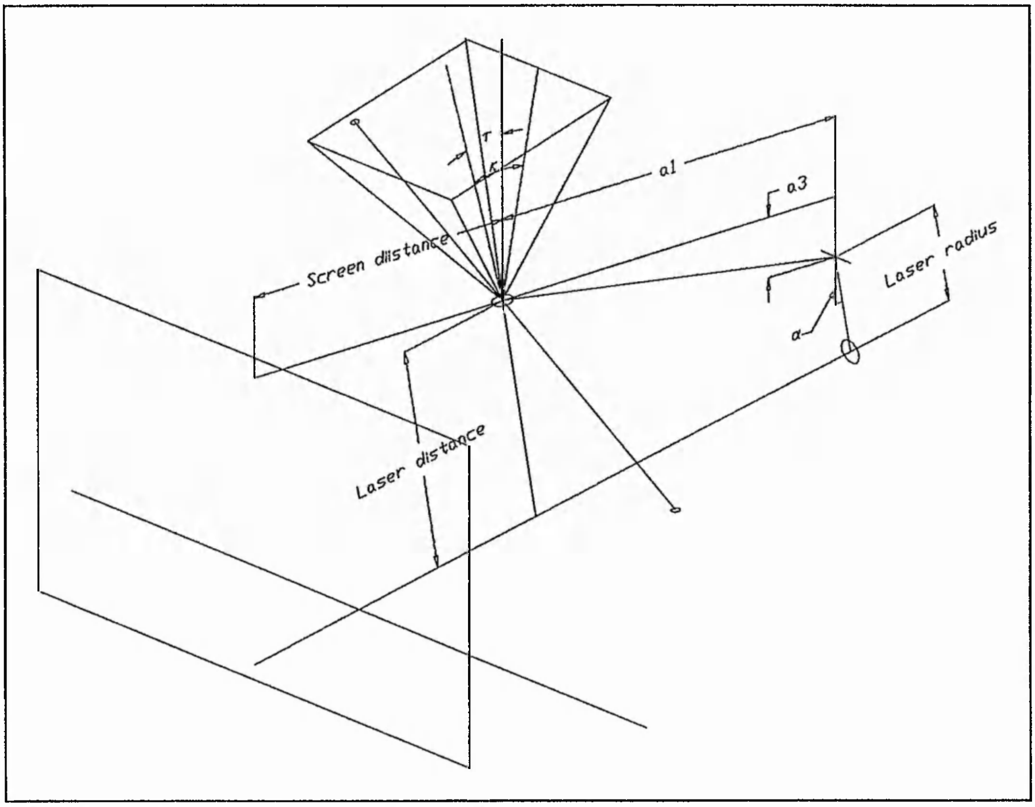


Fig 3.9 Distances and angles

Figure 3.9 indicates the fixed distances and angles involved in a particular inspection set up. All the angles and distances shown remain fixed except for the laser angle α and hence the Laser distance, which is the distance of the laser plane from the panel camera, that may vary for different inspection set ups.

Figure 3.10 attempts to draw relationships between panel traces, screen traces and camera images in terms of vectors. A ray from the fan laser impinges at point **P** on the panel. Most of the light from this ray is directly reflected from the panel and reach the screen **I** at point **S**. A ray of scattered light at **P** also reaches the panel camera and falls on the point **Q** on the image plane of the

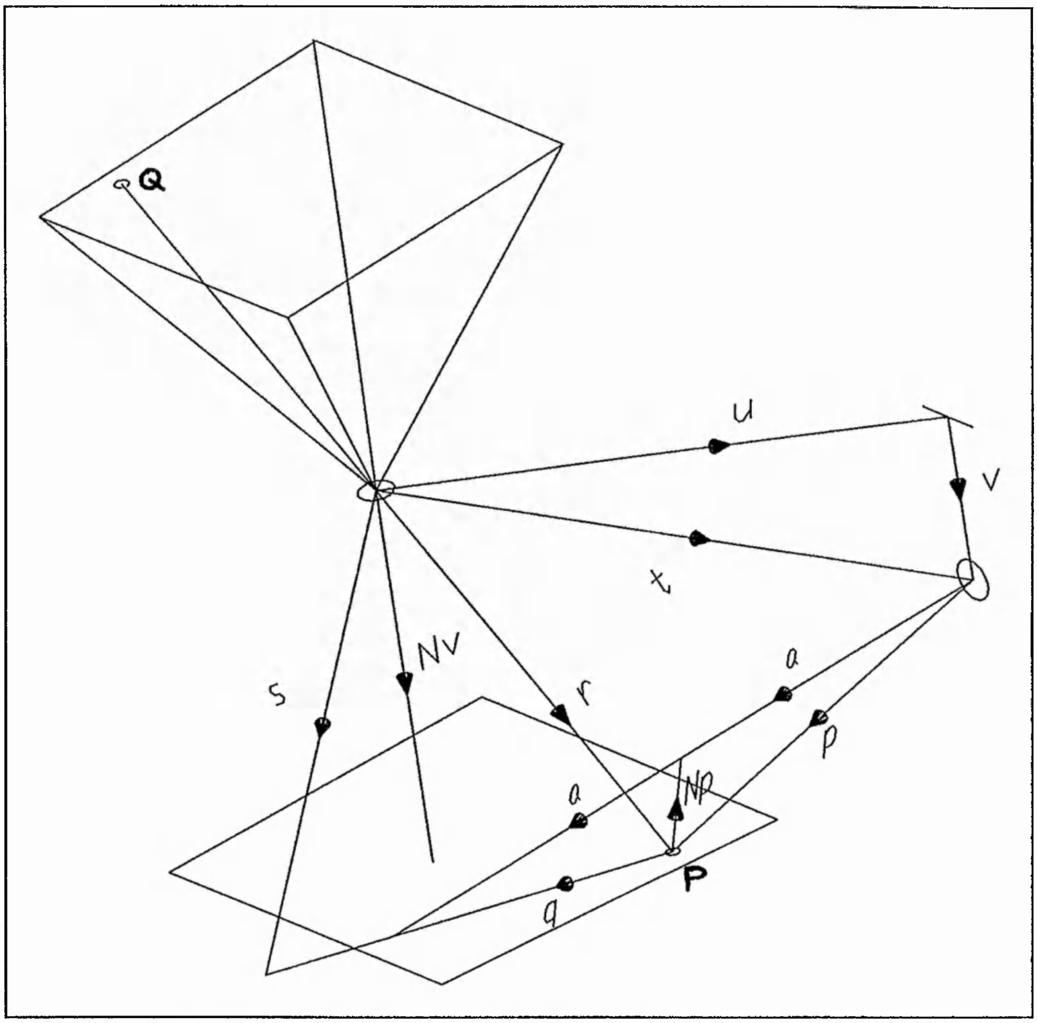


Fig 3.10 Vector representation

panel camera.

Suppose the point **Q** has pixel co-ordinates (x_r, y_r) in the internal camera co-ordinate system {Xr, Yr}. These co-ordinate values are directly available in terms of pixel offsets when the image trace has been identified by suitable algorithms. Thus, its co-ordinates in the 3D axis system {Xa, Ya, Za} are:-

$$x_a = x_r - 256$$

$$y_a = y_r - 256$$

$$z_a = f$$

Thus, the position vector of the point **Q** (in the pixel space {Xa, Ya, Za} of the camera) has components (x_a, y_a, z_a) and may be represented by a vector **w**. Now, imagine the same vector **w** (i.e. with these same components) but in the real space {Xe, Ye, Ze} shown in figure 3.8. Since the axes of the pixel space and the real space are directly opposed, it is clear that this vector **w** will have the same direction as the position vector **r** of the point **P**. Thus,

$$\mathbf{r} = \mu\mathbf{w}$$

where, μ is a scalar constant.

Since we wish to represent the points of the traces in the "world" co-ordinate system {X, Y, Z} (for reasons that may become clear later), rather than in {Xe, Ye, Ze}, both shown in figure 3.8, the components of the vector **w** need to be operated by a transformation matrix. The Y axis of the global co-ordinate system coincides with the Ye axis of the camera external co-ordinate system. The X axis of the global co-ordinate system subtend an angle τ with Xe and is parallel to the direction of motion of the table. The Z axis lies in the central plane and also in a plane passing through the lens perpendicular to the direction of motion of table. That is, a rotation of the co-ordinate system {Xe, Ye, Ze} by an angle τ about the Ye axis would make it coincide with the world axes {X, Y, Z}. This angle is the camera tilt angle referred to in section 3.2. Thus, the vector **w** with components (x_a, y_a, z_a) in the {Xe, Ye, Ze} space can be expressed as follows in the {X, Y, Z} space.

$$w = \begin{bmatrix} \cos\tau & 0 & -\sin\tau \\ 0 & 1 & 0 \\ \sin\tau & 0 & \cos\tau \end{bmatrix} \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix}$$

In the world co-ordinate system {X, Y, Z} with **i**, **j**, **k**, as unit vectors along the X, Y and Z axis respectively, most of the vectors shown in figure 3.10 can be determined. The vectors **Nv**, **u**, **v**, and **t** are fixed for a given inspection setup with laser angle α, and can be determined prior to inspection.

The unit vector **Nv**, which is parallel to the vector **v**, both of which are normal to the plane of the laser light is given by

$$\mathbf{Nv} = \sin(\alpha)\mathbf{i} + 0\mathbf{j} + \cos(\alpha)\mathbf{k}$$

Also, referring to figure 3.9 we have

$$\mathbf{u} = a_1\mathbf{i} + 0\mathbf{j} + a_3\mathbf{k}$$

and

$$\mathbf{v} = (\text{Laser radius})\mathbf{Nv}$$

Also,

$$\mathbf{t} = \mathbf{u} + \mathbf{v}$$

and the distance between the laser plane and the panel camera (which is a measurement along the normal **Nv**) is

$$\text{Laser distance} = \mathbf{t}.\mathbf{Nv}$$

It is clear that this distance too can be determined prior to inspection, knowing only the set up parameters, in particular the value of α.

For every point **Q** on the image trace, a corresponding vector **w**, scalar μ, and the position vector **r** of the corresponding point on the trace of the laser can be found from the above mentioned equation $\mathbf{r} = \mu\mathbf{w}$ where

$$\mu = (\text{Laser distance})/(\mathbf{w}.\mathbf{Nv})$$

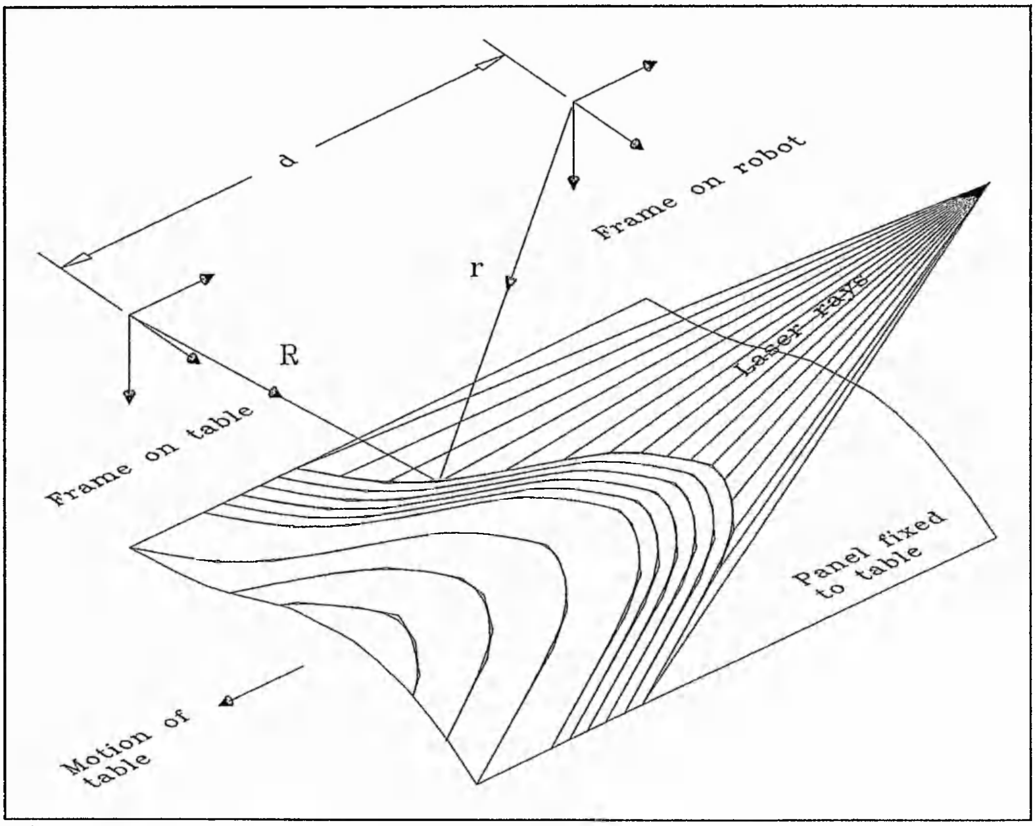


Fig 3.11 Surface data

Now, imagine a new frame of reference coinciding with the world co-ordinate system initially, but fixed to the table as shown in figure 3.11. Since the panel to be inspected lies fixed to the table, this new co-ordinate system can be thought of as being fixed to the panel. Thus, as the table indexes forward for the purpose of scanning the panel, this new co-ordinate system also moves with the table. Thus, when the table has moved a distance d, the position vector **R** of any point **P** on the panel trace with respect to this new system of co-ordinates will be

$$\mathbf{R} = \mathbf{r} - d\mathbf{i}$$

Thus, we obtain sets of 3D co-ordinates of the panel surface along the panel traces, as the table indexes from one position to another relative to this frame of reference fixed to the panel.

3.5 Screen camera trace

If the direction of the surface normal at point **P** of the panel (denoted by the unit vector **Np** in figure 3.10) is assumed to be known, then the position vector **s** of the point **S**, where the reflected laser ray meets the screen can be predicted. Noting that the vectors **p** and **Np** are on the plane of reflection of the laser ray, and also that the vector **a** is perpendicular to **Np**, the following system of equations can be established:-

$$\mathbf{p} = \mathbf{r} - \mathbf{t}$$

$$\mathbf{a} = \mathbf{p} + (\mathbf{p}.\mathbf{Np})\mathbf{Np}$$

$$\mathbf{q} = \mathbf{a} + (\mathbf{p}.\mathbf{Np})\mathbf{Np}$$

that is,

$$\mathbf{q} = \mathbf{p} + 2(\mathbf{p}.\mathbf{Np})\mathbf{Np}$$

$$\mathbf{s} = \mathbf{r} + c\mathbf{q}$$

where, the scalar constant c is given by

$$c = -(\text{Screen distance})/(\mathbf{q.i})$$

The above equations indicate that **s** is dependent on **r** and **Np**. In other words, the trace on the screen is a function of both depth and slope of the panel along the trace on the panel. Thus, it is assumed that the screen trace embodies essential cosmetic information.

3.6 Formulation of cosmetic data

In 3.4 it was shown that 3D surface data of the panel could be gathered from the panel camera image alone. However, this data contains only depth information that is required for cosmetic quality assessment and does not contain the important ingredient - the slope information. Noting that the screen camera trace is a function of both these quantities as described in 3.5, we replace the depth data in the panel camera trace (i.e Z co-ordinates) with data from the screen camera trace, while preserving the other two co-ordinates, thus enabling the mapping (or superposition) of the cosmetic effect on the projected panel surface (on the XY plane).

This requires some assumptions to be made with respect to the correspondence of points in the two traces. If it is supposed that both traces have the same number of pixels then direct correspondence between pixels may reasonably be assumed. The fact that the sequence order of these corresponding pixels may be different and that uniform pixel increments in the Y direction in one trace need not necessarily produce uniform increments in the other (i.e the correspondence may be non-linear) are ignored. (This may become clear if the paths of the individual laser rays are visualised, as it would then be seen why these rays need not be uniformly spaced after reflection, especially when they impinge on a defect. Also the possibility of rays crossing one another before reaching the screen can be imagined). By ignoring these effects what is assumed here is that the non-linear cosmetic effect induced in the Y direction are insignificant and a linear relationship may be used. This assumption is reasonable since the panel surface normals that determine the reflection of the rays vary only by a small angle from the Z-direction even in the regions with defects.

Yet, in practice such direct correspondence cannot be achieved in all situations. With sufficient care taken in locating the screen and cameras, close correspondence of the traces (in the region of 80% or higher) may be may be achieved for objects with slowly varying surfaces. However, in order to obtain full correspondence, a scaling operation is necessary, and the screen trace is divided in a proportionate manner in the Y-direction. For a given inspection pass, the constant ratio involved is evaluated using the set of traces of both screen and panel images. The ratio (r) of the support of the screen trace to the support of the panel trace determines this proportionality. By support is meant the pixel count (or distance in the Y direction) from starting point to ending point of a given trace. Since the individual supports of the same type of images in the set may vary slightly, some central value has to be calculated. For the screen trace, the mean seemed suitable, while for the panel trace, the mode was more appropriate. (Also the mean and mode starting values of the traces - a and b - are calculated for later use). The depth value (i.e. the Z-value) of the panel trace for a given Y-value is replaced by the Z-value of the screen trace for the corresponding to the appropriate Y-value determined using the above proportionality. Noting that these trace data are available as arrays in the computer program, the following equation updates the cosmetic data:

$$Z_{Panel}[i] = Y_{Screen}[a + r(i\text{-}b)]$$

where, a and b are the starting (i.e. the first pixel) indexes of the screen and panel traces respectively mentioned above.

3.7 Cosmetic map

For practical purposes the modified cosmetic data described above may be treated as 3D real space data. The collected data may be used to obtain what may be called the "ideal cosmetic surface" (the use of which may become apparent in the ensuing). The technique of least square fitting of polynomial surfaces [Lancaster 86] is particularly suited for obtaining the ideal surface (and has been used by others with image or range data [Eden 86], [Silverman 86]). However, a prerequisite for this method is the assumption of the degree and form of the polynomial surface that is to be approximated. A bivariate bicubic polynomial is chosen as being sufficient for the representation of the ideal smoothness of the cosmetic surface. Given the slowly varying nature of the surfaces that were considered, this assumption was found to be appropriate. The approximation error with lower order surfaces was high, and higher order surfaces did not significantly reduce the error.

Once the ideal cosmetic surface is determined (as shown in the next section), it is possible to imagine the deviations of the actual data points from this surface. Now, if we *interpolate* surface patches in local regions in a suitable manner, the deviations between the interpolated patch and the (globally approximated) ideal surface can be evaluated (and if necessary shown graphically with different colours for different ranges of deviation). Thus, what we seek is an interpolation scheme for the collected data points. In computer-aided surface design several techniques such as Bezier, B-spline etc. are available for computing such surface patches [Farin 90]. Most of these techniques although computationally expensive are best suited for grid based somewhat sparse data. The data in our problem is dense in the Y-direction (as they are of contiguous pixels of a scan) and relatively sparse in the X-direction (as there is a significant gap between adjacent scans). Exploiting this fact we interpolate the surface by a series of planer piecewise cubic splines with slope continuity - i.e. C^1 - in XZ section planes at valid Y-coordinates of the data points as shown in figure 3.12. During computation of the spline curves the deviations can be integrated for later use in labelling the quality of the panel. However, since minor deviations are unimportant in practice, a user definable upper and lower tolerance from the ideal surface is implemented such that only deviations outside this tolerance band are considered as corresponding to physical panel defects, and as suggested above, are displayed on a computer screen with 15 colours for the visualization of the defects.

3.8 Least square fitting

A bivariate bicubic polynomial of the general form,

$$p(x,y) = a_0+a_1x+a_2y+a_3x^2+a_4xy+a_5y^2+a_6x^3+a_7x^2y+a_8xy^2+a_9y^3$$

is chosen for approximation. Assuming n+1 surface data points (x_i, y_i, z_i), i = 0,1,... n are given, with $n \geq 9$ the difference between the surface elevation and z_i at point i is $p(x_i, y_i) - z_i$. In the least square method the coefficients $a_0,a_1,...a_9$ are to be determined such that the total squared error represented by the function

$$E = \Sigma(p(x_i, y_i) - z_i)^2$$

is minimised. Thus, considering E as a function of the coefficients $a_0,a_1,...a_9$, it will be a minimum only when all the 10 partial derivatives with respect to the coefficients are simultaneously zero. i.e $\partial E/\partial a_i = 0$ for i = 0,...9. This yields 10

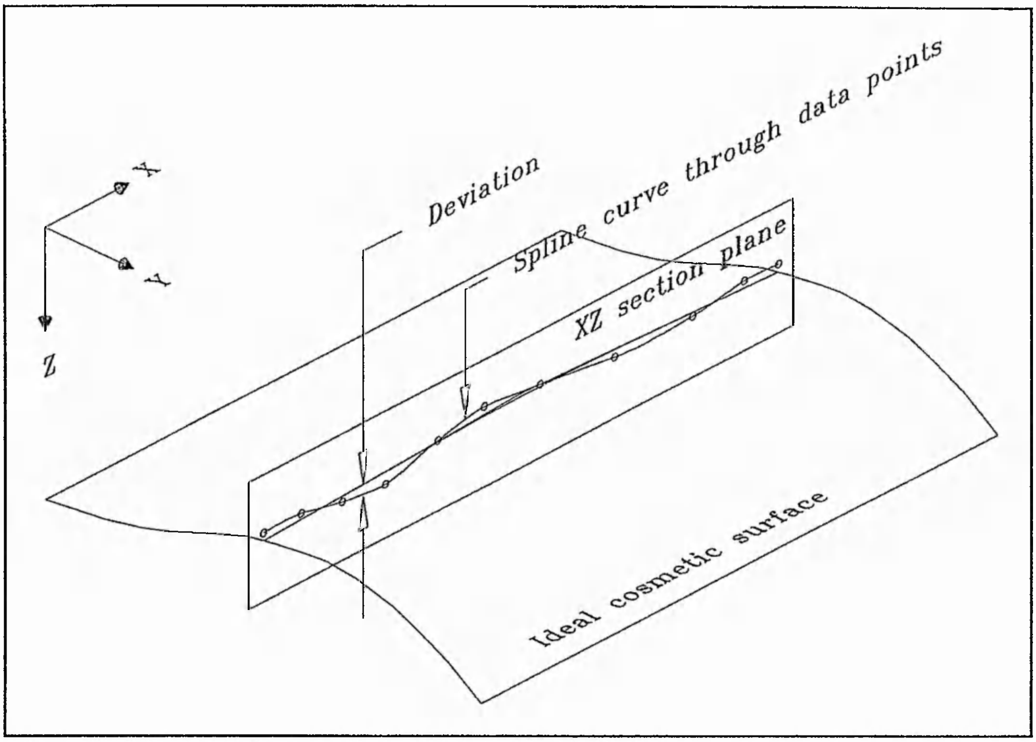


Fig 3.12 Ideal surface and deviations

simultaneous equations for determining the 10 coefficients. These 10 equations can be conveniently represented in matrix form as

$$\begin{bmatrix} v_{00} & v_{01} & \cdots & v_{09} \\ v_{10} & v_{11} & \cdots & v_{19} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ v_{90} & v_{91} & \cdots & v_{99} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_9 \end{bmatrix} = \begin{bmatrix} f_0 \\ f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_9 \end{bmatrix}$$

where the first 10x10 matrix is referred to as the Vandermonde matrix which turns out to be symmetrical due to the polynomial form that was assumed with several other common elements as well. (These properties are exploited in the computational algorithm). The elements of the Vandermonde matrix are as shown below where the summations are over all the n+1 surface data points:

$v_{00} =$	$\Sigma 1$
$v_{01} = v_{10} =$	Σx
$v_{02} = v_{20} =$	Σy
$v_{03} = v_{30} = v_{12} = v_{21} =$	Σxy
$v_{04} = v_{40} = v_{11} =$	Σx^2
$v_{05} = v_{50} = v_{22} =$	Σy^2
$v_{06} = v_{60} = v_{13} = v_{31} = v_{24} = v_{42} =$	Σx^2y
$v_{07} = v_{70} = v_{15} = v_{51} = v_{23} = v_{32} =$	Σxy^2
$v_{08} = v_{80} = v_{14} = v_{41} =$	Σx^3
$v_{09} = v_{90} = v_{25} = v_{52} =$	Σy^3
$v_{16} = v_{61} = v_{28} = v_{82} = v_{34} = v_{43} =$	Σx^3y
$v_{17} = v_{71} = v_{26} = v_{62} = v_{33} = v_{45} = v_{54} =$	Σx^2y^2
$v_{18} = v_{81} = v_{44} =$	Σx^4
$v_{19} = v_{91} = v_{27} = v_{72} = v_{35} = v_{53} =$	Σxy^3
$v_{29} = v_{92} = v_{55} =$	Σy^4
$v_{36} = v_{63} = v_{47} = v_{74} = v_{58} = v_{85} =$	Σx^3y^2
$v_{37} = v_{73} = v_{49} = v_{94} = v_{56} = v_{65} =$	Σx^2y^3
$v_{38} = v_{83} = v_{46} = v_{64} =$	Σx^4y
$v_{39} = v_{93} = v_{57} = v_{75} =$	Σxy^4
$v_{48} = v_{84} =$	Σx^5
$v_{59} = v_{95} =$	Σy^5
$v_{66} = v_{78} = v_{87} =$	Σx^4y^2
$v_{67} = v_{76} = v_{89} = v_{98} =$	Σx^3y^3
$v_{68} = v_{86} =$	Σx^5y
$v_{69} = v_{96} = v_{77} =$	Σx^2y^4
$v_{79} = v_{97} =$	Σxy^5
$v_{88} =$	Σx^6
$v_{99} =$	Σy^6

The elements of the column vector (which shall be referred to as **f**) appearing on the right hand side of the equation are:

$$f_0 = \Sigma z, \quad f_1 = \Sigma xz, \quad f_2 = \Sigma yz, \quad f_3 = \Sigma xyz, \; f_4 = \Sigma x^2 z,$$
$$f_5 = \Sigma y^2 z, \; f_6 = \Sigma x^2 yz, \; f_7 = \Sigma xy^2, \; f_8 = \Sigma x^3 z, \; f_9 = \Sigma y^3 z$$

Using matrix notations the matrix equation can be written as

$$\mathrm{V}\mathbf{a} = \mathbf{f}$$

where **a** is the column vector (of polynomial coefficients), appearing on the left hand side.

The method of lower-upper matrix decomposition was chosen for the solution of the above equation for determining **a**, as this is a computationally stable method than those involving the computation of the inverse of the Vandermonde matrix V.

One of the problems that was encountered during the development phase was the noise in the data due to poor optics. In an attempt to eliminate this the following method was adopted. Noting that a good solution for **a** could be evaluated using only 10 good representative points of the data set, an estimate of E is evaluated. Subsequently, as points are included, **a** is recomputed and the change in E is noted. If this change exceeded a given threshold, the point is discarded as noisy data and its effects in the various summation terms described above are reversed. Though the method performed satisfactorily, the time penalty was significant. As it was felt that the best way to solve the problem was to avoid noise in data, this was the approach that was taken. It was solved by employing suitable filters for the cameras and by setting by software means the sensitivity parameters of the vision hardware.

3.9 System operation and user interface

Some of the broader outlines of usage of the tool are discussed here. The algorithms for the control and functionality of the system were written in Microsoft C. User access to the system consists of a hierarchy of menus with context sensitive help. A menu item may be activated by moving a highlighted menu bar to it, using cursor keys, and selecting it, or by pressing the unique highlighted "hotkey" keyboard character.

3.9.1 Menu structures

Two types of menus are used - one for performing tasks and the other for user input of values used internally by the system. We shall refer to these two types as task menus and edit menus. As the system was to be used by relatively computer illiterate personnel, an intelligent user interface was required. This is particularly important with edit menus where inappropriate inputs are to be disallowed. As such, in the implementation of the edit menus, appropriate allowable ranges of values for each menu item was built into the system, and user input for changing system values was accepted only after checking against this range. If the values were outside the range, an audible warning beep followed by a message reminding the acceptable range was displayed. Control was exercised even at a lower level than this. Whenever the user attempted to change a value in an edit menu, the keystrokes pressed were monitored by checking against a restricted set of permitted keys for the particular menu item, and a message was displayed with warning beep if the key pressed was inappropriate.

The context sensitive help makes use of an ASCII file to provide the help information. This written in a simple but special format in order to locate the relevant help information easily. Such an approach was chosen to allow easy modification of the help information tailored to user requirements.

3.9.2 Start up and safety

On start up an edit menu is displayed requesting for a name of a configuration file. A configuration file is one written by the system at a previous instance, containing all system parameters that are necessary for an inspection such as, robot positions for the various passes during inspection, actuation speeds of its joints, vision parameters such as thresholds, gains offsets etc., of each camera, number of passes required for the complete inspection of a panel, the number of scans per pass, the distance between indexes, system flags etc. (Setting up configuration files requires skill in manoeuvring the robot and sufficient understanding of the system parameters. However, when such files are available, the process of inspection can be carried out by personnel with lower levels of skill). If such a file is not available system defaults are used. The name of a configuration file is easy to remember as it uses the model name of the car and the panel type to be inspected. Once a configuration file is loaded, the system performs checks on the communication links with the controllers and the motor drives of the robot and table and indicates their readiness. Then the main system menu is displayed. (The system also initialises an appropriate data base for storing as well as for providing assistance with decision making on panel quality as discussed in chapter 6). At this stage the system is ready to perform a complete inspection of a panel located on the table without any manual intervention as will be described later. Pressing of any key of the keyboard at any stage of the inspection is interpreted as a panic situation, and the inspection process is halted by stopping the motion of all motors with brakes being applied to them, abandoning the inspection cycle, and then passing control to the user.

The software has proved to be reliable and satisfactory in providing complete control at all times.

3.9.3 Main menu

The task type main menu provides facilities for setting up system parameters either by loading an appropriate configuration file or by modifying values in a hierarchy of edit menus. It also allows saving of these parameters in user named files. As was explained previously, the laser traces provide the basic data of the inspection which is subsequently analyzed for detecting defects. Thus, facility is provided by way of a flag in an edit menu for the user to automatically save if required, the trace data of an inspection, for later analysis with different system parameters such as panel quality threshold (or for experimenting with other algorithms by the system developer).

The main menu also accomplishes tasks such as performing an inspection cycle, graphically displaying any previous inspection results saved by the user, re-evaluating surface quality using previously saved trace data but with different parameter settings, "teaching" the robot new positions that are to be saved in configuration files for later use in inspection, and showing one at a time the live video output of the two cameras (for focusing and assisting in robot positioning during teaching).

3.9.4 Robot teaching

Selecting the robot teaching facility from the main menu initiates a screen with a robot task menu along with a display in another part of the screen of current robot motor positions and other information (such as the total number of robot positions, i.e. configurations, known to the system, and the currently selected position etc). This robot menu allows the robot to be moved on a joint-by-joint basis until it configures to a position that is deemed as suitable for an inspection pass. This is done by selecting a joint, numbered 1 to 4 for the four axis of the robot, and using one of the cursor keys to indicate the direction - whether clockwise or anticlockwise - whereupon the selected drive is fully energised, its solenoid actuated brake is released, and the motion is started. Pressing any key at this stage stops the motion of the drive, applies its brake, and shifts the drive to an idle, low power mode (This is a standard sequence of events programmed for three of the four robot drives; the table drive and the laser drive being the exceptions as they did not require brakes). Importantly, the robot menu also allows saving of this position, referred to as teaching a point, which amounts to saving of four motor step (or co-ordinate) positions, in a sequential list in memory (to be ultimately saved in a configuration file with the order of saved points being the order to be followed later during an inspection cycle). More specifically, a position reached by the robot can be appended to the list, or inserted at a chosen point in the sequence, or can be used to overwrite a previously saved position (now deemed as unsuitable). Positions saved in the list can be selected and deleted with automatic updating of the list. Motor step counts (co-ordinates) of a saved position can also be selected for screen based editing (for fine tuning) if required. A command in the menu also allows the robot to be moved to a previously saved position in the list (which simultaneously actuates all the drives of the robot in order to reach the position), which allows for quick reassessment and if necessary modification of saved points. Here again, in view of safety, any key press would halt the motion of the robot. The menu also allows the accelerations and (move and home) velocities of the robot drives and table drives to be set up via an edit type sub-menu (which are also saved in a configuration file).

Another command in the menu allows hardware reset of the robot. This makes use of hardware limit switches incorporated into the robot to provide signals for setting the datum positions of the drives. Such a requirement may arise when the drives, being stepper motor based without feed back, loose steps due to overload. Since the laser drive did not have a brake, such a reset was also recommended every time the robot was turned on. A much faster but software based reset facility (called Home) is also provided, which simply drives the robot to the position which has zero steps on all four drives (without regard to lost steps).

3.9.5 The Inspection cycle

The inspection process is initiated by selecting the Autoinspect command from the main menu. It is assumed that all system parameters are set to appropriate values (usually by prior loading of a configuration file). The vision system is then initialised with appropriate settings, and the communication links with the robot are checked for readiness. The first position for the robot to move to is selected from the sequentially saved inspection point list. This need not always be the first point of the list. It could be any point in the list and is determined from a user specified index in an edit menu. Perhaps it is worth noting that the last position the robot is to move to also need not be the last point in the list, and is similarly determined from another user specified index. The robot moves from the first to the last specified position taking all the intermediate points in the list in sequence. At any one of these positions an inspection cycle which we shall refer to as a pass is performed. Thus, if for example, there are five robot positions 2 to 6 specified in the list (containing say 8 points), then five inspection passes would be performed, the first with the robot at position number 2, the next at position 3, and so on until the last at position 6.

The events in a typical inspection pass is described here. At the start, the robot is usually at its home position. Once the position the robot is to take is known, the laser axis alone is first moved to its final position. The laser trace at this stage directly falls on the screen as a straight line and is captured by the screen camera. (The orientation of the screen camera is such that this line appears as a vertical line in its image. It may be noted that after the laser axis has reached its final position, the laser head as a whole, including the cameras, are in a fixed configuration, and any motion of the other 3 axis of the robot will not alter the position of the laser line as imaged by the screen camera). This line is used as a datum for calculating the (horizontal) image offsets of future screen laser traces formed due reflection on the panel. This is important in order to separate the reflected part of a screen laser trace, which is of interest, from a possible non-reflected part. The non-reflected part of the trace arises due to some of the laser rays not impinging on the panel surface (as is the case near the edges of a panel or when punched holes are such as door handle apertures are encountered). Once the datum for the screen laser trace has been determined, the other three axes of the robot are actuated in order to move the robot to the selected position and inspection proper starts. This is essentially a process of gathering sets of corresponding laser traces data of screen and panel camera images as the table (and hence the panel) is indexed from one position to another. The number of sets to be gathered and the indexing distance of the table between the sets are common to all passes and user specified. At the end of a pass, cosmetic data is composed, the defects detected are shown (briefly) graphically as a colour map (shown in fig 3.13), synthesized defect information is saved for future use (as explained in chapters 5 and 6) and the robot moves to its home position, in readiness for its next inspection pass.

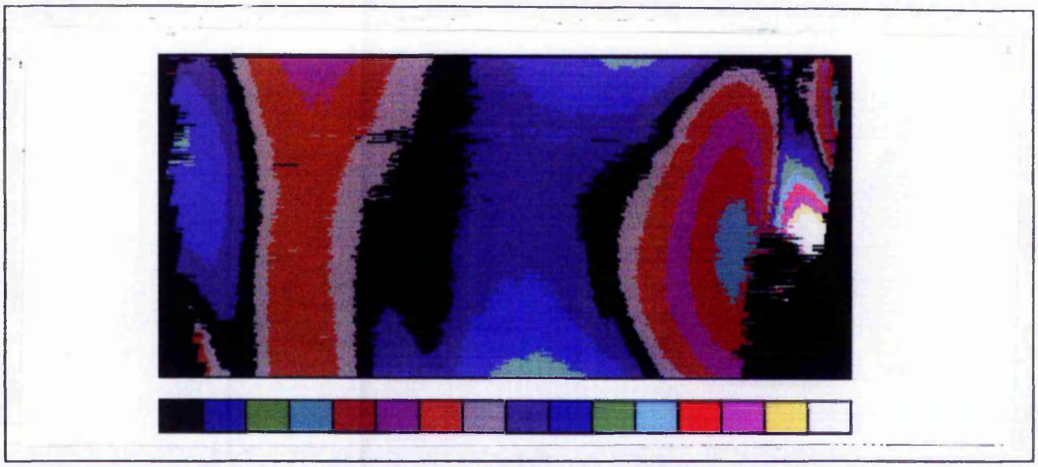


Fig 3.13 Cosmetic map of defects from a pass

When all the specified passes have been performed in sequence, a full graphical representation of the complete cosmetic map of the inspected panel is presented. At this stage, the system generates a special code to classify the quality of the panel (as will be described in later chapters), and prompts where possible what the outcome of the inspection should be - whether accept, rework or reject - and awaits the decision of the user, which is subsequently saved in a database. The system is now ready for inspecting another panel. The complete inspection requires approximately 10 minutes.

3.10 Evaluation

The system met the desired capabilities set out in section 3.1. The repeatability of inspection results was be evidenced by the cosmetic maps generated by the system, and was found to be satisfactory. The portrayal of the cosmetic maps was acknowledged by panel inspectors to be realistic and sufficiently detailed in the description of defects. The accumulated cosmetic data could be used for further analysis such as defect identification, defect quantification, or cosmetic quality assessment by suitable computer algorithms.

An important feature of the inspection system is that it does not require any external data such as design, or CAD data of the surface under inspection, for the detection of defects. (Note that manual inspection also does not require external data). This is because an ideal cosmetic surface is evaluated by the system for every panel inspected. Due to this reason, the positioning of panels on the table is also not critical. In practice, the panels are simply placed on top of a former fitted to the table and the positional error which is judged to be of the order of 5mm maximum, has no noticeable effects on system performance as could be seen from the similar looking cosmetic maps. The system was found to be suitable for inspecting the complete range of door panels of the collaborating car manufacturer. The common feature of these door panels is that their surfaces vary very slowly. Since the system makes use of light reflected from the panel falling on a screen, such slowly varying surfaces are suitable. With highly curved panels (such as the front wing) the reflected light may sometimes fall outside the screen and hence the system can only be used to perform partial inspection.

The robot and linear table systems (designed by the author and manufactured in the workshops of the university) are stepper motor driven and are controlled by individual indexing units, all of which are housed in a control unit. These indexing units which are also individually addressable, are under computer control via a common RS232 communication link. Since positional feedback is not used, the positional accuracy of the system relies on the complete execution of the commanded motor steps. Under overload conditions, as can occur during teaching when the robot accidentally touches the surface of a panel, one or more stepper motors would "loose" steps, and the robot will need to be reset (i.e. hardware reset). Driving the robot very close to panel surfaces, in order to obtain a suitable optical set up, was noted to require a some degree of skill.

Inspectors had some difficulty in acquiring this skill. Some amount of mechanical backlash was observed in the robot. This was not critical since the same amount of backlash was involved both during teaching and execution to a position.

The laser traces were detected only with very low threshold values (typically 15). This problem was sometimes prominent with the panel trace. This is because of the varying levels of reflectivity of panels and angle of incidence of the laser, both of which when high do not promote the scattering the laser light impinging on the panel. As the power of the laser was only 5mW, and as the beam was fanned out at 40 degrees to throw a laser sheet of light, the low threshold values are not surprising. Though it was possible to use the 5mW laser with a low threshold, a more powerful laser is desirable in order to avoid image noise, enhance the robustness of the system, and even simplify some of the imaging processing algorithms. However, this was decided against as the system was only experimental and laser safety considerations outweighed any advantages at this stage.

The time taken by the system for inspecting a panel (10 min) is felt to be excessive. Nevertheless, this was considered to be unimportant for this novel experimental system. Limitations of memory and graphical capabilities of the computer system (operating under DOS) was also noted. This meant that the system could scan only up to 128 scans per pass (due to restrictions in crossing of 64K memory boundaries, which can be overcome if necessary), though this was sufficient for present purposes, and the graphic output to the 16 colour VGA screen was not very informative due to the limited number of available colours.

Chapter 4

Inspection knowledge acquisition

4.1 The need for acquisition of inspection knowledge

The importance of the role of decision making in industrial inspection was mentioned in chapter 1. In the case of panel inspection the complexity of the role in assessing cosmetic quality has been highlighted in section 1.2.1 and the elusive nature of the term cosmetic quality has become apparent. What is known is that it serves as an indicator of rank, the derivation of which is submerged in human perception. That it depends on two levels of subjective assessment - firstly, in the assessment of a single defect, and secondly, in the judgement of overall quality based on a variety of defects in different areas of a panel - has nevertheless been inferred. But further understanding of how these assessments and judgements are made is clearly required, especially if the term cosmetic quality is to be used in a scientific context.

Such an in-depth understanding is also called for if we wish to bring panel inspection within the realms of proper industrial inspection and automation. For, if we attempt to elicit the inspection knowledge with a view to formalising the rules involved in the process, then it may be possible to implement these rules in a knowledge base so that the process of inspection can be fully automated by taking advantage of this knowledge base. Artificial intelligence (AI) which is the study of computer implementation of human abilities considered as 'intelligent', with its implication of abstraction of human abilities may have solutions to offer. Such solutions, it is hoped, would enable the issues of subjectivity to be overcome and the decision making process in cosmetic quality to be automated.

Should this not be possible due to inadequate understanding of the domain of inspection knowledge, it may still be possible to set the lower goal of implementing an "expert assistant" for assisting the inspectors with the decision making process. Here, we may borrow ideas from expert system methodology for possible solutions. In industry, expert systems have been used in this manner as "intelligent operational assistants" to plant operators [White 87],[Kerridge 87]. An expert system in this role serves as a 'coach' to the operators making the expertise of more experienced operators available to all operators in the plant. The benefits are:

- lower training costs,
- better trained operators
- less inconsistent operations across operators and shifts
- fewer errors in judgment,
- fewer processing upsets and
- operational expertise 'captured' by the expert system, can be considered as a company resource, which can be used even if the individuals from whom the knowledge was derived leave the plant.

All the above approaches require analysis of knowledge in the inspection domain. In this connection the tools developed in the fields of knowledge acquisition and knowledge elicitation may be relevant.

In the worst case where an expert assistant cannot be implemented, "examples" could be presented to the mechanical inspection system, as in the case of trainee inspectors, and salient, synthesized output data from the inspection stored in data bases along with inspectors' judgement of the outcome for subsequent "knowledge discovery" by appropriate algorithms. This would then yield the knowledge (patterns, rules or other structures) involved in inspection, enabling full or partial automation. This approach belongs to the study of machine learning (ML) pursued actively by researchers in AI when frustrated by the limitations in conventional methods of acquisition of knowledge.

Alternatively, a neural network approach can be taken with a training set based on the above "examples" to train the network. Such networks have the capacity to learn from examples and represent generalised knowledge in an implicit manner, mimicking the brain.

4.2 Directions in knowledge elicitation

Knowledge acquisition (KA) and knowledge elicitation (KEL) are active areas of research in knowledge engineering(KE). Knowledge elicitation (which is a subset of knowledge acquisition) deals with acquiring information directly from domain experts while knowledge acquisition deals with collection of information from domain experts as well as from other sources leading to the formulation of the basis of a functioning knowledge base [Greenwell 88]. These knowledge bases are often simple data bases which store the facts pertaining to the domain in a declarative format.

Several types of knowledge representation schemes are available for developing knowledge bases [Barr 81], [Waterman 86]. Some of these are production systems [Newell 72], frame [Minsky 75], semantic network [Quillian 75] or logic [Green 69] based systems. There is no ultimate scheme for all purposes, as every scheme has its own advantages and disadvantages. Since the efficiency of an expert system is largely affected by how its knowledge bases have been developed, an important item in an expert system development is the selection of suitable knowledge representation schemes, [Kiyoshi 89],[White 87],[Kerridge 87].

The systems making use of the knowledge bases use one version or other of what is called an inference engine which is supposed to provide the "intelligence" for the system by way of its ability to search and manipulate facts in the knowledge base to suit user requirements. (Inference engines are typified by the techniques they use: forward and backward chaining, Bayesian probabilities, fuzzy logic, etc.). From this description we can infer that such systems at best can only emulate logical behaviour and not internal thought processes of domain experts.

Knowledge acquisition/elicitation fields have lagged behind the implementational aspects of expert systems, which is understandable, given the inherent difficulty in acquiring knowledge from domain experts. Nevertheless, they have established a powerful battery of techniques for acquisition of knowledge in different situations.

Some of the problems with elicitation of expertise are:

- its dependence on critical assumptions which are often implicit [Hawkins 83].
- its inaccessibility due to the expert not being aware of its significance [Collins 85] and
- that much human activity is inaccessible to awareness [Dixon 81].

Verbal data collection by interviewing experts i.e. linguistic transmission of expertise is an important aspect in knowledge elicitation. However, the lack of correlation between verbal reports and mental behaviour has been noted [Bainbri 86].

Knowledge acquisition is often not an exercise performed only once during the development of a knowledge based system. This is because when systems are implemented, failures in performance often arise necessitating system modification which leads to further lengthy processes of knowledge acquisition. This is referred to as the "knowledge acquisition bottle-neck" which Diaper [Diaper 89] identifies as principally due to "apparently intractable problems associated with our ignorance of the real psychology of human knowledge representation". For how experts formulate indescribable conceptual objects for solving some of their problems shall never be known.

The difficulty and high cost of knowledge elicitation has led to the development of automatic systems for knowledge elicitation. Three main approaches are: induction, repertory grids, and documentation systems. So far no useful tools have been produced by any of these approaches. One of the main reasons for this is the inadequacies in the inferential power of these systems which lack or are unable to take into account the common sense element inherent and taken for granted in humans.

Some researchers have attempted to eliminate the knowledge acquisition bottle-neck by employing partially automatic knowledge acquisition processes. For example Mitchel *et al.* [Mitchell 89] have developed interactive Learning Apprentice Systems which they define as "interactive knowledge-based consultants that directly assimilate new knowledge by observing and analyzing the problem solving steps contributed by their users through their normal use of the system". These systems require an initial knowledge base implementation which is refined by adding new rules or by modifying incorrect rules. Some other similar systems are PRODIGY [Carbonell 90], and APT [Nedellec 92]].

These systems integrate knowledge acquisition and machine learning methodologies.

Recent work in knowledge engineering, directed at eliminating the above bottle-neck, has taken a different perspective. The knowledge acquisition process is now viewed as a domain modelling problem rather than a straight forward acquisition problem [Nwana 92], [Heijst 92]. In this respect european effort has concentrated on laying down guidelines for model building. KADS (ESPRIT project 1098) and KADS-II (ESPRIT project 5248) are efforts in this direction. KADS methodology deviates from the common approach of "rapid prototyping" as it insists on thorough analysis of the problem prior to design and coding. The central view in KADS is that KBS development is a modelling process leading to two models; the conceptual model and computational model. The conceptual model consists of four layers; domain layer, inference layer, task layer and strategic layer. (The later three are also referred to as control layer). To aid the knowledge engineer with knowledge acquisition, KADS also provide "interpretation models" in the form of libraries, to serve as templates for the expertise. In this manner it guides the knowledge acquisition process by giving the knowledge engineer some indications (from the abstract interpretation model) which information is still to be collected. In recent literature several projects have been reported based on KADS methodology [Heijst 92], [Jonker 92], [Neubert 92], [Porter 92]. KADS-II project aims to modify, improve and overcome some deficiencies in KADS.

The above model-driven knowledge-base building paradigm is being further extended by authors to include machine learning capabilities [Tsujino 92], [Nedellec 92]. Current expert systems (that are model-driven) are classed as second generation systems. The first generation systems relied purely on heuristic knowledge as expressed by rules. The second generation systems have in addition what is known as a deeper model which gives them an "understanding" of the entire search space over which the heuristic rules operate. The advantages sited by Steels *et al.* [Steels 89] are:

- less brittle systems (i.e. no sudden failure but graceful degradation),
- better (i.e more convincing) explanation capability directly due to deeper model,
- exhibit learning behaviour (in the sense they are able to acquire new heuristics).

They further classify models into;

- anatomical (component/part/whole relationships),
- functional (component function based relationship),
- causal (component property based causal relationships).

Causal models are recent in knowledge engineering literature [Charlet 92]. Bratko *et al.* [Bratko 89] describe two projects they have undertaken to automate the synthesis of knowledge. In one, called KARDIO for the diagnosis and treatment of cardiac arrhythmias, they use a qualitative causal model. The other, called Assistant, is an inductive learning system for use in various areas of medical diagnosis and prognosis, modelled along the lines of Quinlan's ID3 program (with several improvements added), as a decision tree, and learns by examples.

4.3 Psychological perspective

In this field another perspective exists: that of the psychologists. Since its inception, artificial intelligence has had close links with cognitive psychology which has grappled with the problem of expertise elicitation for several decades

[Nisbett 77],[Broadbent 86]. Artificial intelligence attempts to model aspects of human thought processes in a computer program and there are parallels between this and cognitive models. Again, knowledge engineers like psychologists employ identical approaches in reasoning, problem solving strategies and analysis of social situations (such as a typical knowledge elicitation session). Also expertise has a psychological basis. It involves perception, action and language as interfaces with the world; memory, awareness, intentions and emotions as internal phenomena and problem solving skills; all of which exist in a social context. According to Norman [Norman 80], some of these processes go beyond the basic information models of cognitive science, but are nevertheless essential to knowledge acquisition.

Many psychologists feel that verbal reports and data especially in complex situations are of little use. Bainbridge [Bainbri 79][Bainbri 86] supports this view by noting that there is no necessary correlation between verbal reports and mental behaviour. Following Freud, clinical psychologists offer an explanation for this lack of correlation. According to them it is a process of cognitive defence that impedes internal communication. They have developed verbal, interactive techniques to bypass cognitive defences, including those resulting from automization of skilled behaviour, to identify underlying cognitive processes.

4.4 Problematic domains

However, whether all forms of expertise are suitable for analysis by knowledge elicitation techniques in the context of expert systems in its present stage of development is questionable, though most authors on the subject would lead us to believe that this is not the case, and that we should intelligently embark on elicitation armed with expertise in psychology and computer science.

What we seek is the capturing of behaviour of experts (which we assume to be rational). This pre-supposes our ability to represent the behaviour by language constructs. Where a language is inadequate for the description of behaviour in the domain, logically, elicitation has to fail as there is no way for the transmission of knowledge from one person to another. Where an expert can articulate domain behaviour, the knowledge engineer has little role to play. But often this is not the case, as experts do not often examine their domain behaviour to realise the structure of logic, or heuristics they employ when using their expertise. In these instances the knowledge engineer can bring forth the obscure knowledge by the use of his or her special skills. The common element in the success of knowledge elicitation and implementation of knowledge bases must be the discovery of the underlying logical structures that govern the use of the domain knowledge, irrespective of how obscure they may be to the practitioner.

In the context of knowledge elicitation in AI what do we mean by knowledge? Portmann and Easterbrook [Portman 92] address this question and say,

> Many approaches to knowledge acquisition have arisen based on various perspectives on the nature of knowledge (see...for a survey), and experience with these has led to a growing realisation that knowledge is not some objective essence to be mined and refined. Rather, knowledge seems to change its shape depending both on the task to which it is applied and the social setting. An extreme version of this view is expounded by Winograd & Flores (..) who suggest that knowledge is socially constructed, and that it is in the act of communication that knowledge gets formulated.

Surely, one of the aspects it should have is its utility value in AI. This begs the question what then is usable knowledge? This must be the sub-set of propositions of the domain that are useful. Propositions make use of objects of the domain. The objects are complex abstractions or symbols which convey socially accepted meanings of the objects.

We can conceive of domains where mental structures, if they exist at all, are impossible to see with our current understanding. How one judges beauty or balances in riding a bicycle, or swims, or simply pronounces a letter are intractable problems for any science. Take the case of buying a garment form a shop. We pick one amongst many that are available without being aware of any mental processes that compel us towards our behaviour. Most children as well as adults have favourite colours or musical pieces that seem to influence our behaviour at later times. But how and why we make these choices have implications in KEL. Kodratoff *et al.*[Kodrat 89a] cite,

> For example, when does dark grey stop being grey and become black? If the system asks 'is the colour grey?' and the user answers 'no' because he thinks it is black while the expert intended that he answered yes, the system will mis-diagnose the disease (or it will be unable to conclude).

Psychologists are still trying to model these sensory based behaviourial domains which appear to be species dependent and governed by evolution and local conditioning. For example, in the case of visual experiences, psychologists are divided in their opinion. Some are of the opinion that in a cluttered world, selective attention operates to direct our action towards individual objects. They think that the visual field is parsed into objects or groups defined by gestalt principles of organization and that attention is directed to these objects rather than to unparsed regions of space. Others suggest visual attention operates in a manner analogous to a spotlight "illuminating" areas of interest in the visual field [Baylis 93].

Ultimately, all human activities can be analyzed to such sensory behaviour. But mental structures can be abstracted at higher levels of language usage for a useful understanding and representation (in the AI context) of domain behaviour if we do not enter the trap of intensional circularity or semantic regress - which is a recursive search for meaning [Narayanan 91].

Where knowledge elicitation is required of such sensory behaviour, it is beyond the scope of knowledge engineering at present. The question is whether we can model knowledge residing in deeper levels. If knowledge and its evident use in the form of behaviour is embedded in networks as proposed by neuro-biologists and neural computing scientists, we face the problem of finding appropriate tools for the useful understanding of the biological mechanisms of thought at a level deeper than we are able to do at present, for successful computer modelling.

4.5 Inspection knowledge elicitation

At first sight it appears that knowledge engineering tools could be feasibly employed in the analysis of overall quality. Specifically, knowledge elicitation techniques could play a key role in the understanding of the inspection knowledge domain. With a view to rapid prototyping, informal interviewing was chosen as the method for elicitation. This is because the inspectors were not very literate people and formal methods were felt inappropriate as they would have made the inspectors feel uneasy and uncomfortable and even be un-cooperative. A good rapport was established with the inspectors who talked freely (as they knew that we are not from the management and are only trying to do a project study which for some unknown reason they wanted to be successful and wished well). Laddering [Cordingly 89] was the guiding technique in these interviews. The questions asked attempted to span across above and below in the expert domain. But the course of the interviews were not fixed and took a natural course depending on answers to the questions asked. Primarily, we were interested in finding how cosmetic quality was judged. The green room method which is by far the most popular method for panel inspection was chosen as the domain for knowledge elicitation. Some of the questions asked and the replies received are described in **Appendix A**. It is evident that the results were not very informative for rapid prototyping.

Knowledge that was relevant but of low utility value, from a prototyping perspective, that came to light was that in assessing overall quality, inspectors took into account the following:

- severity of individual defects,
- collective effect of defects,
- the type of panel, whether door, boot lid, etc.,
- industry standards and
- location of defects.

The first two items in the above belong a special class of problems with accessing expert's knowledge, which have been articulated well by Gains [Gains 88]. Some of the problems he highlights and are relevant in the context of body panel inspection are (my text within square brackets):

- Expertise may not be expressible in language. An expert may not be able to transmit the expertise explicitly because he is unable to express it. [Panel inspection has no language to describe the process of evaluation of overall cosmetic quality. This is the reason why inspectors resort to what is known in industry the "critique" panel which is taken as the guide for inspection].

- Expertise may not be understandable when expressed in language. An apprentice may not be able to understand the language in which the expertise is expressed. [This is the situation faced by the knowledge engineer eliciting body panel inspection knowledge where a rich jargon exists for description of individual defects],

- Expertise may not be applicable even when expressed in language. An apprentice may not be able to convert verbal comprehension of the basis of a skill into skilled performance. [For example an "air-hole" type of defect which appears on body panels (usually due to trapped air between dies) described by the inspectors as occurring on panels usually in small clusters in sizes of a thumbprint, though conceivable, are nevertheless undetectable to an apprentice such as a knowledge engineer wishing to access expertise by learning the language].

This explains why inspectors are unable to discuss with a knowledge engineer as to how they assess defects individually and collectively. However, it was observed that inspectors were able to communicate amongst themselves mainly using their jargon and qualifications of good, bad etc. which were incomprehensible to the author. As a result it was decided to study how the expertise is used in practice, and how it is transmitted. Another reason for this study was the last three items listed above (panel type, standards and location). It was felt that knowledge specific to particular types of panels, and consideration afforded to management directives could be elicited successfully. Some understanding was also hoped to be gained on the importance of location of defects.

Exposure to the shop floor where this inspection is performed made it evident that no formal methods are used and that this is a process performed by highly skilled inspectors who have been trained for several years. The training they undergo is mainly on-the-job in a group setting, and much of what Gains [Gains 88] mentions as alternative methods of expertise transfer in human society seem evident:

- Expertise may be transmitted by managing the learning environment. A trainer may be able to establish effective conditions for an apprentice to acquire expertise without necessarily understanding the skill or himself being expert in it,

- Expertise may be transmitted by evaluation. A trainer maybe able to induce expertise by indicating correct and incorrect behaviour without necessarily understanding the skill in detail or himself being expert in its performance,

- Expertise may be transmitted by example. An expert may be able to transmit a skill by showing his own performance without necessarily understanding the basis of his expertise.

The inspection knowledge is transmitted by managing the environment, showing examples, evaluation, and in addition most importantly, as a process of learning of the language of inspection where the apprentice is made to realise the symbols of the language by visually perceiving the defects. This seems to allow inspectors to discuss cosmetic quality in terms of these symbols.

Thus, it is clear that gaining a good appreciation of the nature of base classes in cosmetic quality - the defects on panels - would be useful. Many hours were spent by the author on the shop floor (in the green room environment) for this purpose. At the beginning, what inspectors alluded to as defects were totally unobservable by the author in the same environment with approximately the same viewing positions. Subsequently, grossly evident defects were observable as shadowy areas and aberrated reflections, but mastery of the skill, i.e. the ability to detect and discriminate, for example, very faintly visible defects, was not gained. One of the problems that was noted was in the correlation of author's perception of defects with those of the expert inspectors, which proved to be very difficult.

One of the other factors that was confirmed was that the severity of a defect was judged on the visual effect or disharmony that was perceived. The larger the effect the worse the defect was. Another observation was that an adjacent defect in the near vicinity of a defect was considered as part of the same defect.

The question of panel specific knowledge was discussed with the inspectors. It emerged that each type of panel had the potential to generate its own peculiar pattern of defects and therefore had to be dealt with separately. The locations where these defects would appear was predictable. For example door panels always have defects around the holes punched for the door handle etc. The understanding was that they have to be "lived with", provided they are not very severe (as the designers and process planners were aware of the problem from the inception). On the other hand, there are areas of the panel designed to have smoothly varying surfaces. Defects (which are generally process based), appearing in these areas are assessed as more critical, though no light was shed on how they are ranked.

Regarding the importance of the location and distribution of defects what became apparent was that severity ranking of defects also depended on the probability of its detection by a customer, as perceived by the inspectors. (This can be viewed as a bias). Towards this, management too has sometimes given guidelines. For example, when faced with assembly line problems due to shortage of a certain type of door panels, management has directed inspectors to pass panels with defects near the sill area, which would otherwise have been rejected. (This can be viewed as intensionally imposed additional bias).

As an aside, an experiment reported by Goldstone [Goldstone 93] may have some relevance due to the parallels it has with this problem of distribution of defects on panels. In this experiment, subjects were shown two displays; both containing the same number of small black and small white squares. The difference in the two displays was in the way the squares were distributed; one was randomly distributed and the other was specially clustered such that squares of the same colour tend to be close. Subjects systematically overestimated prevalence of features (whether black or white squares) in clustered displays. This perhaps indicates the importance of distribution of defects on panels.

4.6 Usefulness of elicited knowledge

1. Since the cosmetic quality judgements were panel specific, as an initial exercise, the inquiry was directed to a specific type of panel. A door panel was selected for this purpose by the author.

2. Attempt was made to subdivide the panel into several regions on the basis of differential importance indicated by the elicited knowledge.

Referring to fig 4.1, areas **a**, **c**, **d** and **f** often have defects and they appear along edges and around punched areas where that panel has been mechanically worked. These defects do not readily attract attention and are also unavoidable in practice. Hence less weight is given to these defects. Areas **b** and **e** are relatively defect free and defects in these areas are more critical. Defects in area **g** are relatively unimportant unless they are very severe because from the normal vantage points these defects are not usually seen. (However, these criteria can at best be treated as general guidelines).

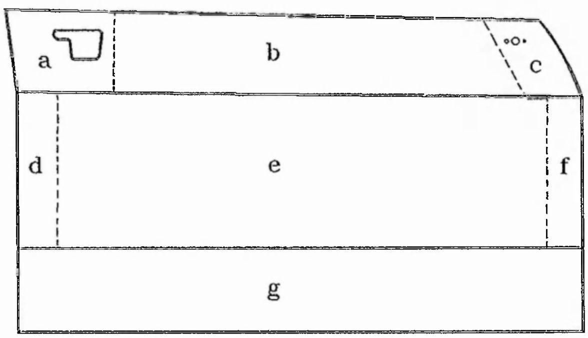


Fig 4.1 Door panel and sub-division

3. General guidelines (as mentioned) above on how judgements were made was observable. But if rules (rather than internal impressions) were involved when evaluating quality with a combination of defects, then they appear to be too complex, non verbalizable and hence difficult to elicit.

4. The differential importance of the regions was conceptualised as acceptability, reworkability and rejectability thresholds of a region.

5. An assumption that the severities of individual defects in any region could be added to determine an effective index of severity for that region was proposed. (The reasonability of this assumption is that it is similar to the reason why inspectors consider a group of individual but neighbouring defects, substantially as one defect. Where this is felt as somewhat inappropriate, the region could be further sub-divided).

4.7 Why knowledge elicitation failed

KEL may be considered a failure because it did not yield high utility knowledge necessary for an effective KBS. The explanation for the inability to extract high utility knowledge could possibly be found in psychology or philosophy. Nevertheless, one clear reason seems to be the fact that we had no idea of the significant mental structures involved in the process of inspection. In all successful knowledge-based implementations reported in literature what is clear to see are these significant structures symbolised and carrying a commonly understood meaning (eg. arrythmia, heart rate, ECG trace etc. for a KBS in the medical domain), and the role of the knowledge engineer was to establish useful relationships amongst such symbols.

In this connection a real life analogy springs to mind. How do judges rank contestants in a beauty contest? It can probably be assumed that ranking is primarily based on the degree of perfection - or viewed in another way on the degree of imperfection (as full perfection is philosophically nonsensical) - of shape and form of various parts of the human body and also the of the body as a whole. How the judges perform this task is a mystery, but the subjectiveness of their performance is obvious, as rarely do a panel of judges unanimously agree on the ranking. The parallels with panel inspection where shape imperfections are judged purely on eye appeal in a similar mysterious and subjective manner is clear.

The analogy does not probably end there. An idea as to what is a perfect form or shape for the human body is probably non -existent. But, what it should not be as well as what a best example should look like can be conceived without much difficulty, by resorting to the memory of past experience in the world. Ideas of what is ugly and what is a best example are realisable due to past experience which serves as background knowledge; but not what is perfect. Inspectors assure us that there is no man made panel that is perfect. The critique panel used in the inspection arena appears to perform the role of providing inspectors the perceptual experience as background knowledge on what is an acceptable standard for defects in order to guide them with their future judgements.

Attempt at elicitation lead to fundamental questions on perception, language, learning, and knowledge. For perception of one or more visual images is at the heart of this inspection. The variety perceived is limitless. Yet, interpretation of them against domain rules, constraints, and acquired inspection skills or "taste", in an apparently mysterious way produce an outcome. In all these processes, "knowledge" is somehow intertwined and used. But verbal description of these perceptions are not forthcoming. Assessment of perceptive experience is perhaps needed prior to interpretation. Abstracting interpretation is the business of elicitation.

Unless we have a language rich enough for the description of subtleties and nuances of our perception, the perception cannot be transmitted for public interpretation. Propositions of the language convey the structure of the perception/thought. Language requires objects. Perceived images must be capable of being analyzed in terms of objects of the language for a description of the perception. A weak language which cannot differentiate different perceptions is of little use.

The objects of a language must be real and distinguishable and sufficiently fundamental for the language to be useful in a social context. (eg. chair or table - their shapes may differ but some quality is constant. So the description of an object must have such a constancy). An object must be knowable - perception of this being the testimony of knowing. The constancy of the object must be learnable and hence knowable.

The problem with panel KEL is that there is no language to describe the nature of a defect in a manner useful in the AI context. Imagine someone describing a scene observed in the past having several hills and valleys. The listener immediately visualises the scene making use of fragments of past experiences that fits the description. The fact that the listener has visualised the scene indicates that some understanding has been gained. But it is questionable whether the degree of understanding attained can ever match that of the person who experienced or is describing the scene. However, whether a useful understanding has been gained is a different question. (For example, if someone says that there are two bricks, then, there is little difficulty is acquiring the useful knowledge of the presence two bricks. Here, the bricks that were described are replaced by the listener's own perception of bricks). If the listeners were aliens from outer space, they probably would have related the description of the scene in a manner fitting with their extra-terrestrial experience. This is the situation of a knowledge engineer who has little or no experience in physically performing panel inspection, when communicating with inspectors. Added to this, the language used by inspectors is not rich enough to convey sufficient details for the knowledge to be useful.

4.8 A philosophical note

Reflection on the problem naturally causes introspection and seems to reveal the general nature of the problem and why it should not have an easy solution. For it makes one inclined to think that sensory-motor systems are autonomous subsystems within us. Seeing, hearing, smelling, tasting, touching, breathing and even thinking seem to occur simultaneously and automatically. It is as if the subsystem concerned behaved like a fully sealed tool, performing its role without ever allowing its internal working to be seen. But at any moment in time we seem to be able to know what one particular system has just done.

The subsystem concerned with motor activity is capable of development; in the form of new physical skills such as walking, writing, driving, balancing a bicycle etc. There is evidence to indicate that even the other sensory subsystems undergo development. But curiously the learning processes involved in these developmental stages never become known. If we ever come to know these learning mechanisms, technology may be revolutionised.

Motor activity succumbs to will. In a technical sense it can be switched "on" or "off". It can also be put on "auto-pilot". Other sensory activities seem to be always "on" or dormant, at different levels of intensity. Their intensity, it appears, can be maximised by bringing them - one at a time - into focus or attention, by will. Even then, knowledge of the processes involved in any of these activities is not available to awareness and hence cannot be elicited. Neisser [Neisser 68] suggests a "main sequence" being in progress corresponding to the ordinary course of consciousness which may or may not be directly influenced by other processes going on simultaneously.

When we focus on one of the senses, the immediate result seems to be an experience, irrespective of the complexity of the input to the sensor. The quality of the experience seems to depend on the quality and intensity level of the sensory system concerned. It is at this level that language and thought processes seems to operate. We can wilfully call on the thinking subsystem to bear on the experience. Psychological investigations are perhaps only possible at this level. Even at this level knowledge of the "whole" experience is unavailable due to the inadequacies of language. What follows from the experience seems complex. An immediate recognition of the essence followed by a reconciliation if known, or learning if unknown seems automatic.

Suppose it is something known that has been experienced. The only positive step we can take with the experience seems to be to query about objects or concepts or symbols known to the thinking subsystem and obtain answers. Let us consider two replies of experts for a question posed during KEL. The first replied, "If the temperature is over 37 deg. I switch on...". The other replied, "If I feel hot and sweaty I switch on ...". In both cases physical senses are used. In the first case, the visual experience of reading a thermometer is inaccessible to the elicitor. But the expert who read the thermometer has applied his/her thinking subsystem on the visual experience gained during reading, and identified the validity of (socially) known symbols. Thus, the reply has high utility in AI. The second case is one of pure physical experiences. The symbols that can be used are hot, cold. sweaty etc. which are subjective and hence have little utility in AI.

The development of the senses was discussed above. In this connection we seem to overlay such things as "taste", "sophistication", "civilised behaviour", etc. during the development. It could even be thought of as a building-in of desirable bias. Thus, we develop a liking to certain music, food, paintings, literature etc. and a dislike to others. Whether this is natural, essential, or can be overcome is not clear. But certainly, it seems cultivatable. In this way the expert panel inspectors may have acquired some degree of sophistication.

Given the above views and the limitations for direct elicitation of useful knowledge, we wish to consider what other approaches can be made for a useful representation of relevant internal activities of an expert in a given situation. We shall pose a different problem, similar in many ways to the panel inspection problem, but which may hold promise towards a general methodology towards our goal. Consider the problem of eliciting the knowledge of a connoisseur of music with respect to his method of grading. We play audio recordings of several different pieces of music, to which the expert listens. Obviously, the expert will not be able to convey the experience in a way that can meaningfully recreate the essence of the experience in others, for it to be useful in AI. But the expert can be requested to grade the music (where we can use powerful mathematical tools) and arrive at a graded list with sufficient confidence. According to Wittgenstein [Wittgen 89] the logical form of the internal relationship is what is common in a gramophone record, the musical idea, the written notes, and the sound-waves (Tractatus 4.014). So the musical ingredients that cause the impact, as it does on the mind of the expert, should be present in the music. The question now is one of pattern discovery. Having given up finding the mental representations of sensory experiences, can we analyze the structures in the pieces of music with whatever tools at our disposal (e.g. in terms of patterns of frequency, loudness, pitch etc.), to create abstract representations and theorise the internal workings of the expert? It is felt that such an approach even though difficult, could be possible. What is suggested is the exposition of a mechanism that explains, not necessarily directly maps, the workings of the mind undergoing an experience. In so doing we are probably taking logical anatomism to a lower level. For, we shall be naming new objects for significant features in musical ways of thinking and making propositions regarding them for conveying our experience.

In this connection Michie's [Michie 86] remarks on future directions of knowledge engineering at three levels seems relevant:

> Level 1
> At first level the process is essentially one of extraction and tabulation of expertise which already exists coded in certain human brains. That in itself is novel and commercially promising.....
>
> Level 2
> Beyond that lies the possibility of automatically constructing new codification of knowledge which are then accepted and used by human professionals, but which did not pre-exist in human brains and hence constitute de novo synthesis. An example due to Bratko and his co-workers .....
>
> Level 3
> Finally, automation can in principle be extended to the synthesis of new knowledge which not only did not pre-exist but could not have pre-existed; that is to say, knowledge which a brain could not possibly synthesize but can assimilate and use if synthesized by some other agency.

Chapter 5

A mathematical basis for surface inspection

5.1 On the need for an algebra

The complex nature of the evaluation of cosmetic quality has been made evident in the previous chapters. It becomes apparent that a mathematical expression for the description of cosmetic quality is a desirable objective. Such an expression is a necessity if the subjective nature of the inspection is to be overcome. A philosophical perspective of inspection in general should be illuminating.

It is observed that any objective inspection process involves:

-a mathematically definable concept for measurement,
-measurement and
-decision making.

For example, the concept for measurement could be a simple distance measure such as a diameter, or temperature, or more complex as pressure, colour, roughness, hardness, loudness, pitch, frequency etc. Before the advent of science, though some of these concepts were named, they did not have the mathematical basis for definition and hence for measurement. The definition of a concept takes into account the ability to measure. Ultimately measurement involves accepted standards for basic concepts such as distance, mass, force, time, temperature etc. For instance, in the case of the concept of pressure, two other concepts - force and area - are involved. Measurement of these two concepts are grounded in the accepted standards to be used in measurement of force and distance.

The concept of area is interesting as it has close parallels with cosmetic quality. This concept is probably as old as the human race and even animals seem to use it. It has been used consciously or unconsciously for the demarkation of territory in early times (mostly for land ownership purposes). The concept of boundary (formed by rivers, mountain ranges or other permanent landmarks) was understood, and the enclosed land was assigned ownership. No objective method existed for its measurement until much later times. Thus we see that even though a definition for its evaluation was not present, the concept has been employed for practical purposes. The situation with cosmetic quality at present is not dissimilar. The concept exists, but not its quantitative measurement; and is used in industry in a subjective manner.

On the subject of area, the idea to use the number obtained by the multiplication of the length by the breadth of a rectangle (both in accepted units of distance measure), as a measure of the concept of enclosed area must have been a breakthrough. Once the algebra involved with addition and subtraction of areas was established, it was then possible to measure any given area (by subdividing into as many rectangles as may be necessary).

The question we face in cosmetic quality is whether we can assign an index to a simple defect and if so whether an algebra can be postulated for dealing with multiple defects to yield the final quality index for a panel; or in simple terms whether we can formulate a mathematical basis for measurement of cosmetic quality.

5.2 Quest for a basis

On the subject of area we see a clear progression in its development. The technological ability to measure "base class" concepts - distance in this case -

leading to a mathematical expression with the formalism required for a definition, as a direct realisation of the concept - the area.

The base class in cosmetic quality is the severity of a defect. With the inspection tool described in chapter 3, we have the capability of measuring a defect, albeit to a standard set by the user of the tool.

The mathematical operations permitted with this entity has to be investigated. Knowledge elicitation has provided cues towards this. The fact that during manual inspection of a panel, inspectors tacitly subdivide it into several regions (of varying importance) has mathematical connotations.

> Firstly, it indicates that an overall value judgement is being made on a subdivided region. We postulate that scalar addition of all severities in a subdivided region is a realisation of the overall severity of that region. i.e if a subdivided region has *m* defects with severities a_1, a_2,... a_m, then, the overall severity A of that region is given by:
>
> $$A = \Sigma a_i \text{ , } \{i=1..m\}$$
>
> This appears to be a reasonable assumption. The assessment of severity of a single defect is not based on the cosmetic effect of a single point but on the totality of effects produced by all of the points enclosed within the defect. Ideas of connectivity and neighbourhood are implicit in the definition of severity. Thus, it is compelling to enlarge the boundaries of a defect to engulf neighbouring defects. Clearly the effect of this is the realisation of a severity value which is the scalar sum of all the individual defects that were engulfed.

Secondly, it indicates that judgements made on the different regions are independent of each other. i.e. the above idea of summation cannot be extended transgressing region boundaries. Thus, we represent all accumulated severities of individual (subdivided) regions A_1, A_2, A_3,... by a multidimensional vector $\{A_1, A_2, A_3, \ldots\}$ with as many dimensions as there are regions.

The notion of representing each region by an axis or dimension in a vector space embodies both these concepts - the permitted scalar summation within a region and the vector summation across the regions. Suppose the region considered above is assigned a unit direction vector $\mathbf{x}$, then a vector

$$A\mathbf{x} = (\Sigma\, a_i)\mathbf{x}\ ,\ \{i=1..m\}$$

can be formed to represent the defect severity of that region. The summation of individual defects conforms to vector summation rules. i.e.

$$A\mathbf{x} = \Sigma\, a_i\mathbf{x}\ ,\quad \{i=1..m\}$$

Thus, individual defects can be thought of as vectors in the direction of the unit vector assigned for the region. For the whole surface, in an *n* dimensional space with unit vectors $\mathbf{x}_1$, $\mathbf{x}_2$,...$\mathbf{x}_3$ etc., the overall surface quality can be represented by the vector

$$\mathbf{Q} = \Sigma\, A_j\mathbf{x}_j\ ,\ \{j=1..n\}.$$

We shall refer to this vector as the cosmetic quality vector. This vector serves as a classifier of cosmetic quality of panels. It is evident from the mathematical ideas used in formulating this vector, that it could be scaled - i.e. each component of the vector multiplied by a given fixed factor - to suit any specific requirement. Again, such scaling need not be uniform; in which case some or all of the components of the vector may be multiplied by different factors. This will be the case, for example, if each component A_j of the vector were to be divided by the total area of the region s_j of the panel to which it belonged. Thus,

$$\mathbf{Q} = \Sigma\ (A_j/s_j)\mathbf{x}_j\ ,\ \{j=1..n\}.$$

Such a scaling would produce a classification based on the density rather than the total content of defects. This non-uniform method of scaling was adopted for the application and is described in detail in chapter 6.

Irrespective of scaling, the cosmetic quality vector is dependent on the regional subdivision used and hence on the surface object itself taken for inspection. Thus, the vector is meaningful only in the context of given vector space. For example, while the vectors evaluated for a number of left hand door panels of a car can be compared meaningfully, those between a left handed door and a right handed door have no basis for comparison as their surfaces that define the regions are not the same. Thus, for the usage of this vector to be meaningful, it must be qualified by the type of surface that is inspected.

However, intuitively we acknowledge that left and right hand panels are comparable. This problem is overcome by including the concepts of symmetry.

The right and left handed panels and their associated properties such as the subdivided regions and their relative importance differ only as much as an object differs from its mirror image. If the assignment of the region unit vectors is made to take account of this fact then we have a basis for comparison.

The above analysis and the implied methodology provides a basis for objective evaluation of cosmetic quality for given type of surface with:

(1) a tool that can quantify individual defects (in a scalar sense) to some standard, and

(2) a knowledge of differential importance of surface regions.

Chapter 6

On-line knowledge acquisition

6.1 On-line Learning as KA

The failure to acquire sufficient, usable, inspection knowledge (due to reasons given in chapter 4.), in a format suitable for implementing a knowledge base, made it clear that a radical change in approach was necessary. A novel on-line automatic knowledge acquisition method was investigated. This method obtains useful knowledge directly from inspectors, without their having to introspect on their actions. It also automatically stores information (or knowledge in AI parlance) in a simple data base, and has the ability to interpret user inquiries, and thus function as an on-line expert assistant. Whenever it makes a suggestion, whether to accept, reject or rework panels, it gives the (statistical) reason as to why it arrived at the decision. (However, the system is not rule-based in the conventional AI sense, as rules of the inspection domain are not known. The external behaviour of the system is similar to the behaviour of typical expert systems).

What it does not have, initially, is inspection domain knowledge. What it does have are the understanding of the framework of inspection knowledge and the ability to learn (while doing), by example, and exhibit its learning when called for (an interpretation). It exhibits only what it has learnt, not what it has not. This is akin to the idea that a person has the capacity (framework) to learn, say, a foreign language, which is learnt in a demonstrative manner (by example), and is able to exhibit the learning when called for. What one has not learnt, one cannot know, and therefore cannot exhibit.

The system may be thought of as a "memorising" system since it lacks the ability to generalise its learning; as such ability comes from "knowing" what is being learnt or by the forming of concepts. (An example is how humans memorise multiplication tables and make use of them. Whether this is learning, or not is not discussed here). In speaking about the logical form of objects in Tractatus (2.02331) , Wittgenstein [Wittgen 89] states, "Either a thing has properties that nothing else has, in which case we can immediately use a description to distinguish it from the others to refer to it; or, on the other hand, there are several things that have the whole set of their properties in common, in which case it is quite impossible to indicate one of them". This in a nutshell highlights the specialisation and generalisation ideas used in machine learning (which is addressed below). The system is more specialised than generalised. It is arguable whether such a system can be classified as a learning system, especially in the light of a lack of accepted definition for machine leaning.

The objects the system deals with are a higher level symbolic representation of inspection "images" (described below), obtained using the tool described in chapter 3 and the algebra described in chapter 5.

6.2 Implementation of learning

The tool described in chapter 3 is used to scan the whole panel and subsequently identify and evaluate individual defects. The evaluation is as described in section 3.7. which is similar to the concept of volumetric deviation of panels from their ideal form based only on 3D surface data, but is in effect the volumetric *cosmetic* deviation. For a given type of panel, the regions of importance are known by the system, using which, it is able to evaluate the cosmetic quality vector described in chapter 5.

However, minor changes are made prior to this evaluation by using a formula for modifying the components of the vector. The reasons for this are:

- convenience of implementation,
- conveyance of the idea of mean defect, and more importantly,
- representing the totality of defects in a region by a symbol, and hence,
- incorporating to some degree the missing aspect of generalisation mentioned above.

The modification done assigns a value between 0 and 9 for each component of the vector. This was thought of as pertinant as what was desirable was an index for every subdivided region of the panel, that would convey some idea of density of defects rather than the cumulative total effect of defects in a region. (This is achieved by making the index the whole number obtained by dividing each component by the area of the sub-divided panel region to which they belong, and multiplying by 10; which if more than 9 was assigned the value 9. Alternative methods are possible, but the sensitivity of these indexes to reflect the grading of defects of the regions was felt as sufficient). Thus, after this operation, the components of the cosmetic vector were guaranteed to be integers between 0 and 9.

The above modification has two important implications. Firstly, the number assigned for each component of the vector (0 to 9) symbolised the state of affairs of the region to which it belonged. Instead of using the numbers, the indexes could have been assigned some different symbols; for example the letters a to z or some other symbols. Secondly, since a range of values of the initial component (prior to its manipulation to convert into indexes) could produce the same symbol, specificity has been abandoned and some degree of generality obtained. This is desirable in the light of what was mentioned in section 6.1, (only it is not clear how to take this concept further). It is surprising now to see that it is possible to talk about the cosmetic quality of a panel in terms of the above mentioned symbols. Even, a string could be formed using all the evaluated symbols for a specific panel, in a manner similar to how words are constructed from letters. (With usage these words could be made to convey appropriate meaning). This is the approach that is taken; and the word formed is referred to in the ensuing as a "key" (in the sense it is used in data base systems).

It will be noted that a common formula was used (for all sub-divided regions of a panel) for the conversion of gross defect values to symbols. This only afforded simplicity, and is not a necessary condition. It would be sufficient to have a unique method (for conversion to symbols) for a specific region of a panel though the methods could be different for the different regions of the panel. What is important is that the key embeds the information, or knowledge, of the inspection, thus serving as a higher level symbolic representation of the results of inspection. For example, for the door panel in fig 4.1, suppose the indices for the regions **a,b,c**..etc are 6,2,4,3,1,2,7 then the key formed is 6243127.

By asserting the outcome of the inspection - whether accepted, rejected or to be reworked - which the inspector inputs (possibly after manual inspection), the system learns. The process is declarative, similar to human learning where internal representations (the keys in this case) are associated with words (the outcomes). This is implemented in the form of a B-Tree data base by storing the outcome against the key. (B-Tree methodology was chosen because of its computational efficiency in key based searching of records).

For any specific key, the data base is designed to store the total number of panels in each category of accepted, rejected, and reworked, in its lifetime. In this manner the system is provided a historical perspective which enables it to predict the best outcome. This aspect is essential because it is known that inspectors do not always agree on outcome. That is, for the same key, different inspectors may assert different outcomes. The proposed implementation overcomes the problem of inappropriate learning, because for any key the probabilities of outcome can be evaluated using historical data which would not be possible if the historical aspect is overlooked. There is also a mode in which the system can be stopped from learning; for example when the system is used by non-experts. In this mode the data base is simply not updated as described above.

The role played as an expert assistant is as follows. When a panel has been completely inspected, the system automatically evaluates the cosmetic quality vector and formulates the key described above and displays it for the benefit of the inspector. If a historical record exists in the data base for the key in question, (which is revealed from its search), then it prompts the inspector with the most probable outcome - one of accept, reject or rework - and the degree of confidence expressed as a percentage probability worked out from historical data in the database. Otherwise, it indicates its lack of knowledge.

After performing the role of assisting it undertakes the role of learning. It prompts the inspector to teach the category to which the inspected panel should belong - one of the above three - or to ignore the results of inspection (which is useful for novice inspectors). If teaching is indicated by the selection a category, it then updates the historical data and hands over control to the inspection tool for inspection of a new panel.

The main advantages of this implementation are:

- ability to learn while doing (on-the-job).
- its ability to function as a specialisation learning system.
- its ability to function as an expert assistant simultaneously.
- the avoidance of the necessity to arbitrate over the judgement of inspectors.
- allows anyone who is deemed an expert to teach, as it is tolerant of disparate outcomes.
- capturing of the consensus amongst experts since it does not reject their opinion.
- enables non-experts to use the system without affecting the learnt knowledge in the system.

The only disadvantage is the inability to generalise, or automatically predict the likely outcome for instances that have not been previously encountered.

6.3 Possible uses of learning

The accumulated historical data contains a wealth of inspection knowledge. When a sufficient number of panels with varying degree of cosmetic quality and a range of defects have been inspected, and the data base updated, it may be possible to analyze the database to determine the heuristics involved for setting up a rule based assistant in the usual AI manner. Note that it is not expected that all possible combinations of keys will occur in practice. Clear subsets of combinations that are always unacceptable or always acceptable may be revealed. This would allow the database to be logically segmented into three regions namely, acceptable, unacceptable and the grey region which has mixed outcome. The heuristics involved in this grey region are bound to be more complex compared with those of the first two regions. If these heuristics can be explicitly identified - at least for a majority of cases - it would be tempting to build a rule based "expert assistant" for the full range of panels [Balendran 91].

Such ideas cannot be fulfilled without finding the generalisations present in the acquired knowledge. Thus we turn to the field of machine learning, which is a branch of AI, to seek a solution to this problem.

6.4 Machine learning (ML)

It is not unbelievable today for one to hear of a learning text editor, a learning database management system, a learning human-computer interface, a learning knowledge acquisition system or a learning scheduling program. The goal set by machine learning, during the past two decades is precisely to introduce such enhanced capabilities into everyday life [Morik 92].

Grasp of some of the central ideas and terms used in ML may be useful in the ensuing. A learning system has to have inputs and produce outputs. The inputs and outputs can be imagined as n-dimensional (n-tuple) vectors with attribute values meaningful in the domain. All input vectors belong to the input space and output vectors to output space. (Some authors use input language and output language instead). A subset of input and corresponding output vectors, are considered as a training set. The aim of the learning system is to acquire new behaviour using the instances of the training set, rather than it being programmed to do so. To effect this, the learning system is given the capability of learning by a learning algorithm. The study of these learning algorithms is the central theme in ML.

Learning algorithms have historically evolved along two approaches; example-based learning, and similarity-based learning [Kodrat 89b]. However, in recent years some general approaches have emerged:

Rule-based approaches:
- inductive methods,
- deductive methods,
- exemplar methods,
- special methods for robotic applications,
- special methods for natural language processing,

Connectionist or neural network approaches, and

Genetic approaches.

It is often difficult to classify the various algorithms presented in the literature due to authors not classifying their algorithms or sometimes using hybrid methods (as in [Kodrat 89b], which uses a combination of explanation based learning and similarity based learning).

A brief description of some of the more important and relevant methods are described below. A good introduction is provided in [Thornton 92].

6.5 Inductive learning

The work horse of machine learning during the past decade has been the inductive learning methodology [Clark 89]. This is also known as empirical or experimental learning as the learning process is data-driven. The derivation of general rules from the instances of the training set, is effected by the algorithm by making inductive leaps from specific to general. Here, learning is viewed as systematic search through the space of possible rules, directed by positive and negative examples of the training set. This is done by examination (of the examples and non-examples) of a concept, to determine by syntactic means alone, which features in the examples led them to be so classified. Here, concepts can be thought of as a partitioning of propositions in a given domain. Version spaces and Focusing are two popular inductive learning methods. They both maintain bounds to indicate the most general and most specific concepts arising from the examples encountered. As more and more examples are encountered, these bounds move closer and when they coincide, the concept is deemed to have been learnt. AQ11 and ID3 are popular algorithms.

6.6 Deductive learning

This is also known as explanation-based learning (EBL), (theorem-proving method), analytic learning and takes into account whatever process caused the classification to be carried out. Hence, this is referred to as knowledge-driven learning. Here, the learning algorithm attempts to build a generalised explanation for the observed instances in the training set. As such, one appropriate training item is sometimes sufficient for the extraction of the concept. However, this method requires sufficient background knowledge to be supplied as a domain theory, which is the primary weakness of this method. LEX is a typical EBL algorithm.

6.7 Exemplar method

Exemplar methods are often characterised by the following properties [Clark 89]:

- Complete descriptions of selected examples are stored in memory.

- For a classification task, a matching algorithm locates -the stored exemplar most similar to the example to be classified. The matching algorithm itself implicitly defines how to generalise from known exemplars.

- the classification of the retrieved exemplar is assigned to the new example.

6.8 Evaluation

The type of input space in panel inspection is the vector space of the cosmetic quality vector. The output vector space is a three dimensional space of accept, reject or rework probabilities. Inductive learning requires sufficient coverage of representative instances in the form of a training set. This may only be possible when a mature knowledge base is available. Hence, the generalization offered by induction has to be viewed with caution. The other problem is the requirement of single valued training pairs. In the panel inspection domain, the output is available as a probability value since all three types of output are possible for the same input vector. Inductive methods have not sufficiently advanced to take into account such probabilistic outputs, though Fisher [Fisher 87] describes such a system. EBL require a domain theory which is lacking in the inspection domain. (This is perhaps what we wish to know too!). In the exemplar method a matching algorithm is required for generalisation; which we are not in a position to propose.

The non-availability of algorithms for our purpose is not surprising, since, according to Wu [Wu 93], the class of problem posed here, seems to be at the frontiers to be attacked by the machine learning community. The frontiers he lists are:

- Constructive learning,
- Incremental learning, and
- learning from databases.

Wu further lists five directions in current intelligent database (IDB) research as:

- object-oriented databases,
- deductive databases,
- expert databases,
- intelligent Man-Machine interfaces, and
- recursive query optimization.

Taking note that the representation of the problem of panel inspection (and also its solution) has always resided only in the biological brains of the inspectors (and never in any other medium) we look at connectionist methods, which promise to replicate brain-processes in computers, for the possibility of generalisation of the acquired knowledge.

Chapter 7

Neural network for generalisation

7.1 Introduction

Neural network is a term used for describing computational topologies of highly interconnected simple processing elements. The study originated from the attempts to model the biophysiology of the brain. It is at its infancy at present, but has established as a branch in AI. Some of its other synonyms are connectionism, parallel distributed processing, neurocomputing, neuromorphic systems, and cyberware. Though much of the early work was done by biophysicists and experimental psychologists, other disciplines such as engineering and statistics, are contributing to its study. The immediate objective of neural research is to model some of the organizational principles of the biological brain in computers. This is not only for gaining more insight into the workings of the brain which has fascinated humans since 3000 B.C (as is evidenced by the written text in Egyptian papyri on the localization of function in the brain [Walsh 78]), but also for exploiting new computational knowledge in the diverse fields of industry, medicine and commerce, by both hardware and software implementations. Considering that the human brain has 10 to 100 million special cells called neurons, each connected to approximately 10 thousand other neurons, the technological limitations of present times for the study of the workings of the brain is clear. However, some researchers wish to stretch the objective of the study even further. Their long-term objective is not only replication low-level brain processes but also of consciousness or spiritual aspects, in material objects thus rekindling the old mind-matter debate. Carling [Carling 92] discussing research in this area which seeks the emergence of consciousness from a network of neurological interactions, quotes Freeman as saying, "Consciousness arises from the interactions of immense numbers of nerve cells that individually act in a thousandth of a second, but that, taken as a whole, produce an evolving sequence of patterns much more slowly, as we well know from introspection". (See also review of the books of I. Rosenfield and G.M. Edelman in [Clancey 93]). The connectionist approach is not without its critics - notably the classicists Fodor and Pylyshyn who argue against connectionism (reviewed in [Narayanan 88]) and Roger Penrose [Penrose 90] who argues against "strong AI" (reviewed in [Sloman 92]).

What is emerging from connectionist experiments in processing information, is a growing realisation that the mere processing of information by appropriate network architectures result in modification of network parameters leading to specific forms of synthesis of information by the hidden layers of the networks. The process of modification of network parameters is a behaviour that is designed into the system and (intuitively) referred to as learning. But the ability of the networks to recognize in the hidden layers, hitherto unknown forms of features in the information; or giving rise to emergent properties as in Hopfield networks, is an unexpected outcome. It is as if the "structure" i.e. the generalized forms of the information has been captured by the networks. Prior to experimentation with learning, it appears impossible to predict in what form the information will be synthesized, especially in hidden layers. It must be said that the significance of this new form of synthesized information for subsequent processing by brain like systems has not been fully understood. Nevertheless, such mysterious behaviour is what has captured the imagination of researchers working in this field.

The ability to implement learning systems mimicking the brain is in itself of immense practical use. Systems can be built to exhibit desirable input output characteristics as a direct consequence of their ability to learn. Also such methods can be employed in solving problems even where it is unclear as to what is to be learnt, but where only the modelling of input output behaviour is of importance. Such approaches are primarily for the exploitation of the connectionist methodology for practical use. They contribute little towards providing better understanding of brain function; which is the main objective of connectionism.

7.2 A brief survey

The skeletal outlines of modern neural networks were discussed in some detail in the nineteenth century [James 1890]. But it was McCulloch and Pitts [McCulloch 43] who laid down the mathematical model for the artificial neuron now known as the McCulloch-Pitts neuron (MCP), and aroused the attention of computer scientists. Their artificial neuron performed simple threshold logic based on the sum of binary input signals, imitating an 'all-or none' law of nervous activity. It is still the main computational element of many modern networks. But, it lacked the ability to learn. (A modern version using the weighted sum of inputs is shown in fig 7.1. Here the biological neuron receiving axonal impulses from several other neurons is modelled as inputs $x_1, x_2, \ldots, x_n$. The efficacy of the synapses is modelled as

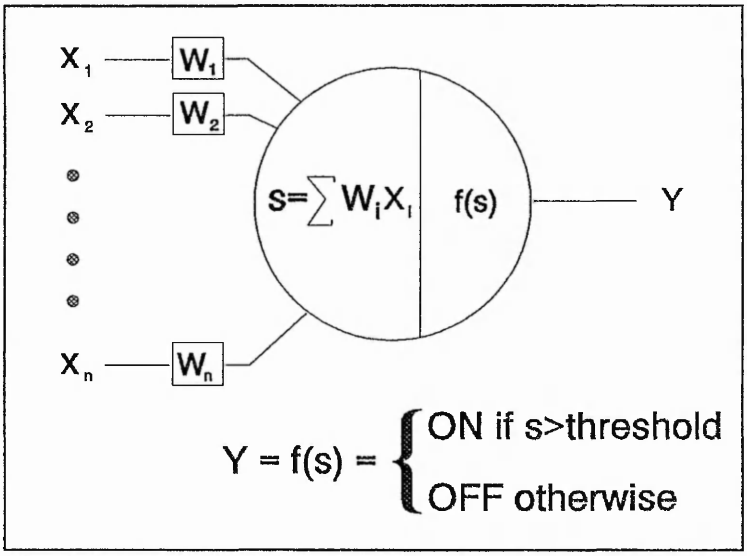


Fig 7.1 McCulloch-Pitt Neuron

interconnection weights $w_1, w_2, \ldots, w_n$. Thus, the effective input to the neuron is the weighted linear sum of inputs. i.e. $s=\Sigma w_i x_i$. The neuron fires by sending an impulse down its own axon - modelled as the output Y - if excitation - modelled as f(s) - exceeds a threshold. The function f(s) could be a linear, non-linear or threshold function and is called the activation function). Though Hebb [Hebb 49] explicitly described a physiological learning rule for synaptic modification, which has recently gained more prominence than in the past, [Hopfield 82], [Grossberg 80], there is no unique interpretation for computational purposes and the rule is often modified to suit computational or other requirements. (Hebb's idea can be traced back to Pavlov).

It was Rosenblatt [Rosenbl 58] who by inventing the perceptron gave the MCP neurons (using linear binary threshold), the ability to learn by simple reinforcement rules and fuelled much media hype that Minsky and Papert [Minsky 69] had to critically analyze and lay to rest the public view that the brain can now be modelled. The central result of their work, which was much delayed, was that the perceptron, being a linear threshold unit, can only discriminate between linearly separable classes. They showed that a large class of interesting patterns are in fact not linearly separable, and hence the perceptron is of limited use. (Though this is true for single layer perceptrons, it is now known that muli-layer perceptrons overcome this difficulty, which Minsky and Papert believed would not be possible). This dampened the enthusiasm for neural networks which entered its 'dark age' for over a decade. However, prior to their publication some significant publications were made. Selfridge [Selfridge 58] suggested how network techniques (of his Pandemonium model) could be used for solving practical problems like Morse code translation. A very useful, and still extensively used, local learning rule for individual neurons, which depended on the input, the output, and what the output should have been, was proposed by Widrow and Hoff [Widrow 60]. This rule is variously known as Widrow-Hoff rule, delta rule, and Least Mean Square (LMS) rule. (A recent review of various learning rules for supervised networks is provided by Widrow *et al.* [Widrow 90]. They also argue that the common concept underlying all current learning algorithms is the "minimal disturbance principle" which is that during training it is advisable to inject new information in a manner that disturbs the stored information in the network by the smallest possible extent).

Despite the criticism [Minsky 69], some researchers continued to explore the subject especially in modelling memory and visual processes, and developed new techniques that clearly showed the applicability of the neural network approach [Kohonen 72], [Anderson 72], [Cooper 73], [Marlsburg 73], [Grossberg 76], [Grossberg 80], [McClell 81]. Perhaps it was John Hopfield who brought back respectability and legitimacy to neural networks with his significant publications [Hopfield 82], [Hopfield 84], with potential practical implementations of his ideas in both hardware and software. He set up an energy function for his network and showed that the evolution of the system with time would lead to a reduction in energy i.e. a stable state. This it was shown would allow content-addressable memory. Kohonen [Kohonen 82] contributed to the ascendency of the study by modelling the topographic self-organization capability (found in the brain) by simple unsupervised learning mechanisms. He applied his method for speech recognition, and built several practical systems. Fukushima *et al.* [Fukushima 83] introduced a multi-layered model called the Neocognitron for visual pattern recognition and demonstrated its ability to recognize handwritten Arabic numerals, even with considerable deformation in shape. (A later version can recognize 35 handwritten alphanumeric characters [Fukushima 91] and a VLSI implementation is discussed in [White 92]). Fukushima *et al.* faced what is called the "credit assignment" problem, i.e., how to change the connection strengths in complex networks in order perform learning. They made the Neocognitron learn in a sequential and directed way by allowing only a specified layer (to be plastic at a time) to learn what it was supposed to learn. A similar problem of how to find the global energy minimum in Hopfield networks (as opposed to local energy minimum which was guaranteed) was addressed by Ackley, Hinton and Sejnowsky [Ackley 85]. They presented a learning system called the Boltzmann machine by introducing a probabilistic component into the system by making the state of individual units stochastic rather than deterministic with characteristics similar to the Boltzmann distribution. This meant that units could assume states that contributed to an increased overall energy, but in conjunction with simulated annealing [Kirkpat 83], the system had a good chance of finding the global energy minimum. However, it was the generalized error correction rule [Rumelhar 86a], [Rumelhar 86b], now known as the back propagation rule that has given a solution to the credit assignment problem in multi-layer networks (though it defies biological plausibility).

The autonomous nature of the sensorimotor subsystems was mentioned in chapter 4. Grossberg's work since the 1960s has been in this direction (for a recent publication with several references see [Grossberg 93]), but of a general nature, for the development of real-time networks, without using external control and making no distinction between learning and processing modes. The ART1's (adaptive resonance theory) self-organising LTM and STM (long and short term memory) which can learn unexpected input patterns very quickly and stably have now been implemented in VLSI circuits [Tsay 91]. Igor Alexander *et al.* [Alexander 90] describes a hardware neural network based on conventional computer RAM memories called WISARD (which stands for

WIlkie, Stonham and Alexander's Recognition Device). Much of the recent work has focused on exploitation of connectionist methods in both hardware and software. Pioneering work of Caver Mead and his co-workers is directed at implementing low level animal nervous systems in electronic VLSI circuitry [Sivilotti 87], [Mead 91], [Delbruck 93]. Hinton (who has made significant contributions to higher-order aspects of cognition by neural networks) describes in a recent paper [Fels 93], a network that maps complex hand movements (using 16 parameters) to speech (with a 203 gesture-to-word vocabulary). Wu *et al.*[Wu 92] describe a neural network based regulator for nonlinear, multivariable turbogenerator control. Such applications of networks for control and systems have gained recognition as a topic in recent years [Warwick 92]. A survey of network applications in manufacturing processes is provided in [Udo 92]. He highlights the tremendous amount of literature that has been generated over the past 15 years on the subject of neural networks and the lack of a single source for locating all due to the multi-disciplinary nature of the field.

7.3 Choice of network

The problem for which a network solution is sought by the author can be stated as follows:

> Given a set of input-output pairs of vectors find a suitable neural network method and an appropriate network architecture that would generalise the input-output characteristics (by learning) so as to produce realistic outputs for hitherto unknown inputs (thus enabling it to perform as an expert assistant).

Typical input-output vector pairs are of the following form:

Input	Output	Class
1. [2,3,4,4,0,1,5],	[1,0,0]	Accept
2. [4,6,5,0,1,1,4],	[0,1,0]	Rework
3. [2,4,8,4,0,5,4],	[0,0,1]	Reject

The components of an input vector are the same as the components of a cosmetic quality vector. These components take integer values in the range of 0 to 9. Cosmetic quality vectors of door panels of the type shown in fig 4.1 are to be used. Since these vectors have 7 components the input vectors also have 7 components. The output vector however can be assigned only one of three possible vectors during learning corresponding the three panel quality classifications - accept, rework and reject. These it is felt are best represented by the vectors [1,0,0], [0,1,0] and [0,0,1] as shown above during learning.

In considering the choice of network for the above problem perhaps the experience in panel inspection may be helpful. It is noted that apprentice inspectors are trained extensively under supervision to acquire mastery of the skill. It appears that the judgement of cosmetic quality is not the result of automatic low-level brain processes (which, if it is the case, can be considered for modelling by unsupervised learning). It is a well meditated judgement based on refined methods of assessment - such refinement being acquired, initially at any rate, by supervised learning. Taking this cue, only network methods involving supervised learning are considered, as in unsupervised learning there is no external force to insure the development of neurons, appropriate for the required mapping.

Hopfield networks originated from the idea that memory states can be constructed as stable states of dynamic physical systems. Of the two models that were proposed, the continuous deterministic model [Hopfield 84] is particularly useful for hardware implementations. But the discrete stochastic model [Hopfield 82] is suitable for software implementations. (For an approachable account see [Alexander 90]). They are also known as associative memories or content-addressable memories. After training with several sets of binary patterns, they are capable of finding the best possible learnt solution for a partial set of information. However, they are inefficient in the number of un-correlated patterns they can store simultaneously - typically about 0.15N for a N neuron network. Thus, is not a suitable choice for our problem as it imposes some restriction on the size of the training set. Further, their is no guaranteed method of reaching the minimum energy state.

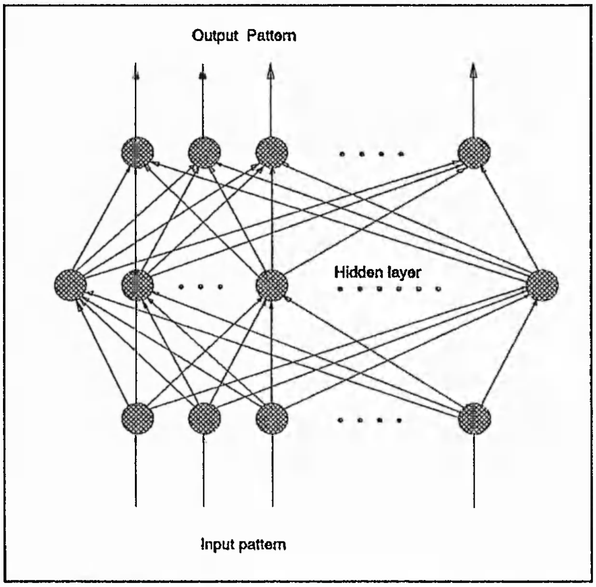


Fig 7.2 A multilayer network for backpropagation

At present backpropagation is the most widely used algorithm for neural network based applications. Though the method was developed by Rumelhart *et al.* [Rumelhar 86a], [Rumelhar 86b] as a training method for neural networks, it was initially introduced by Werbos in his Ph.D. thesis in 1974 [Werbos 90]. An account from a computational perspective is given by Dayhoff [Dayhoff 90]. The attraction of the method is due to the "black-box" like approach that can be taken to mimic generalized behaviour in an input-output modelling situation, using only a training set of input-output pattern pairs - the black-box being a neural network with sufficient neurons and layers (though no strict rules are available for determining the optimum numbers for either). The power of the method is derived from the fact that a neural network (fully connected with sufficient neurons and layers) can represent any desired mapping (existential proof). Apart from the input layer and the output layer, networks using backpropagation technique typically have one or more hidden layers for "internal representation" as shown in fig 7.2. They overcome the credit assignment problem by making use of the steepest gradient descent method (called the generalised delta rule here), for minimising a cost function appropriate for the problem. In neural nets the cost function usually depends on the total error in output for a given training set, and can be viewed as an error surface in a multidimensional space of synaptic weights. The objective is to find the weight vector related to the global minimum of the error surface starting from an arbitrary position on the surface using only local properties (i.e steepest gradient) of that position. (It is like an ant trying to reach the bottom of a valley, which is not in its view, by choosing to follow in the direction of the steepest slope at any position). The method can fail if it gets trapped in a local minimum of the error surface, which is a potential problem with this algorithm, though in practical situations this is rare. The method makes use of the derivative of the activation function f of fig 7.1. Thus, discontinuous threshold functions used by perceptrons will not suffice but the sigmoid function such as $1/(1+e^{-s})$ (which resembles the threshold function and hence receiving physiological validity) is a popular choice. Another (trivial) problem arising in these networks is when initial synaptic weights are equal, but which should be different finally, after training. This cannot be achieved as credit assignment is done in proportion to the existing weights and all weights receive equal increments. The problem is overcome by assigning random values for weights prior to training.

From the above it is clear that the network of choice for the problem posed above is one using the backpropagation algorithm. However, except for the number of neurons in the input and output layers, which must match the dimensions of the input and output vectors, no rules are available to determine the number of hidden layers required and the number of neurons to be used in them. For the work reported here, this was determined experimentally using satisfactory network performance as a criterion. These experiments are described below. The common training and testing sets used in all the experiments are given in **appendix B**.

7.4 Experiment 1

A network topology with a single hidden layer, having 7, 30 (chosen arbitrarily) and 3 neurons in the input, hidden and output layers respectively was investigated. (In the following we shall describe networks by a hyphenated string of numbers; each number being the neurons in each layer in the order from input to the output layers. Thus, the above network shall be described as a 7-30-3 network). The numbers 7 and 3 correspond to the dimensions of the input and output vectors.

It is customary in backpropagation networks to scale the components of the input and output vector to values in the range 0-1. The input vectors of the training set (whose components vary from 0-9) were divided by 10 to meet this requirement. The output vectors were unchanged (as they always met this condition).

The training pairs of the set (in **appendix B**) were presented on a pair by pair basis. When the entire set has been presented, an *epoch* is assumed to have elapsed. If the actual outputs do not match sufficiently the target outputs of the training pairs, the entire training set is presented again for learning in the next epoch. Otherwise, the simulation terminates and the system is considered to have learnt the training set.

The 7-30-3 network was simulated for 10,000 epochs and was found to be unsatisfactory. The learning rate (i.e. the factor used in adjusting weights during backpropagation) was set at 0.75 for all the experiments. The error profiles observed during simulated learning plotted against the epoch number are shown in fig. 7.3. One of the profiles - total error - at output is taken here as the sum of the squares of the difference in output and target values of each neuron of the output layer. It is this value that is to be minimised and made near to zero. (A value of 0.01 was chosen in all the experiments for the maximum permissible total error, which when reached would stop the simulation). The other profile - error rate - shows the percentage of the inputs of the training set that are incorrectly classified. This should ideally converge to zero percent indicating that the whole training set has been learnt.

Considering the total error profile it can be noted that the error increases rapidly and remains at a high level, soon after the start. This behaviour is in contrast with the generally observed behaviour of satisfactory networks employing backpropagation, where the curve drops somewhat sharply at the beginning of the training. If sharp error reduction occurs at the start, it can be argued that it is probably due to the combination of two factors:

1. the presence of a considerable amount of easy generalisations in the training set, and
2. a network topology well suited for learning these easy generalisations.

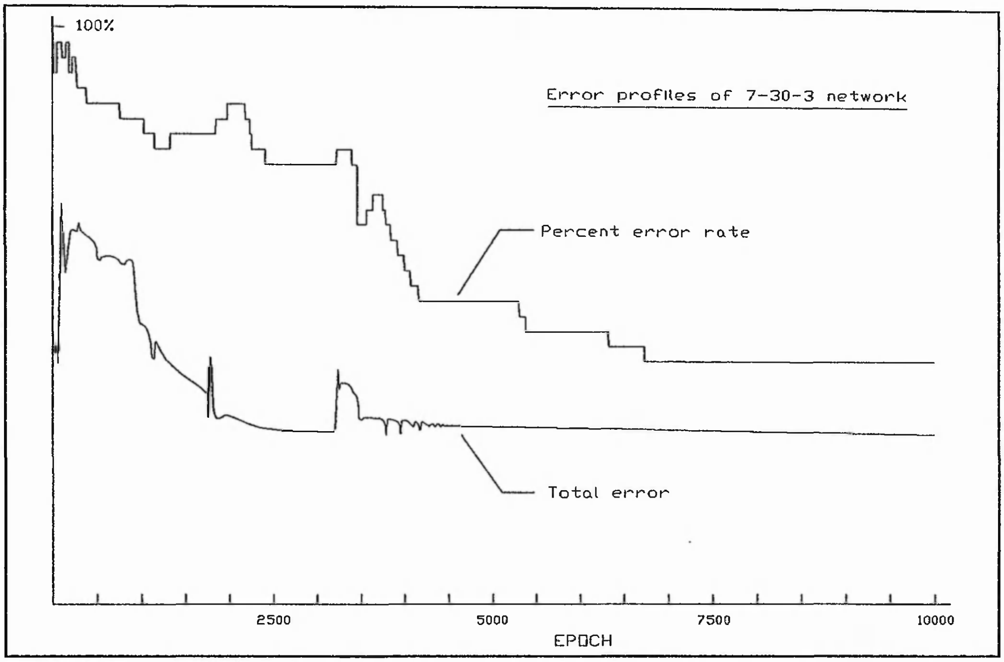


Fig 7.3 Error profiles of 7-30-3 network

It may be that one or the other, or both these factors are not met in this experiment. The slow downward progress in the middle part of the curve appears to be a further indication of the difficulty to find representations for the much 'harder' generalizations in the data. But, the last flat part of the curve (even after 10,000 epochs), which should ideally be very near to the epoch axis, probably indicates that the network topology is unsuitable for the selected training set, which may probably be containing linearly inseparable data. The learning difficulty is also evidenced by the error rate curve which after 3000 epochs still classifies about 75% of the training set incorrectly. At the end of the simulation only approximately 50% of the training set has been correctly classified. (A further experiment with a 7-100-3 network did not produce better results and hence supports the view that the network topology is unsuitable for the problem).

7.5 Experiment 2

If linear separability is the issue that caused the single hidden layer networks to be unsuitable, then a further hidden layer may resolve the difficulty. A 7-16-7-3 network was able to learn the training set successfully. The error profiles

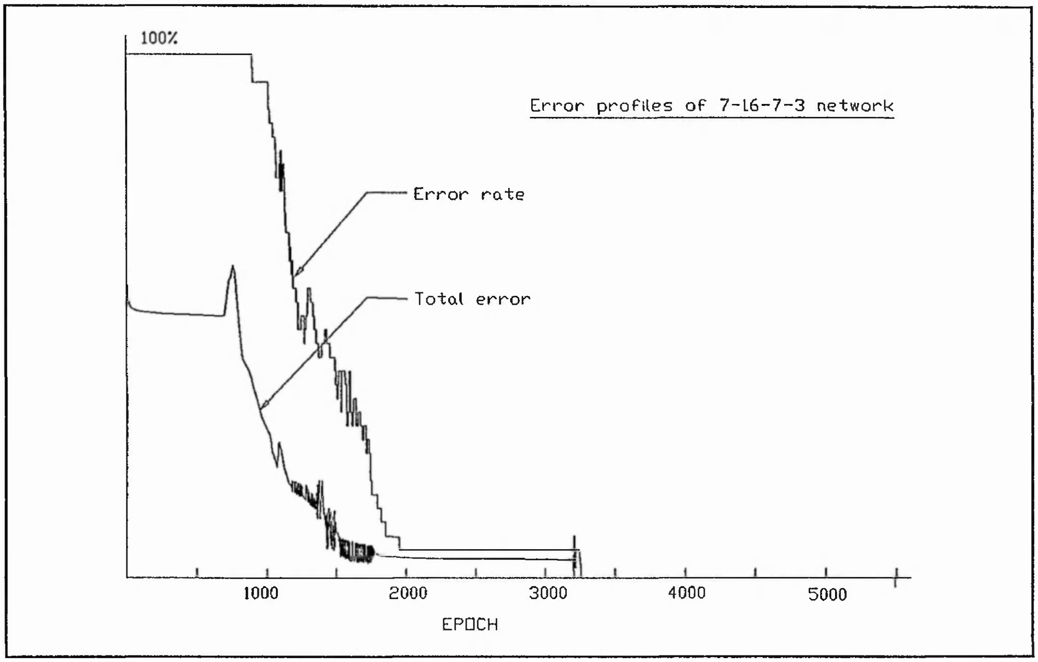


Fig 7.4 Error profiles of 7-16-7-3 network

are shown in fig 7.4. Though the error profiles do not drop down during the first 700 to 900 epochs, the reduction is rapid during the next 1000 epochs (with marked oscillations probably due to the high learning rate of 0.75 employed and also the lack of momentum). By this time over 90% of the training set has been learnt. Approximately a further 1000 epochs are used to learn the remaining 10%. Though the error rate is zero at this time, a further 2000 epochs are needed to bring the total error within 0.01 when the simulation stopped.

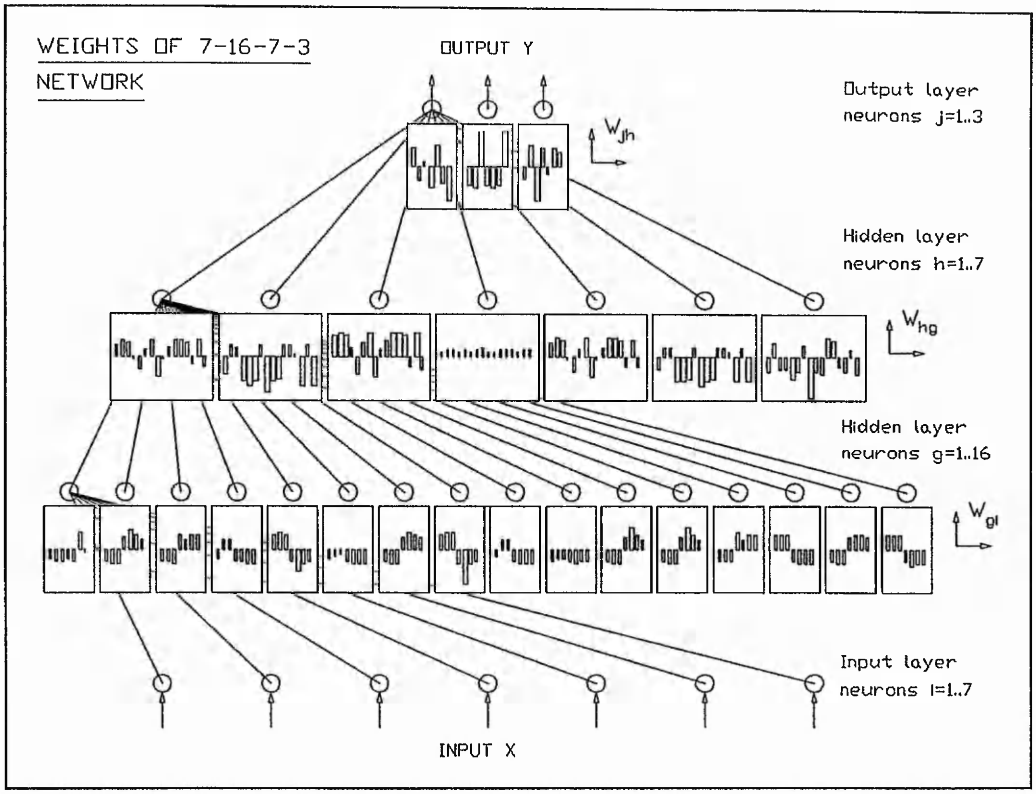


Fig. 7.5 Weights of 7-16-7-3 network

The weights involved with the various layers after learning are shown graphically in fig 7.5. In the first hidden layer, it appears that neurons g = 5, 8, 14, and 16 have become sensitive especially to the first three components of the input vector (i.e. x_1, x_2, and x_3) and inhibit the last four components. (In the ensuing we shall refer to these neurons as 123 neurons). On the other hand, neurons g = 2, 3, 7, 11, 12, 13, and 15 have become sensitive to the last four components but inhibit the first three components of the input vector. The 4th and 9th neurons respond only to x_2, and x_3 and the first neuron only to x_6. Interestingly, neurons 6 and 10 inhibit all inputs.

In the next hidden layer, the fourth neuron (h=4) is the only neuron that responds to the 123 neurons. All the other neurons of this layer inhibit the 123 neurons.

7.6 Experiment 3

The linear inseparability of the data is intriguing and deserves investigation. From a panel inspection point of view, it may be argued that each component of the cosmetic quality vector is composed of two quantities - a decision threshold and a relative intensity value. For example, if one of the components of the vector is 6 and supposing that a value between 4 and 7 would in general cause the panel to be sent for rework, then the decision threshold for rework is 4 and the relative intensity is 2 (=6-4). It may also be inferred that the decision threshold for acceptance could be 0, and for rejection be 8 (one above 7). Note that the three decision thresholds (0, 4 and 8 here) relate to just one component of the quality vector (i.e. for one of the sub-divided area of the panel). The purpose of this experiment is to verify whether such analysis is plausible.

Assuming such thresholds that differ for the different components of the vector are known, then a typical quality code such as 3452085 (which is a 7 dimension vector) could be decoded in two forms as shown below. In the first form each dimension of the original vector is represented by two dimensions; the first dimension indicative of the broad decision category - whether accept, rework or reject (we use below the symbols A, W, and R represent accept, rework and reject respectively) - and the second dimension indicative of the relative intensities above the appropriate decision threshold. In the alternative form each dimension of the original vector is represented by three dimensions corresponding to the three decision categories with the component values indicative of the relative intensities as in the first form. (In this form we lower the threshold values by one in order to have non zero values. Thus, it may be noted that valid vectors in this three dimensional space are always in one of the principal directions, i.e only one dimension can be non zero).

Element	1st Form	2nd Form
3	→ [A,3]	→ [4, 0, 0]
4	→ [W,1]	→ [0, 2, 0]
5	→ [W,1]	→ [0, 2, 0]
2	→ [A,2]	→ [3, 0, 0]
0	→ [A,0]	→ [1, 0, 0]
8	→ [R,1]	→ [0, 0, 2]
5	→ [A,5]	→ [6, 0, 0]

If the second form is chosen, a 7 component input vector such as [3,4,5,2,0,8,5] would now be represented by a 21 component vector:

[4,0,0,0,2,0,0,2,0,3,0,0,1,0,0,0,0,2,6,0,0].

As in the previous experiments the components will be scaled by dividing by 10. It is to be noted that, in practice, the values of the decision thresholds A, W and R for the 7 components of the vector are unknown.

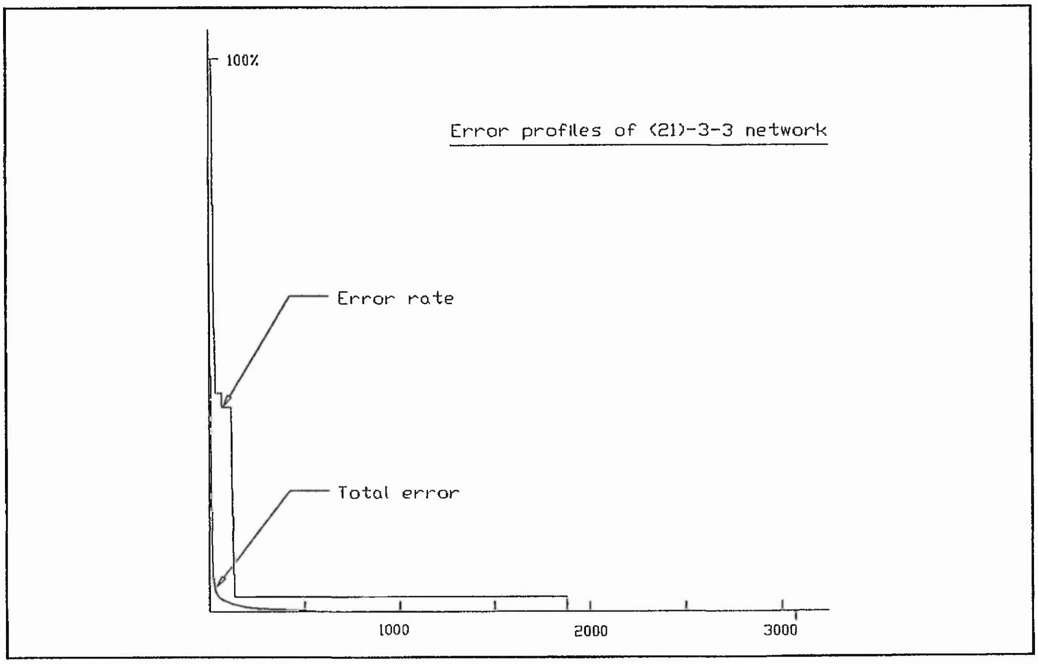


Fig 7.6 Error profiles of (7)-21-3-3 network

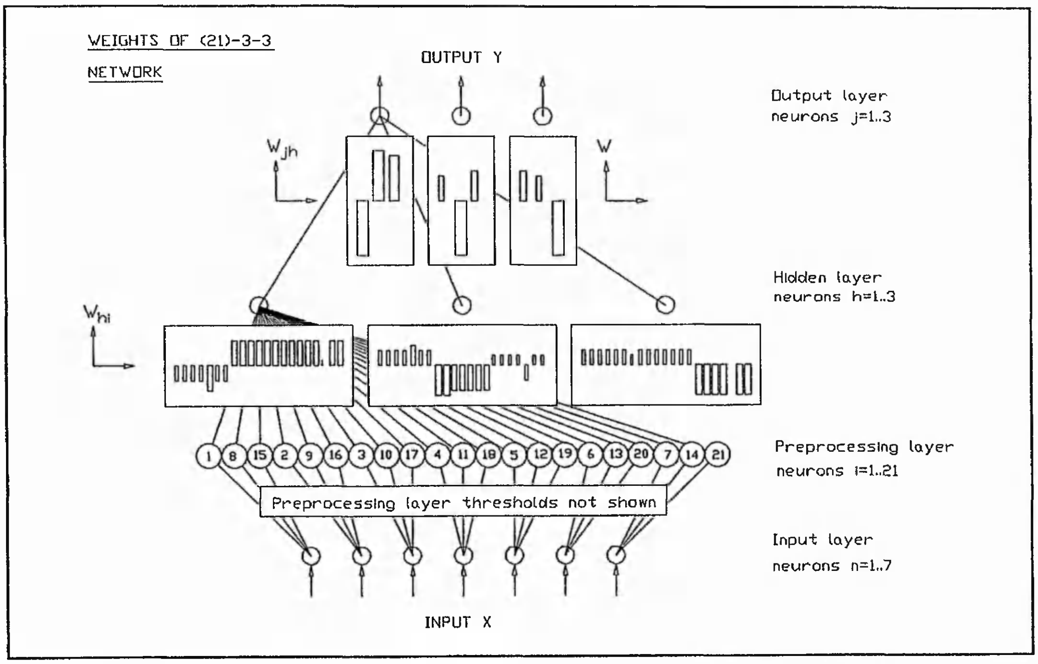


Fig. 7.7 Weights of (7)-21-3-3 network

A neural network with a preprocessing layer, which converts a 7 component input vector into a 21 component vector, that serve as inputs to a 21-3-3 network was investigated. (This shall be referred to as (7)-21-3-3 network). The preprocessing layer is trained separately and learns all the decision thresholds in a single epoch provided that the training set contains sufficient non-contradicting examples. The algorithm used can be described as follows:

We consider the three decision thresholds A, W and R of any region to be monotonically increasing (A<W<R). Thus, for the 7 regions of the panel, we have 7x3=21 threshold values. Prior to "learning", all these ranges are set to an arbitrary negative value. Subsequently, as each input-output pair of the training set is considered, appropriate threshold values are adjusted. This is done as follows. Suppose the output value of a particular input-output pair indicates rework. Then the thresholds to be considered is the rework thresholds of all the 7 regions. From the input pattern the intensity pertaining to any one of the

regions can be directly obtained. If this value is greater than the threshold value (for rework) for the region considered, then this threshold value is adjusted. In a similar manner the other thresholds too are evaluated. Thus, the learning is done in one epoch. For proper functioning as all thresholds must acquire appropriate values, a necessary requirement is that the training set contain sufficient representative instances for the determination of the monotonically increasing thresholds.

The error profiles and the weights are shown graphically in fig 7.6 and 7.7. (Note the non sequential numbering of the 21 neurons in fig 7.7, since they were implemented as a 7x3 array in the simulation program). The quick decaying error profiles (97% correct classification and very small total error within 250 epochs) seem to give credence to the assumption that the components of the cosmetic quality vector are functions of decision thresholds and relative intensities. This was the main motivation for this experiment. In passing we observe that the internal representation in the 3 neurons of the hidden layer are interesting if we note that of the 21 neurons in the preprocessing layer (implemented as 3 rows of 7 neurons/row which we shall refer to as A, W and R rows here), neurons 1 to 7 (A row) deal with "accept" values, 8 to 14 (W row) with "rework" values and 15 to 21 (R row) with "reject" values. Each hidden layer neuron respond to two rows (A,W), (W,R) or (R,A) while clearly inhibiting the third.

7.7 Experiment 4

In experiment 3 it has been tacitly assumed that the decision thresholds are fixed and monotonic for a given type of panel and hence can be obtained by preprocessing. This may not be the case, and complex dependencies may exist with the intensities of the other components of the cosmetic vector. Thus, the strategy followed in experiment 3 for preprocessing is unsatisfactory. Another reason for demeriting this strategy is because of the requirement of non-contradicting examples which do not violate the monotonicity condition. For example if the training set contained {4000000, accept} and {3220000, rework}, then the preprocessing algorithm will not resolve the non monotonic thresholds values that it evaluates for the first component of the input vector. This is because from the first training pair it would be deduced that the rework threshold must be greater than 4, but the second pair (which seems to have taken the adjacent components into consideration) would assert the lower value of 3 for the reject threshold. Therefore, despite the valuable insight provided by the third experiment, it is clearly unsuitable.

A different method of preprocessing was adopted in this experiment. Noting that intensity values ranged from 0 to 9, each component of the cosmetic vector was converted into a 10 dimensional vector with all components set to 0 except the one at the index (starting from zero) equal to the unprocessed intensity, which is set to 1. For example with the cosmetic code 3452089:

$3 \rightarrow [0,0,0,1,0,0,0,0,0,0]$

$4 \rightarrow [0,0,0,0,1,0,0,0,0,0]$

$5 \rightarrow [0,0,0,0,0,1,0,0,0,0]$

$2 \rightarrow [0,0,1,0,0,0,0,0,0,0]$

$0 \rightarrow [1,0,0,0,0,0,0,0,0,0]$

$8 \rightarrow [0,0,0,0,0,0,0,0,1,0]$

$4 \rightarrow [0,0,0,0,1,0,0,0,0,0]$

Such a conversion results in a 70 dimensional input vector. Further, the individual inputs are binary (i.e. 0 or 1) as opposed to the real valued inputs in the range 0 to 1, used in all previous experiments. An obvious choice of representing the components by binary numbers (i.e 3 → 0011, 4 → 0100 etc.) was not selected as it was (intuitively) felt that such a distributed representation would conceal the nature of magnitudes of the intensities, if not ignore such a concept altogether, and add to the complexity and understanding of the internal representation of the network.

A 70-3-3 network was investigated, and the results are shown in fig 7.8 and Fig 7.9. The error profiles have a remarkable similarity with those of experiment 3 but decay much faster.

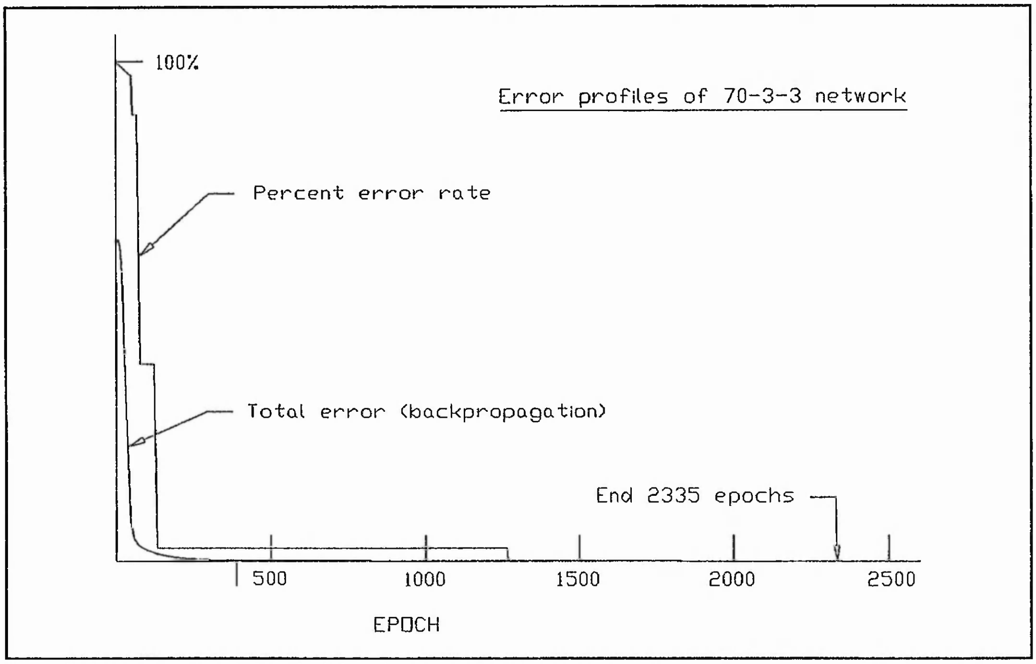


Fig 7.8 Error Profiles of 70-3-3 network

The weights of the hidden layer are interesting. They have captured the decision thresholds discussed in experiment 3. If one of the neurons responds to a particular range of intensities, then, one or both of the other two inhibit this range. The ranges captured by the three hidden neurons are:

accept range: (-,4), (-,3), (-,3), (-,4), (-,1), (-,4), (-,5)

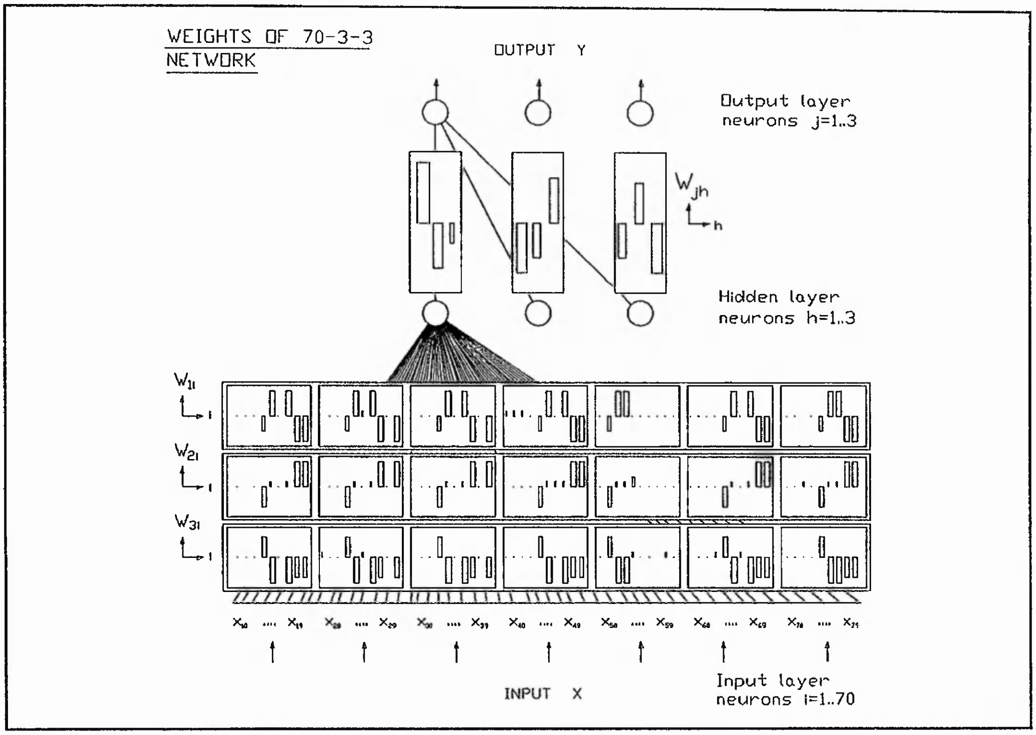


Fig 7.9 Weights of 70-3-3 network

rework range:	(5,7), (4,6), (4,6), (5,7), (2,3), (5,7), (6,7)
reject range:	(8,9), (7,9), (7,9), (8,9), (4,-), (8,9), (8,9)

Such internal representation appears to be credible and understandable.

7.8 Tests and observations

For selecting the best network from the above four experiments, only those used in experiment 2 (the 7-16-7-3 network), and experiment 4 (the 70-3-3 network) need to be studied further, as experiment 1 did not offer a solution and also as the topology used in experiment 4 is superior to that used in experiment 3. A comparison of the error profiles in figures 7.4 and 7.8 clearly favours the 70-3-3 network. Considering the internal representations arrived at by the networks (figures 7.5 and 7.9), that of the 7-16-7-3 is complex and difficult to understand, as opposed to that of the 70-3-3. However, the final choice is to be made by assessing their performance of these two networks with a common testing set given in **appendix B** consisting of 22 testing pairs. The performance of the two networks are summerised in the table below.

Item	Code	Output Desired	Output error of 7-16-7-3	Output error of 70-3-3
1	4334145	accept	error	
2	3223034	accept	error	
3	1010101	accept	error	
4	5445256	rework	error	
5	6556366	rework	error	
6	7667377	rework	error	
7	8778488	reject		
8	5223034	rework	error	
9	8101010	reject		
10	8222122	reject		
11	8223034	reject		
12	9423034	reject		
13	5253034	rework	error	
14	8253034	reject		
15	7364347	rework	error	
16	9394945	reject		
17	9697977	reject		
18	7334145	rework	error	error
19	9334145	reject		error
20	7364145	rework	error	error
21	9394145	reject		error
22	9697377	reject		

The 7-16-7-3 network classifies only 50% of the test set correctly as opposed to 81% of the 70-3-3 network. Clearly, the 70-3-3 network is to be preferred. However, another aspect that is thought to be important, especially with the significant differences in the training and testing set, is to test whether these networks are capable of learning the combined testing and training set. In doing so, it was found that the 7-16-7-3 network was unable to learn the combined set (as the learning error characteristics were similar to those of experiment 1). The 70-3-3 network learnt the combined set without any difficulty as in experiment 4. The error profiles and weights are shown in figures 7.10 and 7.11. Such flexibility is desirable as the intention is to teach the network using actual data (when such data becomes available).

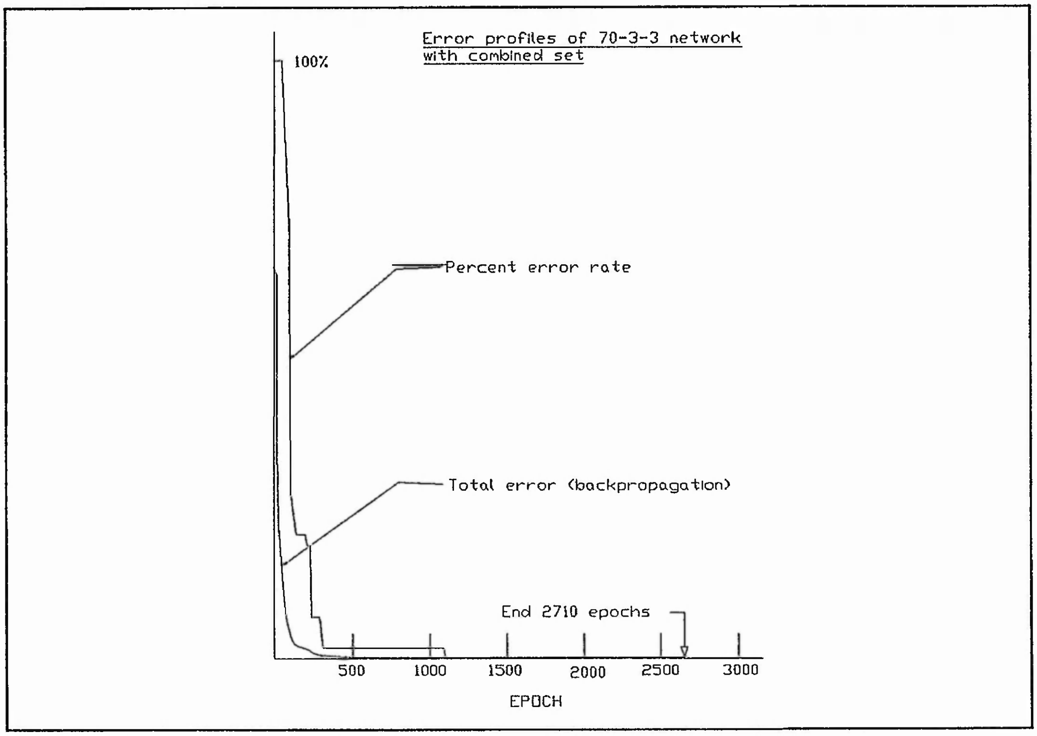


Fig 7.10 Error profiles with combined set

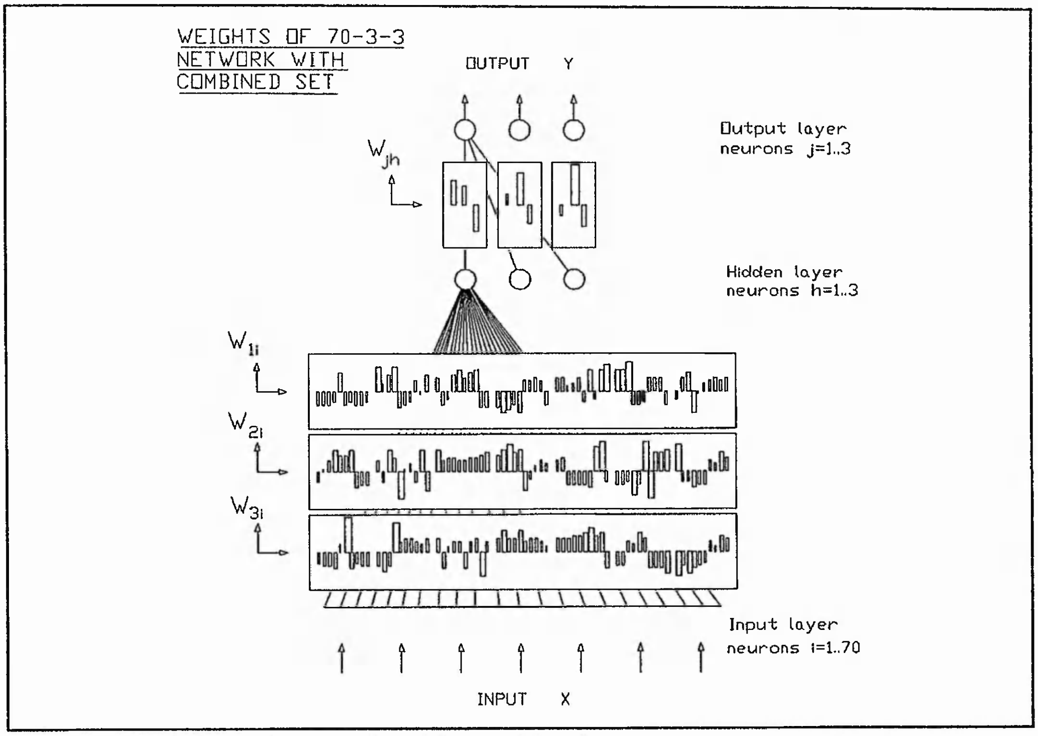


Fig 7.11 Weights of network with combined set

Thus, the clear choice of network for use as an expert assistant is the 70-3-3 network.

Chapter 8

Conclusions

The work reported here is concerned with mechanisation of non-verbalizable, subjective personal experiences and their usage in daily affair in manufacturing industries. The general conclusions that can be drawn from the work are as follows.

The generic problem has been identified as one requiring the solutions to two mapping problems; both of which are concerned with mapping of personal experience to mathematical domain.

The first is a sensory experience while the second is often a decision making process.

In viewing thus, the need for an underlying mathematical basis to perform the role of evaluating sensory experience, and thereby acquiring the potential for its use in decision making or other useful processes has come to light. The requirement of such a new mathematical basis as an essential step towards the solution has been identified.

For the mapping of sensory experience the requirement of a mechanical tool has been noted. The inter-dependence between this tool and any mathematical basis formulated has also been noted.

In the usage of sensory experience, as in making decisions, it was observed that the questions of How? and Why? are intractable. This was identified as possibly due to limitations of language in the problem domain, whose vocabulary (symbols) is not rich enough to convey perceptual experiences or due to the impossibility of gaining awareness of some of the thought processes that are involved. Due to these reasons, it was also observed that conventional knowledge engineering (KE) techniques offer limited approaches for solving such problems.

Nevertheless, it was acknowledged that the insight into the problem, provided by KE techniques, could be successfully utilised in formulating a mathematical basis, which enables new forms of knowledge (inaccessible to conventional KE) to be acquired.

The new forms of knowledge could be used with conventional techniques in either machine learning (when sufficient knowledge has been acquired) or neural networks (even with synthetic data), for satisfactorily solving the problem of mapping the decision making process.

Perhaps more importantly, the new forms of knowledge may contain hitherto unknown conceptual entities (as was evidenced in the internal representations learnt by neural networks) or rules which could be useful in understanding more about the problem.

In experimenting with backpropagating neural networks, the importance of suitably pre-processed inputs, that enable learning of internal representations that are desirable (as they seem reasonable, explainable, and human-like) has been noted. Such representations strengthens the view that networks can behave like humans in their ability to formulate abstract concepts from data.

Discussion

In the application area chosen - cosmetic quality - the sensory experience referred to is a visual one; albeit created artificially with special effects in the "green room" environment. It was shown how this sensory experience can be mathematically mapped as a "cosmetic map" by a newly devised robotic tool that synthesizes data from a computer vision system in a simulated "green room" environment.

It can be argued that some mathematical concept of cosmetic quality is an *a priori* requirement which dictates the capabilities to be provided by the tool. This however, was not the case in this work. When this work was undertaken the full extent of the problem was not realised and the approach taken was simply to design the above robotic tool and leave the decision making role entirely to the operator of the tool. Intuitively, the mapping of the visual inspection experience was felt as important (though theoretically this could be derived from very accurate surface model using techniques similar to ray tracing). The formulation of the mathematical basis, in the form of the cosmetic vector, was a much later development influenced by the attempts at eliciting domain knowledge. Only at this stage, did the broader perspective of panel inspection, and the psychological aspects involved with it, became clear, leading to a formal methodology for dealing with problems of this nature. The mathematical basis adopted yielded hitherto unknown new forms of knowledge, and due to its abstractness prised away the human involvement in knowledge acquisition.

After the system became operational, a difficulty faced was in obtaining representative data covering the full spectrum of defects, even for a specific type of panel. Noting that during a press run, the defects appearing on panels are very similar both in terms of severities and spread, this difficulty was inevitable because of the small volume of cars that were produced by the collaborating car manufacturer, which meant that only a few press runs were required to produce a given type of the quantities required per year. (Over 50 panels from a batch inspected both manually and by machine, indicated nothing more than minor differences). The problem was further aggravated by the economic climate and the changes that followed in organization of the car manufacturer. It was realised that collection of representative data can only be a long term goal and an adequate facility for collection of data should be provided by the system. This was taken into account in the implementation of the B-tree database mentioned in chapter 6.

Further work

Related to the work that has been described, one of the important areas where further work can be done is in attempting to unravel the rules involved in panel inspection. This was mentioned in chapter 6 as a possible area of study with "mature" data bases containing representative inspection examples. Again, towards achieving this same goal, such data bases may be used with neural networks to understand other forms of (internal) representation for the data.

With respect to the hardware used - the measuring tool - attempts can be made to improve the overall speed of inspection of the system (which requires approximately 10 min for the inspection of a complete panel). Here, the indexing of the table with start and stop arrangement has been noted to be time consuming. This may be avoided by giving it uniform motion and obtaining camera images (while the table is in motion) at fixed intervals of time. Some modification to the laser head may be useful to both increase the power of the laser and to reduce the weight of the head by using a diode laser instead of the helium-neon laser. On the other hand, other viable methods of acquiring the mapping of sensory experience could also be pursued. This is to be encouraged to overcome one of the limitations of the tool which is its inability to deal with highly curved surfaces (as opposed to slowly varying surfaces).

There are many situations where the reasons for user preferences or biased choices or leanings cannot be elicited. Examples can be found in the industries of music, fashion, beauty, food, cosmetics etc. These areas have resisted AI approaches. The methodology proposed in this work can be used to explore these problems. More ambitiously taste developed for literature, poems and films too could be studied. But these must await further understanding that has to be gained for mapping of sensory/mental experience of such stimuli (which is being pursued actively by the "strong AI" community).

References

[Ackley 85] D.H. Ackley, G.E. Hinton, T.J. Sejnowsky, 1985, **A learning algorithm for Boltzmann machines**, Cognitive Science, Vol 9, pp 147-169.

[Alexander 90] I. Alexander and H. Morton, 1990, **An Introduction to Neural Computing**, (Chapman and Hall).

[Anderson 72] J.A.Anderson, 1972, **A simple neural network generating an interactive memory**, Mathematical Biosciences, Vol 14, pp 197-220.

[Anon 74] Anon, 1974, **Lasers in glass industry - I. Detecting defects in glass sheet.** Glass pp 309-314, September 1974.

[Bainbri 79] L. Bainbridge, 1979, **Verbal reports as evidence of the process operator's knowledge**, Int. Journal of Man Machine Studies, Vol. 11 pp 411-436.

[Bainbri 86] L. Bainbridge, 1986, **Asking questions and accessing knowledge**, Future Computing Systems, Vol. 1, pp 143-149.

[Balendran 89] V.Balendran, K. Sivayoganathan and D. Al-Dabass, 1989, **Detection of flaws on slowly varying surfaces**, Proc 5th

Natl. Conf. on Prod. Res. (Ed) J.Chandler, (London: Kogan Page), pp 82-85.

[Balendran 91] V.Balendran, K. Sivayoganathan, D. Al-Dabass and L.Evett, 1991, **An expert assistant for inspection of car body panels**, Proc 7th Natl. Conf. on Prod. Res. (Eds) D. Spurgeon and O. Apampa, pp 183-187.

[Ballard 82] D.H. Ballard and C.M. Brown, 1982, **Computer Vision**, (Prentice Hall).

[Barr 81] A. Barr and E.A. Feigenbaum, 1981, **The Handbook of Artificial Intelligence**, Vol 1, (Los Altos, CA:William Kaufmann).-Brachman, R.J., and Levesque, H.J. (Eds), 1985,Readings in Knowledge Representation, (Los Altos CA:Moegan Kaufmann).

[Bastusch 86] C.M. Bastuscheck and J.T. Schwartz, 1986, **Experimental implementation of a ratio image depth sensor**, in Techniques for 3-D machine perception (ed) A. Rosenfeld, (Amsterdam: North-Holland), pp 1-12.

[Batchelor 85] B.G.Batchelor, D.A. Hill & D.C. Hodgson (eds), 1985, **Automated visual inspection**, IFS (Publications) Ltd, UK, North-Holland

[Baylis 93] G.C. Baylis and J. Driver, 1993, **Visual attention and objects: Evidence for hierarchical coding of location,**

Journal of Experimental Psychology: Human perception and performance, Vol 19, No. 3, pp 451-470.

[Bezier 90] P. Bezier, 1990, **Style, Mathematics and NC**, Computer-aided design, Vol 22, No. 9, November 1990, pp 524-526.

[Boyle 88] R.D. Boyle and R.C. Thomas, 1988, **Computer vision: A first course**, (Blackwell Scientific Publications).

[Bratko 89] I. Bratko, I. Kononenko, N. Lavrac, I. Mozetic and E. Roskar, 1989, **Automatic Synthesis of Knowledge: Ljubljana Research**, in Machine and Human Learning, Advances in european research, (Eds) Y. Kodratoff and A. Hutchinson, (London: Kogan Page) pp 25-33.

[Broadbent 86] D.E. Broadbent, P. FitzGerald P & M.H.P. Broadbend, 1986, **Implicit and explicit knowledge in the control of complex systems**, British Journal of Psychology, Vol. 77, pp 33-50.

[Brook 76] R.A. Brook, 1976, **Automatic inspection of flat strip in the steel industry**. Proc VII IMEKO Congress.

[Carbonell 90] J.G. Carbonell, C.A. Knoblock & S. Minton, 1990, **PRODIGY: An integrated architecture for planning and learning**, in Architectures for Intelligence, (ed) K. VanLehn.

[Carling 92] A. Carling, 1992, **Introducing neural networks**, (Wilmslow: Sigma).

[Charlet 92] J.Charlet, J.-P. Krivine and C.Reynaud, 1992, **Causal Model-Based Knowledge Acquisition Tools: Discussion of Experiments**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 319-336.

[Chin 82] R.T. Chin, C.A. Harlow, 1982, **Automated Visual Inspection: A Survey**, IEEE Trans. on Pattern Analysis & Machine Intelligence, Vol. PAMI-4 No 6, pp 557-573.

[Clancey 93] W.J. Clancey, 1993, **Book Review**, Artificial Intelligence, Vol 60, pp 313-356.

[Clark 89] P. Clark, 1989, **Knowledge representation in machine learning**, in Machine and Human Learning, Advances in european research, (Eds) Y. Kodratoff and A. Hutchinson, (London: Kogan Page) pp 35-49.

[Collins 85] H.M. Collins, 1985, **Changing Order:Replication and Induction in Scientific Practice**. (London: Sage).

[Cooper 73] L.N. Cooper, 1973, **A possible organization of animal memory and learning**, Proc. of the Nobel Symp. on Collective Properties of Physical Systems, (Eds) B.

Lundquist and S. Lundquist, (New York: Academic Press), pp 252-264.

[Cordingly 89] E.S. Cordingley, 1989, **Knowledge elicitation techniques for knowledge-based systems**, Knowledge elicitation: Principles, Techniques and Applications, (ed) D. Diaper, (Chichester: Ellis Horwood), pp 89-175.

[Creath 88] K. Creath and J.C. Wyant, 1988, **Comparison of interferometric contouring techniques**, SPIE Vol. 954 Optical Testing and Metrology II, pp 174-182.

[Diaper 89] D. Diaper, 1989, **Designing expert systems: from Dan to Beersheba**, Knowledge elicitation: Principles, Techniques and Applications, (ed) D. Diaper, (Chichester: Ellis Horwood), pp 17-46.

[Dayhoff 90] J. Dayhoff, 1990, **Neural network architectures: An introduction**, (New York: Van Nostrand Reinhold).

[Delbruck 93] T. Delbruck, 1993, **Silicon retina with correlation-based velocity-tuned pixels**, IEEE Trans on Neural Networks, Vol.4, No.3, pp 529-541.

[Dixon 81] N.Dixon, 1981, **Preconscious Processing** (Chichester: John Wiley)

[Eden 86] M. Edem, M. Unser and R. Leonardi, 1993, **Polynomial representation of pictures**, Signal Processing, Vol 10, pp 385-393.

[Farin 90] G. Farin, 1990, **Curves and Surfaces for Computer Aided Geometric Design: a practical guide**, (Academic Press).

[Fels 93] S.S.Fels and G.E. Hinton, 1993, **Glove-Talk: A neural network interface between a Data-Glove and a speech synthesizer**, IEEE Trans. on Neural Networks, Vol 4, No 1, pp 2-8.

[Fisher 87] D. Fisher, 1987, **Knowledge Acquisition via Incremental Conceptual Clustering**, Machine learning, Vol 2, No. 2, pp 139-172.

[Flanagan 69] T.P. Flanagan, 1969, **Automatic inspection technology and its development in manufacturing industry.** Conf of SIRA on Automated Inspection for defects and dimensions, May 69, Eastbourne, UK, (Adam Hilgar Ltd: London), pp 1-7

[Fukushima 83] K. Fukushima, S. Miyake, and T. Ito, 1983, **Neocognitron: a neural network model for a mechanism of visual pattern recognition**, IEEE Trans. on Systems, Man and Cybernetics, SMC-13, pp 826-834.

[Fukushima 91] K. Fukushima and N. Wake, 1991, **Handwritten alphanumeric character recognition by the Neocognitron**, IEEE Trans. on Neural Networks, Vol 2, No 2, pp 355-365.

[Gains 88] B.R. Gaines, 1988, **An overview of Knowledge - acquisition and transfer**, in Knowledge Acquisition for Knowledge-Based systems (eds) Gains.B & Boose J, (Academic Press), Vol 1, pp 3-22.

[Gasvik 87] K.J. Gasvik, 1987, **Optical Metrology**, (Chichester: John Wiley)

[German 85] D.German, 1985, **Vision Aided Inspection Techniques for the Automotive Industry**, Vision '85 Conf, SME, March 1985.

[Goldstone 93] R.L. Goldstone, 1993, **Feature Distribution and Biased Estimation of Visual Displays**, Journal of Experimental Psychology: Human perception and performance, Vol 19, No. 3, pp 564-579.

[Green 69] C.C. Green, 1969, **The application of Theorem Proving to Question-Answering Systems**, Proc of IJCAI-69, p 219

[Greenwell 88] M. Greenwell, 1988, **Knowledge Engineering for expert systems**, Ellis Horwood (England).

[Grossberg 76] S. Grossberg, 1976, **Adaptive pattern classification and universal recording: I. Parallel development and coding of neural feature detectors,** Biological Cybernetics, Vol 23, pp 121-134.

[Grossberg 80] S.Grossberg, 1980, **How does a brain build a cognitive code?**, Psychological Review, Vol 87, pp 1-51.

[Grossberg 93] S. Grossberg, 1993, **A Solution of the Figure-Ground Problem for Biological Vision,** Neural Networks, Vol 6, pp 463-483.

[Hariharan 92] P. Hariharan, 1992, **Basics of interferometry,** (San Diego: Academic Press)

[Hawkins 83] D. Hawkins, 1983, **An analysis of expert thinking,** International Journal of Man Machine Studies,Vol 18, pp 1-47.

[Hebb 49] D.O. Hebb, 1949, **The Organization of Behaviour**, (New York: Wiley), Chapter 4, "The first stage of perception: growth of the assembly", pp 60-78.

[Heijst 92] G. van Heijst, P. Terpstra, B. Wielinga, N. Shadbolt, 1992, **Using Generalised Directive Models in Knowledge Acquisition,** in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 113-132.

[Hopfield 82] J.J. Hopfield, 1982, **Neural networks and physical systems with emergent collective computational abilities**, Proc. of the National Academy of Sciences, Vol 79, pp 2554-2558.

[Hopfield 84] J.J. Hopfield, 1984, **Neurons with graded response have collective computational properties like those of two-state neurons**, Proc. of the National Academy of Sciences, Vol 81, pp 3088-3092.

[Jagan 92] S. Jagannathan, et al., 1992, **Intelligent Inspection of wave-soldered joints**, Jnl. Manufacturing Systems, 11(2) pp 137-143.

[James 1890] W. James,1980, **Psychology (Brief Course)**, (New York: Holt), Chapter XVI, "Associations" pp 253-279.

[Jonker 92] W. Jonker and J.W. Spee, 1992, **Yet another formalisation of KADS conceptual models**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag) pp 211-229.

[Kerridge 87]	A.E. Kerridge, 1987, **Operators can use expert systems**. Hydrocarbon Processing, 66, pp 97-105.
[Kirkpat 83]	S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, 1983, **Optimization by simulated annealing**, Science, Vol 220, pp 671-680.
[Kiyoshi 89]	N. Kiyoshi, 1989, **Comparison and Selection of Knowledge Representation Schemes**, in Knowledge-Based Systems in Manufacturing, (ed) A. Kusiak (London: Taylor & Francis)
[Kohonen 72]	T. Kohonen, 1972, **Correlation Matrix Memories**, IEEE Trans. on Computers, C-21, pp 353-359.
[Kohonen 82]	T. Kohonen, 1982, **Self-organized formation of topologically correct feature maps**, Biological Cybernetics, Vol 43, pp 59-69.
[Kodrat 89a]	Y. Kodratoff, M. Mango and J. Blyth, 1989, **Generalization and noise**, in Knowledge Acquisition for Knowledge-Based systems (eds) Gains.B & Boose J, (Academic Press), Vol 1, pp 301-324.
[Kodrat 89b]	Y. Kodratoff and G. Tecuci, 1989, **Rule Learning in DISCIPLE**, in Machine and Human Learning, Advances in european research, (Eds) Y. Kodratoff and A. Hutchinson, (London: Kogan Page) pp 265-277.

[Kopard 93] P. Kopardekar, A. Mital and S. Anand, 1993, **Manual, Hybrid and Automated Inspection Literature and Current Research**, Integrated Manufacturing Systems, Vol 4 No 1, pp 18-29.

[Lancaster 86] P. Lancaster, 1986, **Curve and surface fitting: An introduction**, (London: Academic Press)

[Marlsburg 73] Chr. von der Marlsburg, 1973, **Self-organization of orientation sensitive cells in the striate cortex**, Kybernetik, Vol 14, pp 85-100.

[Marr 76] D. Marr and T. Poggio, 1976, **Cooperative computation of stereo disparity**, Science, Vol 194, pp 283-287.

[Marr 82] D. Marr, 1982, **Vision**, (New York: W.H. Freeman).

[Mason 90] F. Mason, 1990, **3D surface mapping aids programming**, American Machinist, October 1990, pp 91-93.

[Mead 91] C.A.Mead, X. Arreguit and J. Lazzaro, 1991, **Analog VLSI model of Binaural Hearing**, IEEE Trans. on Neural Networks, Vol.2, No.2, pp 230-236.

[Menq 92] C. Menq, H. Yau and C. Wong, 1992, **An intelligent environment for automated dimensional inspection using coordinate measuring machines**, Trans of the ASME, Vol 114, pp 222-230.

[McClell 81] J.L. McClelland and D.E. Rumelhart, 1981, **An interactive activation model of context effects in letter perception: part 1. An account of basic findings**, Psychological Review, Vol 88, pp 375-407.

[McCulloch 43] W.S. McCulloch and W. Pitts, 1943, **A logical calculus of the ideas immanent in nervous activity**, Bulletin of Mathematical Biophysics, Vol 5, pp 115-133.

[Michie 86] D. Michie, 1986, **On machine intelligence**, (Chichester: Ellis Horwood).

[Minsky 69] M. Minsky and S. Papert, 1969, **Perceptrons**, (Cambridge MA: MIT press) Expanded Edition, 1988.

[Minsky 75] M. Minsky, 1975, **A Framework for Representing Knowledge**, in The Psychology of Computer Vision, ed P. Winston, (New York: McGraw-Hill).

[Mitchell 89] T.M. Mitchell, S. Mahadevan & L.I. Steinberg, 1989, **LEAP: a learning apprentice for VLSI design**, in Machine Learning III:An Artificial Intelligence Approach, (eds) Y. Kodratoff & R. Michalski, pp 271-289, (Morgan Kaufmann)

[Morik 92] K. Morik, 1992, **Applications of Machine Learning**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D.

Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 9-13.

[Narayanan 88] A. Narayanan, 1988, **Fodor and Pylyshyn on connectionism: an extended review and brief critique**, Artificial Intelligence Review, Vol 2, pp 195-213.

[Narayanan 91] A. Narayanan, 1991, Review of **Wittgenstein's Early Philosophy: Three sides of the mirror**, Artificial Intelligence Review, Vol 5, No. 3, pp 262-263.

[Nedellec 92] C. Nedellec and K. Causse, 1992, **Knowledge Refinement Using Knowledge Acquisition and Machine Learning Methods**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 171-190.

[Neisser 68] U. Neisser, 1968, Excerpt from **The multiplicity of thought**, Brit. J. Psychol., vol 54 (1963) pp 1-14, in Thinking and Reasoning: selected readings, (Eds) P.C. Wason & P.N. Jhonson-Laird (Middlesex: Penguin Books).

[Neubert 92] S. Neubert & R.Studer, 1992, **The KEEP Model: A Knowledge Engineering Process Model**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th

European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag) pp 230-249.

[Newell 72] A. Newell and H.A. Simon, 1972, **Human Problem Solving** (Englewood Clifs, NJ: Prentice-Hall).

[Nisbett 77] R.E. Nisbett & T.D. Wilson, 1977, **Telling more than we can know: verbal reports on mental processes"** Psychological Review Vol. 84, pp 23-259.

[Norman 80] D. Norman, 1980, **Twelve issues for cognitive science,** Cognitive Science, Vol 4, pp 1-32.

[Norton 77] L. Norton-Wayne et al., 1977, **Automatic visual inspection of moving steel surfaces**, British Journal of Non-destructive Testing, September 1977, pp 242-248.

[Nwana 92] H.S. Nwana, R.C. Paton, M.J.R. Shave and T.J.M. Bench-Capon, 1992, **Domain-Driven Knowledge Modelling: Mediating and Intermediate Representations for Knowledge Acquisition**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 250-263.

[Osten 87] W. Osten, R. Hofling and J. Saedler, 1988, **Two computer-aided methods for data reduction from interferograms**, SPIE Vol 863 Industrial Optoelectronic Measurement Systems Using Coherent Light, pp 105-113.

[Penrose 90] R. Penrose, 1990, **The Emperor's New Mind**, (London: Oxford Univ. Press)

[Perry 69] G. Perry, 1969, **Inspection of glass containers**, Conf of SIRA on Automated Inspection for defects and dimensions, May 69, Eastbourne, UK, (Adam Hilgar Ltd: London), pp 75-80.

[Pham 92] B. Pham, 1992, **Offset curves and surfaces: a brief survey**, Computer-aided design, Vol 24, No. 4, pp 223-229.

[Poggio 88] T. Poggio, H. Voorhees and A. Yuille, 1988, **A regularized solution to edge detection**, Journal of Complexity, Vol 4, pp 106-123.

[Porter 92] D. Porter, 1992, **Reusable Analysis and Design Components for Knowledge-Based System Development**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 373-391.

[Portmann 92] M.-M. Portmann and S.M. Easterbrook, 1992, **PMI: Knowledge Elicitation and De Bono's Thinking Tools**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 264-283.

[Quillian 75] M.R. Quillian, 1975, **Semantic Memory**, in Semantic Information Processing, ed M. Minsky (Cambridge, MA: Addison-Wesley).

[Quinlan 91] P. Quinlan, 1991, **Connectionism and Psychology: a psychological perspective on new connectionist research**, (Herts. UK: Harvester Wheatsheaf)

[Reid 84] G.T. Reid, 1984, **Moire fringes in metrology**, Optics and Lasers in Engineering, Vol 5, pp 63-93.

[Reid 86] G.T. Reid, 1986, **Automatic fringe pattern analysis: A review**, Optics and Lasers in Engineering, Vol 7, pp 37-68.

[Reynolds 88] R.L. Reynolds and O.L. Hageniers, 1988, **Optical enhancement of surface contour variations for sheet metal and plastic panel inspection**, SPIE Vol.954, Optical Testing and Metrology II, pp 208-216

[Reynolds 90] R.L. Reynolds and O.L. Hageniers, 1990, **A rigorous optical theory of the D Sight phenomenon**, SPIE Vol.1332, Optical Testing and Metrology III: Recent advances in industrial optical inspection, pp 208-216

[Rosenfeld 82] A. Rosenfeld and A.C. Kak, 1982, **Digital picture processing**, Vols. 1 and 2, (Academic press).

[Rosenbl 58] F. Rosenblatt, 1958, **The perceptron: a probabilistic model for information storage and organization in the brain**, Psychological Review, Vol 65, pp 386-408.

[Rumelhar 86a] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, 1986, **Learning internal representations by error propagation,** in Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol I, (Eds) D.E. Rumelhart and J.L. McClelland, (Cambridge MA: MIT Press) pp 318-362.

[Rumelhar 86a] D.E. Rumelhart, G.E. Hinton, R.J. Williams, 1986, **Learning representations by back-propagating errors,** Nature, Vol 323, pp 533-536.

[Scot 82] A.J. Scott, 1982, **Why Should I Use Automatic Inspection?** Proceedings of the 6th International Conference on Automated Inspection and Product Control, pp 1-6.

[Selfridge 58] O.G. Selfridge, 1958, **Pandemonium: a paradigm for learning**, Mechanisation of Thought Processes: Proc. of a symp held at National Physical Laboratory, Nov. 1958. (London: HMSO) pp 513-526.

[Silverman 86] J.F. Silverman and D.B. Cooper, 1986, **Unsupervised estimation of polynomial approximations to smooth surfaces in images or range data**, IEEE 1st Int. Conf. on Robotics and Automation, April 1986, San Fransisco, Vol 1, pp 299-304.

[Sivilotti 87] M.A. Sivilotti, M.A. Mahowald and C.A. Mead, 1987, **Real-time visual computations using analog CMOS processing arrays**, Advanced Research in VLSI: Proc of the 1987 Stanford Conf. (Ed) P. Losleben, (Cambridge, MA: MIT Press) pp. 295-312.

[Sloman 92] A.Sloman, 1993, **Book Review**, Artificial Intelligence, Vol 56, pp 355-396.

[Steels 89] L. Steels and W Van De Velde, 1989, **Learning in second generation expert systems**, in Machine and Human Learning, Advances in european research, (Eds) Y. Kodratoff and A. Hutchinson, (London: Kogan Page), pp 53-75.

[Su 88] X. Su, J.Li and L. Guo, 1988, **An improved fourier transform profilometry**, SPIE Vol 954 Optical Testing and Metrology II, pp 241-244.

[Thornton 92] C.J. Thornton, 1992, **Techniques in Computational Learning: An introduction**, (London: Chapman & Hall).

[Tsay 91] W. Tsay and R.W. Newcomb, 1991, **VLSI implementation of ART1 memories**, IEEE Trans. on Neural Networks, Vol.2, No.2, pp 214-221.

[Tsujino 92] K.Tsujino, V.G. Dabija and S. Nishida, 1992, **Knowledge Acquisition Driven by Constructive and Interactive Induction**, in Current Developments in Knowledge Acquisition - EKAW'92, 6th European Knowledge Acquisition Workshop, Germany, May 1992, (Eds) Th.Wetter, K.-D. Althoff, J.Boose, B.R. Gains, M.Linster and F.Schmalhofer, (Springer-Verlag), pp 153-170.

[Udo 92] G.J. Udo, 1992, **Neural networks applications in manufacturing processes**, Computers and Industrial Engineering, Vol 23, Nos 1-4, pp 97-100.

[Walsh 78] K.W. Walsh, 1978, **Neuropsychology**, (Churchill Livingstone).

[Warwick 92] K. Warwick, G.W. Irwin and K.J. Hunt (Eds) 1992, **Neural networks for control and systems**, (London UK: Peter Peregrinus).

[Waterman 86] D.A. Waterman, 1986, **A Guide to Expert Systems**, Reading MA: Addison Wesley.

[Werbos 90] P.J. Werbos, 1990, **Backpropagation through time: What it does and how to do it?**, Proc. of IEEE, Vol 78, pp 1550-1560.

[West 77] R.N.West and W.J. Stocker, 1977, **Automatic inspection of cylinder bores**, Metrology and Inspection, July 1977.

[West 82] R. N. West, 1982, **Laser scanners for automatic inspection of strip products**, Proc. 6th Int. Conf. on Automated Inspection and Product Control, (UK:IFS Publications Ltd), pp 287-295.

[White 87] G.R. White, 1987, Factory of the future:II.Artificial intelligence in the factory of the future. Proc of the 1987 Control Exposition, pp 627-633.

[White 92] B.A. White and M.I. Elmasry, 1992, **The Digi-Neocognitron: A digital Neocognitron neural network model for VLSI**, IEEE Trans. on Neural Networks, Vol 3, No 1, pp 73-85.

[Widrow 60] B. Widrow and M.E. Hoff, 1960, **Adaptive switching circuits**, 1960 IRE WESCON Convention Record: Part 4, (New York: IRE), pp 96-104.

[Widrow 90] B. Widrow and M.A. Lehr, 1990, **30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation**, Proc of IEEE, Vol 78, pp 1415-1492.

[Wittgen 89] L. Wittgenstein, 1989, **Tractatus Logico-Philosophicus**, translated by D.F.Pears and B.F. McGuinness, (London: Routledge).

[Wu 92] Q.H. Wu, B.W. Hogg and G.W. Irwin, 1992, **A neural network regulator for turbogenerators**, IEEE Trans. on Neural Networks, Vol 3, No 1, pp 95-100.

[Wu 93] X. Wu, 1993, **Inductive learning: Algorithms and Frontiers**, Artificial Intelligence Review, Vol 7, No. 2, pp 93-108.

[Zuech 88] N. Zuech, 1988, **Applying machine vision**, (New York: John Wiley).

Appendix A

Some of the questions put to expert inspectors during informal interviews and the answers received are given below.

Q: You say you look for defects in panels - what are these defects like? A: There are several types of defects ... {already known jargon}.

Q: How do you classify them? This is from experience since we know what causes the problem, for example, ...

Q: How do you grade defects? A:It is hard to explain, you have to see it.

Q:What influences the decisions? A: When you highlight the panel and view you are able to see as some kind os shadowy and light regions. (Some mentioned dark and light areas, some mentioned abberations).

Q: Do they always look similar?. No.

Q: How dissimilar are they? Difficult to say, but you may get similar looking defects in a particular batch of panels but a different batch may have entirely different looking defects.

Q: How do you know that you have assessed them correctly?

A: We use a critique panel.

Q: How does that help you? A: Any condition that is worse than in the critique panel has to be rejected.

Q: Why do you use the critique panel? A: It sets the quality standard for inspection.

Q: What other use can a critique panel be put to? when we have problems with supplied panels we have to sometimes show the supplier how the supplied panels deviate from the critique panel.

Q: How do you judge these deviations? we look for the conditions in a panel and if they are worse than in the critique panel then we can't accept.

Q: Is a single bad defect sufficient to reject a panel? Sometimes yes if it is very bad, sometimes no. It depends where this defect is and also you have to take account of the other defects on the panel.

Q: In making this decision do you rely only on the critique panel? A: Most of the time, yes. But management also gives us certain guidelines from time to time on to be lenient or stricter in the inspection.

Q: How is this done? A: Usually by a different critique panel according to which we may, for example, overlook defects in some areas.

Q: Does that mean you completely ignore some defects? Not at all, it is only that we accommodate certain serious defects which we would otherwise have marked as bad.

Q: Does it matter where the defects are? Yes. The defects that are common near the door handle area which we may overlook if not too critical will certainly not be acceptable in the middle part of a door panel; not that they occur that way.

Q: Can you give some guide lines or broad rules on what combinations of defects are acceptable? A: No panel is perfect. Each type of panel has its own specific class of possible defects. In the door panels defects appear near the hole punched as the door handle aperture and all around the edges that are bent at right angles to accommodate the inner panels. Again defects appear near the holes punched for fixing side mirrors etc. These defects are in a sense unavoidable. Then there are other defects such as recoils, air holes splits etc., that are caused by problems with the pressing process. We have to take all these into account in determining the quality.

Q: Do all the inspectors agree on the outcome? A: usually yes, we don't work in tight compartments. Sometimes differences arise, but these are very minor.

Q: Can you describe a situation where such difference became an issue? A: It is not within our group. In the body building area they sometimes send panels back for rework. Since we do not inspect all the panels, they may or may not be ones we had inspected.

Q: If I want to learn your skill what do you think I should do? A: You have to put your gloves on and start looking at panels.

Q: Can't you teach me a thing or two about your skill without my getting my hands dirty? A: This is hard because it is like telling you how to appreciate a picture.

Q: Don't you have any rules for inspection? A: There are. But you have to learn them the hard way.

Q: Supposing I give you an inspection report of a panel what would you understand from it? A: The inspection report contains the types of defects that are on the panel, which are the reason for the outcome mentioned in the report.

Q: Can you get an idea of how serious the defects are in the report? A: Only the person who inspected can realistically have an idea. But the report tries to convey why he decided to pass or fail the panel which we can follow from the description.

Appendix B

Training set used in chapter 7

Since the inspection tool described in chapter 3 is a new innovation, representative actual data is not available. Therefore, the only way forward is to compose realistic synthetic data as shown below for the three output categories, namely accept, rework and reject. The door panel of fig 4.1 (which has 7 sub-divided regions) is considered here and the quality codes are to be used as input. In formulating synthetic data one of the main concerns would be the avoidance of arguable or uncertain instances of synthetic patterns. Therefore, we use only patterns with extreme conditions occurring in the 7 regions (i.e. clearly acceptable, reworkable or rejectable conditions in only one of the regions at a time). Further, a general requirement for any training set is that it be as representative as possible in the coverage of the domain. This it is believed to be met if all the possible extreme conditions of all regions, i.e. lower and upper ranges, are included, as in the following. In formulating the set, the index ranges (of accept, rework, and reject) for each region of the panel (**a** to **g**) are assigned shown in the figure below. It is believed that such a partitioning, even though arbitrarily made, is sufficiently realistic. Also it must be noted that the training set composed is only for the purpose of evaluating experimental networks. When an appropriate network has been chosen, it is suggested that a training set be built with actual representative data when such data becomes available.

"Accept" quality codes:

0000000

4000000

0300000

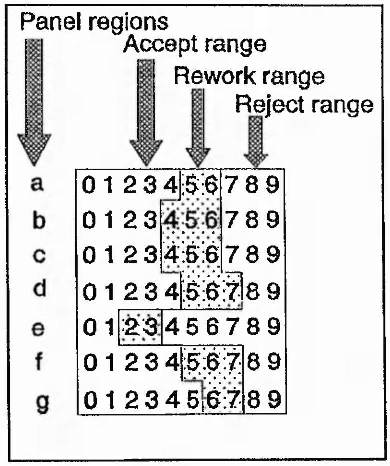
0030000

0004000

0000100

0000040

0000005



Index ranges

<u>"Rework" quality codes:</u>

5000000

7000000

0400000

0600000

0040000

0060000

0005000

0007000

0000200

0000300

0000050

0000070

0000006

0000007

<u>"Reject" quality codes:</u>

8000000

9000000

0700000

0900000

0070000

0090000

0008000

0009000

0000080

0000090

0000008

0000009

Testing set used in chapter 7

The testing set is required for testing experimental neural networks that were trained using a training set. If actual data is available, this is usually done by partitioning the actual data into training and testing sets. However, since real data is not available, synthetic data is composed, that can be best described in terms of categories to which they belong. The categorisation is based on the index ranges to which the individual indices belong. Since there are 10^7 possible combinations of indices for forming an input code, the aim is to formulate a representative subset of input codes (including those which are envisaged to be generated in practice) for which output can be predicted with some justification. Twenty two testing pairs are listed below. It is believed that they are sufficient for network testing purposes. In the ensuing it is convenient to discuss in terms of higher or lower ranges by noting that the accept range is the lowest range, the reject range is the highest range, and the rework range is in between these two.

Category A

A type of range (i.e. one of accept, rework, or reject) is first selected and input codes are formulated using indices from all the 7 index ranges of this selected type. It is believed that the output should be the same as the selected type of range. With this restriction, some codes are composed using only the upper indices of the ranges or only the lower indices of the ranges, and others are composed using intermediate indices.

1.	{4334145, accept}	- upper accept range indices
2.	{3223034, accept}	- mid accept range indices
3.	{1010101, accept}	- mid accept range indices
4.	{5445256, rework}	- lower rework range indices
5.	{6556366, rework}	- lower rework range indices
6.	{7667377, rework}	- lower rework range indices
7.	{8778488, reject}	- lower reject range indices

Category B

This is made up of mid range indices of a given type with one index belonging to a higher range. It is expected that the output should be the type of this exception.

8.	{5223034, rework}	- all mid range accept except one rework
9.	{8101010, reject}	- all mid range accept except one reject
10.	{8222122, reject}	- all mid range accept except one reject
11.	{8223034, reject}	- all mid range accept except one reject
12.	{9423034, reject}	- all mid range accept except one reject

Category C

This is similar to category A but with 2 indices belonging to a higher type range. The highest type range is expected to be the output.

13. {5253034, rework} - all mid range accept except two; both rework

14. {8253034, reject} - all mid range accept except two; highest reject

Category D

This category is made up using the upper indices of a chosen range with one or more exceptions; the exceptions belonging to ranges higher than the chosen range. The expected output is the highest type of range of the exceptions.

15. {7364347, rework} - upper accept range; 4 exceptions, all rework

16. {9394945, reject} - upper accept range; 3 exceptions, all reject

17. {9697977, reject} - upper rework range; 3 exceptions, all reject

18. {7334145, rework} - upper accept range; 1 exception; rework

19. {9334145, reject} - upper accept range; 1 exception; reject

20. {7364145, rework} - upper accept range; 2 exceptions; all rework

21. {9394145, reject} - upper accept range; 2 exceptions; all reject

22. {9697377, reject} - upper rework range; 2 exceptions; all reject

Appendix C

Excerpts from the first report

2.2 Scene Optics

This involved the formulating of a method of presenting a scene with a skin panel, to the vision system camera, in a manner suitable for detection of flaws, if any, on the panel. The method adopted makes use of basic principles of reflection of light to detect flaws on a panel, by treating the panel as a reflector and observing aberrations, and hence changes in width of the image of a uniform object, caused by flaws on the panel.

The highlighting fluid presently used in manual inspection was used to make the skin panel sufficiently reflective. The image of a black rectangular object that was approximately 4" wide and long enough to cover the span of the panel was presented to the camera, at an acute angle to the plane of reflection as shown in fig.2. The black object was chosen in order to provide good contrast against the white background. (The black object in the figure is the interior of a channel section which as it avoids direct external illumination appears black. This effect is enhanced by painting the interior of the channel black). Flaws in panels generated aberrations (and hence non-uniform variation in the width of the image stripe), that were detectable by the vision system depending on the acuteness of the angle of view. Illumination of the scene though important in order to provide good contrast between the skin panel and the black image, did not seem to be critical in this study. However, for the next stage of development controlled illumination is vital. The possibility of using two parallel sheet lasers to simulate the edges of the black object and "light sectioning" is not ruled out in industrial conditions or flaw detection thresholds demand.

2.3 Image Analysis

For this analysis only door panels were chosen as this afforded the simplification resulting from the fact that any horizontal section of a panel (when mounted on a car) is almost a straight line (and convexity only in one plane). The image of the above black object is a black stripe of uniform width if no flaws are present. The scene is digitized by the system and edge pixel data are available in the data structure CPLINE. The algorithm developed "windows" on the black image stripe, rewrites this data structure to include only the pixels within the window, (thus avoiding all other unnecessary edge pixel data) and evaluated the width of the stripe at each pixel of an edge. This stripe width data (shown as a width curve in fig 3) is then analyzed using the polygonal approximation method (which is a method of approximating a curve by a number of connected straight lines so that the error between the curve and the approximated line at any point is within a preset tolerance). Gig 4 shows the approximation of a curve by different sets of lines depending on the setting of the allowable error. The error allowable on the width of the stripe is variable under software control (and at a +/- pixel proved quite satisfactory). If the approximation yielded only one straight line then the stripe is of uniform width (or of constant tapering width), which indicates a flawless area. If the approximation yielded more than one straight line, then each break point is a point of discontinuity of width pattern, and points to the position of the flaw. By setting the allowable error to a higher limit, i.e. thresholding (for example to +/-3 pixels), it is possible to make the system ignore a flaw.

3. Discussion

The study indicated that

(a) at acute angles of view, flaws in panel can be readily ascertained by the vision system,

(b) such a system can be used as an accept/reject decision making element, and integrated to the present production line,

(c) the quality of inspection can be set to desired levels by a combination of settings of software parameters and camera angle of view.

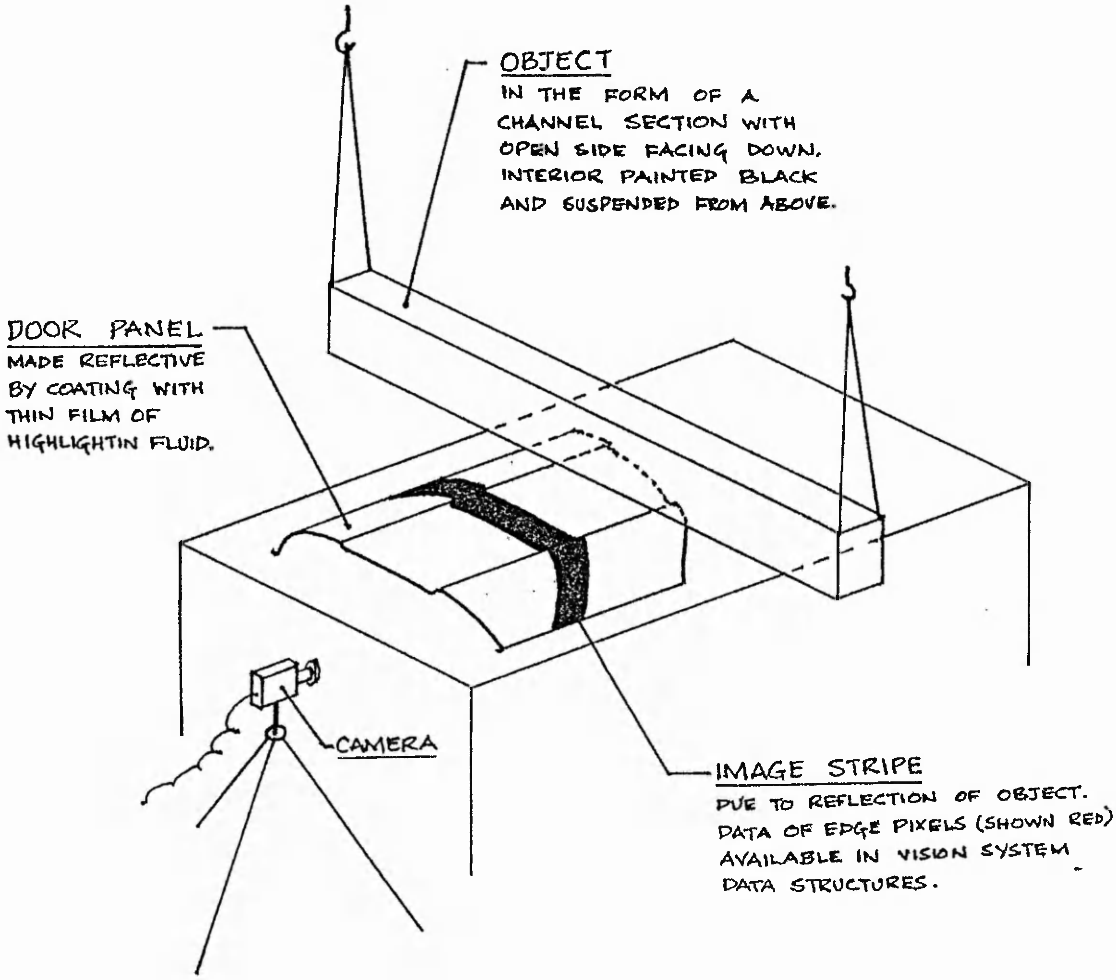


FIG. 2. EXPERIMENTAL SETUP

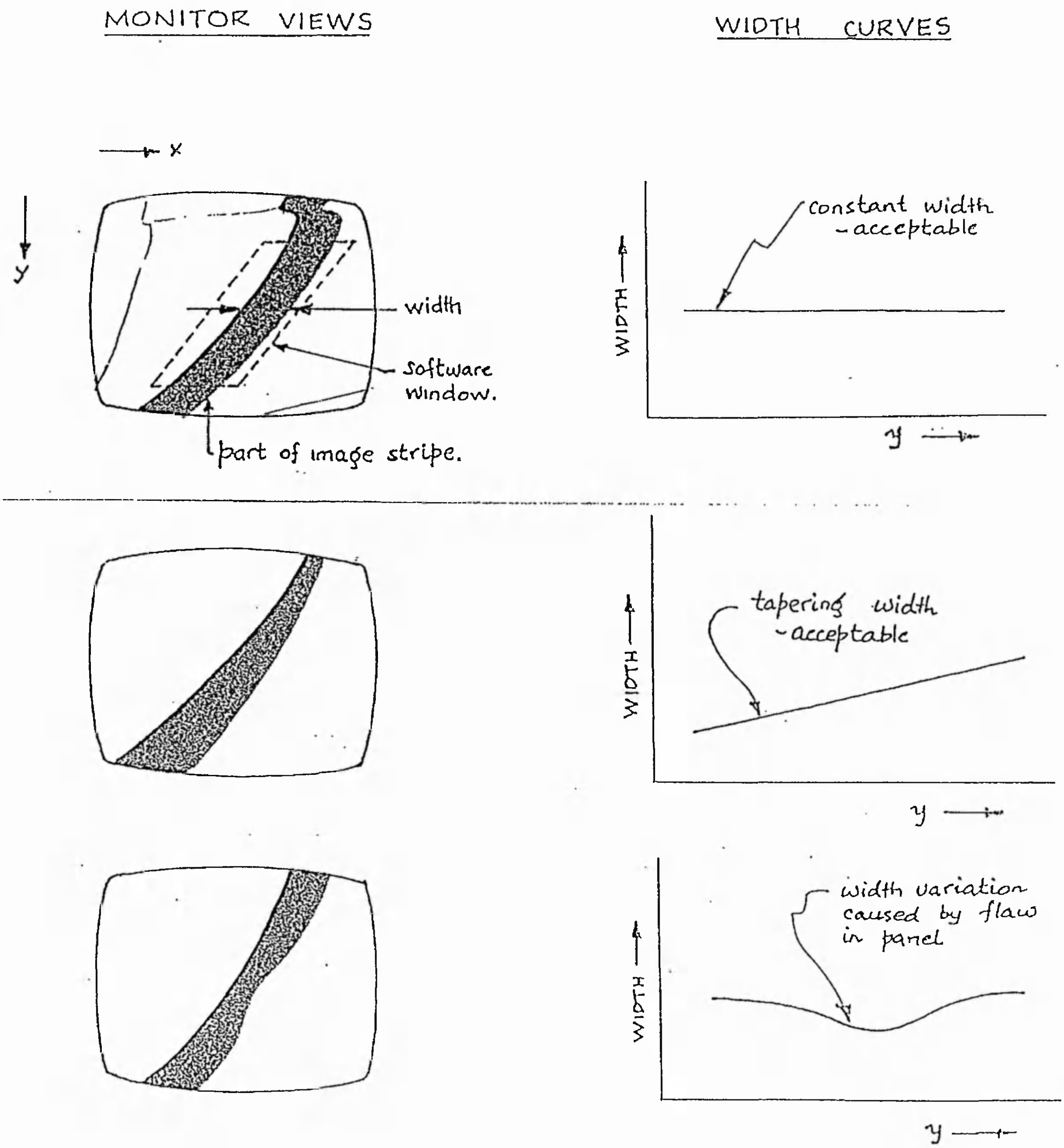
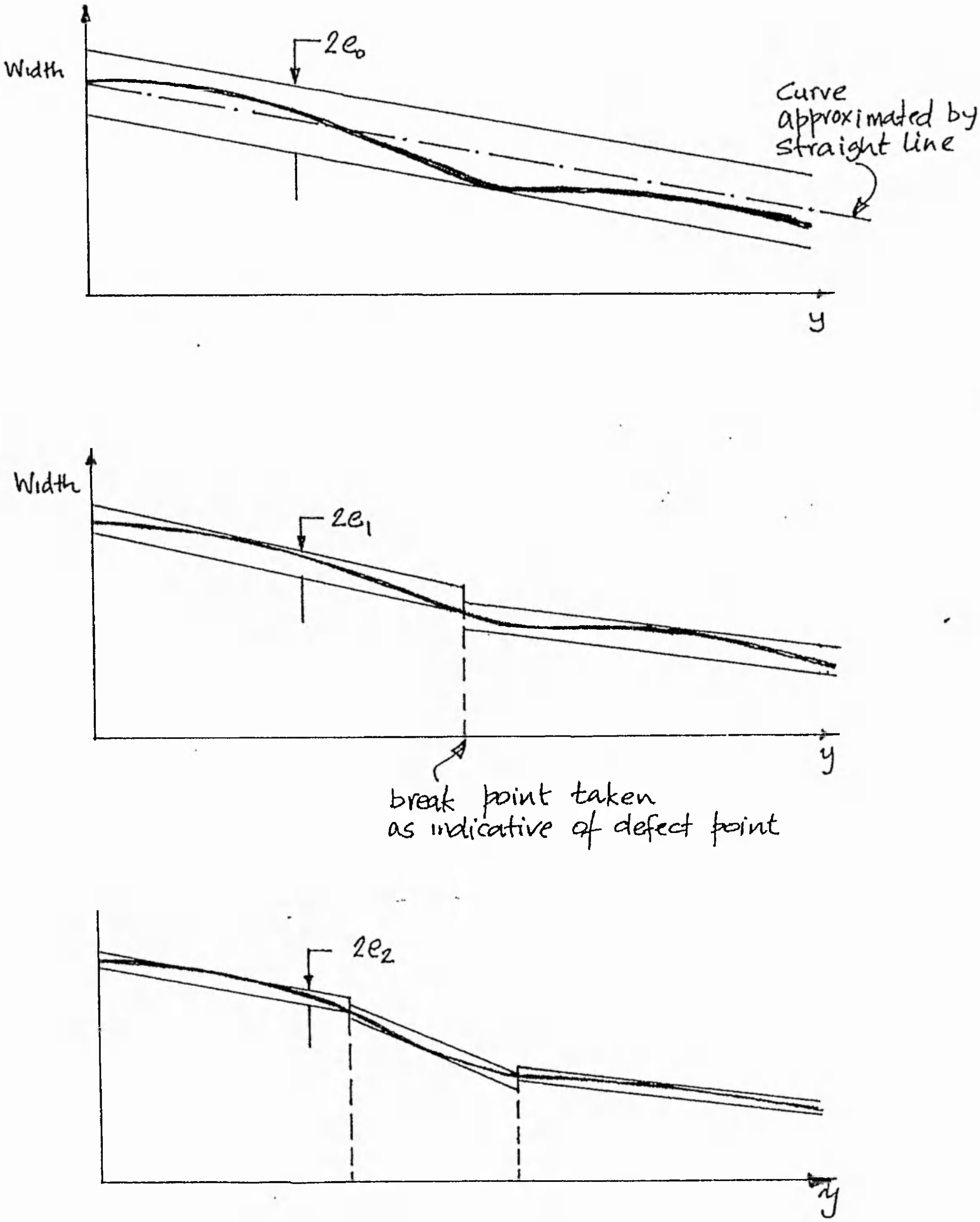


FIG. 3. MONITOR VIEWS OF IMAGE STRIPE AND SOFTWARE INTERPRETATION

Fig 4. POLYGONAL APPROXIMATION OF A WIDTH CURVE AT DIFFERENT THRESHOLD VALUES



Excerpts from the second report

Introduction

This study was planned as a continuation of the study carried out earlier at .... The machine vision work reported here was carried out on the Gaydon Technology premises ..... In order to facilitate the study a rig was designed and manufactured in the Department of Industrial and Production Engineering, Trent Polytechnic. The rig enabled the mounting of the camera and positioning of the skin panel for the study. This rig was used to inspect six car skin panels using machine vision reported earlier.....

Methodology

1. From the starting position the panels were indexed forward by 100mm and defects, if any, were ascertained using the machine vision system as described in the earlier work carried out at Castle Bromwich.

2. Small areas close to the edges and feature lines of the panels were not investigated as software needs to be modified to do so. (It was deemed not necessary at this stage of the study).

3. Locations of the defects were measured with respect to the edges or feature lines of the panels and recorded.

4. Nature of defects were classified as 'high', 'low' or 'ripples'. Assessment of the nature of the defect was done visually, as this feature is not available in the software.

5. All panels were inspected at the software settable resolution of 15. However, some areas of a few panels were also studied at a resolution of 10, for the comparison.

Comparison of results

This was carried out in the presence of inspection personnel from .... Castle Bromwich, who presented the manual inspection report on the six car panels.

1. Exact locations are not given in the ... inspection reports. Thus closer comparison is not possible. However globally most of the defects listed ... have been detected by the machine vision system. In fact the machine vision system finds even more defects not listed in the manual inspection reports, but was agreed to by the ... inspection representative as present on the panels on closer scrutiny.

2. It appears that the .. inspection report covers the middle section of the lower quarter of the top section of the door panel. Thus these defects have been compared in these areas although it was accepted that there were defects in the other areas as detected by the machine vision system. The comparison was found to be favourable in general for the machine vision system. ....

3. 'Slip lines' ... have not been detected by the machine system. This could be partly due to the fact that approximately 18mm wide area near the feature lines (near where these slip lines have been located) have been not investigated by the machine vision system. (Software should be modified to take care of this).

Appendix D

Panel camera pixel resolution

The x and y co-ordinates used for constructing cosmetic maps are obtained from the images of the camera that views the panel. For the usual configuration the robot takes during inspection, the approximate scene size of this view can be considered to be a square of 175mm x 175mm. The frame buffer of the vision system stores this scene as a 512 x 512 pixel array. This results in a square pixel resolution of approximately 0.35mm/pixel.

Adequacy of resolution for inspection

The defects appearing on panels vary in size and shape. The coverage of the smallest defect on the surface of a panel is larger than 10mm diameter. With a pixel resolution of 35mm/pixel, this is equivalent to a diameter of approximately 30 pixels. This resolution was found to be more than adequate for detection of all defects that are of significance. Experiments done at half the resolution in the y-direction by ignoring every other raster line of the image did not produce any deterioration in the performance of defect detection. Since such an approach was computationally attractive in terms of memory requirements and speed, it was implemented in the final algorithm.