



City Research Online

City St George's, University of London

Citation: Norton-Wayne, L. (1982). The Detection of Defects in Automated Visual Inspection. (Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/35384/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

THE DETECTION OF DEFECTS IN AUTOMATED VISUAL INSPECTION

Thesis Submitted for the Degree of Doctor of Philosophy

Leonard Norton-Wayne

Department of Systems Science

The City University, London

September 1982

TABLE OF CONTENTS

GLOSSARY	xii
ACKNOWLEDGEMENT	xv
SUMMARY	xvii
CHAPTER 1 INTRODUCTION	
1.1 Automation of visual decision making- purpose of this thesis	1
1.2 Nature and specification of visual inspection requirement	13
1.3 Previous work on instrumentation for detection of visible defects.	17
1.4 Previous work on signal processing for automatic detection.	37
1.5 Inspection of cold rolled steel strip.	44
1.6 Plan and structure for this thesis.	50
CHAPTER 2 SIGNAL ACQUISITION AND PROPERTIES	
2.1 Introduction.	53
2.2 Scanners.	56
2.3 Theoretical properties of noise.	71
2.4 Measured properties of noise.	85
2.5 Mathematical model for noise generation.	90
2.6 Properties of defect signals.	92
2.7 Conclusions.	99
CHAPTER 3 DETECTION-FUNDAMENTAL CONCEPTS	
3.1 Introduction.	103

3.2 Detection as a decision process.	106
3.3 Decision taking in detection- the maximum likelihood detector.	112
3.4 Alternative approaches to decision in detection.	118
3.5 Ideal detectors.	123
3.6 Concluding remarks.	128

CHAPTER 4 DETECTION-DECISION AND CONTRAST ENHANCEMENT

4.1 Introduction.	130
4.2 Detection using decision only.	133
4.3 Contrast enhancement using filters.	138
4.4 Contrast enhancement using channel combination.	154
4.5 Detection of non-local defects.	163
4.6 Conclusions.	166

CHAPTER 5 ELIMINATION OF FALSE ALARM TRIGGERS USING ASSOCIATION

5.1 Introduction.	168
5.2 Association in a box.	172
5.3 Association using 'nearest neighbour' cells.	182
5.4 Association using 'add,dump and threshold' (ADT) filters.	186
5.5 Comparison of methods using theory.	193
5.6 Analysis of resulting improvement in detection performance.	195

5.7 Evaluation using simulation analysis.	203
5.8 Conclusions.	206
CHAPTER 6 COMPOSITE DETECTION SYSTEM	
6.1 Introduction.	208
6.2 Trade-offs in channel combination.	210
6.3 Recommended configuration.	217
6.4 Adaptation.	223
6.5 Hardware implementation.	228
6.6 Conclusions.	250
CHAPTER 7 CONCLUSIONS	
7.1 Introduction.	253
7.2 Conclusions regarding signal processing methodology.	254
7.3 Conclusions regarding systems aspects and hardware implementation.	260
7.4 Suggestions regarding further work on automated visual inspection.	262
APPENDIX 'A' DATA BASE AND PROGRAM PACKAGE	269
APPENDIX 'B' NATURE AND PROPERTIES OF SIGNALS	287
REFERENCES	304
LISTINGS OF COMPUTER PROGRAMS (MICROFICHE)inside back cover.	

FIGURES

Chap.1

- 1.1 Strip divided into inspection bands.

Chap.2

- 2.1 Signal describing a moving surface.
- 2.2 Flying field scanner.
- 2.3 Flying spot scanner.
- 2.4 Variation of surface noise with detector aperture
(white and laser light).
- 2.5 Probability density functions for speckle compared
- 2.6 Variation in surface noise with angle of beam incidence
- 2.7 PDF estimates for filtered surface noise,
showing Gaussian tendency.
- 2.8 Spectra for individual scans, averaged over a sheet
- 2.9 PDF for clean sheet.
- 2.10 Bi-modal PDF from bent sheet.
- 2.11 Histograms of sample levels for sheet divided into
regions.
- 2.12 Autocorrelation of surface noise in the Y-direction.

- 2.13 Scatter diagrams of surface noise.
- 2.14 Correlations between successive pairs of scans.
- 2.15 Surface microprofile for cold rolled steel strip.
- 2.16 (a)Synthesis and (b) Analysis of correlated random noise signals.
- 2.17 Message signals in terms of optical interaction.
- 2.18 Reflected intensity vs. angle from specular.

Chap.3

- 3.1 Triggers generated by a simple detector.
- 3.2 Canonic form for detection system.
- 3.3 Detection described as a statistical decision process.
- 3.4 Receiver operating characteristic-form(1)
- 3.5 Receiver operating characteristic-form(2)
- 3.6 Payoff matrix for 'Game Theoretic' detector.
- 3.7 'Perfect' detector.
- 3.8 'Optimal' detector.

Chap.4

- 4.1 SIRA Detector-principle of operation.
- 4.2 Response of SIRA detector to different views.
- 4.3 Detection of edge strain.
- 4.4 Matched Filter Bank(MFB) detector-block schematic
- 4.5 Matched filter bank detector-analogue outputs.
- 4.6 Point wrench sensed with standard processing.
- 4.7 Pinch marks sensed with standard processing.
- 4.8 Matched filter described in signal space.
- 4.9 Seams undetected using SIRA detector.
- 4.10 Two-dimensional filter for enhancing seams.
- 4.11 Two-dimensional filter for enhancing point wrench
- 4.12 Seams revealed by a two-dimensional matched filter
- 4.13 Contrast maximisation using weighted linear combination.
of signal components.
- 4.14 Detection of sticker wrench analysed in signal space
- 4.15 Sticker wrench-normal contrast.
- 4.16 Sticker wrench-videoprints using standard detector.
- 4.17 Sticker wrench-videoprint from optimally processed
signal.

- 4.18 Processing for detection of sticker wrench.
- 4.19 Sticker wrench-reversed contrast.
- 4.20 Sticker wrench-sensed by CCD array.
- 4.21 Chatter marks detected using spectral analysis.
- 4.22 Spectral analysis of sheet without chatter marks.

Chap 5

- 5.1 Principle of noise elimination using trigger association.
- 5.2 Staggered association regions.
- 5.3 Principle of two-stage association.
- 5.4 Principle of 'nearest neighbour' filter.
- 5.5 Sheet divided into strips for 'add,dump and threshold' (ADT) filter.
- 5.6 'ADT' filter as a finite state machine.
- 5.7 Performance improvement following trigger association-continuous Gaussian signals.
- 5.8 Probability of false alarm vs.threshold level.
- 5.9 Performance improvement using trigger association

negative exponential noise.

- 5.10 Trigger association in cells 10 samples by 10,

5.11 Trigger association in cells 5 samples by 12.

5.12 Effect of one pass through 'AJ' filter.

5.13 Effect of three passes through 'AJ' filter.

5.14 Effect of 10 passes through 'AJ' filter.

5.15 Effect of 'ADT' filter, one sample per band.

5.16 Effect of 'ADT' filter, six samples per band.

Chap.6

6.1 Channels combined before filter.

6.2 Channels combined immediately after filter.

6.3 Channels combined immediately after threshold.

6.4 Channels combined after trigger association.

6.5 Regions of signal space contributing noise
triggers for three thresholding schemes.

6.6 Some more realistic message signals.

6.7 Degradation in detection performance when
all forms of message must be detected.

6.8 Quadratic detector -two signal components .

6.9 Variation in trigger count measured in on-line trials.

6.10 System for detecting defects on cold rolled steel
strip.

6.11 Two-dimensional delay line filter.

Appendix 'A'

A.1 Log sheet showing defect locations.

A.2 Subroutines integrated to form software package.

A.3 Perspective plot and videoprint compared.

A.4 Grey scale plot and videoprint compared.

Appendix 'B'

B.1 Orthogonal representation described in signal space.

TABLES

- 1.1 DEFECT TYPES CONSIDERED IN SIMULATION
- 2.1 COMPARISON OF SCANNER TYPES
- 2.2 COVARIANCE RELATIONSHIPS BETWEEN SIGNAL COMPONENTS
- 4.1 PERFORMANCE OF SIRA DETECTOR
- 4.2 DEGRADATION IN MATCHED FILTER GAIN DUE TO MISMATCH
- 5.1 PROCESSING GAIN, SINGLE LEVEL CELL ASSOCIATION
- 5.2 PROCESSING GAIN, TWO LEVEL CELL ASSOCIATION
- 5.3 PROCESSING GAIN, ADJACENCY DETECTOR
- 5.4 PROCESSING GAIN, 'ADD, DUMP and THRESHOLD' FILTER
- 5.5 GAINS OF ASSOCIATION METHODS COMPARED
- A.1 COMPOSITION OF DATA BASE

GLOSSARY

ADT	'Add,dump and threshold' filter.
AJ	Adjacency filter.
BSC	British Steel Corporation
CCD	Charge coupled device
DFT	Discrete Fourier transform
FFT	Fast Fourier transform
Hooghovens	Dutch Steel Manufacturing Firm
l	signal(generally)
m	the message component of a signal
message	signal component giving useful information
MFB	Matched Filter Bank
n	the noise component of a signal
noise	signal component not giving useful info.
ORSID	French steelmaking research organisation
P(cd)	probability of correct detection
P(fa)	probability of false alarm
P(md)	probability of missed detection
P(na)	probability of correct 'noise alone'
PI	the constant π
pdf	probability density function
s	standard deviation
ROC	receiver operating characteristic

signal	physical observable conveying information
SIRA	SIRA Institute, Chislehurst.
U	processing gain using trigger association
VDEH	German steelmaking research organisation.
W	distortion resulting from trig. association
X	probability of detection improvement resulting from trig. association.

SYMBOLS DENOTING OPERATIONS

$\langle A \rangle$ indicates mean value of A

$A * B$ indicates A multiplied by B

$A ** B$ indicates A raised to power B

$A \sim B$ indicates A approximately equal to B

$\text{SQRT}(A)$ indicates square root of A

ACKNOWLEDGEMENT

I should like to thank Professor L.Finkelstein Professor of Measurement and Instrumentation at The City University and Manager of the Instrument Systems Centre, who has directed the programme of research on which this thesis reports, and who supervised its production. His faith in the ultimate success of the work in automated visual inspection, in its early days amid a universe of scepticism, is particularly appreciated.

Many provocative (and acrimonious!!) discussions with my colleague [REDACTED] have proved invaluable in defining and illuminating the detection problem. Financial support from the Science and Engineering Research Council, European Coal and Steel Community (via the British Steel Corporation), and SIRA Institute are gratefully acknowledged. Collaboration with the latter organisation through [REDACTED] [REDACTED] has proved outstandingly valuable.

DECLARATION

I grant powers to the University librarian to allow this thesis to be copied in whole or in part, without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

SUMMARY

This thesis is concerned with the application of modern concepts in instrumentation, and particularly signal processing, to systems for intelligent industrial automation. Its chief objective is to provide a signal processing methodology for defect detection in automated visual inspection.

Previous achievement in the field is surveyed in the introduction. Since the properties of signals used in surface inspection depend critically on the opto-electronic scanning devices used for signal acquisition, these receive first consideration. Then, from concepts generated originally for target detection in radar and sonar, a canonic form is developed for the detection process. This comprises the three consecutive stages, contrast enhancement, decision, and trigger association. Alternative forms of processing for the three stages are proposed, and analysed theoretically using statistical concepts. Computational experiment, in which the processing methods are implemented as computer programs operating on stored data, is then used, to confirm the practical effectiveness of the methods, and to compare alternatives. A distinction is made between local and global defects, and methods are discussed for detecting both kinds. Alternative methods of combining the various methods together are described and compared. Finally, a complete system is proposed, for the detection of visually perceivable defects on cold rolled steel strip, which is the material used during the simulation experiments. It is concluded that the suggested processing will yield inspection systems whose performance meets or exceeds operational requirements, and is consistently superior to that of a human operative in speed of operation, consistency, objectivity, and in ability to detect defects of poor contrast.

CHAP.1 INTRODUCTION

1.1 Automation of Visual Decision Making Purpose of this Thesis

The objective of this thesis is the development and validation of a methodology for the detection of defects in automated visual inspection. This involves study of sensors for obtaining an electronic signal characteristic of the optical properties of a rough surface which may move rapidly, but is concerned principally with the provision of signal processing methods capable of distinguishing the message signals generated by the defects from noise generated by surface roughness. The aim is to provide an optimal processing strategy considered to have general applicability. The validity of the processing methods is established by applying them in computer simulation to a real inspection problem which is demanding and significant. This is the detection of surface defects on cold rolled steel strip.

Conventionally, the decision process inevitable in visual inspection is performed by the human brain, using information extracted by the eye-brain visual processing

system. Visual inspection by human operatives is unfortunately slow, subjective and prone to error, fatiguing to perform and expensive, and is consequently used only because nothing better is available. It is highly desirable therefore that it be replaced by automatic instrumentation involving electronic signal processing. The viewpoint taken is that since human vision is capable of detecting (and, if trained, of identifying) all classes of surface defect provided they are presented slowly enough, then it must, at least in principle, be possible to devise electronic processing methods having the same capability, and to implement these as programs on a man-made computer. The human brain is simply an alternative form of computer, different from electronic computers made by human engineers in its method of storing and transmitting information and in its organisation, but not necessarily superior in its ultimate capability.

The nature of the processing whereby a defect is perceived and selectively identified by a human observer is not properly understood. Defect detection certainly involves a statistical decision process, (performed subjectively and unconsciously by the human observer)

which operates upon noisy data. To automate the process, this must be imitated electronically, using instrumentation. Since the contrast of the signals representative of the defects is often poor relative to the accompanying noise, errors are unavoidable. It is necessary to give careful attention to the camera device which transduces the optical description of the inspected sample to an electronic signal, but the critical problem is the provision of adequate methods for the signal processing. Fortunately, there exists a substantial body of work on the related problem of detecting targets in radar and sonar against a background of noise at low contrast. This is exploited in the thesis as a starting point, from which methods applicable to automated defect detection are developed.

It is, further, essential that any methodology proposed be clearly shown to be effective. This is achieved using a simulation investigation for a particular task (namely, the detection of surface defects on cold rolled steel strip). In the work to be described, signal processing methods are implemented as computer programs, and applied to stored data, with satisfying results. It is demonstrated that automated

systems are possible whose performance is significantly better than the human operatives they replace.

The possibility that machines might someday take over from human operatives many of the more trivial and uncongenial activities required in human civilisation has long fascinated philosophers. As a source of motive power, machines have replaced muscle for perhaps 150 years. Mechanisation of activities involving thought, particularly those in which the human eye-brain system makes a critical contribution, is however much more difficult.

Recently, however, electronic hardware has become available which provides information processing which is cheap, reliable, versatile (programmable and adaptive), and compact. Whilst integrated electronic circuits, particularly microprocessors and cheap semiconductor memory, have been the most important innovations, opto-electronic devices such as laser scanners and CCD array cameras have also been significant. There has simultaneously been an explosion in theoretical understanding of the nature of information, and of its application in

control,sensing,detection and identification, the sorting and coding of data,communication between man and machine,and so on.

The result is that replacement of human activity in many of the routine aspects of industrial production has now become practicable. Application of the sophisticated processing termed artificial intelligence to mechanise activities hitherto restricted to the human eye-brain system produces robot instrumentation systems. These are generally faster,more objective and consistant,less expensive and more tractable than the human operatives they replace. The result is improved industrial efficiency.

Robots are of two types; PASSIVE robots,which may see,hear and take decisions, and ACTIVE devices which can in addition move,to pick and assemble for example, and must hence cope with a much less constrained environment. The latter thus require a substantially higher degree of intelligence, i.e.,information processing which is faster and much more complex. The distinction between passive and active robots is hence almost one between plants and animals. In this work,we shall consider only the

simpler, passive kind of robot, whose vision is used for automatic inspection, i.e., for the automatic examination of material produced by an industrial manufacturing process, to establish that the quality of the material meets a specification. It may be argued that in comparison with more sophisticated 'animal' robots which may interact with their environment and adapt their behavior and form plans, this form is insignificant. It is known, however, that automatic surface inspection comprises a problem of real economic importance in industry. Further, unlike most problems in active robotics, widespread application of passive robots to real industrial problems is much closer to realisation. All that is required is for powerful methods of signal processing theory, developed for use in other fields to a high degree of sophistication, to be brought to bear, with extension and development where necessary. Since many other aspects such as the physics of the interaction between radiation and material, and the performance of human beings under stress, and operation to a specification suitable as a basis for commercial negotiation, must be taken into account, this is a true systems problem.

The social and economic consequences of robotics, in terms of what shall be done with the operatives it displaces or the extra wealth it creates, is the concern of politics not engineering, and will not be further considered here. In fact, we shall hereafter concern ourselves not with robots but with INSTRUMENTATION and so eliminate an emotional subjectivity which might be disturbing and controversial.

In this thesis we study in detail one particular robot activity which involves mechanised vision, namely, inspection for visually perceivable defects. This requires the development of a methodology for the detection of irregularities on surfaces which are moving rapidly. The latter are generally noisy in producing additional information which is unwanted, and which tends to conceal the defects. This involves the development of a statistically-based theory applicable to the analysis of electronic signals. The validity of this theory is then established by applying it to a specific real life inspection problem (in simulation), using a general purpose computer processing stored data. The problem is the detection of surface defects on cold rolled steel strip. However, the applicability of the

methodology is quite general, though modification may be necessary to accomodate variation in properties between different materials.

A principal motivation for the introduction of automated visual inspection lies in the increased speeds possible in lines manufacturing strip material such as sheet steel, as a result of the incorporation of automated control in which a computer is used to optimise operating parameters in real time. The inspection which is an indispensable final stage in the manufacturing process cannot however be speeded up accordingly, without use of automated vision of the type to be described.

We define a DEFECT as being some property of an article which is not normally present (i.e., which represents a departure from the 'design specification') and which further is unwanted or undesirable, in making the article less acceptable to a purchaser or user. One requirement in industrial inspection for quality control is an ability to detect defects, i.e. to generate an indication that a defect is present, so that the defective material may be repaired, directed to a less demanding application, or so that a fault condition in a

manufacturing process may be identified and corrected.

In practice, detection instrumentation is never perfect, thus some defects will be missed. Also, 'false alarm' indications will be generated, in which a good product is designated as being defective. Generally a trade-off is involved between the two types of error, decided in the end by their consequences, for example, in goodwill lost from a customer sold a poor product, or of damage caused to a machine in a subsequent process, or, alternatively, of wasted good material. Particularly with the current imperfect state of the art of automated visual inspection, it is often useful to follow an initial machine inspection stage by a second stage of inspection by a human operative. In this hierarchical approach, the machine first takes the easy decisions, cheaply and quickly. The more difficult decisions remaining are then taken by the human. His workload is consequently much reduced, leading to decreased strain and fatigue. The decisions taken automatically are generally boring and trivial, requiring little exercise of judgement. Those remaining are more challenging and hence interesting. The operative's efficiency and happiness, and the cost effectiveness of

the overall system,are increased in consequence.

There may be a further requirement for the location and extent of defects to be specified,or their severity,or even,ultimately,for them to be identified as being of some particular kind,so that appropriate corrective action may be taken,or so that the article or material may be directed to the particular use or customer who can tolerate whatever defects are present, but possibly none more severe.

The theory for the detection of useful message information buried in competing noise information is well developed,because of its applicability in the well funded defence activities of radar and sonar. However,most of the basic ideas were developed originally for use in industrial quality control,and were then adopted for use in electronic warfare,around the end of the second world war. The standard theory of target detection in more or less its current state of development has been well summarised by Root(1970).Though there have subsequently been further developments,notably in treating the problem as one of time sequence analysis,their advance is chiefly in ease of implementation in digital hardware. There has

also been some progress in development of ability to adapt to changing conditions, and to estimate parameters which are unknown to optimise performance. However, the characteristics of the signals found in surface inspection differ sufficiently from those in radar and sonar (notably in being multi-component vectors describing a two-dimensional phenomenon) that some extension of the general theory is possible. Unlike those in radar and sonar, the messages in surface inspection are made by God rather than by man. They cannot be designed; their form is 'a priori' unknown, and is generally highly variable.

The first stage in any system for automated visual inspection is a camera for transducing the optical description of a surface, (which is sensed visually by a human) into an electrical signal for processing electronically. Thus, this thesis starts by analysing suitable cameras, and defines parameters which enable alternative cameras to be compared. Then, the properties of the noise and message (defect) signals as predicted by theory are discussed, and are compared with the results of experimental measurement. A model is proposed to account for the properties of the noise. The theory of the

detection of message signals in noise is then discussed in detail, since this is crucial to our problem. Arising from this, we develop a canonic form to describe detection systems (several alternatives are possible), and subject this to theoretical analysis followed by verification using computer simulation. Considerable improvement in performance over the capability of naive processing will be demonstrated, to the point where further improvement becomes pointless. This occurs when there are present on the surface innocuous marks of higher contrast than the defects of interest, and a completely different philosophy must be adopted to improve detection performance. This involves classification of the surface marks, rather than merely sensing their existence.

The simulation confirms the importance of selecting the best possible camera device for signal acquisition, since this determines the contrast of the defects. However powerful the signal processing may be, it can use only the information provided in the camera output. It is further necessary to establish that the processing methods investigated can be integrated into a system capable of operating economically, in a manufacturing plant, given knowledge of the properties of

message signals which is inevitably incomplete and may vary with time. The high rate at which information must be processed makes hardware implementation particularly challenging.

Finally, conclusions are presented, regarding processing methods, systems, hardware, scanning devices, shortcomings in the state-of the -art, and directions for further investigation.

1.2 Nature and Specification of the Inspection Requirement

The objective of automated visual inspection is to determine whether the quality of a product meets its specification, using visual information only. In surface inspection, quality is determined by the number, location, type, extent and severity of defects visible on the surface. Although (Norton-Wayne, Hill and Watts, 1980) it is really desirable that defects be identified by type, detection of their presence constitutes a useful first step. The parameter defining surface quality is then the proportion of surface which is contaminated.

Conventional surface inspection is performed visually by a human operative and is hence subjective. However, although a human inspector is inconsistent and unreliable (Moraal, 1972, 1974), he can exercise judgement and interpret what he sees with some flexibility. Quality designations can thus be varied to accommodate for example variation in supply and demand for a product. Instrumentation is in contrast objective and inflexible. A series of measurements (rather than estimates) is made, from which a quality designation is obtained by algorithm. This must be consistent with the quality designation produced by a human visual assessment.

The area and contrast of defects may easily be measured instrumentally. Further quantitative information is available from the shape of the defect signal, and from its effect on consecutive scans. From measures of this kind, a quality designation may be obtained.

For example, calling the measured properties $v(1), v(2), \dots, v(i), \dots, v(N)$, we can form a 'quality' parameter Q as a weighted linear sum of the N measures $v(i)$, i.e.:-

$$Q = \sum_{i=1}^N w(i) \cdot v(i) \quad \dots\dots \quad 1.1$$

Here the weights $w(i)$ are chosen to make Q compatible with the quality designation desired, e.g. that produced by a human inspector.

Most of the simulation work to be described was performed in collaboration with the British Steel Corporation (BSC), whose quality requirement for the surface of strip products was in terms of the area of strip contaminated by defects. The presence, position and type of the defects was of more interest than their size or severity. To enable the inspection instrumentation to provide an indication as to quality consistent with the BSC requirement and current inspection methods, the following specification has thus been adopted for surface quality, in the simulation work. It accords with commercial practice and is compatible with the designation produced by a human inspector.

The steel strip is assumed divided (fig.1.1) into bands 12.5 cm. long in the rolling direction, which extend the full width of the strip. The material within a band

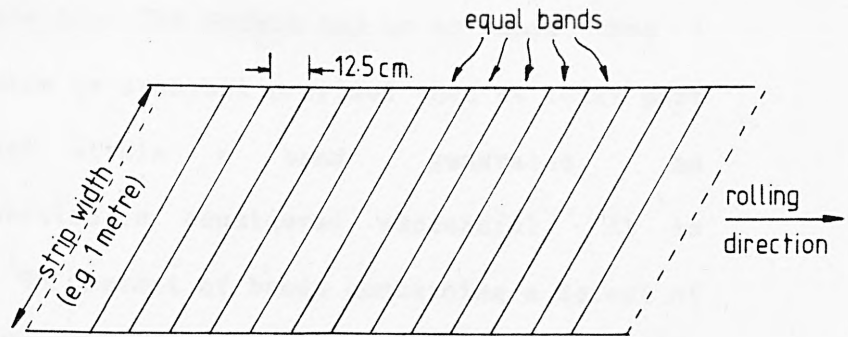


FIG 1.1 STRIP DIVIDED INTO INSPECTION BANDS

is considered defective if one or more defects are indicated inside it. The defect may be no more than 1 square millimetre in area, but provided that at least part of the defect within a band generates an indication, detection is considered successful. It is required that 95 percent of bands containing a defect of any kind be indicated.

Detection instruments sometimes generate 'false alarm' indications, in which defect free bands are designated as containing defects. The result is waste of good material. The inspection requirement insists that no more than 2 in every 1000 clean bands be thus incorrectly designated. For a strip 1 metre wide, with scans at 1 mm. intervals and samples spaced by 0.4 mm. along each scan, the probability of false alarm $\{P(fa)\}$ per sample cannot exceed 6.4×10^{-9} if this requirement is to be met. This specification of the detection requirement is used in chapters 4 and 5 to provide quantitative parameters for the signal processing. Most of the systems described in the review of previous work provided in the next section (1.3) use alternative (or unstated) definitions for the inspection requirement, making performance figures hard to compare.

1.3 Previous Work on Instrumentation for Surface Defect Detection

We consider here previous work on the automated detection of visually perceivable defects present on the surface of a moving strip of material, and in particular on detection of defects on metal surfaces, where defects must be sensed in the presence of a high noise background.

The earliest instrumentation for automated inspection for visually perceivable defects on moving strip material appears to have been used for paper inspection (Stapely 1965, Brook 1971b), but this application is relatively undemanding, since the defects normally extend right through the material which can thus be illuminated from behind, so that defect contrast is very high.

The earliest reported attempt to apply instrumentation to facilitate the inspection of moving surfaces (rather than translucent material) for visually perceivable defects started around 1960. A survey of

quality control methods for sheet steel products performed by Thomas (1961) contains no mention of automated visual inspection. The first known publications which touch the subject, by Sykes and Burns (1963) describe instrumentation using vidicon TV cameras as scanners on sheet tinplate. The investigators made a careful examination of the way in which defects interact with incident light, and concluded that viewing slightly off-specular giving a bright defect on a dark field gave best results. Drums of mirrors were used to freeze the motion of the surface. The actual decision as to whether or not a defect was present was made by a human inspector, thus the system was really only an operator aid. It was evaluated on-line for five months, but does not seem to have been adopted.

The first system in which the detection decision was completely automatic was reported by Stapely (1965). This had been developed from a system designed for use on a paper mill. It used flying field scanners each covering a width of 50 cm to examine a total width of 150 cm. The area element size was 12 mm in the rolling direction by 6mm in the scanning direction ;thus, although the system was tuned to respond preferentially to

longitudinal defects and achieved 100% coverage of the surface, the resolution was poor by current standards. The scanner could however accomodate strip flutter of plus or minus 6mm.

Stapely's paper does not provide a detailed description of the signal processing used, but it may be inferred that his system was sensitive to changes in signal level rather than to the level itself, (i.e. that the signal was differentiated before being thresholded), and that the system was thus adaptive to variation in surface reflectivity but insensitive to large area defects of low contrast. A special detector was also available optionally for defects whose contrast change was transverse to the scan direction.

The most noteworthy feature of Stapely's work is the thoroughness with which its performance was evaluated. The first evaluation involved 302 samples of defects in tinplate (86% detected successfully), 198 defects for cold rolled steel strip (78% successful), and 126 for stainless steel (61% successful). For the cold rolled strip, stickers, edge strain and chatter marks (amongst others) were found difficult to detect. The

significance of the other defects mentioned as being difficult is hard to assess, because defects with the same name are not reported in steel manufacture in the U.K. The performance overall was much below the 95% required by BSC, so it is not surprising that at the time of publication, no system had been installed in a steelworks.

The paper contains little discussion or detailed explanation of the signal processing methods used. Most important, there is no definition as to exactly what is meant by detection.

A little later, around 1970, Allnutt and Brook at the SIRA Institute (Brook 1971a) were performing laboratory investigations on defect detection systems. They used flying field scanners developed originally for inspecting paper. The work resulted eventually in a system used 'on line' in a tinplate works (Brook et al. 1978), but is most significant for its origination of concepts and methodologies much used in subsequent research. Most important, the need for improved scanners was realised, leading to the SIRA laser scanners, which have much increased speed and less noise.

Bowers(1970) and Hill(1972) ,also Webber (1970) and Donnelly (1971),attempted to apply theoretical analysis to the signals generated by SIRA scanners in the hope of improving performance. All failed since they did not appreciate the nature of the detection process,as a statistical hypothesis test for determining the presence of an essentially LOCAL message buried in noise. However,both used stored digitised data,processed off-line in a general purpose digital computer,using programs written in high level language. This was a major fundamental advance in the methodology of research in surface defect detection,and was important in establishing the procedure used invariably in later more successful work. The advantages of working with stored data in simulation cannot easily be oversold; the ability to evaluate many methods in succession over the same data to obtain quantitative comparison,the availability of graphic output (particularly on microfilm) to provide irrefutable specification of detector efficiency,the facility thereby to compare sample material with the results of the processing (on line, the sample material is evanescent,and correlation of indications with samples is almost impossible),and the utility of the computer for keeping quantitative and

reliable records of its own findings are indispensable. Unlike the original material, digital data describing surfaces does not deteriorate with time providing reasonable precautions are taken. The ease with which complex processing methods may be implemented is remarkable, particularly when library subroutines are used.

In 1974 workers at Glasgow University (Logan and McLeod, 1974, Logan, 1974) published results of an investigation on surface inspection using data acquired by a SIRA flying spot scanner. They had seen the initial stages of the work reported herein two years previously. Their work concentrated on identifying defects by type, and the detection problem was not realised, chiefly because their data base was unrepresentative. Their work has nevertheless introduced many useful concepts and methodologies for defect identification.

By 1972 laser scanners were becoming available for surface inspection applications, promising much increased operating speed and sensitivity. The noise component of the output of these scanners is dominated by the

roughness of the scanned surface rather than by self noise generated within the scanner.

The first laser-based inspection system for cold rolled strip to be described in the open literature was developed at the Axel Johnsson Research Institute in Sweden (Sjolin,1972,Nordquist and Millgard,1974). Its operation is based on advanced principles.

The property is exploited that the far-field scatter pattern of light returned from the surface is the Fourier transform of the complex reflectivity of the spot of surface instantaneously illuminated, and that when brought to a focus, the position in space of the distribution of energy is independent of the location of the illuminating spot within the field of view. Nordquist and Millgard had recognised that poor contrast was the reason for poor detection performance in many cases, and they used masks applied to the scattered radiation to increase the contrast. This is a form of two-dimensional spatial filtering, implemented optically rather than electronically. One can assume, for example, that for isotropically rough sheet, the distribution of energy in the transform plane will have

cylindrical symmetry about the origin, and will decay in intensity radially with distance from the origin. The rougher the surface, the gentler will be the decay. If however a scratch is present along the 'y' direction, then in the Fourier plane a corresponding bright line will appear in the 'x' direction. If a disc shaped diaphragm is placed in the Fourier plane to reject light energy distant less than a threshold R from the origin, then the contrast, i.e. the change in received light energy due to the presence of a defect, will be enhanced considerably. The optimal form for the diaphragm will depend on the shape of the defect; by using a slit open in the 'y' direction (and with the center portion blocked), high sensitivity can be achieved to scratches in the 'x' direction, and so on.

As Nordquist and Millgard (N and M) point out, the resulting system can thus be tuned to respond preferentially to defects of interest, but to ignore other kinds, and hence to provide information regarding defect type.

The N and M system covers the strip completely using a line of heads each of which scans a width of

25cm. Six such heads will thus be required to cover the 1.5 metre strip width usual in the steel industry. It is claimed in the laboratory to detect reliably on cold rolled strip scratches 20 microns wide and 2 microns deep, and dark spots with a diameter of 400 microns. The system when evaluated 'on line' in a steel works is claimed to detect 93% of occurrences of defects from 13 (unspecified) groups. Presumably these cover all types occurring on cold rolled strip. This falls only slightly short of the exacting performance requirement set by BSC, though the strip speed is much lower.

Further, the importance of achieving compatibility between the assignments as to quality made for a particular material by man and by machine was appreciated fully. The performance of the automated inspection was measured to be more consistent than different human inspectors, and the surface was divided into regions (1cm.squares) with grade assigned according to the count of defects within the square. The system is claimed to be operable at line speeds up to 20 metres/second.

However, the masks would seem to be optimal for one kind of defect only, and it is hard to see how several

kinds of defect having different forms (lines parallel to rolling direction, lines perpendicular, spots, etc) can be accommodated with one set of scanning heads, since light, unlike an electronic signal, cannot easily be amplified to permit the same information to be examined simultaneously by several Fourier plane processors. How the system deals with area defects like chatter marks, or whether it can detect sticker wrench, are unknown. Most important, many defects merely absorb incident light, and could not be detected with a masked Fourier plane detector, which blanks out the centre spot in the intensity distribution in the transform.

Thus, although the performance claimed for the system is impressive, its significance is hard to assess. Its high technical sophistication is beyond dispute, but it must be extremely expensive.

The Japanese worker Akutsu and his collaborators (1976) have described a system based on a flying spot scanner which used white light. Again a battery of scanners was required to inspect the full width of the strip. Each covered a width of 182 mm with a rectangular spot 2mm in the scan direction by 30 mm in the rolling

direction. A signal/noise ratio of 3 for a 0.1mm black line on good quality rolled strip was claimed. However, the maximum line speed acceptable was only 2.5 metres/second (only 10% of that required by BSC), presumably because of the difficulty of obtaining a sufficiently intense white spot. The paper goes on to suggest that the scanner should be developed to incorporate a laser, presumably to give a more intense spot, and a high speed mirror drum to give faster scanning. Why the mirror drum cannot be used with the white light source is not explained.

The Akutsu scanner also has a separate detector for flaws in the cross-rolling direction such as coil break.

The details of the signal processing used are again not explained. However, this seems to be simple, probably consisting only of thresholding the high pass filtered signals. Hence the insensitivity to cross-roll defects. The processing is noteworthy for including a periodicity detector which responds to defects separated along the strip by the circumference of the roller. Also, the width and depth of the flaw signals, and the number of flaws in an 'inspection area', (6cm. by 100 cm.) were

recorded, for quality designation.

The system in its hardware implementation was evaluated on-line by comparison with a human inspector working slowly, and by repeated examination of the same strip. Performance consistently above 97 percent was reported, compared with a human operator working at the normal linespeed of 1.7 metres/sec. It is not surprising therefore to read that 2 systems of the type described were already working on a tinplate line, and that 5 units working at the faster speed of 6 metres/second were to be operational on cold strip lines by May, 1976.

On the other hand, the types of defect on which the system was evaluated are not specified; it is doubtful whether the system described could detect sticker wrench, seams or chatter marks and its speed is quite inadequate.

The appearance of laser scanners in the early 1970s marks a major advance in the development of surface inspection technique. A typical example is the SIRA type 1500 laser scanner shown in figure 2.3. This can cover a strip 1.5 metres wide with a 1mm spot in 200

microseconds. The rapid coverage is achieved using a mirror drum having 12 facets, driven at 24,000 rpm using an air-bearing motor. The most sensitive transducer available, a photomultiplier, is used as sensor, with the additional advantage of sensing over a 5 inch diameter disc to ease the design.

The SIRA scanner provides complete coverage of sheet moving at 15 m/s with the 1.0 mm. diameter spot. To achieve the maximum web speed of 25 m/s, the spot has to be elongated in the direction of sheet movement. Although this produces a slight loss of resolution, it renders the system sensitive preferentially to defects oriented along the direction of sheet movement, which is desirable. The angles of viewing and illumination remain virtually unaltered over the whole width of the scan. Most important, the SIRA scanner has several sensors accepting information simultaneously from the same area of strip, but from different viewpoints. Thus, it generates a vector signal. This provides more information than a single sensor, permitting more complex kinds of signal processing to be attempted, and offering the opportunity for increased sensitivity in detection.

Subsequently, other laser scanners have been constructed. Whereas the SIRA scanner uses lenses (for scanning widths of up to 50 cm) and mirrors (for widths greater than 50 cm) to gather the returned light, scanners by the Erwin Sick company and Bendix Corporation use perspex rods which direct the light to a sensor at one end using total internal reflection. Attempts have been made to incorporate multiple scanners in this kind of sensor, but have not been completely successful. Ferranti (Clarke and Bedford 1977) have produced a particularly simple scanner which gathers the returned light in a simple light box (the Feldmuhle Company makes a similar device), which is cheaper but has inferior performance. The VDEH Institute in Dusseldorf has compared the three basic designs of laser scanner for tinplate inspection (Gries and Hoerster 1977), and has found the SIRA design to give best performance.

Finally, Chittineni (1980, 1982) has described a surface inspection instrument which it is claimed can detect surface defects on magnetic tape and sandpaper and identify them as being as one of four possible types. In his system, signals from successive scans of a laser scanner are passed through a bank of matched filters,

whose outputs constitute a feature set. These are identified in a linear classifier. Combining tentative identifications from consecutive scans in sequence is claimed to improve the fraction of defects detected from 82% to 91%, and to reduce the probability of false alarm from 17% to 3%. No results are available concerning the performance of his system for sheet metals.

Developments in solid state electronics had meanwhile provided yet another kind of optical sensor, based on arrays of solid state photodiodes. This provides low-noise scanners which have no mechanical moving parts, vacuua, filaments or high voltages. Moreover, non-coherent illumination may be used. A solid state scanner whose performance is roughly equivalent to the SIRA 1500 costs only one-quarter the price. Barker and Brook (1978) have cast doubt as to whether solid state scanners can compete with laser scanners at the highest operating speeds because of the difficulty of providing adequate illumination. This controversial problem is explored in detail in chapter 2, section 2.

Modern solid state scanners use line arrays of photosensitive sites diffused in a single silicon

chip, but application of solid state scanners to steel surface inspection began before these became available. A scanner containing 10 arrays each having 100 discrete sensors was produced at great expense (£900,000) by the Dutch steelmaker Hooghovens (Alderink and DeJonge, 1976, DeJonge 1979), which could operate at a linespeed of 28 metres/second.

Laboratory experiments by the Hooghovens workers showed a resolution spot 2mm in diameter to be optimal for spots, and another 2mm by 20 mm (elongated in the direction of line movement) to be best for line defects. A third 10 by 20 mm was found best for stains. A lower level of contrast was felt to be desirable for lines than for stains, and the processing of data from the line windows had to be configured so as not to be disturbed by stains. Specular observation was used, with diffuse illumination obtained from tungsten filament lamps. These consumed 7.2 kw of input power, and had to be cooled by air blast.

The scanner was housed with the processing electronics in a protective metal housing; it dissipated 1kW and was so large and heavy that it had to be moved by

a special crane. High speed operation (and tolerance of variations between individual diodes, which would otherwise mask faint defects) was achieved ingeniously but expensively by operating in parallel. This involved providing a large number of processors, one for each diode or array of diodes, depending on the viewing window size desired. The rather complex processing used is well described in the report by DeJonge . With a minimum window width of 2mm and a maximum linespeed of 28 m/sec, the bandwidth required per channel was only about 15kHz., compared with say 15MHz for a serial scanning system. The processing for defect detection was completely hardwired. Nothing was programmable since there was no computer for signal processing.

Construction of this system was effectively complete by autumn 1975, and a thorough assessment was undertaken during the following year. It is difficult to compare the results of this as described by DeJonge (1979) with those for other systems, since, though the performance of the machine was compared directly with that of a human inspector operating slowly over the same material, the criterion used was quality of the coil overall rather than whether specific defects had been

detected. A reliability of 84% (perhaps consistency would be a better term) was reported, but the system definitely could not detect sticker wrench. Further, there were difficulties with non-fatal contamination which could not be distinguished from significant defects.

The Hooghovens system was expensive and inflexible. It could not easily be modified to incorporate improvements in processing methodology, and used only a scalar signal acquired at the specular angle. The theoretical aspects of its operation were investigated with increasing thoroughness as shortcomings became apparent, but the Hooghovens team lacked computing facilities powerful enough to investigate their ideas properly in simulation, thus their pioneering efforts with two-dimensional matched filters were sterile.

The Hooghovens system has consequently not been adopted by the plant, and a new system is being prepared.

Subsequently Edy (1977) and Yaxley (1979) have experimented at the City University with solid state cameras for cold strip inspection, with encouraging results. These used as sensors line arrays of CCD

photodiodes containing 1728 elements on one chip in the focal plane of a high quality camera lens. Illumination was provided by a 1000 watt tungsten filament floodlamp, but the light was not concentrated into a slit. Using transverse illumination and vertical sensing, sticker wrench was detected successfully using a scalar signal. The full capabilities of solid state cameras for surface defect detection have so far as we know not yet been investigated thoroughly.

TV cameras are still worth considering in some applications. A Hooghovens report (DeJonge 1979) mentions a Philco-Ford Tv scanner based system seen in operation in Dearborn in 1971, but we can find no published description of the device.

Saridis and Brandin (1976,1979) have described a system for defect detection and identification researched using simulation techniques, using photographs of defects as input data. The intention was to use TV cameras or CCD array scanners as input devices in an eventual hardware implementation, viewing the surface vertically. Their work consists largely of a detailed analysis of processing for defect identification, with the systems

aspect largely ignored. Detection seems to have been taken for granted. Only eight kinds of defect were considered for cold rolled strip. Defects known to be difficult such as sticker wrench and chatter marks were not included.

One attempt to use TV cameras for defect detection described by Lacroix, Pons and Pinard (1979) has however been notably successful. Although intended for hot strip rather than cold, the principles are the same. Their system used an Isochon camera as sensor, after early attempts using a Nuvicon (a superior form of vidicon) had been unsuccessful. The Isochon is designed for low light level viewing, and works on the Orthicon principle. (VanDaele, 1977) Strip motion was frozen using a shutter having an aperture time of 100 microseconds. High intensity, constant illumination was used. The blurring was only 1.5mm (equal to the size of resolution cell used) at a linespeed of 17 metres/second. The angle of view was specular.

In preliminary experiments, defects of area greater than 10 cm^2 were reliably detected. Difficulty was experienced only with long, low contrast lines less than

0.5mm wide, and with flaws having a minimum dimension 6mm by 3mm. The 'Add, Dump and Threshold' binary filter described in section 5.4 was therefore incorporated. This permitted a detection threshold to be used on the original analog signal which was close to the noise since it disposed very effectively of the large number of false alarms. The system is of course sensitive preferentially to lines along the rolling direction, which is consistent with the properties of a rolled product.

The equipment classified the defects into four kinds based on length and intensity.

1.4 Previous Work on Signal Processing for Automated Detection

A mathematical theory for handling ordered sequences of random quantities (i.e. stochastic processes) has been evolving since a publication by Pascal in 1654 opened up the theory of probability. The motivation originally was to provide a theory for gambling. The fundamental idea of inverse probability was introduced by Bayes in 1763, inspired presumably by lectures Bayes had attended given by Demoivre. Following contributions by

Laplace in 1812, the theory of probability was almost completely formulated.

The growth of Natural Science around the beginning of the 19th century provided applications for probability theory less frivolous than gambling. In particular, the related science of statistics developed, which deduces the properties of whole populations from measurements on small but representative samples. In natural science, this was applied first to astronomy (hence the contributions of Gauss), then in statistical mechanics, anthropomorphics and finally even to engineering. In 1875 Schols introduced the concept of mutual correlation between variables, from consideration of the impact of projectiles on a target. He was able to decorrelate variables using a geometrical approach, which worked provided the distributions were Gaussian. These concepts were re-discovered by Galton and Pearson in applications in biology. Schuster (1898) used Fourier analysis to look for periodicities in apparently random sequences. The various concepts were eventually integrated into a sound theory for signal description by Wiener (1930).

Efficient methods for computational analysis of

stochastic processes based on the fast algorithm (FFT) for computing the discrete Fourier transform of a data sequence were originated by Cooley and Tukey in 1965, and have been developed since notably by Gold, Oppenheim, Rabiner and Schafer. Their work fortunately coincided with digital computers becoming widely available, so that the methods could be used easily. Blackman (1958) and Bartlett (1955) have provided methods for estimating the statistical properties of time sequences efficiently and reliably.

Analyses in which decisions are made between two alternative hypotheses from sequences of observations originated from the application of statistics to quality control in industry around 1926. This led to measures such as Operating Characteristic (OC) which enable the decision efficiencies of alternative sampling schemes to be compared quantitatively. The methodology was adapted for analysis of target detection in radar and sonar between 1940 and 1945 leading to the use by Marcum and Swerling of Receiver Operating Characteristic ('ROC') for comparing different schemes for detection, accepting the inevitability that some decisions will be made incorrectly, but minimising the consequences

of these incorrect decisions in the Maximum Likelihood Detector.

D.O.North had in 1938 described the matched filter which maximises the contrast of a message of known form immersed in white noise. His analysis based on Lagrangian multipliers has since been superseded.

In 1943 Wald described (in a memo confidential because of the war) his concept of sequential detection, which had once again been devised to facilitate quality control by sample analysis. This was eventually published (Wald, 1947), and applied to the detection of targets immersed in noise by Bussgang (1955).

In 1970 a special issue of IEEE Proceedings appeared (IEEE 1970) devoted to presenting an overall picture of the contemporary state of the art in signal processing for detecting message signals immersed in noise. Applications in radar, sonar and geophysics were considered, but not as yet in automated visual inspection. Subsequent progress in the fundamentals of the subject has largely involved development of digital methods (forms of time series analysis), to exploit the

explosive growth in digital signal processing.

Although the detection of surface defects is a problem somewhat different from that of detecting radar targets, particularly in that in defect detection the messages signals cannot be 'designed' to have desirable and known characteristics, that the noise in inspection problems is rarely Gaussian, and that the signal obtained from a surface is a vector function of two spatial variables, the key to successful signal processing in automated visual inspection is to recognise the analogy with target detection. Bowers (1970) and Hill (1972) did not recognise this analogy, De Jonge (1973) did recognise it, but failed to carry his analysis very far, partially because he lacked the necessary support and computational facilities.

Published works dealing with signal processing directed specifically towards the detection of surface defects are hard to find. Dainty (1971) has analysed the detection of messages with a laser scanner when the dominant cause of noise is laser speckle. He shows that since in this case the noise is not additive to the message, then adjusting the spot size to maximise

message/noise ratio does not yield optimal performance, and further that uniform illumination within the laser spot gives significantly better performance than the radial Gaussian distribution possessed by existing scanners.

Sawatari (1972) shows that the defect contrast for a coherent system using a central diaphragm to increase contrast (i.e., a system similar to the Axel Johnson system referred to earlier) may be maximised by selection of an appropriate angle of incidence.

Attempts by Bowers and by Hill already referred to to apply statistical concepts to defect detection did not appreciate that defects are (almost invariably) LOCAL phenomena, in which the properties of the surface must be considered within small areas only, otherwise the variations indicating the existence of defects will be swamped. In a report by Norton-Wayne (1972) suggestions were made for improving the performance of surface inspection instruments by using banks of filters to improve defect contrast, but it was assumed that identification rather than detection was the dominant problem.

Obray (1973a,b) investigated for defect detection a number of time series analysis approaches ,including that developed by Box and Jenkins. He concluded that this form of analysis was not superior to the maximum likelihood detector with adaptive mean (the SIRA detector) currently being used. However, his conclusions were based on analysis of a very limited selection of data.

The inability of the Hooghovens system to detect sticker wrench provoked the first attempt to analyse the surface defect detection problem as a two-dimensional (though scalar) process. De Jonge(1973)suggested adding samples of signal before thresholding to increase contrast,but his method was too cumbersome to be applied practically.

The trigger association suggested by investigators at ORSID (Lacroix et al.,1977) was the first signal processing method for performance improvement to be both effective and implemented in hardware. It has been used on-line with impressive results,but was not investigated theoretically by its originators. It was left to Norton-Wayne (1980) to provide a theoretical analysis,to

propose and investigate other methods for trigger association advantageous in being economical in implementation and having less distortion, and to compare the methods using theory and computational simulation.

Finally Munday (1980) has devised ingenious processing whereby the component parts of the scanner output signal due to surface profile and to absorption may be separated. This unfortunately requires that the surface be scanned simultaneously with two beams.

1.5 Cold Rolled Strip Inspection

By 1973 early experience with laser scanners had generated the belief that the detection problem for defects on rough surfaces was effectively solved, and that consequently research on automated visual inspection should concentrate on the more difficult task of identifying defects by type. Work by Hill (1977) on tinplate inspection in particular had shown that there remained shortcomings in the detection aspects of visual surface inspection, notably for large area defects of low contrast. However, only when in 1976 a determined attempt was made to solve the inspection problem for a specific

material (cold rolled steel strip) in a thorough and systematic manner, were the shortcomings of the existing methodology properly appreciated. Cold rolled steel strip has a particularly rough surface giving poor defect contrast, and a high linespeed is required. Other materials such as textiles share this problem. Thus, signal processing methods more complex and powerful than those in common use were seen to be required for detection. Hence the present work, which aims to provide a general methodology.

Although the defect detection problem has been investigated for many different materials in the programme of work which has generated this thesis (e.g. aluminium, tinplate, photolithographic plate, textiles, timber etc.), generally there has been time to consider each of these only superficially. Thus, a concentrated effort has been mounted on one material, which has been investigated in great detail. The results are considered nevertheless to be of general validity and significance. The material selected was cold rolled steel strip, largely because its inspection currently presents a serious problem, in that progress in improving production linespeeds cannot be properly exploited because of the

low speed and erratic performance of existing visual inspection methods involving human operatives.

Steel strip is produced by a rolling process at linespeeds up to 25 metres /second, with material up to 1.5 metres wide. Defects such as rust spots, scratches and scale present on the strip surface may render it unsuitable for some purposes (e.g. the outsides of car bodies) but still usable for others such as oil drum manufacture. Currently, the strip has to be inspected visually to detect the defects. When the highest surface quality is required, the strip has to be uncoiled and re-inspected at low speed (1.5 metres/sec) ,which is costly and time consuming. Owing to boredom, fatigue, inconsistency, and a long training time, visual inspection by human inspectors is generally undesirable. Investigations by the psychologist Moraal (Moraal, 1971, 1974) have quantified the inadequacy of human inspectors for moving strip. Consideration of the physical properties of the human eye-brain system (i.e., that the eye can examine in detail a region only 2 degrees wide, which at a working distance of 1 metre covers 3 centimetres, and has a response time of about 100 millisecond) help explain this poor performance.

To provide instrumentation for automating the inspection, a joint programme of research has been pursued, involving the Instrument Systems Centre at The City University, the Welsh divisional Laboratory of the British Steel Corporation at Port Talbot, and the SIRA Institute. This has developed steadily since 1971; early results of the investigation were disclosed by Norton-Wayne and Hill (1974), and by Norton-Wayne, Hill and Brook (1977). The City University contribution was to provide and evaluate signal processing methods, the SIRA Institute to provide scanners and an eventual hardware implementation, whilst the British Steel Corporation has provided samples and specifications, conducted on-line trials and co-ordinated the project overall (Norton-Wayne, Hill and Watts, 1980). The objective has been to detect 95 percent and ultimately identify 85 percent of defect occurrences, at the maximum linespeed (25 metres per second), without wasting more than 0.2 percent of good material. The first objective (of good detection) has of course to be achieved before significant progress can be made on the second.

Although more than 50 distinct kinds of surface defect occur on cold rolled strip, this investigation has

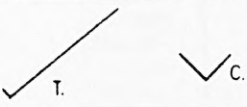
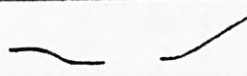
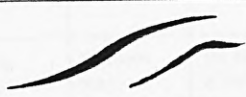
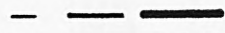
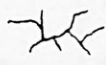
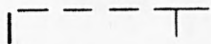
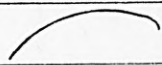
concentrated on 36 which are representative of the full range. BSC have gathered 600 sheets of steel each about 12 inches square (for ease of handling), and containing (nominally) defects of one kind only. To ensure statistical validity, a minimum of 10 samples is considered necessary for each defect type, taken from different lines at different times. For defects exhibiting wide variation in form, 50 samples are preferable. Table 1.1 (adapted from Watts [1977a]) gives the types of defect considered, and ^{table A.1} the number of sample sheets included in the data base for each defect type.

The sheets were then scanned at the SIRA Institute using a laser scanner containing a 3 milliwatt helium-neon laser with a 1 millimeter diameter spot. The beam intercepted the surface 15 degrees from the normal. The scanner had three photomultiplier sensors, each accepting light within a cone 2 degrees wide in the plane of incidence. The sensors viewed at the specular angle, and at 7 degrees and 30 degrees on the remote side of the specular angle. The signals from the three sensors were digitised to 8 bits and recorded on industry standard magnetic tape, together with identification and the output pulses from a moving average (SIRA) detector.

Table 1.1 Defect Types Considered in the Simulation Work

Defect	Rough Physical Description	Origin and Cause	Distribution Thro' Coil			Size Range		Scatter Absorb or Deflect
			Distribution	Location across Width	Frequency along Length	In Rolling Direction	Transverse to Roll Direction	
1. <u>Skin Lamination</u>	Elongated broken line or narrow band with intermittent surface rupture	Steel defect. Exposed Blowholes on slab fail to weld up in rolling or incompletely scarfed cracks in slab	Single or multiple	Anywhere	Random	Any length from a few cm to whole length	Up to about 3 cm. Usually 2-6 mm	A, D
2. <u>Seams</u>	Longitudinal marks in rolling direction varying from series of dark and light fine lines to wider band with peeling surface	Steel. Non-metallic inclusions. Exposed blowholes or cracks in ingot, or mechanically induced at the slab stage	Single or multiple	Anywhere	Random	Any length up to several metres	1-2 cm	
3. <u>Slivers</u>	Elongated sliver of steel which is fused to the base metal and one end of which is normally detached (broken sliver)	Molten metal splashed on ingot mould wall or due to surface discontinuities on the ingot. Steel defect	Single or multiple	Anywhere	Random	Up to 15 cm. Typically about 7 cm	Up to about 25 mm. Typically 10-15 mm	D Broken part
4. <u>Fleck Scale</u>	Fine fleck pattern usually darker but sometimes lighter than background strip	Hot mill. Worn work roll surfaces. Pick-up secondary scale and impress this into strip surface	Multiple spots in large patches 100's or 1000's per sq metre	Anywhere	Random but continuous once started	1-4 mm	1-2 mm	A
5. <u>Jet Scale</u>	Pattern of small dark patches in linear or arrow-head formation	Hot Mill scale formed in reheating furnaces and between roughing stands and not removed by water sprays due to blocked jet or low pressure	Varies from a few patches to dense coverage	Anywhere usually in a band	Random usually continuous	Up to 10 cm. Usually 0.5-3 cm	Up to 2 cm (bands can be any width depending on which jets were faulty)	A
6. <u>Holes</u>	Surface discontinuities passing right through the strip	Any surface or internal defect which causes localised tearing of surface on rolling. Hot mill defect	Usually single isolated or a small group	Anywhere	Random	Normally less than 3 cm (Over about 10 cm Most likely would break strip)	Less than 3 cm. Usually about 1 cm x 1 cm	A
7. <u>Hot Mill and Pickle Line Scratches</u> (Hot Strip Scratches)	Short broken dark lines in the rolling direction, which may take the form of small gouges with 'tails'	Mechanical scoring due to steel fragments, seized rolls, etc. or by scraping of folds in P. L. looping pit	Usually multiple (in localised patches in case of P. L. scratches)	Anywhere	Random	Usually less than about 20 cm	Up to 4 mm	A, D, S
8. <u>Scale Pits</u>	Pattern of short scratches in rolling direction usually in a band several cm wide	Hot Mill defect. Caused by small lumps of scale not washed off between Hot Mill stands	Irregular pattern of multiple spots or short scratches	Anywhere	Random	Few mm	Up to 1 mm	A, D, S

9. <u>Serrated Edge</u>	Break-up of strip edge into 'saw-tooth' pattern	Usually caused by faulty side trimming but can be due to wrong chemical composition or poor rolling practice	Multiple serrations	Edges	Continuous or Intermittent	Up to about 4 mm	Up to about 4 mm	A
10. <u>Steel Pits and Digs</u>	Small clean depressions in sheet surface	Tandem Mill. Rolling in of small particles from strip edges, surfaces and re-circulated coolant	Usually single for pits and multiple in groups for digs	Anywhere but usually near centre	Random but digs normally occur at coil ends	Up to 3 or 4 mm	Up to 3 or 4 mm	D, S
11. <u>Scrap Marks</u>	Relatively large and deep impression on the sheet surface which can take irregular shapes *	Tandem Mill. Rolling in of edge scrap or surface scrap	Usually single or small cluster	Anywhere	Random	Up to about 1 cm	Up to about 1 cm	D, S at boundary
12. <u>Sticker Wrench</u>	Irregular arced lines transverse to rolling direction	Sticking of adjacent laps of coil during annealing which snatch apart when uncoiling	Multiple at irregular intervals of a few cm	Central Part of strip (does not occur at edges)	Continuous on middle part of coil (not usually 1st or last 2-3 tonne)	Up to 4 mm	15-500 mm	S
13. <u>Point Wrench</u>	Localised pattern of lines in shape of concentric semi-circles	Similar to sticker wrench, but more localised	Localised single or small group of 4 or 5	Anywhere	Usually on middle part of coil	Up to a few mm (Take form of concentric semi-circles)	Up to about 10 cm	S
14. <u>Oxidised Edges</u>	Discolouration at edges. Sometimes with large 'scallops' extending further into strip, varying from light brown to dark blue	Annealing defect. Oxidation of surface due to air coming into contact with the metal at an elevated temperature	Continuous Band	Strip edges	Usually affects whole coil	Full length or long intermittent patches	Usually up to about 25 cm	A
15. <u>Sand Pits</u>	Irregular matte spots or indents often in patches. Lighter in colour than sheet before oiling, darker after oiling	Sand entrapped between coil laps during annealing becoming rolled into surface on temper rolling	Multiple spots, widespread over surface	Usually strip edges	Normally affects coil ends	Up to a few mm	Up to a few mm	A
16. <u>Roll *</u> Tandem Mill and Temper Mill	Mark left in light, relief due to defect on work roll surface. Recur at regular intervals depending on roll dia. and subsequent ext.	Tail or corner of strip, detached slivers, scabs or pieces of scrap, bruising work roll which causes imprint on strip	Usually single or small cluster depending on which type	Anywhere	Repeating at fixed interval, which is different for Tandem Mill and Temper Mill	(See details on final page)		

Type	Cause	Description and/or Diagram	Size Range	S, A or D
16A Tail mark and corner mark	Bad tail end of coil passing through mill - sometimes occurring on break caused by bad edges or welds etc.		Tail mark - any length Corner mark - several cm	A
16B Draw-out	Strip squeezing out of rolls at tail end, with rolls hard down - (sometimes nips the T/end)		Up to full width	A
16C Bruise	Proud inclusion in or on strip contacting roll	Very light shiny area	2-4 cm	A, S
16D High Spot	Indentation in roll due to edge scrap	Small spots in relief on strip surface	2-3 mm	D
16E Pinch Mark	Extra thickness of strip due to crimping or doubling on passing through rolls leaves corresponding mark on the roll		Any length usually in 20-40 cm range. 1-2 cm wide	A
16F Drag Mark	Caused by rolls making contact with each other during a roll		5 mm x 20 cm approx.	
16G Fracture Mark & Firecrack	Crazing of roll due to stresses in the roll	 Thin marks proud of strip surface	Up to several cm normally but can be larger	S & D
16H Bite Mark	Usually caused when entering strip into rolls which sometimes snatch the leading edge and drags small bits of metal onto the roll		Any size	
16J Fen Mark	Usually caused by rolled in side scrap from pickle line		Any length	
16K Chain Mark	Line of high spots-equi-distant, around circumference of roll		3 mm spots	S, D
16L Shot Blast Marks	Marked rolls leaving roll shop	Small pits	Less than 1 mm	A, S

17. <u>Chatter Marks</u>	Light and dark bands across strip surfaces at regular intervals, except spiral chatter when bands are longitudinal	Fault on roll shop grinder, loose or worn bearings, or shudder	Regular Bands	Full width except spiral chatter where bands repeat across width	Continuous and repeating (spiral-cont)	4-5 cm Except for work roll chatter which is very narrow	Full width except spiral 2-3 cm	A/D
18. <u>Cold Mill Scratches</u>	Continuous scoring of strip surface along rolling direction	Strip rubbing on sharp corner or hard particles embedded in guides	Usually single	Anywhere	Usually continuous but sometimes with random breaks	Full length or very long	Usually less than a mm	S, D
19. <u>Pick-Up</u>	Shallow (but sometimes deep) depressions in strip surface. Recurring at fixed interval, smoother and brighter than normal surface	Extraneous matter e. g. oil and grease picked up on temper mill rolls which leave impression on strip	Single spot, multiple spots in linear array or mass pattern	Anywhere	Repeating at fixed interval but may affect only part of coil if cause removed	Up to 5 cm usually less than 1 cm	Up to 5 cm usually less than 1 cm	D, (S) - A (Brighter)
20. <u>Carbon P. U.</u>	Irregular pattern in light relief covering large areas of strip surface	Residue remaining on strip surface after annealing, adhering to temper mill work roll, imprints back onto surface	Large irregular band(s) in rolling direction	Usually affects one side of coil but can be full width	Continuous Tends to get progressively heavier	Indeterminate Pattern Bands 10-40 cm		D or S at boundary

21. <u>Coil Digs</u> Temper Mill digs or Tension digs	Small marks in form of bright scratches, pits or holes (in ex- treme case)	Slippage of coil laps against adjacent laps due to failure of man- drel to grip coil or insufficient tandem tension or extreme back tension	Multiple	Anywhere	Limited lengths (several laps of coil) usu- ally at front or back ends*	3-8 mm	About 1 mm	
22. <u>Coil Break</u>	Irregular trans- verse creases on strip surface	Local yielding of material due to insufficient temper rolling. Can be caused by damage to coil before temper rolling	Multiple irregular creases	Usually full width	Random with inter- vals of a few cm to metres	1-2 mm	Up to whole width	D, S
23. <u>Edge Strain</u> (Stretcher) Strain	Lines of reduced cross sectional area transverse to rolling direction	Local under skin passing (insuffi- cient temper rolling) caused by excessive edge gauge taper or by unsuitable roll crowns	Multiple at close intervals usually 5 mm or less	Strip edges	Random may affect whole length	Up to 2 mm	Usually 15-50 mm	
24. <u>Indents</u>	Small depres- sions in strip surface	Small particles e.g. loose scrap or wire adhering to pinch roll leaving impres- sion each revolution	Single or multiple	Anywhere	Repeating at a fixed distance	Less than 1 mm	Less than 1 mm	D, S
25. <u>Rust Spots</u>	Small light or dark patches usually rosette shaped	Hygroscopic con- tamination e.g. acid carry over from pickle lines, moisture in com- pressed air on tandem mill, bacteria in rolling oils	Single or multiple spots usu- ally mult. with separation of several cm	Anywhere	Random	Normally up to 3 mm but can occur in larger patches	Up to 3 mm normally	A
26. <u>Feather Marks</u> <i>wing bone</i>	Parallel shadow marks diagonal to rolling direction (connected by a curved spine when fully developed)	Uneven work on the strip during temper rolling or too much centre in the sheet from the tandem mill	Multiple Parallel lines with separation of about 3-5 cm	Usually away from strip centre	Intermittent	Diagonal streaks with width of up to 1 cm and length up to about 8 cm		A
27. <u>Water Stain</u>	Dark irregular pattern	Water or coolant (soluble oil) trapped between laps during re- coiling which is carbonised on annealing	Large patches	Anywhere but usually at edges	Random	10-30 cm	10-50 cm	A

The resulting data was processed in the CDC7600 computer at the University of London Computer Centre (ULCC). Processing algorithms were implemented as Fortran subroutines and combined to form alternative processing schemes for the overall system, whose performance has then been evaluated quantitatively. Microfilm plots generated on a Calcomp 1670 plotter were used both to validate the data and to evaluate the effectiveness of the signal processing methods. The processes involved in gathering and validating the data base and the program package and overall philosophy used in the processing are described in further detail in appendix 'A'.

The simulation investigation proceeded in the following stages. A program was produced for automatically measuring detection performance. Data from about 150 sheets were processed by this using two standard algorithms, the moving average (SIRA) detector, and the Matched Filter Bank (MFB) detector. It was found that only 6 of the 36 defect types were difficult to detect, and the remainder of the simulation work concentrated on providing methods capable of

detecting these.

1.6 Plan and Structure for this Thesis

The objective of this thesis is to investigate systematically methodologies of signal processing of potential application in automated visual inspection, and particularly in the detection of defects on rough surfaces. A further objective is then to propose an optimal processing strategy considered to have general applicability. The validity of the processing methods is established by applying them in computer simulation to a real inspection problem which is demanding and significant. This is the detection of surface defects on cold rolled steel strip.

The viewpoint taken is that since human vision is capable of detecting (and, if trained, of identifying) all of the defect classes considered provided they are presented slowly enough, then it must in principle be possible to devise signal processing methods having the same capability, and to implement these as programs on a man-made computer. The human brain is simply an alternative form of computer, different from electronic

computers made by human engineers in its method of storing and transmitting information and in its overall organisation, but not necessarily superior in its ultimate capability. As understanding grows in signal processing methodology, and as faster, more compact and reliable, less expensive and less power hungry hardware becomes available, the possibility emerges that the performance of instrumentation systems for automated visual inspection may even exceed that of humans. Intelligent instrumentation will be able to detect defects that are more subtle and of lower contrast than human beings can detect, as well as being faster, more consistent and more reliable.

If all signal processing methods available to the investigators failed to detect a particular kind of visible defect, it was assumed that the fault must lie in the camera, which would consequently have to be improved.

In many cases, more than one signal processing method is capable of adequate detection. That easiest to implement in hardware is preferred, to ensure cost effective instrumentation. However, sufficient information regarding alternative methods must be

provided, to allow the end user to exercise judgement and make a selection. An inexpensive though slightly imperfect system may on commercial grounds be preferable to one which fulfills exactly an exacting specification, particularly when the latter is extremely expensive.

A final objective of the work was to propose a system (i.e. a combination of processing methods plus a camera design) considered to be optimal for inspecting cold rolled steel strip, implemented using the best available hardware. This aspect is covered in chapter 6.

CHAP.2 .THE INSPECTION SIGNAL - ITS ACQUISITION AND

PROPERTIES

2.1 Introduction

The first stage in any system for the automated visual inspection of surfaces is a camera which converts an optical signal (a distribution of light energy) into an electrical signal for subsequent processing electronically. The signal provides information describing the interaction between the surface and the incident illumination. This indicates the presence, extent and nature of defects, and may possibly also yield a statistical description of the roughness properties of the surface. However, the signal is invariably contaminated by noise which tends to mask defect indications. The contrast between message signals arising from defects and the accompanying noise determines the ease with which defects can be detected. The magnitude of the contrast depends critically on the design of the camera. Also, the rate at which the camera can acquire information limits the speed at which inspection can proceed. It is thus vital to use the best camera available.

In this chapter, we first (section 2.2) discuss four alternative electronic cameras (usually called scanners in surface inspection) which are used to obtain, at high speed and resolution, an electronic signal characteristic of the local optical properties of a moving surface. Properties which theory indicates scanner output signals should possess are discussed in chapter 2.3 for the noise component and 2.5 for the message component, whilst 2.4 discusses properties of the noise component measured experimentally from a laser scanner viewing a rough steel surface. In 2.5, a model is postulated to account for the correlation properties measured in this signal, for use later in maximising defect contrast. Section 2.6 discusses the properties of message signals due to defects, and section 2.7 presents conclusions and recommendations regarding scanner selection and design.

The signal (fig. 2.1) from a typical scanner contains information at any instant from a particular small area of surface whose position in the direction perpendicular to that in which the sheet is moving (called the scan direction) is specified by the co-ordinate x , and whose location in the direction of sheet movement is specified by the co-ordinate y . The

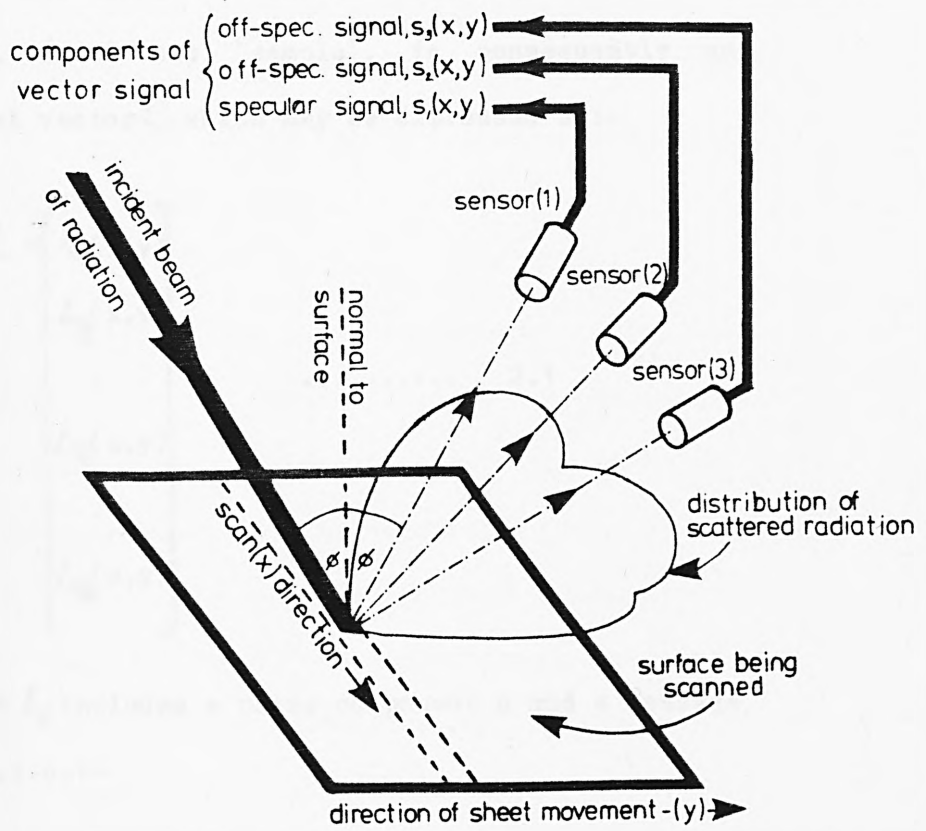


FIG.2:1 SIGNAL DESCRIBING A MOVING SURFACE

scanner will generally have several sensors which extract simultaneously different information from the same small area of surface. Thus, the scanner output signal L , may be described mathematically as an N-dimensional function of the orthogonal spatial variables x and y . The signal describing the area of surface located about the point, (x,y) , termed the 'sample', is consequently an N-dimensional vector L , which may be expressed as:-

$$L = \begin{pmatrix} l_1(x,y) \\ l_2(x,y) \\ . \\ l_i(x,y) \\ . \\ l_N(x,y) \end{pmatrix} \quad \dots\dots\dots 2.1$$

Each element l_i includes a noise component n and a message component m , i.e.:-

$$l_i(x,y) = n_i(x,y) + m_i(x,y) \quad \dots\dots 2.2$$

The noise component (which is an unwanted nuisance) is always present. It is chiefly stochastic, but may be in

part deterministic. The message component $m(i)$ (which carries useful information indicating the presence of defects) is present only occasionally. Though really stochastic, it is for convenience in analysis considered to be deterministic, though its form is rarely known explicitly. The noise components are in general mutually correlated; correlation may exist both between different components of the same vector, and between corresponding components of different vectors.

The properties of the noise component n and the message component m are both dependent on the scanner used for signal acquisition, in addition to the properties of surface and defect. It is desirable to use the scanner type which maximises the contrast between defect message and noise. Other considerations are also important in selecting a scanner, such as the rate at which data is acquired, cost and reliability.

2.2 Scanners

Four types of scanner have been used for defect detection on moving strip material. These were generally designed empirically, without reference to theory either

Table 2.2 Relationships between signal components

(a) covariance Matrix for sheet C12.24:

$$\begin{pmatrix} 9.68 & 0.49 & -1.72 \\ 0.49 & 3.23 & 0.11 \\ -1.72 & 0.11 & 8.33 \end{pmatrix}$$

(b) correlation Matrix for sheet C12.24:

$$\begin{pmatrix} 1 & 0.09 & -0.19 \\ 0.09 & 1 & 0.02 \\ -0.19 & 0.02 & 1 \end{pmatrix}$$

(c) covariance Matrix for sheet C99.14:

$$\begin{pmatrix} 24.2 & 2.11 & -0.54 \\ 2.11 & 15.31 & 0.19 \\ -0.54 & 0.19 & 1.58 \end{pmatrix}$$

(d) correlation Matrix for sheet C99.14:

$$\begin{pmatrix} 1 & 0.11 & -0.09 \\ 0.11 & 1 & 0.04 \\ -0.09 & 0.04 & 1 \end{pmatrix}$$

for minimising noise, or for maximising defect contrast. Thus, even the best existing designs are not necessarily optimal.

The FLYING FIELD scanner, (fig.2.2), (Brook, 1971a, 1971b), was the first device to be used for the automated visual inspection of moving sheet material, originally paper. In this camera, the surface is illuminated uniformly over a narrow strip perpendicular to its direction of movement using diffuse wideband visible radiation, obtained for example from a fluorescent tube driven by D.C or high frequency A.C. to eliminate flicker. A rotating drum with lenses mounted at equal intervals about its periphery is used to scan a slit window (elongated in the direction of sheet movement) across the strip. A system of lenses and mirrors images the strip onto a photomultiplier tube.

The great bulk of the drum limits its speed of rotation to about 2000 RPM. For a drum containing 8 lenses, the maximum scan rate is thus 250 scans/sec. The illumination is used wastefully. Measurements by Hill (1972) have shown that the output self-noise is white and Gaussian. It is generated presumably by shot effects in

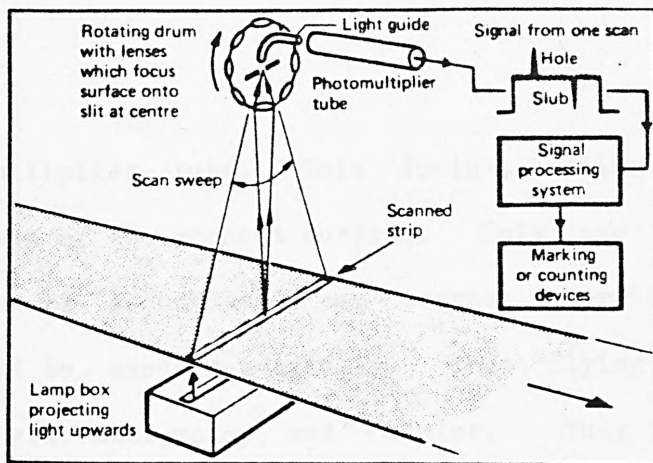


FIG. 2.2 FLYING FIELD SCANNER

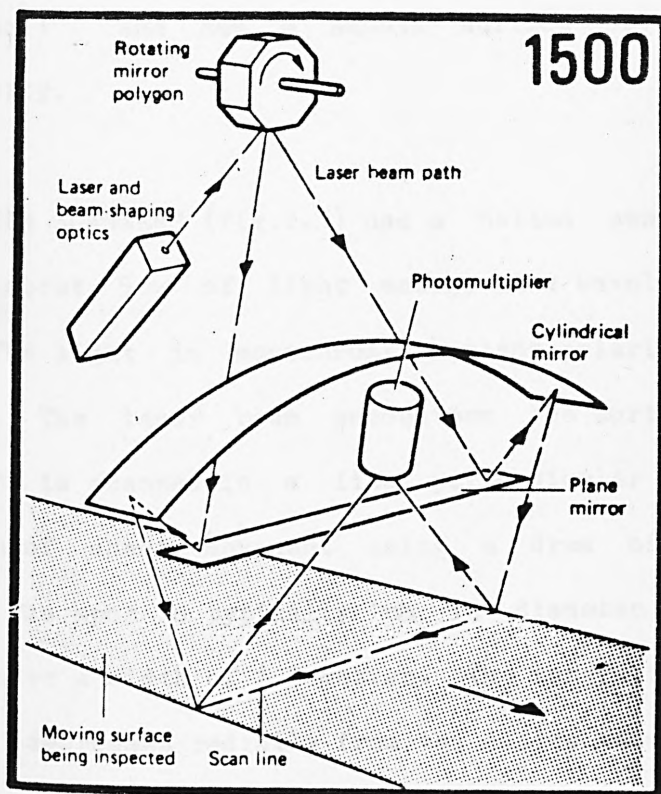


FIG. 2.3 FLYING SPOT (LASER) SCANNER

the photomultiplier tube. This dominates noise due to the roughness of the scanned surface. Only one sensor channel can be accommodated per scanner, otherwise the lenses would be used non-axially. Thus flying field scanners are slow, noisy and scalar. They are consequently obsolete, despite having been used successfully by Brook et al., (1977), to inspect tinplate which is manufactured on a low speed line (2.5 metres/sec.) and has a smooth surface of high reflectivity.

LASER SCANNERS (fig.2.3) use a helium neon laser emitting about 5mw of light energy at a wavelength of $6328\text{\AA}$. The light is monochromatic, plane-polarised and coherent. The laser beam encounters the surface as a spot, which is scanned in a line perpendicular to the direction of sheet movement using a drum of mirror facets. The spot is typically 1mm in diameter and is scanned over a width of 1.5 metres. The intensity within the spot decreases radially from the centre according to a Gaussian distribution. The mirror drum in the fastest designs contains typically 12 facets and can be driven at up to 24,000 rpm using an air bearing motor. Thus, 4 microsecond scans are possible. (Barker and Brook, 1978)

The light returned from the surface at a selected angle is directed by a system of lenses and mirrors to a photomultiplier tube. This has a sensitive photocathode which may be 125mm. in diameter, and is (at room temperature) the most sensitive photo-electronic transducer currently available. Devices having cut-off frequencies up to 30 Mhz. are readily obtainable. Several photomultipliers accepting light at different angles may be incorporated in one scanner, to yield a vector output signal (fig.2.1). In fact, scanners are available sensing over a two-dimensional array using fibre optics .

Thus, laser scanners are fast, sensitive and can provide a vector signal. They have the highest resolution of any kind of scanner, with upwards of 20,000 points being resolvable in a device manufactured by the Bofors company (Antonsson et al., 1979). They have thus been used in almost all the experimental work to be described. However, laser scanners are cumbersome (mechanical scanning and high voltage vacuum tube devices are used), are troubled by speckle noise, and are expensive.

Scanners using line arrays of photodiodes as sensors are much simpler. A scanner using arrays of discrete photodiodes was developed by DeJonge (1979), described in chapter 1 section 3, but is now obsolete. Currently sensors using lines of photosites diffused into a single chip of silicon and separated by as little as 13 microns are available. The electronic signal is read out by transferring charges accumulated by the photosites in parallel to a CCD delay line, which is then clocked so that the charges are transferred serially to a readout point. The sensor chip is mounted in the focal plane of a high quality macro lens, i.e., a lens designed to give best image quality when the object is comparatively (e.g. 50 cm.) close. Illumination is provided by arrays of tungsten filament lamps with diffusers to achieve uniformity, or by fluorescent tubes. The distribution of emitted energy with wavelength for tungsten lamps is well matched to the response characteristics of silicon sensors, but filament lamps are unstable and inefficient, and have too much of their output in the infra-red where the imaging characteristics of CCD arrays are inferior due to cross-talk. A good discussion of self-scanned arrays of photodiode sensors has been provided in a book by Beynon and Lamb (1979).

Photodiode array cameras are simple (no moving parts, high voltages or high vacuua), and are relatively inexpensive. Although one camera generates a scalar signal, several may be used to view the same portion of strip simultaneously from different angles, and synchronised to obtain a vector signal. A 2000 element width may be scanned in 100 microseconds with the best devices currently available commercially. The width covered by the scan (and correspondingly the size of the samples) is set by the magnification of the lens system.

The chief problem with photodiode arrays lies in non-uniformity of response. Individual elements may vary in sensitivity by up to 10 percent, and there may be a few (less than 0.5%) elements having a substantial leakage current in the absence of incident illumination (noise spikes). The effect of these is to introduce false excursions into the signal which can be mistaken for defects. Further, there may be crosstalk between the elements which reduces the effective resolution.

Norton-Wayne (1980) has provided a mathematical model which may be inverted to yield a correction filter. Provided the chip is operated at a stable

temperature, excellent correction is possible at least for non-uniformities. Cameras having built-in correction for fixed pattern noise have just (Anon, 1981) become available commercially.

TV cameras differ from the scanners previously described in imaging at any instant a two-dimensional region of surface rather than merely a line. The cheapest TV cameras use as sensors vidicon tubes, in which an image is formed on a photoconductive surface which is scanned in a raster pattern by an electron beam. The vidicon camera is cheap but extremely slow (the photoconductive material has a response time of the order of milliseconds), is non-linear, has poor sensitivity and a low resolution (about 500 points in each direction). Dimensional accuracy is generally poor, being typically 2 percent for inexpensive devices. However, other, more expensive tubes are available which give better performance. Van Daele (1977) provides a good critical survey. Nevertheless, TV cameras have been used successfully in surface inspection, by freezing the motion of the surface using rotating drums of mirrors, electronic shutters or synchronised stroboscopic lights. When in addition the vidicon tube is replaced by the Isochon

which is faster and more sensitive by three orders of magnitude (having been developed for low-light level viewing and not depending on photoconductivity), then automated defect inspection using TV cameras becomes viable (Lacroix et al., 1979). However, several cameras would have to be used in synchronism to obtain a vector signal, and an area (rather than merely a line) of surface would have to be illuminated, uniformly and intensely. The cost advantage of vidicon cameras tends to disappear when their output must be processed by a digital computer, since a sampling interface must be provided which can interpolate over several frames. This matches the data acquisition rate of the camera to that of the computer interface, which is generally an order of magnitude lower.

The properties of the various scanners are compared briefly in table 2.1 below (the drum scanner has been omitted since it is obsolete):-

TABLE 2.1 COMPARISON OF SCANNER TYPES.

	Vidicon TV Camera.	CCD-Line Array Camera	Laser Scanner
Resolution in X-Direction (Points)	400	2048	20,000
Shortest Scan Time (Microsec)	40	200	200
Upper Cut-Off Frequency (MHZ)	4.5	5	30
Self Noise (dB WRT.Saturation Signal)	-45	-50	>-60
Geometrical Linearity	1%	>0.05%	prop. to scan semi-angle

Electrical Linearity	Needs Gamma Correction	Linear	Linear
Cost	£450	£1,500	c.£20,000
Cost of Interface to Digital Processor	£3,500	£500	c. £2,000
Life	Life of Vidicon Tube	inf.	Laser lasts c.1000 hours
Compatability with Moving Surface	Requires flash or Shutter	OK	OK

This table must be interpreted carefully, since the signal/noise ratio is for all types dominated by self-noise in the input transducer (e.g. the photomultiplier), if the input light level is not great enough. Otherwise, noise generated by the roughness of the surface dominates. Thus, ALL scanners may be operated so as to make self-noise negligible, though obtaining illumination which is sufficiently intense may be unacceptably expensive.

The performance of all scanners is limited ultimately by shot noise which becomes significant when the level of sensed light is so small that individual quanta may be distinguished. We now investigate this limit. However, only scanners having photomultipliers as sensors (and probably also CCD arrays but definitely not vidicon TV cameras) have self noise sufficiently low for the analysis to be relevant. Flying spot scanners have an advantage here in that the whole of the energy of illumination is concentrated on the region being viewed at any instant.

In devices such as photomultipliers and photodiodes which perceive individual quanta with efficiency η , the

output current i is given by:-

$$i = \eta e \{W(s) + W(b)\} / h\nu \quad \dots 2.3$$

Here, e is the charge on the electron, $W(s)$ the signal power, $W(b)$ the background noise power, h is Planck's constant, and ν is the frequency of the incident light. As ν decreases, the number of quanta available from a given light energy decreases, so sensitivity decreases accordingly. The efficiency η , the number of carrier pairs per photon, is about 0.5 for typical semiconductor devices. For sensors having electron multipliers, η is greater than one.

For line scanners, a strip of width say 1.5 metres long by 1 cm. wide must be illuminated. Assuming a luminous efficiency for the source of 1%, an electrical input of 15 Kw will be required to achieve a surface illumination of 1 mw per square millimetre. Barker and Brook state (1978) that the spot from the SIRA laser scanner has an average luminous density of 3 mw per square millimeter.

Suppose an element of surface of arbitrary area is illuminated with power W milliwatts, and that some fraction f of this reaches the sensor, due e.g. to losses in lenses and mirrors, and to off-specular sensing. Measurements by the VDEH Institute in Duesseldorf (Gries and Hoerster, 1977) shown in figure 2.18. show that 10^{-2} is a reasonable value for f . In high-speed operation, energy from a particular element of surface will reach the sensor during a time interval T seconds, which may be of the order of 10^{-7} . (NB - for laser scanners the illumination is convolved with the surface whereas for CCD arrays discrete sampling is involved. The effect of this distinction in sensitivity analysis is trivial, and will be ignored.) The energy E reaching the sensor within the interval T is thus given by:-

$$E = W \cdot T \cdot f \quad \dots\dots\dots 2.4$$

If W is 1 milliwatt, then E will be 10^{-12} watt with the values suggested.

To determining the number of quanta required to supply E , we invoke Einstein's equation, e.g.:-

$$E=h\nu \quad \dots\dots\dots 2.5$$

which expresses E (the energy per quantum) in terms of the frequency of the radiation and of Planck's constant h. The latter is 6.66×10^{-34} joules.second.

For the visible light used, $\lambda = 5 \times 10^{-7}$ metres, so that using the relation $c = \nu\lambda$ which relates the velocity c of a wave to its wavelength and frequency, we obtain:-

$$\nu = 6 \times 10^{14} \text{ Hz.} \quad \dots\dots\dots 2.6$$

The energy per quantum in Joules is thus given by:-

$$E = 6 \times 10^{14} \times 6.62 \times 10^{-34} = 3.97 \times 10^{-19} \\ \approx 4 \times 10^{-19} \quad \dots 2.7$$

and the mean number N of quanta received per element of surface is then given by:-

$$N = 10^{-12}/4 \times 10^{-19} = 2.5 \times 10^7 \quad \dots 2.8$$

The number of quanta arriving per unit interval is a statistical quantity and obeys a Poisson distribution, for which the variance is equal to the mean. The RMS variation in N, which forms the noise, will thus be:-

$$N^{**}(1/2) = 5 \times 10^{**3} \quad \dots\dots\dots 2.9$$

and the ratio of signal mean to noise rms variation is then:-

$$(2.5 \times 10^{**7}) / (5 \times 10^{**3}) = 5 \times 10^{**3} \quad \dots\dots 2.10$$

i.e., 74 dB. Since a defect causing the mean to vary by 3 dB must be considered of good contrast, the noise due to distinguishable quanta is thus negligibly small compared with that due to surface roughness, and CCD scanners should be efficient as sensors at the highest linespeeds. This result contradicts that of Barker and Brook (1978), who concluded that CCD arrays would be unsuitable for inspecting steel strip at the higher linespeeds because adequate illumination would be too expensive.

Even taking into account the 1.5 order of magnitude decrease in mean signal level for observation 30 degrees away from specular, and inefficiency in sensing quanta, the signal/noise ratio falls to only 58dB.

When noise due to surface roughness dominates, any scanner using incoherent illumination should be about twice as sensitive as an equivalent using coherent illumination, because the latter contains noise due to speckle. (see section 2.3)

2.3 Theory of Generation of Noise due to Surface Roughness

In this section, we use theory describing the interaction of beams of radiation with rough surfaces to predict the statistical properties of the scanner output noise. These comprise its probability density function, correlation properties and relative amplitude, C . This last quantity is the ratio of the mean squared variation in scanner output signal due to the noise to the squared mean level of the signal.

$$C = \langle I^2 \rangle / \langle I \rangle^2 \quad \dots\dots\dots 2.11$$

Here, I is the instantaneous amplitude of the scanner output signal, and the expression $\langle . \rangle$ denotes 'mean'. Variation in I in the absence of a defect constitutes the noise. The contrast of a defect message relative to the noise implied by the variance of I is evidently inversely proportional to C .

When the pdf of I is a one-parameter function (see appendix 'B'), like the negative exponential, C depends only on the form of the pdf. For pdfs such as the Gaussian which are described by more than one parameter, a particular form for the pdf gives rise to a range of values of C , according to the relative values of the various parameters.

Two distinct kinds of illumination are commonly used in surface inspection applications, coherent monochromatic, plane-polarised beams obtained from lasers, and non-coherent, randomly-polarised, broadband radiation, obtained from tungsten filament lamps or fluorescent tubes. The latter are commonly also uncollimated. Somewhat surprisingly (Barker et al., 1976), the latter illumination generally gives a lower

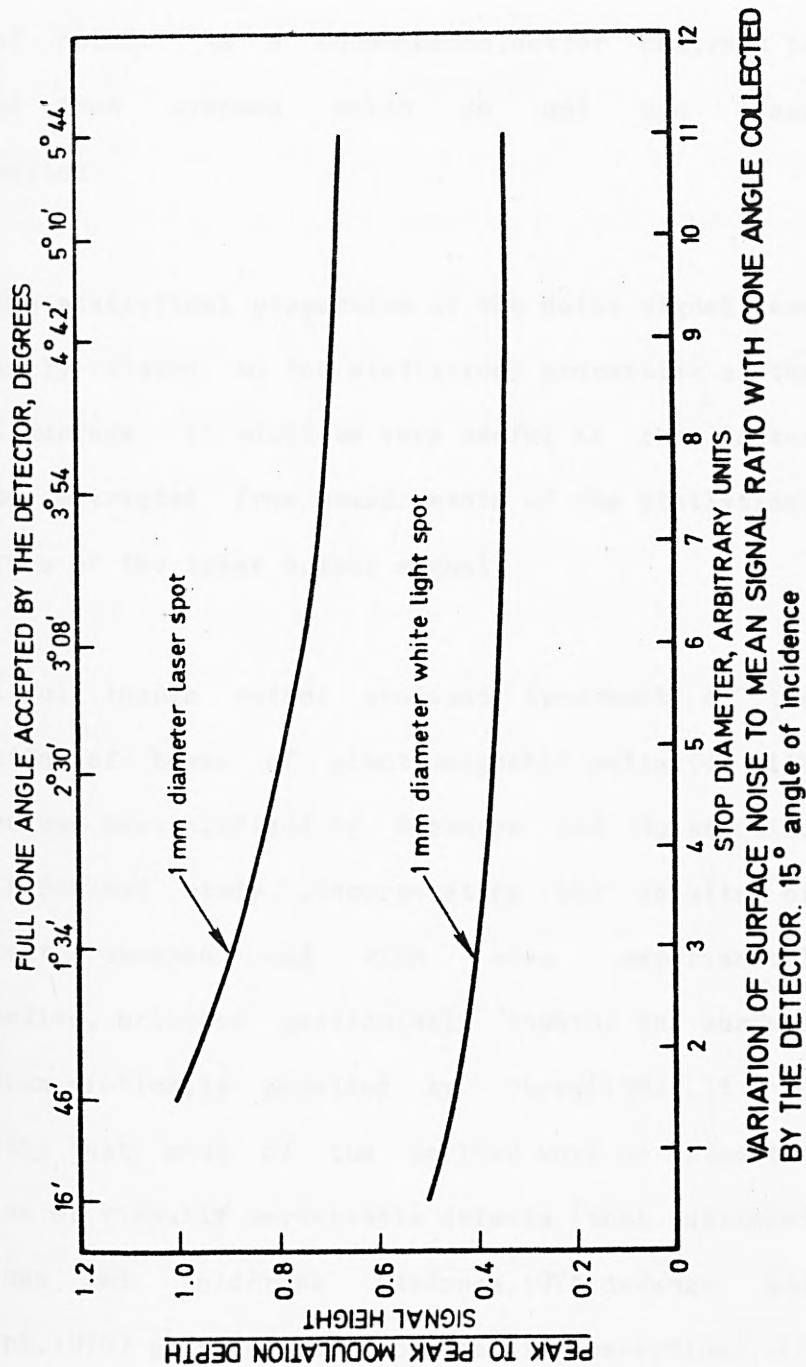


FIGURE 2.4

level of noise. As a consequence, better contrast is obtained from systems which do not use laser illumination.

The statistical properties of the noise signal are evidently related to the statistical properties of the scanned surface. It would be very useful if the latter could be extracted from measurements of the statistical properties of the laser output signal.

A full though rather profound treatment of the interaction of beams of electromagnetic radiation with surfaces has been provided by Beckmann and Spizzichino (1963). A revised study, incorporating the results of subsequent research and with some experimental confirmation, oriented particularly towards the surface inspection problem, is provided by O'Bray (1982). It is noteworthy that most of the earlier work on automated detection of visually perceivable defects (that published by deJonge and Aalderink (deJonge, 1973, deJonge and Aalderink, 1979) provides the outstanding exception), did not use theoretical analysis to optimise the design of either system or scanner, although the necessary theory was available.

Beckmann and Spizzichino, and O Bray, postulate that two alternative theories may be used to account for the interaction of radiation and surfaces. A rigorous approach involving use of a Kirchhoff's integral to solve Maxwell's equations, with the surface profile supplying boundary conditions, is always valid. A much simpler approach which assumes the surface to comprise a two-dimensional array of mirror facets is valid provided the structure of the surface is large compared with the wavelength of the radiation so that a scalar (ray) theory may be used. The only restrictions on the Kirchhoff vector theory are that the surface contain no edges whose radius is small compared with the wavelength of the radiation, ('sharp'), and that there is no shadowing.

If the radiation is coherent and the region instantaneously illuminated is large compared with the dimension of the surface structure, then additional noise is produced due to the presence of speckle. This was first explained by Von Laue in 1919. The modern classical theory for speckle has been well summarised by Dainty and his collaborators (1975). The speckle arises from the summation of a large number of wavelets of equal amplitude, whose phase is distributed uniformly in the

interval $(-\pi, +\pi)$. Although the expectation of the sum of the amplitudes is zero, the expectation of the quantity actually sensed (i.e. the intensity, which is amplitude squared), is rather surprisingly not zero, but varies rapidly as the viewpoint is altered. Thus, as a spot is scanned over the surface, the first order probability density for the intensity of the returned signal is a negative exponential (app'B'):-

$$p(I) = \frac{1}{\langle I \rangle} \exp(-I/\langle I \rangle) \quad \dots I > 0 \quad \dots 2.12$$

$$= 0$$

$$\dots \text{otherwise}$$

On substituting the values for $\langle I^2 \rangle$ and $\langle I \rangle^2$ implied by equation 2.12 into 2.11, the value unity is obtained for the speckle contrast. $\langle I \rangle$ is the expected level of the signal (its mean). Equation 2.12 holds provided that the average departure of the surface profile from its mean value is large compared with the wavelength of the light (i.e., that the surface is rough), and that the size of the spot is large compared with the correlation length of the surface profile. These properties ensure that the distribution of phase for the combining wavelets is

uniform. If the surface height variation H is not larger than the wavelength of the illumination the noise amplitude decreases monotonically and almost linearly with H . Chandley and Escammilla (1979) showed that the effect on I of decreasing spot size is highly complex, and provided an experimental determination.

It is usually assumed (nominally on account of the central limit theorem, though more practically to facilitate analysis) that the surface profile is Gaussian. In practice this may not hold since the form of a surface is influenced by the process which created it. For example, stylus measurements on isotropic ground surfaces by Milana and Rasello (1981) suggest that a composite distribution formed as a superposition of Gaussians is more accurate. It seems that sharp negative excursions tend to occur, the corresponding excursions in the positive direction having been removed by the grinding operation, leading to an asymmetric distribution. They recommend the Gaussian function nevertheless as an adequate first approximation. Computer simulation by Fujii, Uozumi and Akamura (1976), later confirmed by experiment by Fujii and Asakura (1977), shows that in fact I really is remarkably

independent of the surface profile, provided this is not so regular as to constitute a grating.

If coherent light is present at two or more wavelengths, the speckle patterns for the individual wavelengths combine to produce a new pattern having reduced contrast. Fujii and Lit showed (1978) that its magnitude I then remains dependent on H for much larger H .

As already mentioned, the probability density function of the intensity of the speckle is inherently negative exponential but in practical scanners, the viewing aperture is not small compared with the size of the speckle and the variations are consequently smoothed somewhat. One important consequence is a decrease in the speckle contrast C , which shows clearly in curves measured by the SIRA Institute (Barker et al., 1976), shown in figure 2.4. Another is that the shape of the pdf is modified. An approximate solution provided by Goodman (1975) shows the pdf of the speckle then to be close to a gamma distribution, i.e.:-

$$p(I) = \frac{M^{(b+1)}}{\Gamma(b+1)} \exp(-mI) \quad \dots 2.13$$

in which the parameters M and b are selected to obtain a best match to experimental observation. Recently, Stansberg (1981) has provided an exact solution for an arbitrary aperture, which shows incidentally that the gamma distribution is a good approximation for very large and for very small apertures. His equation giving the exact form for the pdf is too complex to be worth reproducing here, but predicts a distribution which is closer to negative exponential than to gamma, i.e., is more skewed. His theoretical prediction was supported by experimental measurements.

The various distributions may be compared using figure 2.5 (adapted from Stansberg, 1981). This shows negative exponential, gamma, one-dimensional exact, two-dimensional exact and a gaussian pdf having the same variance as the negative exponential example, plotted on the same axes.

Movement of the spot over the surface also modifies the form of the pdf (McKechnie, 1975) ; the general effect is to make it incline towards the Gaussian.

The polarisation of the beam is also degraded by interaction with a metal surface. The general effect was shown by Hariharan (1977) to reduce the contrast in the speckle pattern. Recent measurements by Obray (1982) have shown the intensity of light returned from steel strip to vary by 16% when the plane of polarisation was rotated through 180 degrees; incidence was normal but the surface structure was not isotropic.

The size of the speckle at the point of observation is also of interest, since the viewing aperture smooths out speckle variations if it is large compared with speckle size. It may be determined from the autocorrelation of the speckle ; Goodman (1975) shows that for a uniform square spot of dimension L by L, illuminated with light of wavelength W, the average width of the speckle Δl is given by:-

$$\Delta l = \frac{W \cdot z}{L} \quad \dots 2.14$$

where z is the separation between the surface and the speckle, and L is the dimension of the spot. Ennos(1975)

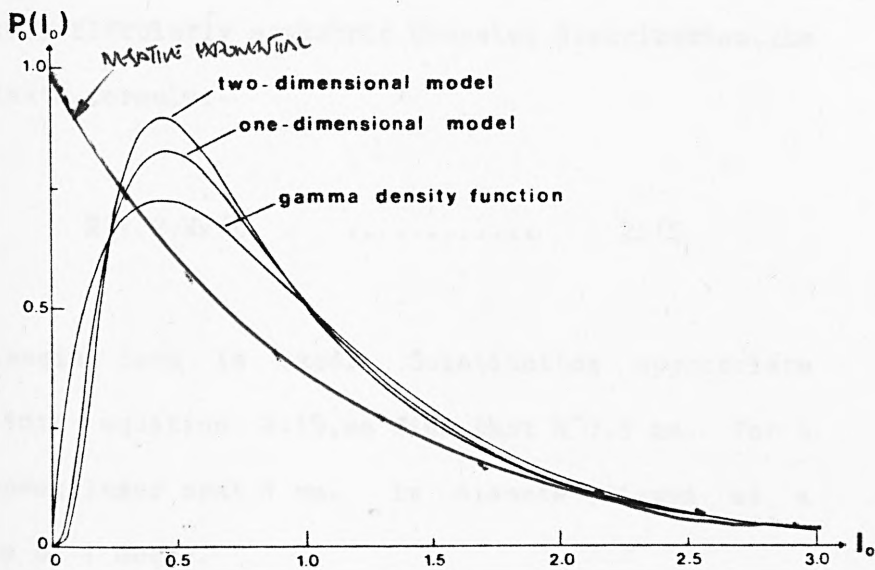


FIGURE 2.5 PROBABILITY DENSITY FUNCTIONS FOR SPECKLE
COMPARED

suggests for the distribution normally present in a laser spot, i.e. a circularly symmetric Gaussian distribution, the approximate formula:-

$$R \sim 1.2 W_z / L \quad \dots\dots\dots 2.15$$

if no imaging lens is used. Substituting appropriate values into equation 2.15, we find that $R \sim 0.5$ mm. for a helium-neon laser spot 1 mm. in diameter, viewed at a distance of 1 metre.

As the angle of incidence (θ) moves away from the normal, we would expect V to fall since the effective height H of the surface structure will decrease due to geometry, provided that $H \cos(\theta)$ is not large cf. W . But, even for a surface so rough that the latter does not hold, V will still decrease because the mean amplitude decreases. This behavior has been confirmed by experimental measurements by Barker et al., (1976), shown in figure 2.6

The correlation properties of speckle noise have been investigated by Iwai et al., (1981). They show both theoretically and experimentally that the autocorrelation

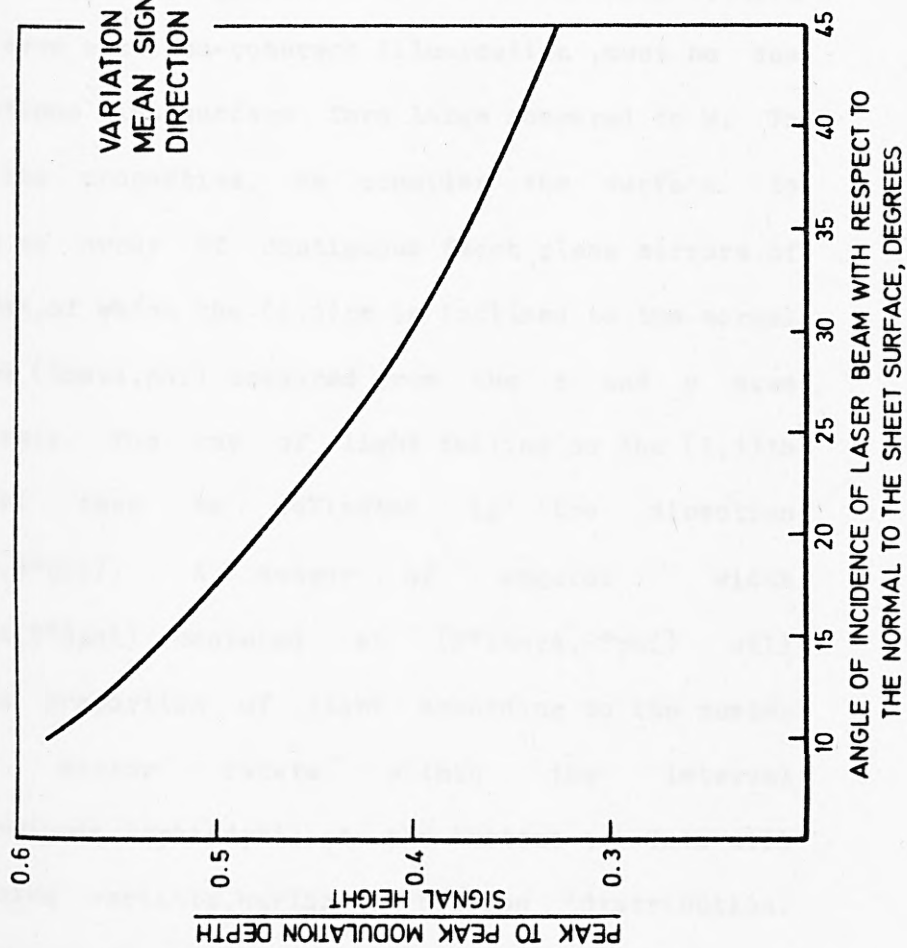


FIGURE 2.6

of the speckle falls to close to zero, as the lag approaches one millimeter, for a scanning system with parameters similar to those of the laser scanners used in this work.

Speckle is certainly not the only cause of noise in the output signal from a laser scanner. The physical mechanism generating this other contribution, which appears even with non-coherent illumination, must be due to variations in surface form large compared to λ . To analyse its properties, we consider the surface to comprise an array of contiguous facet plane mirrors of equal area, of which the (i,j) th is inclined to the normal at angles (θ, ϕ) measured from the x and y axes respectively. The ray of light falling on the (i,j) th facet will then be deflected in the direction $(2\theta, 2\phi)$. A sensor of angular width $(2d\theta, 2d\phi)$ centered at $(2\theta, 2\phi)$ will receive a proportion of light according to the number $N(t)$ of mirror facets within the interval $(2\theta + d\theta, 2\phi + d\phi)$ at the instant t . This will be a random variable, having a Poisson distribution. However, the intensity I will be proportional to $N(t)^2$, and will hence obey the more complex (unnamed)

discrete distribution derived by Obray (1982):-

$$p(I) = \frac{\mu^{NI} e^{-\mu}}{(NI)!} \quad \dots 2.16$$

$$I = 0, 1, 2, 3, \dots$$

Here, μ is the mean value of the intensity. When N is so large that the distribution is effectively continuous, then $p(I)$ will converge to the Gamma distribution:-

$$p(I) = \frac{\mu^{(b+1)} \exp(-\mu I)}{\Gamma(b+1)} \quad \dots 2.17$$

which has mean and variance both μ .

This theory predicts for the returned light a variance which is equal to its mean, in contrast to the speckle component for which the variance (and hence the observed noise signal) is proportional to the square of the mean. As the point of observation moves away from the specular, the mean will decrease and the variance (representing the 'apparent' surface roughness) will decrease more than proportionally. This accords with experimental measurements (fig.2.6).

The pdf of the noise signal from a uniform, flat rough surface for a scanner using coherent illumination is thus the resultant of two distributions which lie somewhere between the Gaussian and negative exponential. Since the component distributions are uncorrelated, one expects from the central limit theorem that their resultant will be even more closely Gaussian.

When white noise of zero mean having any distribution is passed through a linear filter, its pdf at the output will approach the Gaussian form. In its transversal interpretation, the filter is effecting a linear summation of noise sources which are independent, thus the central limit theorem will apply. Experimental measurements by Hill (Norton-Wayne and Hill, 1979) for a band pass filter applied to noise signals obtained during the simulation work show the noise pdf at the filter output to be very close to Gaussian for several sizes of filter, triangular in the time domain. (figure 2.7). Thus, use of a Gaussian form in the analysis is sound for filtered noise.

The statistical properties of the noise signal are evidently related to the statistical properties of the

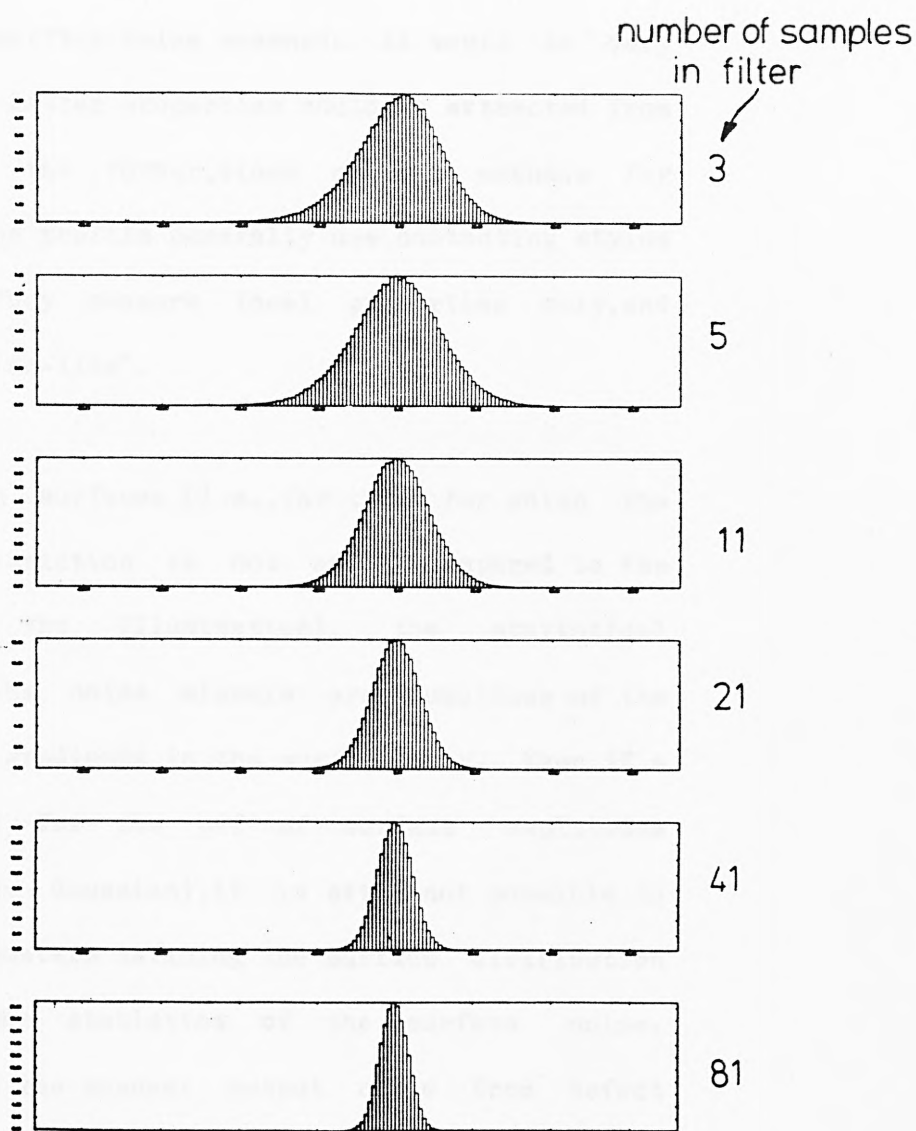


FIG. 2.7 PDF ESTIMATES FOR FILTERED NOISE, SHOWING GAUSSIAN TENDENCY

(sheet C16-04, specular view)

profile of the surface being scanned. It would be very useful if the latter properties could be extracted from measurements of the former, since current methods for measuring surface profile generally use contacting stylus instruments. They measure local properties only, and cannot be used 'on-line'.

For 'rough' surfaces (i.e., for those for which the surface light variation is not small compared to the wavelength of the illumination), the statistical properties of the noise signals are functions of the distribution of gradients in the surface only. Even if a form is assumed for the pdf of surface amplitudes (e.g., that it is Gaussian), it is still not possible to recover the parameters defining the surface distribution solely from the statistics of the surface noise. Thus, analysis of the scanner output noise from defect free surface cannot for any form of scanner yet considered be used to provide absolute parameters defining the distribution of surface profile.

For example, a recent paper by Milana and Rasello (1981) has used Beckmann and Spizzichino's vector approach to derive an equation describing the variance of

the surface profile sigma in terms of the variance $\langle I^{**2} \rangle$ of the output of a scanned laser beam:-

$$C = T \lambda / 4 \sqrt{\pi} L \cdot \sigma \quad \dots \quad 2.18$$

here, L is the extent of the variation in surface height. The parameter T is the correlation distance for the surface (another unknown), thus the method does not provide an absolute measurement, merely a comparative one. Experimental verification showed that with 0.6328 micron light, vertical illumination and specular viewing, the relationship is valid for surface variances from 0.6 to 1.8 microns, confirming the empirical predictions of Fujii and his collaborators mentioned earlier.

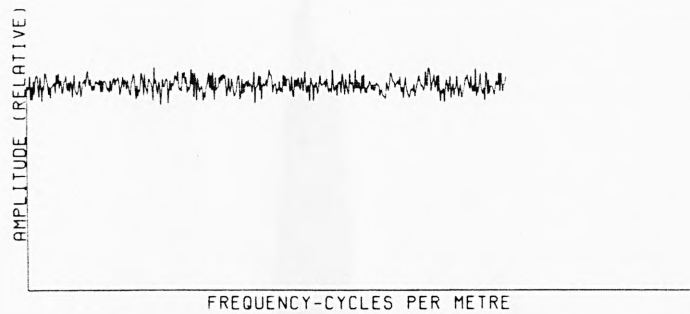
2.4 Measured Properties of Noise

In this section we describe the properties of the noise component of the signal considered from another point of view, i.e. as actually measured using a laser scanner examining cold rolled steel strip. The statistical concepts used in describing and analysing the measurements are explained in appendix B. The properties

of the laser scanner used are described in more detail by Barker and Brook (1978). As a consequence of these measurements a model is postulated to account for the generation of the noise. This model is exploited later (chap.4) to obtain improved defect contrast and hence better detection performance.

Direct measurement of the power spectral density in many individual scans using the Fast Fourier Transform shows that the noise is white. A typical spectrum is shown in figure 2.8 . Estimation of the probability density function for amplitude using histograms gives for flat sheet having a defect free surface a unimodal distribution. This is skewed towards small amplitudes as would be formed by smoothing the negative exponential distribution as predicted in section 2.3 (fig.2.9). No significant difference is evident between distributions measured for different sensors. For bent sheet a wide range of distributions occur ;in extreme cases these may even be bimodal,as in figure 2.10. Dividing the sheet into regions and computing a separate histogram for each (fig.2.11) shows a diversity too great to be caused by sampling error (app.'B'),confirming that the perceived properties of a surface may vary substantially even

FOURIER SPECTRUM
COLD ROLLED STEEL STRIP



SHEET:- C17.02

DEFECT TYPE:- CHATTER

PROCESSING:-

DETECTOR:-SPECULAR

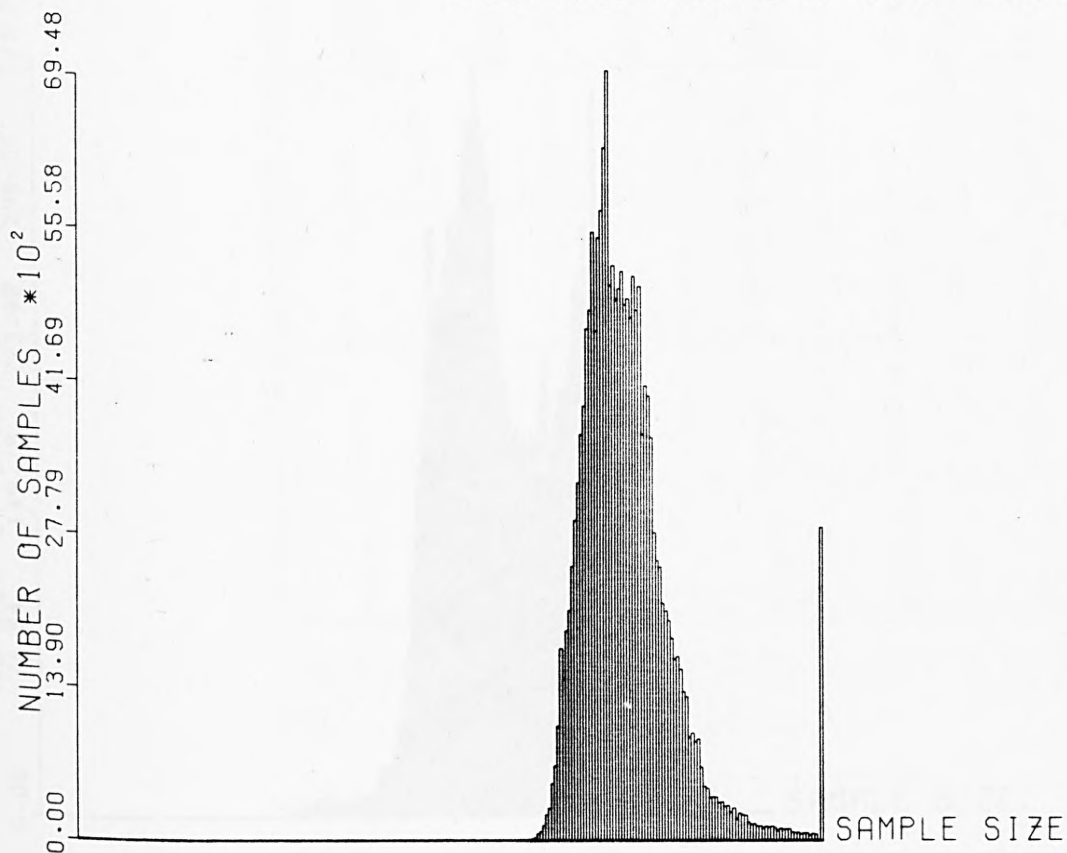
FIGURE 2.8 SPECTRA FOR INDIVIDUAL SCANS, AVERAGED OVER A SHEET

FIGURE 2.9 PDF FOR GOOD SHEET

HISTOGRAM OF SAMPLE LEVELS

MODAL LEVEL IS:-183

MODE CELL CONTAINS 6948 SAMPLES.



C21.26

DEFECT TYPE:- ROLL MARKS

PROCESSING:- NONE

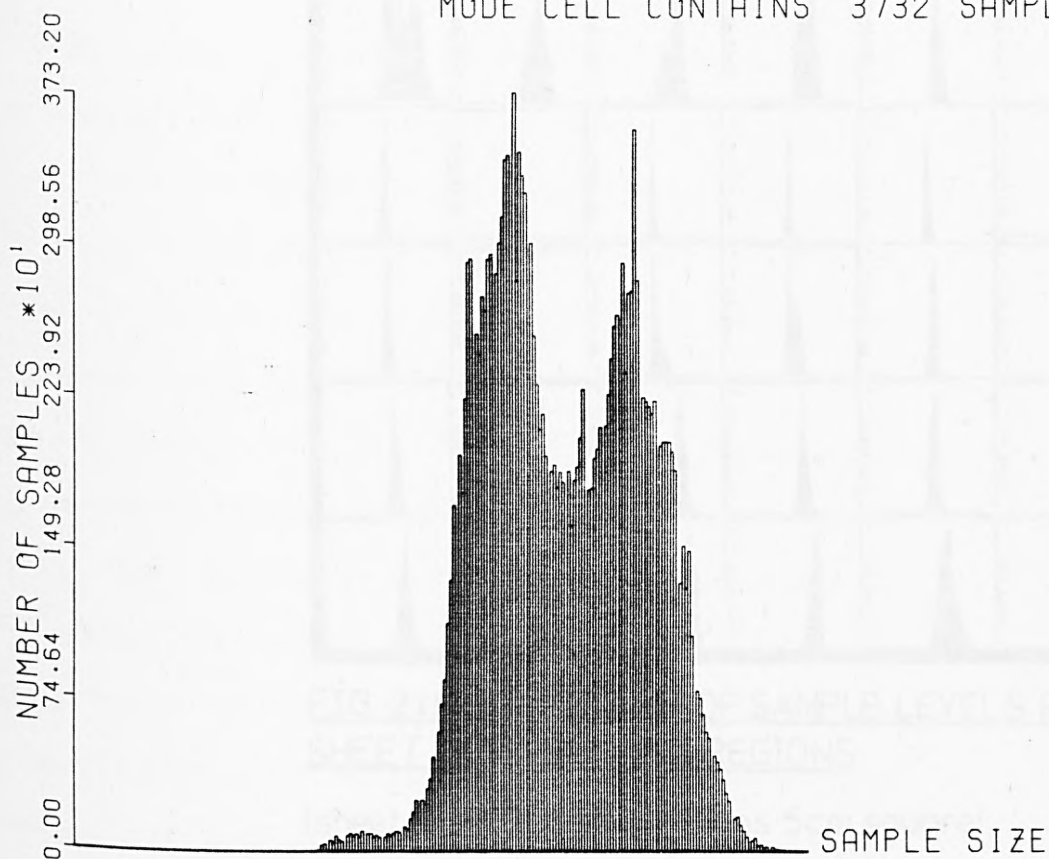
DETECTOR:- 30 DEG OFF-SPEC

FIG. 2.10 Bi-Modal p.d.f. from Bent Sheet

HISTOGRAM OF SAMPLE LEVELS

MODAL LEVEL IS:-154

MODE CELL CONTAINS 3732 SAMPLES.



C16.08

DEFECT TYPE:- ROLL MARKS

PROCESSING:- NONE

DETECTOR:- SPECULAR

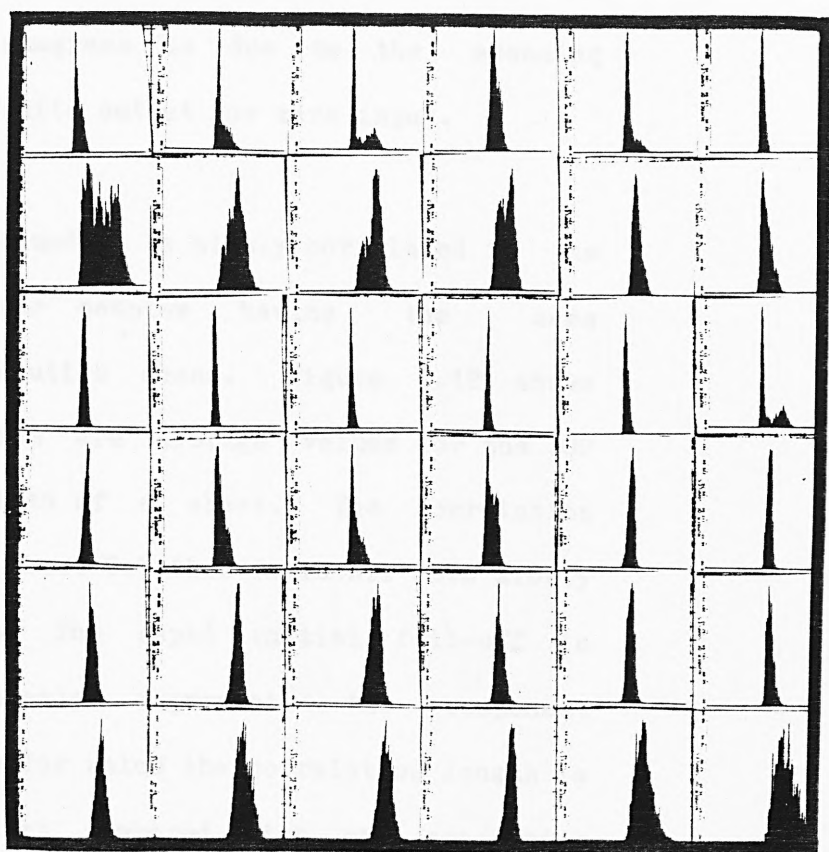


FIG. 2-11 HISTOGRAMS OF SAMPLE LEVELS FOR
SHEET DIVIDED INTO REGIONS

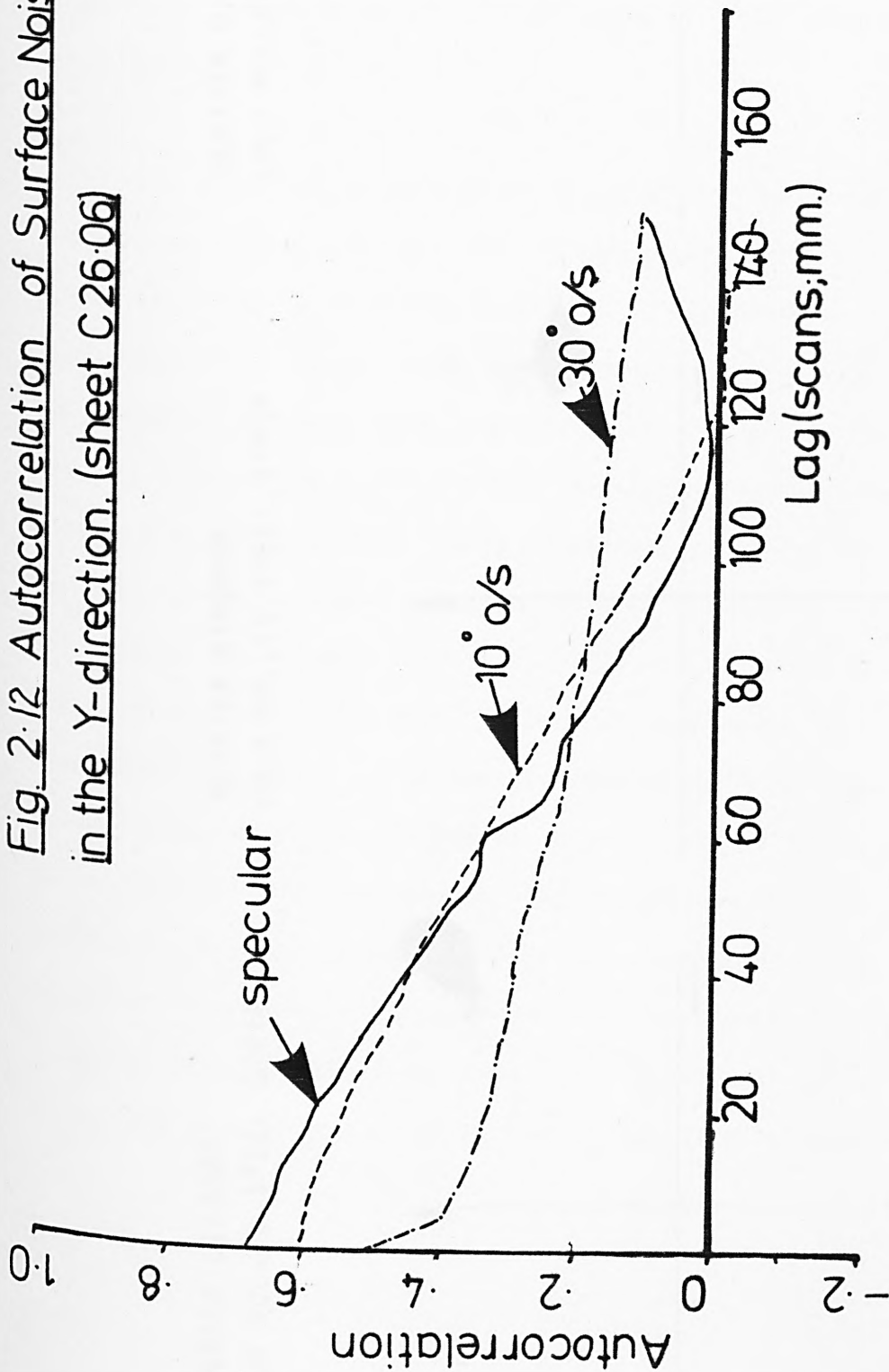
(sheet 16-10, 70/s view, regions 5cm.square)

within a 12inch square if the material is bent. The offset on the histograms is due to the scanning system, which has a finite output for zero input.

The signal is found to be highly correlated in the y-direction, i.e., for samples having the same x-coordinate in consecutive scans. Figure 2.12 shows some measurements, which are average values for the 700 samples across the width of a sheet. The correlation drops immediately to about 0.7, then falls off more slowly and may even rise. The rapid initial fall-off in autocorrelation presumably represents the component arising from speckle, for which the correlation length is (Iwai et al., 1981) short compared with the separation between scans of 1 millimeter. The slow fall-off which is superimposed then represents the component due to gross surface roughness, for which the mirror facet model applies.

Scatter diagrams showing the relationships between the pairs of components (1,2), (1,3) and (2,3) are plotted in figure 2.13., for sheet C12.24 whose processing is discussed in considerable detail in chapter 4. Each point represents one sample of signal. The D.C. term

Fig. 2.12 Autocorrelation of Surface Noise
in the Y-direction, (sheet C26.06)

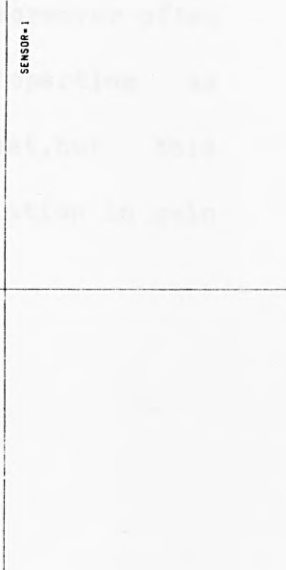


SCATTER DIAGRAM
COLD ROLLED STEEL STRIP

SENSOR-2



SENSOR-1



SHEET :- C12.24

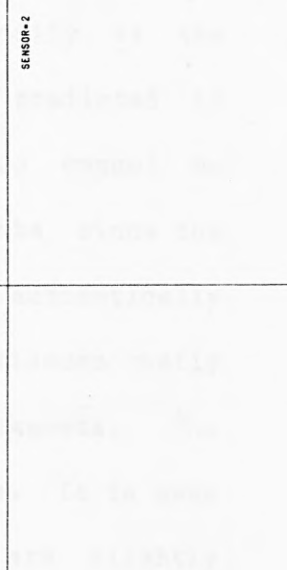
DEFECT TYPE :- STICKERWRE

SCATTER DIAGRAM
COLD ROLLED STEEL STRIP

SENSOR-2



SENSOR-1



SHEET :- C12.24

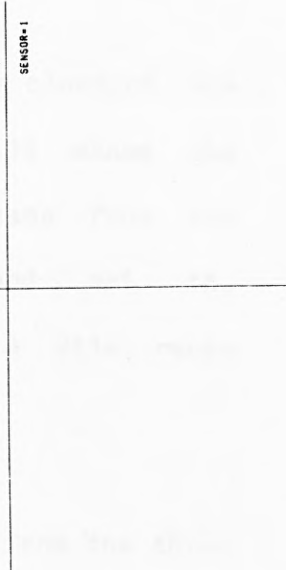
DEFECT TYPE :- STICKERWRE

SCATTER DIAGRAM
COLD ROLLED STEEL STRIP

SENSOR-2



SENSOR-1



SHEET :- C12.24

DEFECT TYPE :- STICKERWRE

FIGURE 2.13 SCATTER DIAGRAMS

(d.c. level not removed)

has incidentally not been removed, hence the clusters are not centred on the origin. Figure 2.14 shows the correlations between successive pairs of scans from the same sensor, i.e., between the n th and $n+1$ th. Again, D.C. levels have not been removed. A wide range of variation is evident.

The correlations between the signals from the three sensors are indicated by the covariance (app.'B') matrices in table 2.2. The variance of the noise as a fraction of mean signal level decreases steadily as the viewpoint moves away from specular, as predicted in section 2.3, but the size of this variation cannot be obtained accurately from these measurements since the gain in the three signal channels adjusts automatically to changes in signal level. Thus, the variances really drop off more severely than table 2.2 suggests. The quantitative picture is of use nevertheless. It is seen that the specular and 7 degree signals are slightly correlated, but that the correlation between these and the 30 degree channels is close to zero, and is moreover often slightly negative. The correlation properties as measured vary widely from sheet to sheet, but this variation is undoubtedly due in part to variation in gain

Table 2.2 Relationships between signal components

(a) covariance Matrix for sheet C12.24:

$$\begin{pmatrix} 9.68 & 0.49 & -1.72 \\ 0.49 & 3.23 & 0.11 \\ -1.72 & 0.11 & 8.33 \end{pmatrix}$$

(b) correlation Matrix for sheet C12.24:

$$\begin{pmatrix} 1 & 0.09 & -0.19 \\ 0.09 & 1 & 0.02 \\ -0.19 & 0.02 & 1 \end{pmatrix}$$

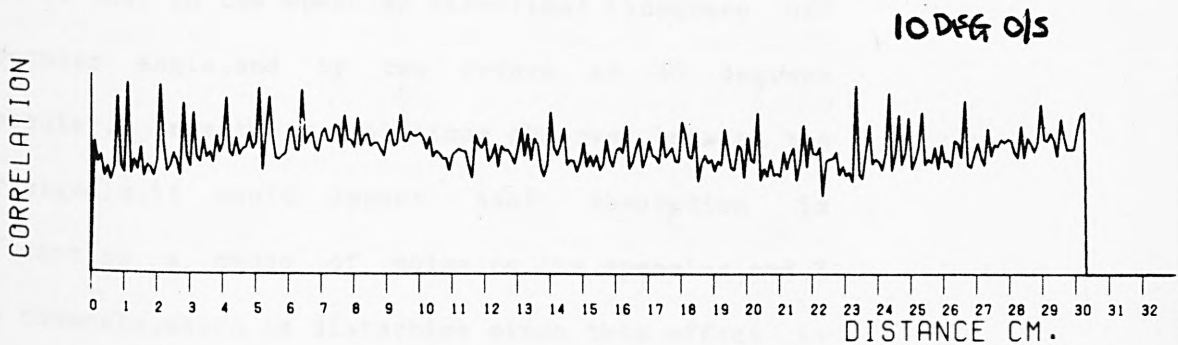
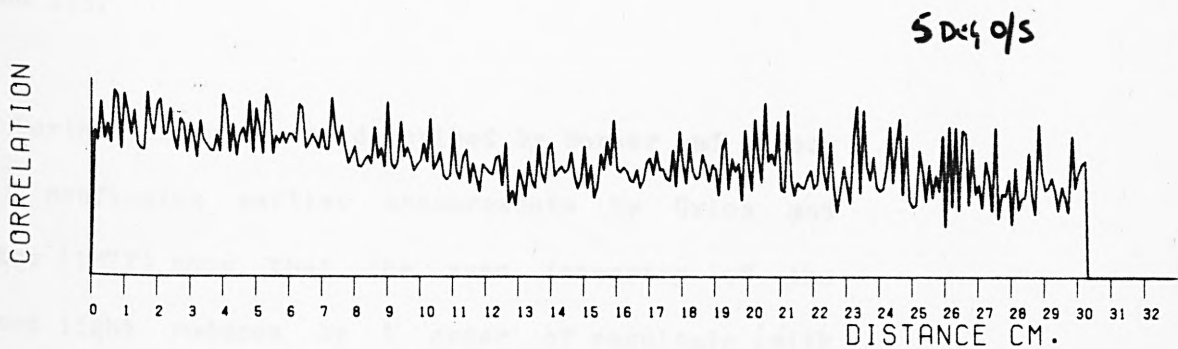
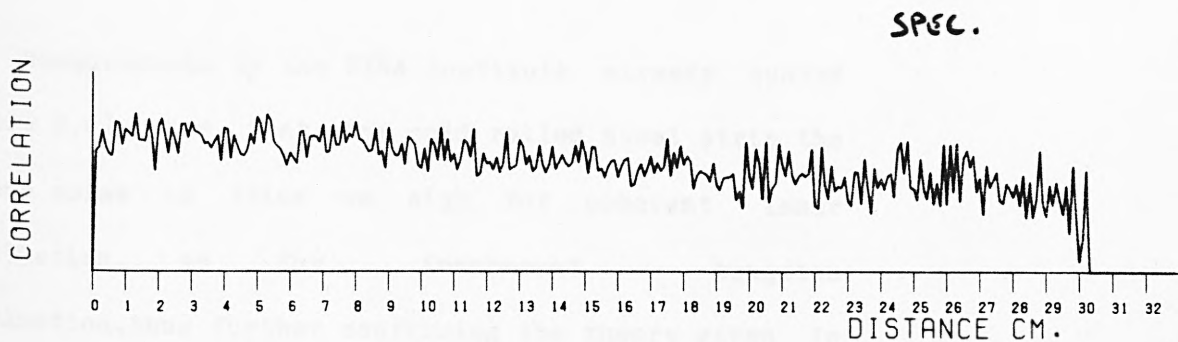
(c) covariance Matrix for sheet C99.14:

$$\begin{pmatrix} 24.2 & 2.11 & -0.54 \\ 2.11 & 15.31 & 0.19 \\ -0.54 & 0.19 & 1.58 \end{pmatrix}$$

(d) correlation Matrix for sheet C99.14:

$$\begin{pmatrix} 1 & 0.11 & -0.09 \\ 0.11 & 1 & 0.04 \\ -0.09 & 0.04 & 1 \end{pmatrix}$$

FIGURE 2.14 CORRELATIONS BETWEEN SUCCESSIVE PAIRS OF SCANS



SHEET:- C12.24

DEFECT TYPE:- STICKERWRE

and offset in the sensor channels.

Measurements by the SIRA Institute already quoted (figure 2.4) show that for cold rolled steel strip the surface noise is twice as high for coherent laser illumination, as for incoherent tungsten illumination, thus further confirming the theory given in section 2.3.

Further measurements described by Barker and Brook (1978) confirming earlier measurements by Gries and Hoerster (1977) show that the mean intensity of the returned light reduces by 1 order of magnitude (with respect to that in the specular direction) 15degrees off the specular angle, and by two orders at 40 degrees off-specular. From the correlations observed between the sensor signals, it would appear that absorption is significant as a cause of noise on the specular and 7 degree channels, which is disturbing since this effect is ignored in all models considered for noise generation. The slight negative correlation observed with the 30 degree channel is reassuringly consistent with the mirror facet scattering model; energy which has disappeared at position 'a' must reappear at some other position, say

'b'.

2.5 Model for Noise Generation

The measured correlation properties of the noise signal may be used to generate a model which helps account for the physical origin of the noise, and which is of potential use in devising processing methods for improving defect contrast.

The correlations observed between successive scans on the same channel may be explained in terms of the way in which the surface profile was generated. This is assumed to occur as follows. At each pass through the rollers, a random variation in height (due to the shot blasted surface of the roller) is impressed onto the surface of the sheet. Subsequent passes elongate this in the rolling (y) direction, but not in the perpendicular (x) direction. Imperfections in the original billet (e.g. air or slag) are also elongated. If severe enough, they appear in the final material as visually perceivable line defects such as seams or laminations. Sometimes defects of this kind not visible to the naked eye are enhanced to visibility during processing for

contrast improvement (section 4.5). Jones (1977) has suggested an alternative explanation, that the longitudinal correlation is due to inadequate shot blasting of the ground surface of the rollers. However, the success of experiments to increase the contrast of certain defects using longitudinal matched filters to be described in chapter 4 section 5 tend to support the first conclusion.

This model explains the observation that the x-direction noise (i.e. that within a scan) is white, but that noise for given sample positions along a scan (y-direction noise) is highly correlated between scans. Direct measurements by Obray (1982) using a stylus instrument add support to this interpretation. A typical surface microprofile thus measured is shown in figure 2.15.

To put the model in precise terms such that it may be used to improve defect contrast, we first assume the signals to be Gaussian. With this assumption, the noises observed may be regarded as a weighted linear combination of uncorrelated white Gaussian noise sources having specified variances, as described by the signal flow graph

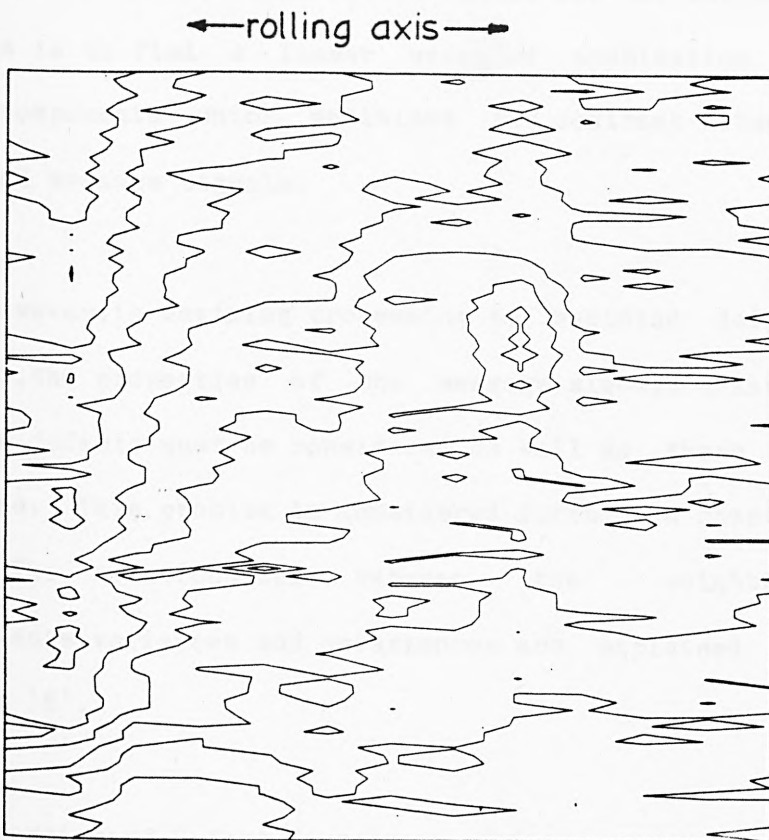


FIG. 2.15 SURFACE MICROPROFILE FOR COLD ROLLED STEEL STRIP

region shown is 3mm. square, sampled on a 25 micron grid. contour interval is 0.1 micron.

in figure 2.16(b). Figure 2.16(a) shows the inverse process, of analysis, which yields values for the weights. The idea is to find a linear weighted combination of signal components which maximises the contrast between noise and message signals.

However, in devising processing to maximise defect contrast, the properties of the message signals arising from the defects must be considered as well as those of the noise. This problem is considered further in chapter 4.4. The relationships between the weighting coefficients, variances and covariances are explained in appendix 'B'.

2.6 Properties of Defect Signals

Defects on a surface may absorb, scatter or deflect the incident light. In practice, most kinds of defect do all of these to some extent, though generally one effect predominates for each type of defect. There have been several systematic studies of the types of defect found on cold rolled strip and their properties (e.g. (British Steel Corp., 1975), (Gruenhofer et al., 1967) and (Watts, 1977a)). Table 1.1 from (Watts, 1977a) is

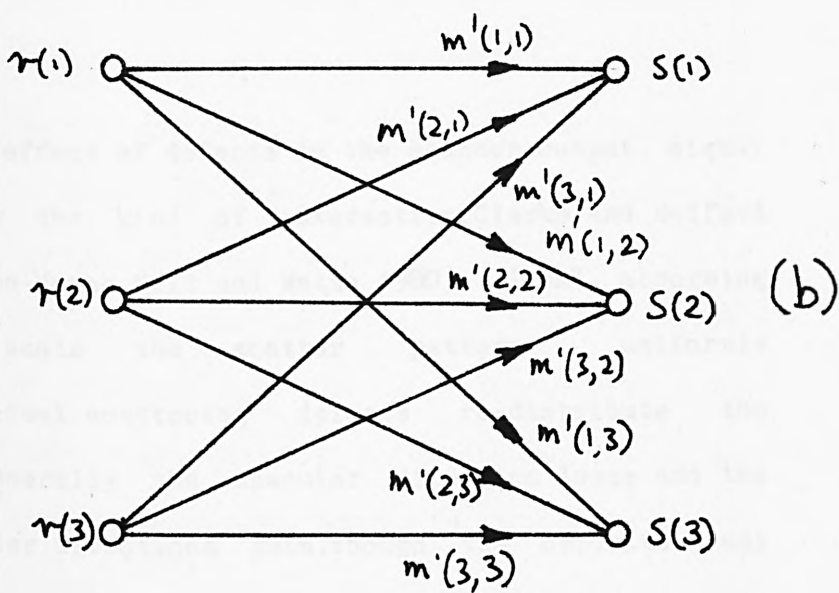
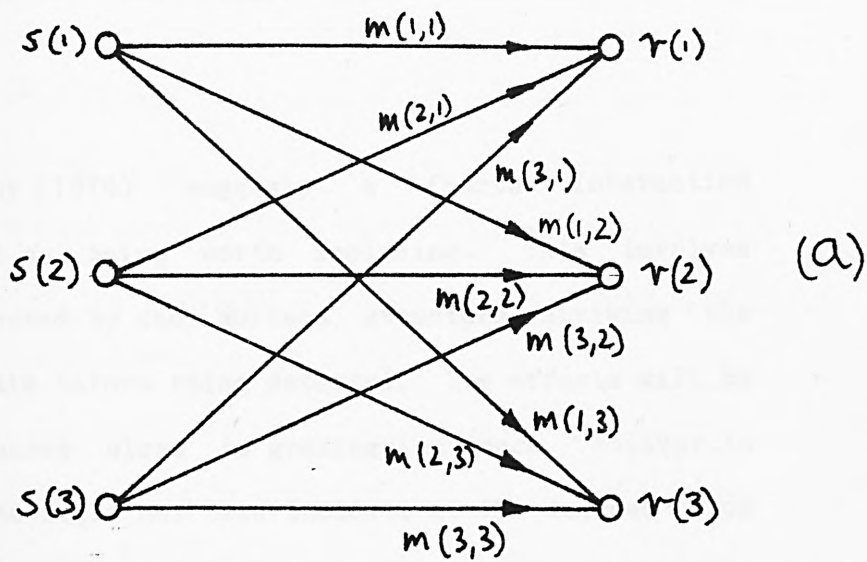


FIGURE 2.16 (a) SYNTHESIS AND (b) ANALYSIS OF CORRELATED
RANDOM NOISE SIGNALS

considered most authoritative and describes the interactions for the defect types considered in this work.

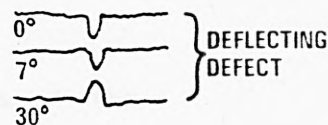
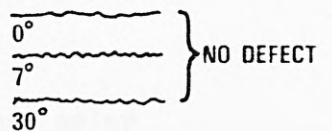
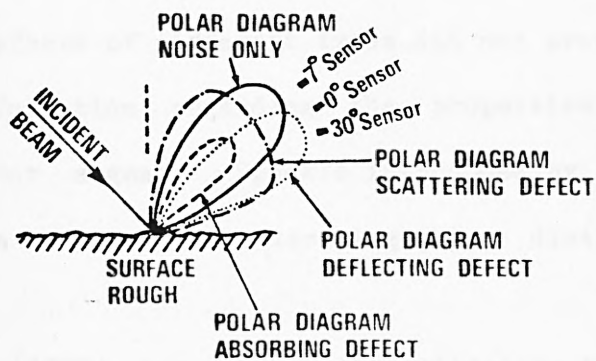
Munday (1979) suggests a fourth interaction (shadowing) as being worth including. This involves light deflected by the surface structure striking the surface again before being detected. Its effects will be most pronounced close to grazing incidence. However, in our work the light has been incident at 15 degrees from the normal so shadowing should not be significant. If present, its effect on the vector output signal would be to reduce off-specular signals more than specular signals, creating an impression of locally decreased roughness.

The effect of defects on the scanner output signal depends on the kind of interaction (Clarke and Bedford 1976, Norton-Wayne, Hill and Watts 1980). Ideal absorbing defects scale the scatter pattern uniformly downward, ideal scattering defects re-distribute the energy (generally the specular direction loses and the off-specular directions gain, though the opposite may occur), and ideal deflecting defects rotate the lobe

pattern bodily. Figure 2.17 illustrates these interactions for a vector scanner having one sensor at the specular angle, and an off-specular sensor at each side of the specular. The effect on the output signals is shown. With sensors on both sides of specular it is often possible to distinguish deflecting defects from defects which scatter, but this distinction is not guaranteed. The scanner used in the experimental work reported had both its off-specular sensors on the same side of the specular sensor, thus some potentially useful information was thrown away.

Measurements at the SIRA Institute (Barker et al., 1976), figs. 2.4 and 2.6, show the ratio R between the peak of the noise modulation and the mean of the signal to decrease (a) as the angle of incidence increases, and (b) as the cone of acceptance (i.e., the numerical aperture, f) of the detector increases. The parameter R is in fact proportional to the quantity $C = \langle I^2 \rangle / \langle I \rangle^2$ introduced in section 2.3 for quantifying noise. The contrast of defect messages will (other parameters of the scanner being unchanged) be inversely proportional to R , but the measurements give by themselves no further information regarding defect contrast. Figure 2.4 shows

FIG. 2-17 MESSAGE SIGNALS IN
TERMS OF OPTICAL INTERACTION



incidentally very clearly the superiority of incoherent compared with coherent illumination.

A study by Jones (1977) of the effect on the polar distribution of returned radiation measured for a large number of defects of different types did not provide much definite information regarding the properties of the message vector signal. Speckle introduced by the laser illumination obscured the underlying polar distribution.

Obray (1982) has shown theoretically that the disturbance on the lobe pattern due to the presence of the defect is maximum, and is moreover most sensitive to defect shape in three dimensions, when the defect and the wavelength of the illuminating radiation are comparable in dimension. In this circumstance (Beckmann and Spizzichino, 1963) the lobe patterns resulting from pillbox shaped protrusions and hemispherical caps of equal size on a conducting surface (for example) are quite different. On cold rolled strip, defects are rarely less than 5 micron in scale, thus the 0.5 micron radiation from the Helium-Neon laser is much too short. The 10.6 micron radiation from a Carbon Dioxide laser would almost certainly yield improved sensitivity. It would also

improve contrast by reducing noise due to speckle, because the surface would no longer be 'rough' compared with the wavelength of the illumination.

As we have already remarked, two different mechanisms seem to operate simultaneously, in determining the lobe pattern of returned radiation. That due to structure comparable with or smaller than the wavelength of the illumination provides in the far field a distribution of radiation which is the Fourier transform of the structure of the surface. Thus, scratches in the x-direction in the strip result in bright lines in the y-direction in the far field. This property has been exploited to increase defect contrast (Sjolin, 1972). The component due to large structural irregularities yields speckle if the incident radiation is coherent.

Gries and Hoerster (1977) quote the results of an attempt to measure defect contrast (as a scalar) on cold rolled steel strip. They show (fig. 2.13) a contrast of one order of magnitude for a 'bright' defect, and a contrast of about 2 for a 'rough' defect. So little information is provided regarding the severity, extent, and type of the defects considered that their results are

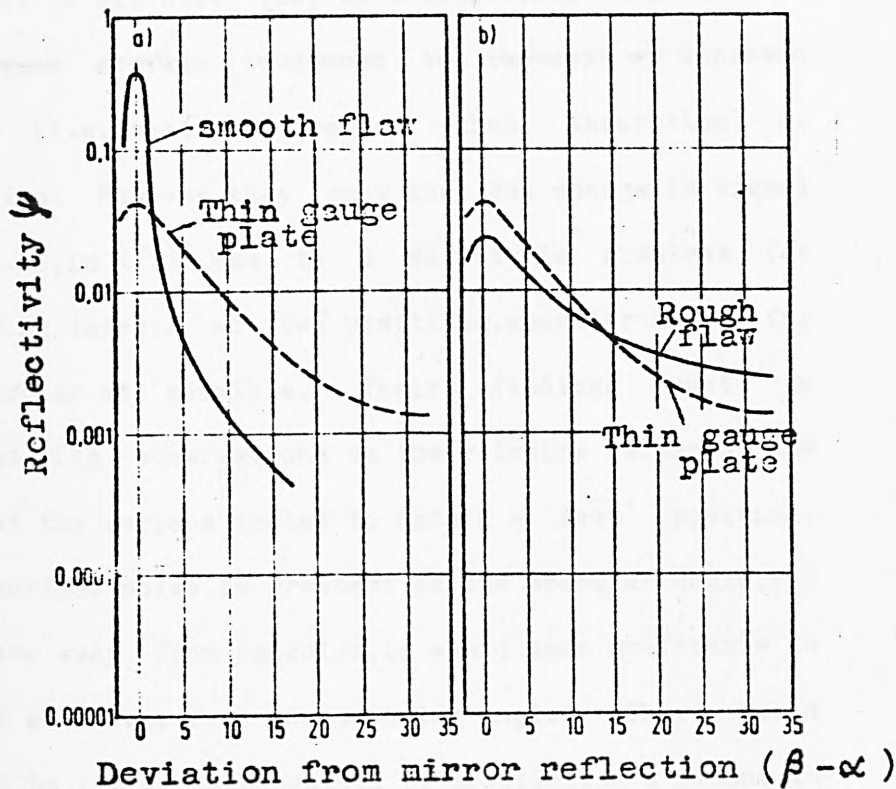


FIGURE 2.18 REFLECTED INTENSITY VS. ANGLE FROM SPECULAR
(ROUGH AND SMOOTH FLAWS)

difficult to exploit. They have considered only defects which cause surface roughness to increase or decrease locally (i.e. scattering), rather than absorption or deflection. However, they show that the change in signal mean (i.e., in $\langle I \rangle$) due to a defect is greatest for scattering defects at two positions, specular and as far off-specular as possible. Their findings must be combined with observations on the relative values of the noise at the various angles to obtain a 'best' position. Since surface noise is greatest at the specular angle, and decreases away from specular, it would seem preferable to operate well away from the specular angle. There would seem to be little to be gained by positioning a sensor in an intermediate position, such as small-angle off-specular.

Defect contrast can also be improved by selecting an appropriate size of scanned spot. Although contrast would at first sight seem to be maximised by having spot and defect of the same size, this does not hold when speckle is significant as a cause of noise, or indeed in any circumstance when the noise and message signals are not additive i.e. when the probability densities of noise and the combination of message and noise are not

the same. Dainty (1971) shows that scanner performance depends critically on the intensity distribution within the scanned spot. A 'top hat' distribution produces a false alarm rate smaller by two orders of magnitude than the radially symmetric Gaussian distribution found in the laser scanner used in most of our experimental work.

However, the most important property of defect signals in practical investigations is that their form is unknown 'a priori', is almost impossible to measure, and is highly variable even within one class, even for simple parameters such as spatial extent.

The defect (message) and noise components of scanner output signals may be represented by a density distribution of vectors in signal space (appendix 'b'). The forms of these distributions must be known, before even the sub-optimal (but attractive) detection scheme involving the linear detector may be used. The optimal likelihood ratio detector, to be described in section 3.5, requires a complete specification of the probability densities at all points in the signal space, which is in practice impossible to obtain.

The information concerning defect signals available at the beginning of the project is summarised in table 1.1. Because sensitivity to defects at the scanner output depends on the relationship between spot size and defect size, and the latter varies widely, the normal approach is to use a spot matched to the smallest defect of interest, and use electronic processing (chap.4) to improve sensitivity for the larger defects.

2.7 Conclusions

From this analysis, we conclude firstly that the pdf of the noise component of the scanner output signals which arises from surface roughness cannot be regarded as taking a specific form. It lies however somewhere between the Gaussian and negative exponential distributions, which thus provide bounds for use when systems are analysed quantitatively. For filtered noise, the Gaussian distribution is a good approximation.

The properties of the output noise measured from laser scanners in the simulation agree well with theoretical prediction.

Little information is available regarding the nature of message signals due to defects, particularly their behavior as vectors in signal space. Detection methodologies, to be introduced in the next chapter, will have to be chosen to accommodate this shortcoming.

Two principal considerations must be followed in designing scanners to maximise defect contrast:-

(1) The ratio of noise to mean signal level must be minimised.

This suggests use of incoherent illumination whose wavelength is long compared with surface roughness, but is comparable in size with typical defects. The sensors must accept radiation over an angle large enough to smooth out noise variations, and several sensors should be used, at the specular angle, and on both sides. The off-specular sensors should be positioned as far from the specular direction as is compatible with requirement (2).

(2) The sensors must be positioned to view the segment of the lobe pattern which undergoes the greatest proportional change due to the occurrence of a defect.

Unfortunately, the position for the sensors which

best satisfies (2) will generally not simultaneously satisfy (1), and a compromise setting will be required. This can be discovered only from experimentation.

Although the best ratio of surface noise to mean signal level is obtained using illumination at grazing incidence, experiments tend to show that this configuration does not yield optimal contrast. The variation in $\langle I \rangle$ due to a defect evidently decreases more quickly, as grazing incidence is approached, than the noise contrast C .

Considerations important in scanner selection were summarised in table 2.1. For convenience in signal processing, it is also useful if the scanner output signal has a mean level which is constant over a scan and has no geometrical distortion. This condition is fulfilled for CCD line scanners, but for laser scanners only provided the angle of scan is small. All forms of scanner may be operated so that self noise is small compared with noise generated by surface roughness, though the laser scanner is best in this respect.

It is concluded that on overall merit CCD-based

cameras are preferable to the laser scanners, TV cameras and flying field scanners also considered in section 2.2. However, to obtain best contrast using CCD cameras, it would be necessary to use at least two, viewing the same line of strip from different angles, synchronised to provide a vector signal. The information obtainable from the noise component of the output signal would still not provide elementary characteristics of surface contour such as RMS roughness.

CHAP.3 .DETECTION:FUNDAMENTAL CONCEPTS

3.1 Introduction

In this section,we discuss the general problem of detecting a message in the presence of noise. Our purpose is to develop and extend the methodology developed for target detection in radar and sonar,in which a useful message signal must be discriminated from unwanted background noise,and to recommend a strategy for the detection of defects in automated inspection.

We start by providing a definition of detection,and then reduce the standard system for statistically-based detection to a simple canonic form.(sect.3.2). The maximum-likelihood decision process is analysed in detail in section 3.3,with some alternatives then considered in 3.4. Some detectors which promise better performance than our canonic form are introduced in 3.5. Finally, 3.6 concludes by recommending a strategy for the detection of defects on rough surfaces. More detailed analysis of the individual components,including the development of criteria for selecting parameter values is provided in chapters 4 to 6. This covers both

theoretical and experimental aspects in detail.

The detection problem in its most general form may be stated as follows. We are given a signal, which always contains a noise component, and which sometimes, though rarely, also contains a message component. We are then required to state, as a result of observations (measurements) on the signal, whether noise only is present in the signal, or whether both noise and message are present. Thus, detection involves a decision between two alternative hypotheses, termed 'states of nature'.

The message component is regarded as carrying useful information; in this work, its presence indicates the occurrence of a defect. The noise component, on the other hand, provides no useful information, and further tends to conceal the message. It comprises an unavoidable nuisance.

Message and noise are assumed uncorrelated and, initially, mutually additive.

The noise is generally stochastic, although it may include a deterministic component. It is also, particularly

in surface inspection problems, generally non-stationary, though in most of our analysis stationarity will be assumed for simplicity. The message is also generally stochastic; the only constraint on its properties is that it alter the properties of the signal. However, in surface inspection applications, it is both convenient and realistic to assume messages which have (or, are similar to) some deterministic form.

In the simple case of a one-dimensional scalar system, we designate the signal as:-

$$l = n \quad \text{.....} \quad \text{when noise only is present,}$$

and,

$$l' = n + m \quad \text{....} \quad \text{when a message accompanies the noise.}$$

More generally (and certainly in most automatic inspection problems) the signal is multidimensional. In surface defect detection, the signal is two-dimensional, so we must write

$$l = l(x, y)$$

Additionally (in surface inspection applications, this describes the output of a scanner

having multiple sensors, as explained in section 2.1) the signal may be an N -dimensional vector, in which the signal describing the surface about a point x, y comprises N components $\ell_1(x, y), \ell_2(x, y), \dots, \ell_N(x, y)$ etc. These will in general be correlated, partially but not completely.

3.2 Detection as a Decision Process

Detection is fundamentally a process of decision, which has normally an analog signal (i.e. many 'bits' of information) as input, and a binary signal (one bit, indicating message present or absent) as output. In practice, statistical decisions almost invariably have some erroneous outcomes. An efficient detector minimises the probability of error, or, better, minimises the unfavorable consequences of the errors, where these can be quantified.

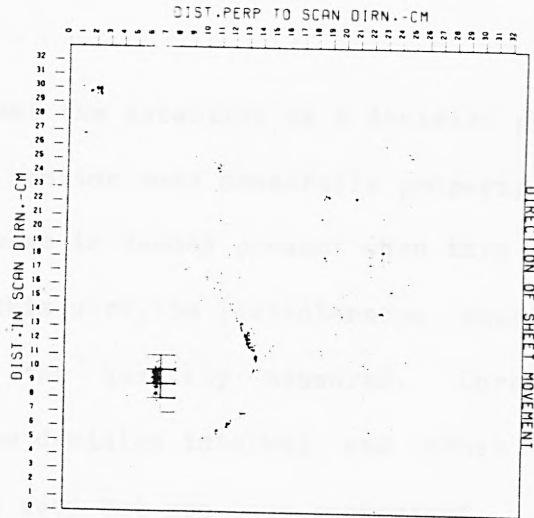
It is necessary at this stage to define the terms 'detection' and 'sensing' as applicable in surface inspection, and to distinguish clearly between them. A detection system has a binary output, for which a pulse at level '1' indicates that the window of observation

contains a defect, whilst a '0' indicates that it contains noise only. The system is considered merely to SENSE defects if the '1' level pulse appears not only when a defect is present, but also frequently when there is no defect. This distinction is illustrated by figure 3.1, in which trigger clusters arising from a defect are clearly present, but there are so many false alarm triggers that the system could not be used for automatic on-line inspection. The ability to distinguish defects from noise, i.e. to DIFFERENTIATE between them, is clearly absent. Sensing is nevertheless an essential prerequisite to detection. A system which cannot sense defects cannot hope to detect them.

To implement the detector, we divide the signal into cells termed 'intervals of observation', and perform an analysis on the information within each, as a basis for the decision for that particular cell. The size of the cell is chosen to be equal to the minimum message length expected, to maximise the contrast, defined below. This makes best use of the information provided in the message.

In this work, signals are assumed to consist of

VIDEOPRINT
COLD ROLLED STEEL STRIP



SHEET:- C16.11 DEFECT TYPE:- ROLL MARKS
PROCESSING:-NONE
DETECTOR:-5 DEG OFF-SPEC

RUN ON:- 06/10/77

FIGURE 3.1 TRIGGERS GENERATED BY A SIMPLE
DETECTOR (for a two-dimensional signal)

sequences of discrete samples, to accord with analysis by digital computer. The samples must not be confused with intervals of observation, each of which generally includes many samples.

To implement the detection as a decision process, it is necessary to monitor some measurable property of the signal. A message is deemed present when this crosses a threshold. In this work, the instantaneous amplitude of the signal is the quantity measured. Current, total energy within the decision interval and other measures might equally be used but are less convenient.

Thus, we detect by noting when the signal amplitude v exceeds some threshold $v(t)$. In essence, we are looking for a 'bump' in the signal. If the message is manifest in some other way, such as by a change in texture implying a change in the spectral distribution of signal energy without variation in its total amount, then some device is required which will convert this into a change in signal amplitude.

The signal may cross the detection threshold due to large fluctuations in the random noise, even when no

message signal is present (fig. 3.3). The larger the 'bump' due to the message compared with the variance of the noise, the further may the threshold be set from levels the noise is likely to reach, and the more reliable the detection. The size of the 'bump' due to the message relative to the expected value of the noise is termed the message contrast, and is characterised herein by the parameter Z (Schwartz and Shaw 1975), defined as:-

$$Z = \{E(\ell) - E(\ell')\}^2 / E(\ell'^2) \dots\dots \text{eqn. 3.1}$$

The designation $E(.)$ indicates expectation, and the use of squared quantities ensures that contrast is always positive, besides expressing it in terms of variance which is the relevant parameter for describing the magnitude of stochastic noise signals.

The performance of the decision process is optimised by making the contrast as large as possible. This usually involves placing an appropriate filter before the decision stage, to reject signal frequencies containing little message energy, but pass those containing much. The filter providing maximum contrast when the noise is white and Gaussian, and message

and noise are mutually uncorrelated, is termed the 'matched filter'. Provision of this filter requires explicit knowledge of the properties of both message and noise. In surface inspection applications, the former is often not available.

If the signal is a vector, or is multidimensional, or both, then message contrast may often be improved by combining components of signal between which the noise is uncorrelated but the message is correlated. Considered naively, the contrast then increases because the noise expectation rises as the square root of the sum of squares, whereas the contributions from the correlated messages add directly. Further, if some component in the noise is deterministic, it may be subtracted directly from the signal. All methods of contrast improvement may in principle be applied together, without interaction leading to mutual degradation.

In the decision process which is fundamental in detection, the signal will sometimes cross the threshold due to noise alone. The resulting erroneous detections are called 'false alarms'. Also, it may fail to cross the threshold when a message is present, resulting in a

'missed detection'. The threshold must be selected to minimise these errors, or better, to minimise their consequences. Since lowering the detection threshold $v(t)$ to reduce the number of missed detections will inevitably increase the number of false alarms due to noise, selection of the optimal threshold involves a compromise.

In most practical applications, false alarm triggers generated by noise tend to be distributed uniformly and at random over the signal, whereas message triggers tend to occur in clusters. Thus, triggers may often be rejected as being most probably due to noise if they are isolated. The local density of triggers may thus be used as a second decision parameter. This is particularly applicable in inspection applications, in which the signal is two-dimensional.

Thus, a detection system may be considered to comprise three basic components, acting on the signal in sequence as indicated in figure 3.2. Block(2) takes the essential first decision, based on signal amplitude. It is the one component in the system which is indispensable. It is non-linear and irreversible;

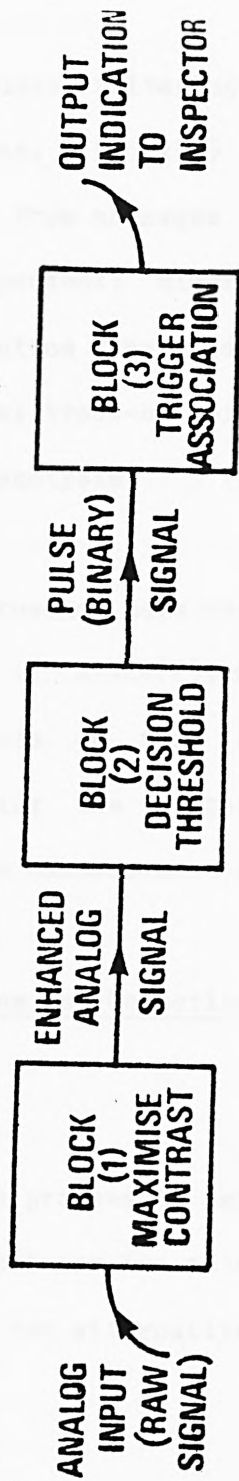


FIGURE 3.2 CANONIC FORM FOR DETECTION SYSTEM

information is thrown away on thresholding which cannot later be recovered.

Block (1) uses filtering and/or correlation to maximise contrast. Block(3) associates triggers, to distinguish those from messages from those due to noise. They are not independent; block (1) may introduce a loss in spatial resolution that reduces the effectiveness of block (3). Optimal trade-off of the various parameters requires careful analysis.

Several approaches have evolved for designing the decision process in block(2), which we now examine. The best choice depends on the 'a priori' information available concerning the properties of the noise, of the message, and of the consequences of the various errors.

3.3 Decision Taking in Detection-the Maximum Likelihood Detector

The decision process in detection is essentially a statistical test between two alternative hypotheses. The signal can occupy two alternative 'states of nature'; in

state I, noise only is present, whereas in state II, both message and noise occur together. Detection involves determining the correct state of nature during each interval of observation.

When the 'a priori' probabilities of noise and the combination 'message plus noise' are known, then the maximum likelihood ratio test (Hoel, 1971) may be used as a basis for the decision. Maximum likelihood is the best decision approach possible (in terms of minimising average overall loss L , defined later). However, its use requires complete information concerning the statistics of the decision, including quantification of the costs of the errors. When this is not available, inferior methods such as Neyman-Pearson have to be used.

The noise component is described by an 'a priori' probability density function, (the 'pdf'), $p(v, n)$. This specifies the probability that the signal amplitude takes the particular value v , for all values of v , given that noise only is present. The combination of message and noise is described by a second probability density function $p(v, m)$. Both pdfs are assumed to be stationary. To implement the detection, we compare the signal level v

with a threshold $v(t)$. If v exceeds $v(t)$, then a message is deemed to be present. Otherwise, there is no message. Since the distributions $p(v,n)$ and $p(v,m)$ will in general overlap (fig.3.3), some errors in the decision are inevitable. Thus, there are four possible outcomes, as follows:-

(1) Correct Detection, in which a trigger is generated when a message is present. The probability of correct decision, $P(cd)$, is given by:-

$$P(cd) = \int_{v(t)}^{\infty} p(v,m).dv \quad \dots\dots\dots \text{eqn.3.2}$$

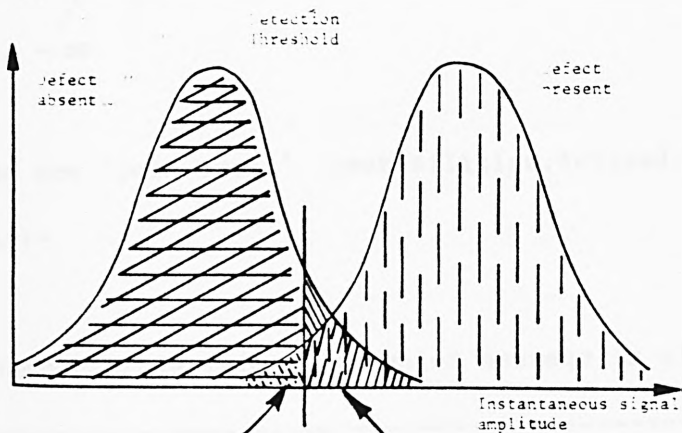
(2) Missed Detection, in which a message is present but no trigger is generated. The probability $P(md)$ of missed detection is given by:-

$$P(md) = \int_{-\infty}^{v(t)} p(v,m).dv \quad \dots\dots\dots \text{eqn.3.3}$$

(3) False Alarm, in which a defect trigger is generated although noise alone is present. Its probability $P(fa)$ is given by:-



(a)



(b)

Statistics of taking a decision with 'noisy' data

(a) When signal crosses threshold, defect is signalled as being present

(b) Statistical description of situation presented

FIG. 3.3 DETECTION DESCRIBED AS A STATISTICAL DECISION PROCESS

$$P(fa) = \int_{v(t)}^{\infty} p(v,n).dv \quad \dots\dots\dots \text{eqn.3.4}$$

(4) Correct decision that noise alone is present.

Here, no trigger is generated. Its probability of occurrence $P(na)$ is given by:-

$$P(na) = \int_{-\infty}^{v(t)} p(v,n).dv \quad \dots\dots\dots \text{eqn.3.5}$$

These are 'per event' probabilities, defined (for example) as:-

$$P(md) = \frac{\text{number of decisions for which message is missed}}{\text{number of decisions for which message is present}}$$

The effect on detector performance as the threshold $v(t)$ is varied is illustrated graphically using the receiver operating characteristic (ROC), shown in figure 3.4. This comprises a family of curves. Each curve is for a particular false alarm probability (and hence, for a

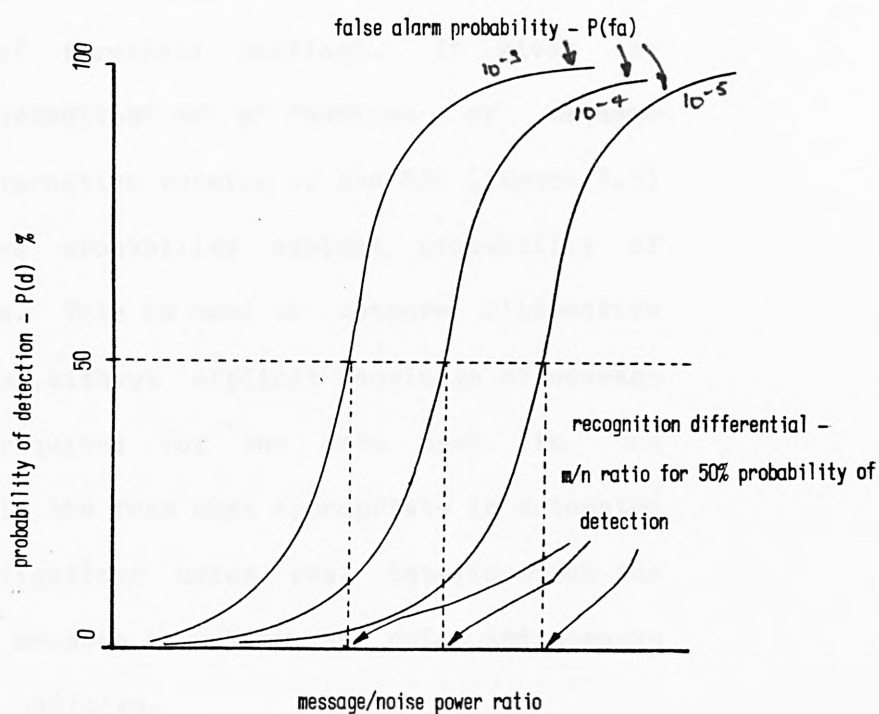


FIGURE 3.4 - RECEIVER OPERATING CHARACTERISTIC

Form (1) - message/noise ratio known

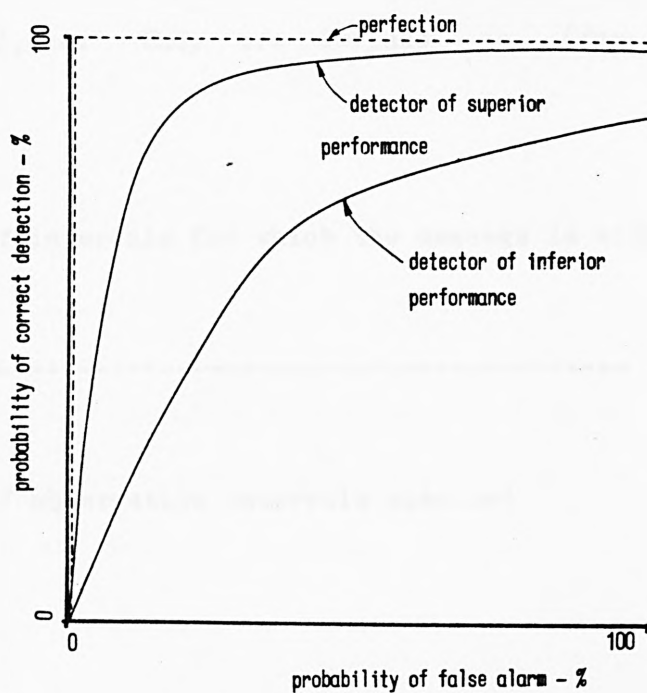


FIGURE 3.5 RECEIVER OPERATING CHARACTERISTIC

Form (2) - message/noise ratio not required to be known

simple detector, of threshold setting). It gives the probability of detection as a function of message contrast. An alternative version of the ROC (figure 3.5) plots false alarm probability against probability of correct detection. This is used to compare alternative processing schemes, without explicit knowledge of message contrast being required for the data used in the comparison. It is the form most appropriate in automated inspection investigations using real data, in which the amplitude of the message is unknown, and noise and message are probably not additive.

In order to properly characterise the decision process, however, we really need to specify the probabilities 'per observation interval examined'. These we designate $Q(md)$, etc. They are defined as (for example):-

number of intervals for which the message is missed

$$Q(md) = \frac{\text{number of intervals for which the message is missed}}{\text{number of observation intervals examined}}$$

number of observation intervals examined

The coefficients are evidently related by (for example):-

$$Q(md) = R(m).P(md) \dots\dots\dots \text{eqn.3.6}$$

Here, $R(m)$ is the overall probability of an observation interval containing a message. Similarly, we specify by $R(n)$ ($=1 - R(m)$) the probability of an observation interval containing noise alone.

Thus, the basic decision process is specified by the following four equations:-

$$(1) \dots Q(cd) = R(m).P(cd)$$

$$(2) \dots Q(md) = R(m).P(md)$$

$$(3) \dots Q(fa) = R(n).P(fa)$$

$$(4) \dots Q(na) = R(n).P(na)$$

Cases (2) and (3) constitute errors which are to be avoided. Some of each are however inevitable because of the nature of the statistical decision. Each kind of

error incurs an economic or operational penalty which can be determined quantitatively only by examination of the overall system of which the decision process forms part. We quantify the penalty by specifying $C(fa)$ as the cost of a false alarm, and $C(md)$ as the cost of a missed detection. The following equation then specifies the complete process:-

$$L = C(fa).(1 - R(m)).P(fa) + C(md).R(m).P(md) \quad \dots\dots \quad 3.7$$

Here, L is the overall loss which must be minimised. Minimisation is achieved by suitable choice of the decision threshold $v(t)$. Raising $v(t)$ reduces $P(fa)$ but increases $P(md)$.

3.4 Alternative Approaches to Decision in Detection

Maximum likelihood is not the only approach possible for decision taking in detection. Others are available, which require less 'a priori' information, or which promise better performance in particular circumstances. We now describe the most important of

these.

The SQUARE LAW detector, in which the decision threshold is imposed on squared samples of signal, gives slightly better performance (i.e., better $P(d)$ at given $P(fa)$) when the message/noise power ratio is small. The improvement thus obtained is however trivial, being equivalent to increasing message/noise ratio by about 0.38 dB (Whalen 1971). The extra complication introduced by the squaring would thus not seem to be justified.

Often, all parameters required for describing a statistical detection process in analytic terms are not available. The parameters most frequently missing are the probability of message occurrence $R(m)$ and the probability density function $p(v,m)$ for the message. The cost penalties $C(fa)$ and $C(md)$ may also be unavailable. The surface defect detection problem is of this kind. Thus, an alternative approach must be made to design of the decision process, for example the NEYMAN-PEARSON criterion (Skolnik, 1962). The overall performance can then be expected to be somewhat inferior to that for the Maximum Likelihood detector just discussed.

In the Neyman-Pearson detector, the threshold $v(t)$ is adjusted to obtain a specified false alarm rate, and the remaining parameters then chosen to maximise the probability of detection.

In some visual inspection applications, such as the automated screening of pap smears for cancer cells, and the examination of PC boards for defects, missed detections cannot be tolerated. In these applications, an INVERSE NEYMAN-PEARSON criterion is adopted, in which the detection threshold is set as low as is required to be triggered by all defect messages. The resulting high false alarm rate is accepted, and further examination of the data, possibly by a human operative, is used then to distinguish the false alarms. The overall cost effectiveness of the inspection process is thereby much improved.

Wald (1947) showed that improved performance could be obtained in a likelihood based decision by breaking it up into a series of decisions, each of which considers only part of the data available. Each decision in the sequence has now three possible outcomes, instead of only two. These are: message present, message absent or

'insufficient evidence, look further'. It is applied to message detection by examining a particular physical region as often as necessary, until sufficient information has been accumulated to reach a definite decision, with specified statistical reliability, that a message is present or absent. Wald showed mathematically that the amount of data (or time) required to reach a decision with a given degree of statistical reliability is thus reduced, or the reliability with a given quantity of data improved, compared with an examination of fixed length. This can be applied to message detection only when it is possible to examine the signal arising from a given region of sample repeatedly, as often as is required. This is practicable for example in radar systems, in which a beam may be steered under computer control, and can be allowed to dwell on a region of space containing a doubtful target until sufficient data has been accumulated for a sound statistical decision. It is difficult to apply the approach to defect detection, since with all scanners in common use the sample material can be examined once only. Further, the noise component is determined by the properties of the surface in the neighborhood of the defect, and will thus be the same each time it is scanned (giving no improvement from repeated

scannings), unless the properties of the scanner are modified each time. Application of sequential analysis to practical radar systems typically provides an improvement in performance equivalent to an increase in message contrast of 3dB (Bussgang and Middleton, 1955).

The GAME THEORETIC detector is an alternative and radically different approach for detector design, in which 'a priori' probabilities are not required. This regards detection as being a game between an experimenter and an uncooperative Nature. The game theoretic approach developed originally by Von Neumann is used to describe the decision process. This tabulates the consequences of the decision in terms of a payoff matrix (fig. 3.6), in which the (i, j) th entry specifies the benefit to A (the experimenter, in this case), consequent upon A taking the course of action i and B (nature) taking the action j . The benefits correspond to the inverses of the weighted costs, L , in the Bayesian detector.

The criterion most frequently recommended for A (Hancock and Wintz, 1965) is the MINIMAX approach, in which A selects his decisions to minimise his maximum risk. The cost of a minimax decision is always worse than the Bayes

Fig. 3.6 Payoff Matrix for
'Game Theoretic' Detector

Action by 'B' – true 'state of nature'
for signal

	I	II
I	x_{11}	x_{12}
II	x_{21}	x_{22}

Action by 'A' –
assignment of
signal

cost, and if the 'a priori' probabilities of the two states of nature differ widely, the minimax criterion is desperately inefficient. It is too conservative; bolder tactics would generally bring about better performance. The approach is thus of little use in defect detection, when 'a priori' probabilities are widely different for the alternatives, 'defect' and 'no defect'.

3.5 Ideal Detectors

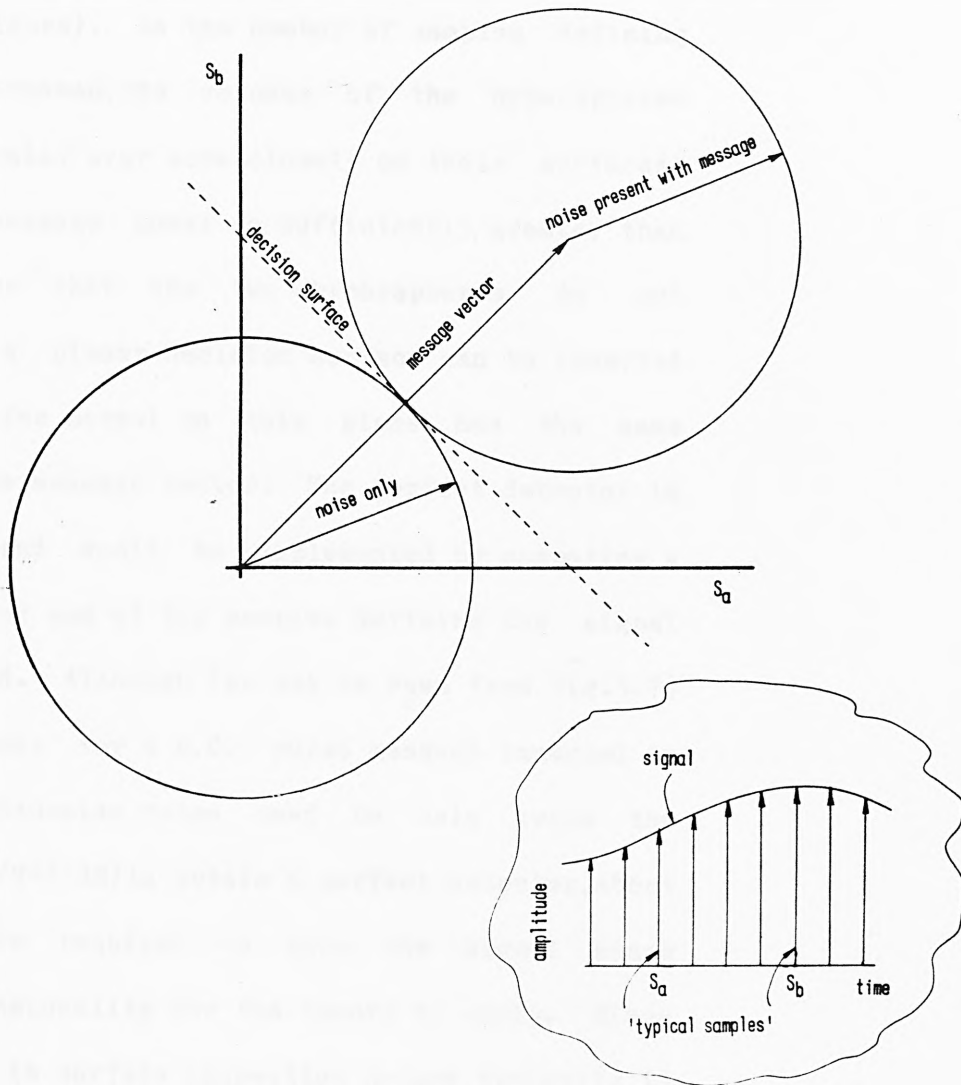
In figure 3.2 we assumed a particular canonic form for the detection system, in which decision (a non-linear process involving an irreversible loss of information) is preceded by contrast enhancement and followed by association of the outcomes of the decisions. This is the most common form but is certainly not the only one. Alternatives may exist providing improved performance. To get the best performance possible with a given signal, one must include the maximum amount of information in the decision. We now describe two detectors which utilise this principle. The first (the PERFECT detector) generates no false alarms or missed detections provided the message contrast exceeds 3db. The second, the OPTIMAL

detector, provides the best possible maximum likelihood performance using given signal information.

In certain rather ideal circumstances, a detector can be obtained in which all decisions are correct ($P(m_d)$ and $P(f_a)$ both zero) despite the presence of noise in the signal. This is the PERFECT detector.

This operates as follows. Consider a signal defined by very many samples, such that the signal space (app. 'B') required to describe it is of high dimensionality. The situation is that considered in the derivation of Shannon's second law (Raisbeck, 1963) quantifying channel capacity, except that for the perfect detector there are only two signals, noise only, and message plus noise. In the signal space (fig. 3.7), the possible 'noise only' signals occupy the surface of a hypersphere of radius $R = (2.W.T.P)^{1/2}$ centered on the origin, whilst 'message plus noise' signals lie on the surface of a second hypersphere also of radius R whose centre is the tip of the message vector. The signal space illustrated in figure 3.7 is two dimensional, representing a signal specified by the two typical samples $S(a)$ and $S(b)$ only (explained in the

FIGURE 3.7 - PERFECT DETECTOR
(operation explained in signal space)



inset to the figure). As the number of samples defining the signal increases, the volumes of the hyperspheres become concentrated ever more closely on their surfaces. Provided the message power is sufficiently greater than the noise power that the two hyperspheres do not intersect, then a planar decision surface can be inserted between them. The normal to this plane has the same direction as the message vector. The perfect detector is hence linear, and would be implemented by comparing a linearly weighted sum of the samples defining the signal with a threshold. Although (as may be seen from fig.3.7) the message power for a D.C. pulse message immersed in additive white Gaussian noise need be only twice the noise power ($S/N=3$ dB) to obtain a perfect detector, about 10^{**5} samples are required to give the signal space sufficient dimensionality for the theory to apply. Since defect signals in surface inspection occupy typically 10 samples with the system parameters (particularly resolution) normally used, windows of observation large enough for perfect detection would be poorly matched to defect signals, giving a poor message/noise ratio.

This analysis suggests, however, that if the system bandwidth could be increased by several orders of

magnitude, and if both noise and message powers increased with bandwidth in the same proportion, then perfect detection could be exploited to improve system performance. This approach regards detection as a communication process, in which low contrast can be compensated by increased message bandwidth.

The OPTIMAL DETECTOR examines the signal within a grid of dimension N 'windows of observation' by M . If the signal is an L component vector, we consider the signal within L such windows simultaneously. The total number of windows - $N \times M \times L$ - is designated K .

The signal amplitude measured within each window provides one axis of a K dimensional hyperspace. Certain combinations of amplitude values are found to be characteristic (on the basis of maximum likelihood ratio) of the presence of a message, the remainder indicate noise only. Thus, the space can be partitioned into disjoint regions, some indicative of noise only, the others of message and noise (fig. 3.8). Root (1970) in describing this detector, assumes the regions to have linear boundaries, but this restriction is unnecessary. Assignment of regions is based on maximum likelihood

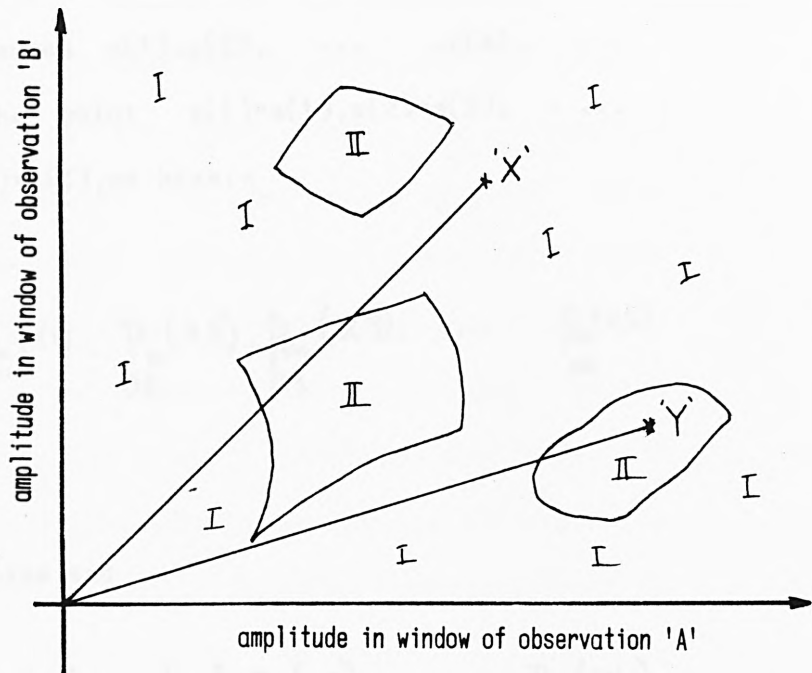


FIGURE 3.8 OPTIMAL DETECTOR two dimensions only shown, for clarity

Assignment of regions in space:-

noise only I

message plus noise II

measured vector 'X' attributed to noise only.

measured vector 'Y' attributed to message.

(better, on minimisation of overall loss). If we designate the measures $x(1), x(2), \dots, x(k), \dots, x(K)$, then for the point $x(1)=a(1), x(2)=a(2), \dots, x(k)=a(k), \dots, x(K)=a(K)$, we have:-

$$P(m) = P_{a1}^m(x_1) \cdot P_{a2}^m(x_2) \cdot P_{a3}^m(x_3) \cdot \dots \cdot P_{ak}^m(x_k)$$

for message and noise, and

$$P(n) = P_{a1}^n(x_1) \cdot P_{a2}^n(x_2) \cdot P_{a3}^n(x_3) \cdot \dots \cdot P_{ak}^n(x_k)$$

for noise only. The measures are assumed mutually independent. We detect by deciding that the signal contains noise only if $P(n) > P(m)$, otherwise a message is present also.

This is the ultimate form of detector, since no information is lost prior to the one 'final' decision, based on the location within the space of the observed vector of signal amplitudes. All possible forms of signal are provided for, and all available information is utilised.

This approach is seldom used in practice, because of the difficulty of obtaining explicitly the 'a priori' probability density functions for the various combinations of measures. If however the pdfs of noise and message are known analytically (easy for radar systems, difficult for defect detection), or may be assumed SAFELY to have some known form, its use may become practicable. McDonough (1971) provides for example a complete analytic description of an optimal detector for a vector signal comprising Gaussian messages added to Gaussian noise. The optimal detector is formally equivalent to a feature space pattern classifier (Duda and Hart, 1973).

3.6 Concluding Remarks

In this chapter, we have discussed detection as a decision process in which a choice must be made, from analysis of finite segments of signal, as to which of two alternative 'states of nature' has generated the signal. The standard theory of message detection for radar and sonar has been used as a basis. Several alternative approaches have been considered, and a canonic form has been proposed which covers many of these, making them

simple to understand and to analyse. We must now recommend a preferred approach for use in surface defect detection.

In the presence of noise, erroneous decisions are unavoidable with signals of finite length, and we must minimise their consequences. The best performance possible in this respect is obtained using the optimal version of the maximum-likelihood detector, but the 'a priori' information required for its design is in surface inspection normally not available. Game theoretic detectors give poor performance (information available even in surface inspection problems is thrown away) and sequential detection is inapplicable since repeated scanning is not possible. The perfect detector would require signals ('windows of observation') perhaps 100,000 samples long, and anyway offers no advantage when contrast is poor. Thus, we recommend the three-block canonic detector shown in fig.3.2, with the Neyman-Pearson criterion used to set the decision threshold. This form is assumed in the analysis provided in the following chapters.

CHAP.4 DETECTION-DECISION AND CONTRAST ENHANCEMENT

4.1 Introduction As a result of the analysis presented in chapter 3, a preferred system has been proposed for detecting defects in automated visual surface inspection. This is a canonic form, comprising a sequence of components which can be analysed (to a first approximation) separately. Of these, one only (block(2), the decision stage) is indispensable, and the Neyman-Pearson criterion was shown to be the best strategy for the design of this stage.

Experiments show however that a single stage system involving decision alone is in general not sufficient to give adequate detection performance. This is certainly true in the specific case studied using simulation to demonstrate the applicability of the methodology. Inclusion of other stages proposed in the canonic form is essential.

Chapter 4 considers the first block in the canonic form, which comprises processing applied to the analog signal to increase defect contrast before a first decision. For local defects, i.e. those whose message

energy is concentrated within a small area of surface, two methods are proposed for contrast enhancement, namely matched filtering, and linear weighted summation. The two methods are incidentally mutually independent, and may be applied simultaneously to the same vector signal, without interacting to produce a mutual degradation. For non-local defects (those whose message energy is dispersed over a considerable area of surface), a different approach to contrast enhancement is necessary. Processing for enhancing the contrast of non-local defects tends to be specific to the particular kind of defect.

Discussion of processing methods in block(3), which are applied after the first decision, is provided in chapter 5.

The analysis in this chapter is arranged as follows. Section 4.2 discusses simple detection involving decision (block(2)) only, and uses a simulation performed using data from cold rolled steel strip to demonstrate its shortcomings. Section 4.3 then considers the first approach for contrast enhancement, matched filtering. It provides a theoretical analysis of the methodology using a discrete approach, and shows how an

optimal set of match waveforms may be chosen, in the absence of definite 'a priori' information concerning the shapes of defect signals. Simulation analysis is then invoked to demonstrate that matched filtering enables defects to be detected which the system would otherwise miss. Section 4.4 then considers the alternative method for enhancement involving linear weighted combination of several segments of signal containing message information from the same defect. A theoretical analysis indicating the improvement in performance obtainable is again provided, together with a procedure for obtaining optimal values for the various parameters involved in the combination. This is followed by a demonstration using simulation that the approach is effective in practice.

Discussion of methods for enhancing non-local defects, i.e., those whose message energy is distributed over a substantial area of surface, is provided in section 4.5. A completely different approach is required in this case, and it is impossible to provide a single analysis covering all types of non-local defect, so attention is concentrated on one specific type. A methodology for enhancing this defect is presented. This is followed by an analysis indicating the magnitude of the increase in

contrast thus obtained. Simulation analysis over cold rolled steel strip is then used to confirm that the methodology is effective. Finally, general conclusions regarding enhancement prior to decision are presented in section 4.6.

The specific problem used for demonstrating the applicability of the various processing methods by simulation (i.e., the visual inspection of the surface of cold rolled steel strip) was introduced in section 1.5. A detailed description of the data base and package of computer programs used in the simulation is provided in appendix 'A'.

4.2 Detection Using Decision Only

The first task was to determine the limitations of detection using decision only, using simulation applied to a scalar signal. This was carried out over the steel strip data base. A particular form of 'decision only' detector, the SIRA (or moving average) detector originally described by Brook (1971b) was used. This is noteworthy in having been implemented in hardware in fully

engineered form, and having been used on-line in several installations (Brook et al., 1977). It compares the signal with an attenuated (the attenuation may be greater than unity) (fig 4.1), low-pass filtered version of itself, and generates a defect trigger whenever the unprocessed signal crosses this threshold. It thus adapts to local variations in signal mean. The low-pass filter delays the reference signal relative to the raw signal. When this delay is not removed, the SIRA detector is adaptive only to the mean of 'past' signal amplitude; future values are not considered. This combination of low pass filtering followed by subtraction generates a high pass filter, which is sensitive preferentially to edge defects, and responds badly to large area defects, particularly when these have poor contrast.

Figure 4.2 illustrates the wide variation in response of the SIRA detector. The variations in trigger clusters derived from the different signal components show that different information is available in the various components. It shows triggers generated from the specular, 7deg. off-specular and 30 deg. off-specular views of the same sheet, with both positive going and

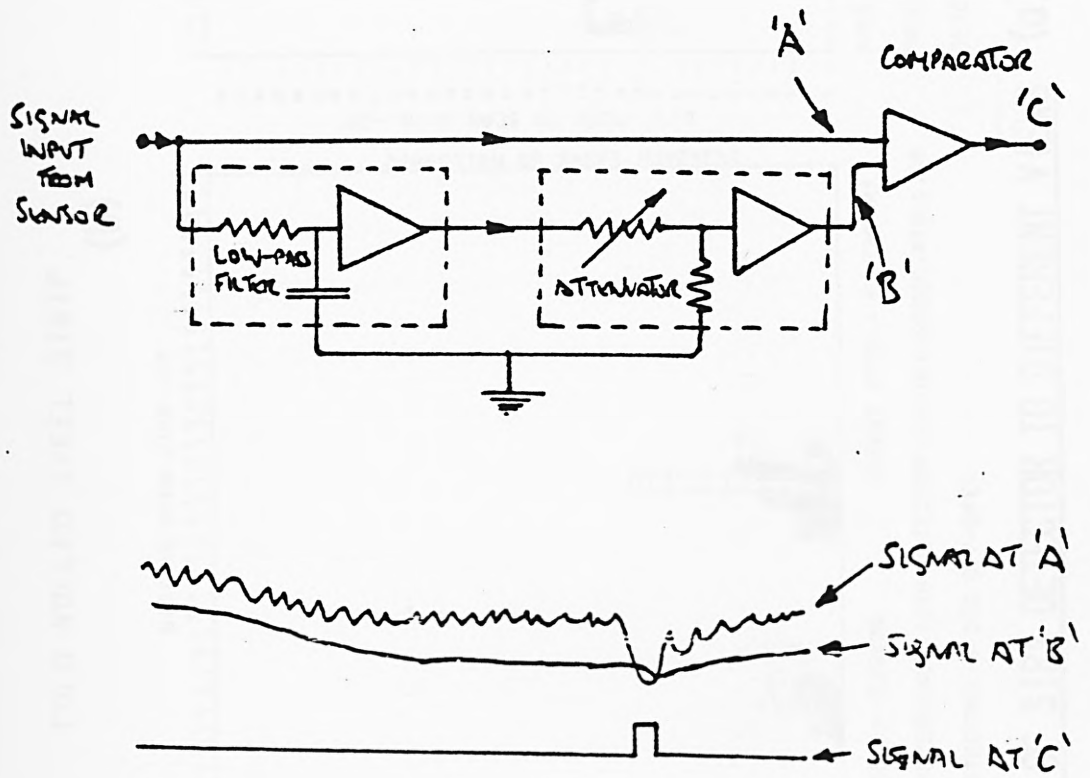
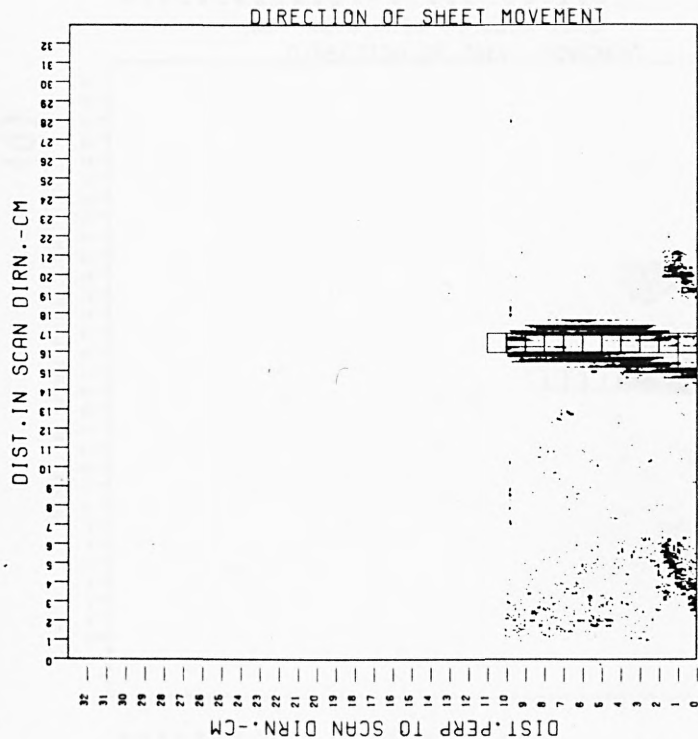


FIGURE 4.1 SIRA DETECTOR - PRINCIPLE OF OPERATION
(negative going message signal)

COLD ROLLED STEEL STRIP

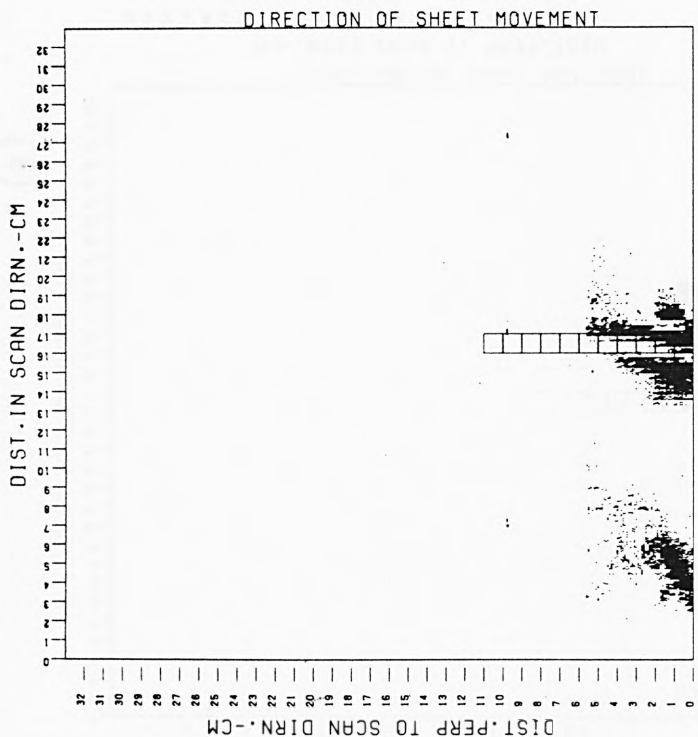
(a)



SHEET:- E16.35 DEFECT TYPE:- PINCHMARK
PROCESSING:-SIRA DETECTOR,CUT-OFF=20000.ATN=1.08
DETECTOR:-SPECTULAR

COLD ROLLED STEEL STRIP

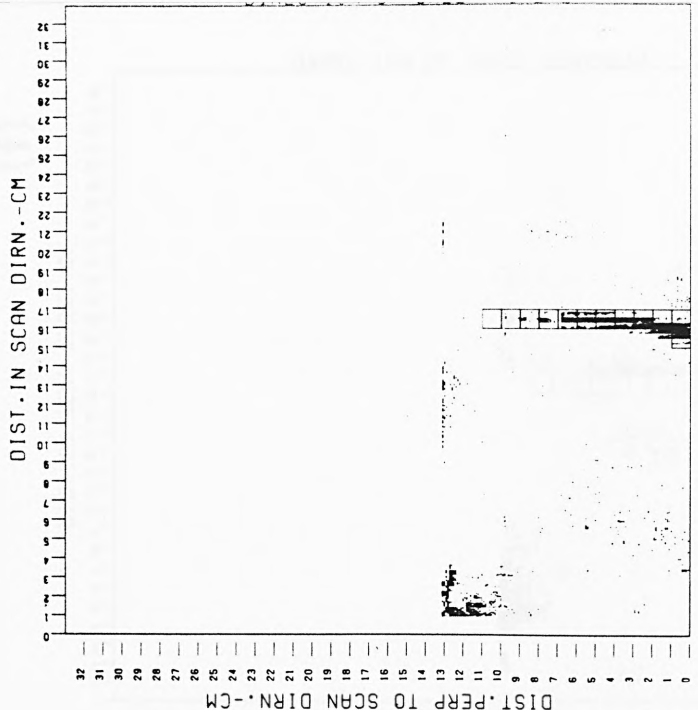
(b)



SHEET:- E16.35 DEFECT TYPE:- PINCHMARK
PROCESSING:-SIRA DETECTOR,CUT-OFF=20000.ATN=1.08
DETECTOR:-5 DEG OFF-SPEC

COLD ROLLED STEEL STRIP

(c)



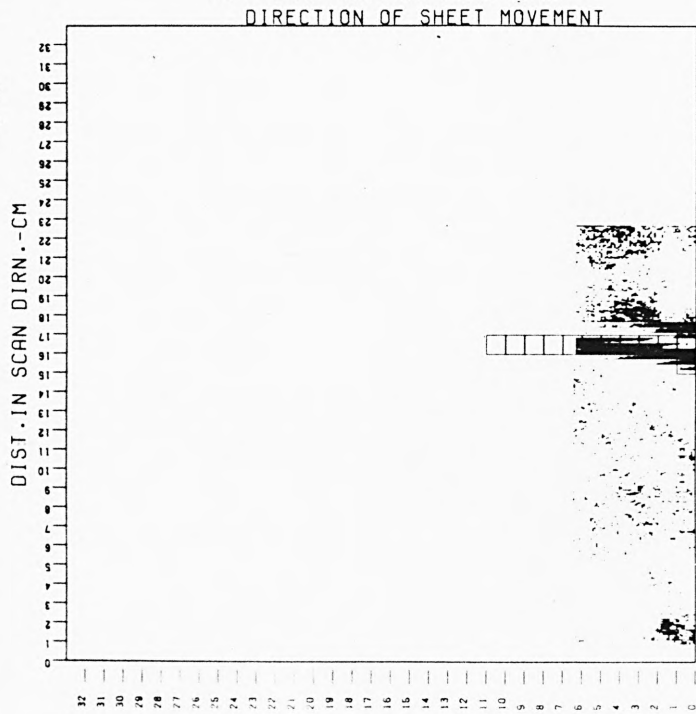
SHEET:- E16.35 DEFECT TYPE:- PINCHMARK
PROCESSING:-SIRA DETECTOR,CUT-OFF=20000.ATN=1.08
DETECTOR:-30 DEG OFF-SPEC

FIGURE 4.2 RESPONSE OF SIRA DETECTOR TO DIFFERENT VIEWS

(a)-(c)

COLD ROLLED STEEL STRIP

(d)



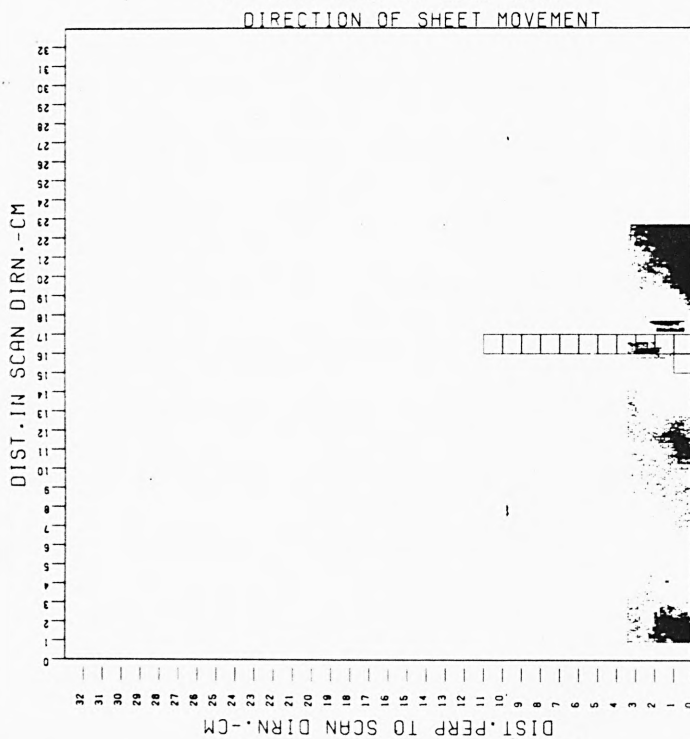
SHEET:- E16.35 DEFECT TYPE:- PINCHMARK

PROCESSING:-SIRA DETECTOR.CUT-OFF=20000.ATN=0.95

DETECTOR:-SPECTULAR

COLD ROLLED STEEL STRIP

(e)



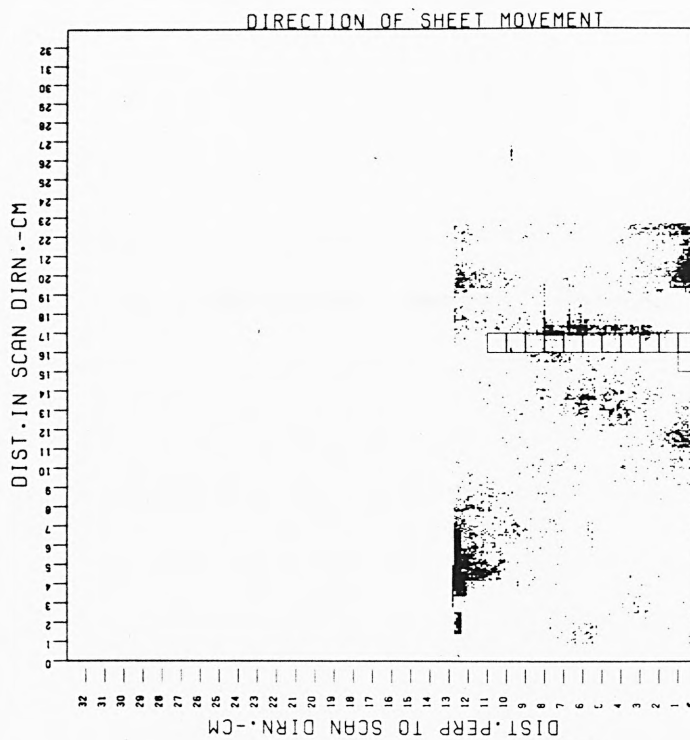
SHEET:- E16.35 DEFECT TYPE:- PINCHMARK

PROCESSING:-SIRA DETECTOR.CUT-OFF=20000.ATN=0.95

DETECTOR:-5 DEG OFF-SPEC

COLD ROLLED STEEL STRIP

(f)



SHEET:- E16.35

DEFECT TYPE:- PINCHMARK

PROCESSING:-SIRA DETECTOR.CUT-OFF=20000.ATN=0.95

DETECTOR:-30 DEG OFF-SPEC

FIGURE 4.2 RESPONSE OF SIRA DETECTOR TO DIFFERENT VIEWS (d)-(f)

negative going thresholds. The variation in the area of triggers plotted in the videoprints arises because detection stops when 4000 triggers have been gathered, for economy. The parameter ATN specifies the attenuation applied to the reference signal. If ATN is less than unity, negative going pulses generate triggers, if ATN is greater than unity, the triggers are generated by positive going pulses.

The first stage in evaluating the SIRA detector in simulation was to select values for the attenuation and cut-off frequency to give best detection whilst satisfying the false alarm rate requirement. This involved, firstly, processing sheet known to be free of defects and varying the threshold level until the required false alarm probability was obtained. Sheets containing selected marginal defects were then examined, and the cut-off and delay adjusted to achieve best detection. Optimal parameters were selected for both positive and negative going messages.

Initially, the delay introduced by the low pass filter was not removed. To investigate its effect, the delay was halved, then reduced to zero. This produced no

improvement in detection performance, so the delay was retained (removing it would complicate the hardware). Since the characteristics of the signal may differ between the three sensors, a separate optimisation was performed for each. ^{Best parameters for machine according to the Nyquist-Person} criterion are as follows:-

cut-off frequency-20,000 Hz.

attenuation, 1.12 for +ve going and 0.95 for -ve going signals.

Cut-off frequency had little effect on performance between 2000 and 20000 Hz. The evaluation was performed on sheets C99.04, C99.05, C99.06, C99.07.

To determine the efficiency of detectors using decision taking from threshold comparison only, data from 140 steel sheets containing defects (app'B') was analysed computationally. Detection performance was determined by noting the locations of triggers produced by the detector, relative to squares indicating the position on

the sheet at which defect markings are known to be present, using a videoprint (app'A'). To minimise the cost and labour involved in evaluating detector performance over a vast data base, a computer program (called EPERF) was produced which examined the locations of triggers on videoprints, without actually plotting them on microfilm. Detection was considered successful when defect triggers were detected within the squares, but not outside them.

The simulation investigation showed the SIRA detector applied in the 'x' direction to be adequate for 30 of the 36 kinds of defect considered. It was however inadequate for detecting edge strain, point wrench, pinch marks, sticker wrench, seams and chatter marks.

For some defects, the direction along which the threshold is applied is significant. On cold rolled steel strip, edge strain and sticker wrench generate marks for which the signal changes much more rapidly in the direction of sheet movement (y-direction) rather than in the scan direction which is perpendicular. Merely turning the SIRA detector round through 90 degrees, so that it operates in the y-direction (between scans)

rather than within individual scans in the x-direction was found to be sufficient to detect one kind of defect (edge strain) adequately. The optimal settings for the SIRA detector used were found by experiment to be:-cut-off frequency, 0.75 cycles per cm., attenuation, 0.87, excursions, negative going, and signal component, specular. Figure 4.3 (a) shows sheet C23.04 containing edge strain viewed specularly, with triggers produced by the standard SIRA detector. No trace of the defect is discernable. Figure 4.3 (b) shows the same sheet viewed at the same angle, but using the rotated SIRA detector with optimal parameters. The shape of the marks due to edge strain is clearly evident.

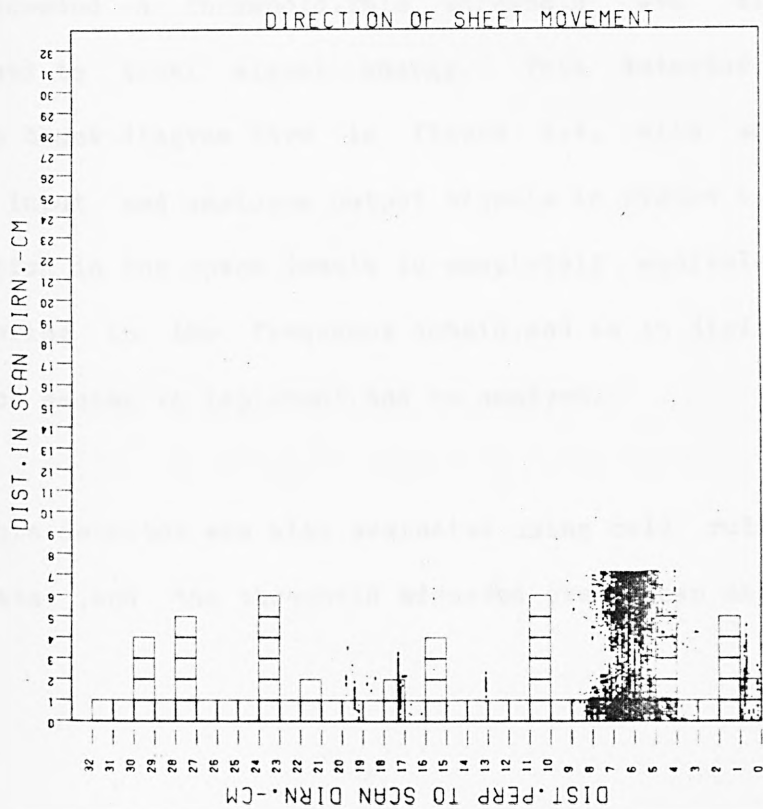
This approach did not however work for sticker wrench, for which the defect contrast appears to be too small.

4.3 Contrast Enhancement Using Filters

The first elaboration of the detection system to be investigated for enhancing message contrast is filtering (block (1) in figure 3.1). In practical application, this poses difficulties because the defects produce message

COLD ROLLED STEEL STRIP

(a)



COLD ROLLED STEEL STRIP

(b)

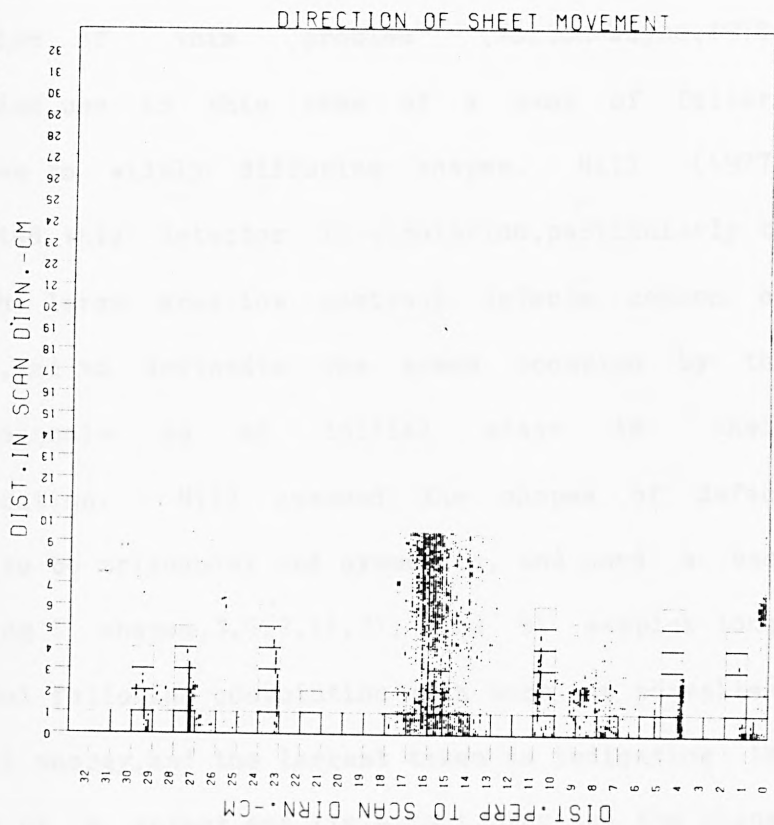


FIGURE 4.3 DETECTION OF EDGE STRAIN

signals varying widely in size and shape. An early examination of this problem (Norton-Wayne,1972) recommended use in this case of a bank of filters responsive to widely differing shapes. Hill (1977) implemented this detector in simulation, particularly to detect the large area, low contrast defects common on tinplate, and to delineate the areas occupied by the defect signals as an initial stage in their identification. Hill assumed the shapes of defect signals to be triangular and symmetric, and used a bank containing 6 shapes, 3, 5, 7, 11, 21, and 31 samples long. The signal following convolution with each was normalised for equal energy, and the largest taken as indicating the presence of a defect, and its extent, provided the signal level exceeded a threshold. This threshold was also normalised to total signal energy. This detector is shown in block diagram form in figure 4.4, with some typical input and analogue output signals in figure 4.5. Convolution in the space domain is completely equivalent to filtering in the frequency domain, and is in digital form much easier to implement and to analyse.

This detector was also evaluated using cold rolled strip data, and the threshold adjusted over clean sheet

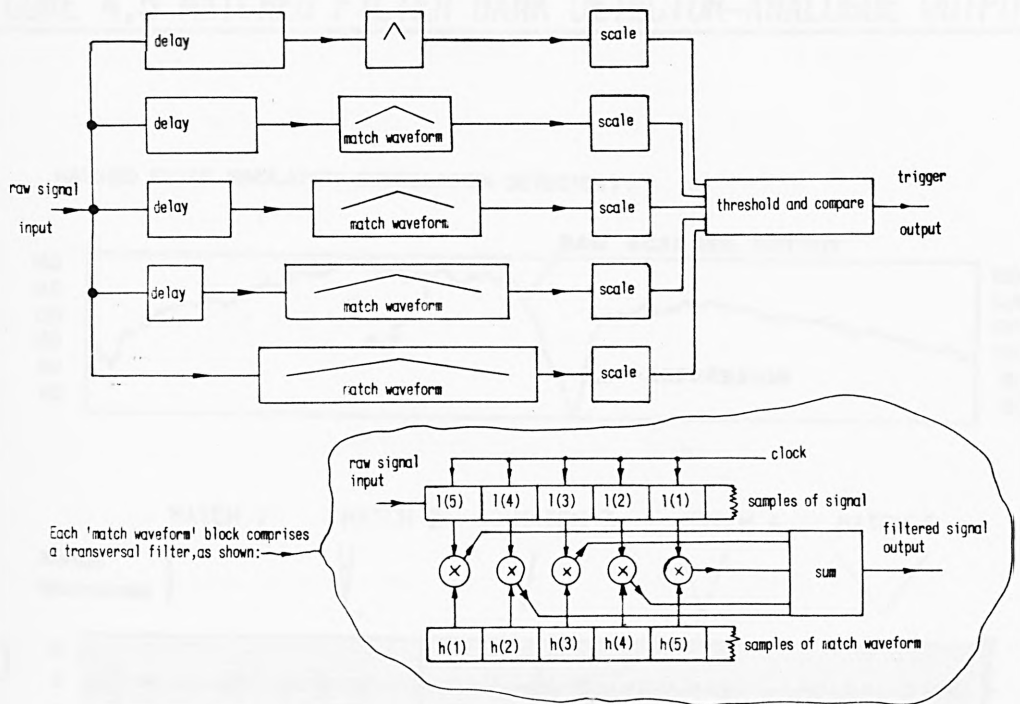
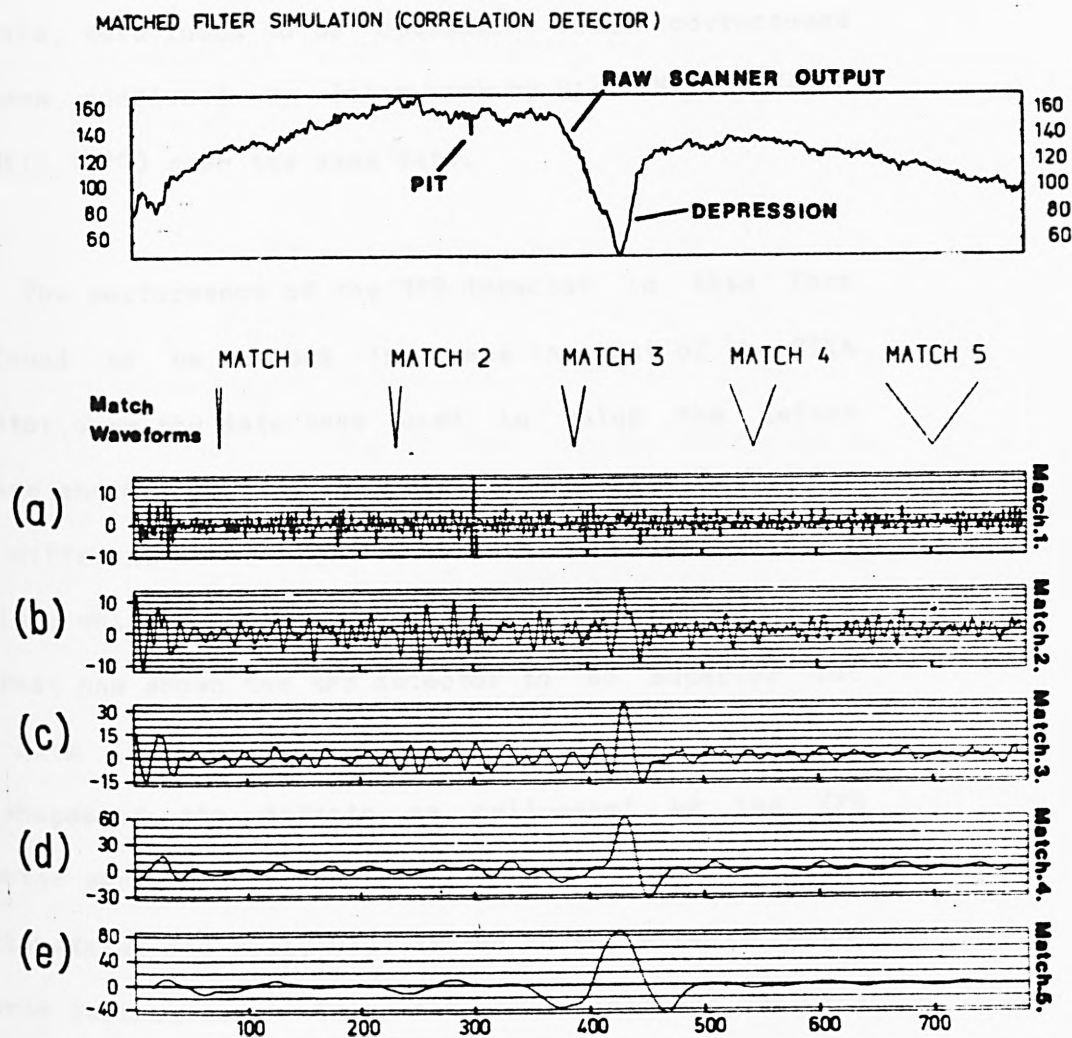


FIGURE 4.4 MATCHED FILTER BANK (MFB) DETECTOR
 block schematic showing principle of operation

FIGURE 4.5 MATCHED FILTER BANK DETECTOR-ANALOGUE OUTPUTS



(signals (a) to (e) respectively are outputs of transversal filters 1 to 5 respectively)

to optimise performance. It is difficult with the MFB detector to adjust to a specified false alarm rate, since there are in effect 6 different windows of observation. Thus, the same significance must be attached to each output pulse, irrespective of its length. Values of 30 samples for the specular, and 20 for the two off-specular channels, were found to be optimal. Their correctness has been confirmed by later work by Hill (Norton-Wayne and Hill, 1979) over the same data.

The performance of the MFB detector in this form was found to be almost the same as that of the SIRA detector, over the data base used in which the defect message shapes rarely resemble triangles. Evaluation over different data, e.g. that obtained from scanning tinplate which contains many large area defects of low contrast has shown the MFB detector to be superior for that data (Hill, 1977). However, for cold rolled strip, the shapes of the defects as delineated by the MFB detector were significantly inferior. This was later (Norton-Wayne and Hill, 1979) traced to an effect termed reverse delineation. This can be eliminated, but the processing required is expensive.

By lowering the detection threshold to increase sensitivity, it was possible to sense point wrench and pinch marks, but the number of false alarm triggers then obtained was much too high (figs. 4.6 and 4.7). However, detection can be achieved using this low threshold if a means can be found for rejecting false alarm triggers whilst retaining triggers due to defects. This is considered in chapter 5.

Although the MFB detector incorporates filters intended to be matched, their form had been selected empirically and without theoretical analysis. The MFB filter is further one-dimensional whereas most defects are inherently two-dimensional. Thus, much improved contrast enhancement may be possible, provided an appropriate form of filter can be devised. The first step in obtaining such an optimal filter is theoretical analysis. Hill (1977) attempted to apply theoretical analysis to matched filters for surface defect detection, but used a cumbersome analogue theory. We use an alternative discrete analysis, which is equally rigorous but is far more comprehensible. It extends an approach described by Schwartz and Shaw (1975).

FIGURE 4.6 POINT WRENCH SENSED WITH STANDARD PROCESSING

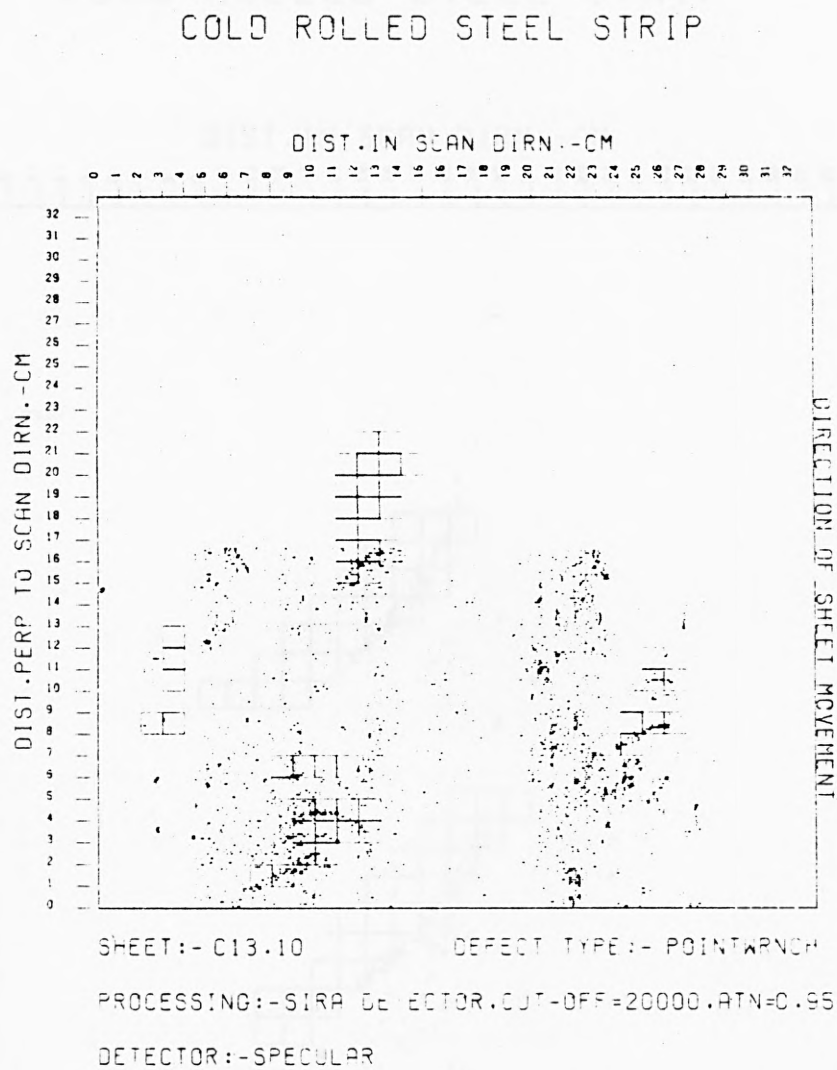
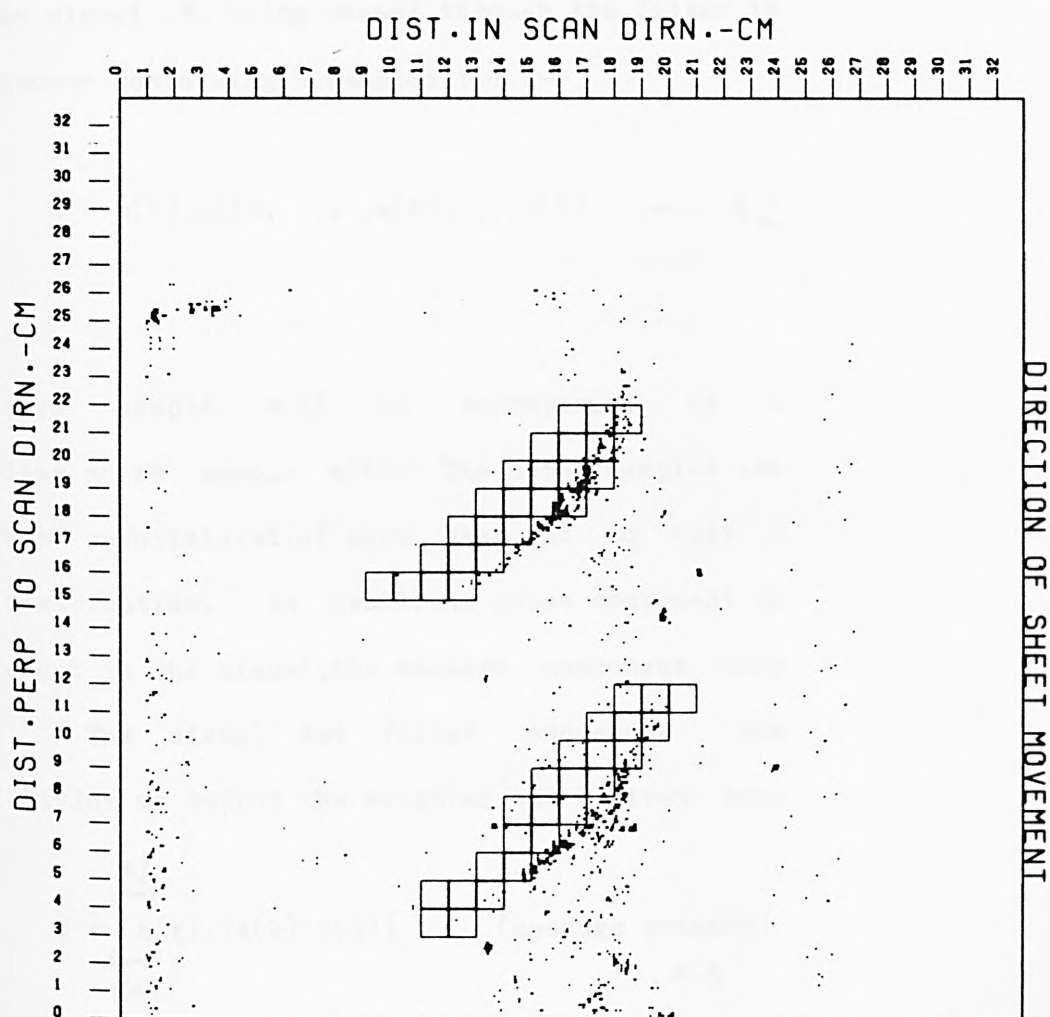


FIGURE 4.7 PINCH MARKS SENSED WITH STANDARD PROCESSING

COLD ROLLED STEEL STRIP



SHEET:- E16.26

DEFECT TYPE:- PINCH MARK

PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.12

DETECTOR:-SPECULAR

A non-recursive filter, H , (and ANY filter may be implemented non-recursively, see fig.4.4), consists of a sequence of N weights, i.e.:-

$$H = h(1), h(2), \dots, h(k), \dots, h(N) \quad \dots 4.1$$

The message signal, M , being passed through the filter is also a sequence containing N samples, i.e.:-

$$M = m(1), m(2), \dots, m(k), \dots, m(N) \quad \dots 4.2$$

Each message sample $m(k)$ is accompanied by a corresponding noise sample $n(k)$. The noise samples are assumed to be uncorrelated, of zero mean, and to obey a Gaussian distribution. As usual, the noise component is always present in the signal, the message component only sometimes. The signal and filter sequences are convolved, giving as output the weighted sum S , given by:-

$$S = \sum_{k=1}^N h(k) \cdot \{m(k) + n(k)\} \quad \text{(message present)} \quad \dots 4.3$$

$$S' = \sum_{k=1}^N h(k) \cdot n(k) \quad (\text{message absent}) \quad 4.4$$

To find the optimal filter sequence H' i.e. that which maximises the probability of detection for given false alarm rate, we maximise the contrast parameter Z , defined by:-

$$Z = \{E(S) - E(S')\}^2 / E(S'^2) \quad \dots\dots 4.5$$

Z is considered as being measured when the sequences H and M coincide exactly, and $E(.)$ denotes expectation. It may be shown (Skolnik, 1962) that if the noise is both white and Gaussian, then maximising Z will maximise $P(d)$ at given $P(fa)$, but that this is not necessarily true if the noise, though white, is non-Gaussian.

Since the noise samples are independent and have zero mean, $E(S')$ is zero and $E(S)$ is given by :-

$$E(S) = \sum_{k=1}^N m(k) \cdot h(k) \quad \dots\dots\dots 4.6$$

Further,

$$E(S'^2) = s^2 \sum_{k=1}^N h^2(k) \dots\dots\dots 4.7$$

where s is the standard deviation of the noise.

Thus, the expression for Z becomes:-

$$Z = \left\{ \sum_{k=1}^N h(k) \cdot m(k) \right\}^2 / s^2 \sum_{k=1}^N h^2(k) \dots\dots 4.8$$

To determine values for the coefficients H' which maximise Z , we use the Schwartz inequality, which states in its discrete form (Ayres, 1962) that the squared length of the product of two vectors X and Y is equal to or less than the product of their lengths squared, and that equality is achieved only when one vector is a linearly scaled version of the other, i.e., they both point in the same direction in signal space (Appendix B). That is, $\|X \cdot Y\| = \|X\| \cdot \|Y\|$, with equality when $X = cY$, with c a scalar. Thus, dividing equation 4.8 by the RHS of equation 4.6, we obtain:-

$$Z/S = \left\{ \sum_{k=1}^N h(k) \cdot m(k) \right\}^2 / s^2 \sum_{k=1}^N m(k)^2 \sum_{k=1}^N h(k)^2 \quad \dots \quad 4.9$$

which is maximum when each filter sample $h(k)$ is some constant times the corresponding message sample $m(k)$. The height of the 'blip' due to a message, relative to the surrounding noise, is seen to be maximised by convolving the signal with a scaled replica of the message. The optimal or 'matched' filter is thus (for a real signal) identical to the message. It is, however, time-inverted (Skolnik, 1962), as is seen by recalling that the first stage in a convolution process is an inversion in time.

The analysis can be extended to cover the case when the noise is Gaussian but non-white. This involves passing the signal through a pre-whitening filter, which suppresses and enhances appropriate frequencies to make the noise white. The message is also modified by the pre-whitening filter, and the enhancing filter must be designed to match this modified message.

The magnitude of the contrast improvement consequent upon matched filtering is obtained by

substituting $m(k)$ for $h(k)$ in the formula for Z (eqn.4.8) giving:-

$$Z(\max) = \sum_{k=1}^N m(k)^2 / s^2 \dots\dots\dots 4.10$$

When filter and signal are matched exactly, their contrast is (Skolnik, 1962) $2E(0)/N$ where $E(0)$ is the message energy within the window of observation, and N the noise energy per cycle of bandwidth.

In surface inspection applications, the message sequence is generally not known explicitly, and may vary widely for the different classes of defect within a particular task, and even within a given defect class. Thus, an 'optimal' M cannot be found, and an estimate M' has to be used instead. This provides some gain in contrast, but the match is not exact. The performance achieved is not truly optimal. To determine the extent of the sub-optimality consequent upon mis-match, with the objective of selecting a 'best' waveform for the imperfect match, we extend the analysis as follows.

Let the contrast parameter correct for the message being filtered be Z , but only Z' with the sub-optimal

weighting sequence being used. Z' will be smaller than Z , in the ratio $D=Z'/Z$. Z' is obtained by substituting the sub-optimum sequence m' for the correct sequence m in equation 4.9. On making this substitution, we obtain for D the following expression:-

$$D = \frac{\sum_{k=1}^N \{m'(k) \cdot m(k)\}^{**2} / \sum_{k=1}^N \{m'(k)\}^{**2}}{\sum_{k=1}^N \{m(k) \cdot m(k)\}^{**2} / \sum_{k=1}^N \{m(k)\}^{**2}} \dots\dots\dots 4.11$$

and since the denominator equals $\{m(k)\}^{**2}$, we have for D :-

$$D = \frac{\sum_{k=1}^N \{m'(k) \cdot m(k)\}^2}{\sum_{k=1}^N \{m(k)\}^2} \cdot \frac{\sum_{k=1}^N \{m'(k)\}^2}{\sum_{k=1}^N \{m(k)\}^2} \dots\dots\dots 4.12$$

in which $m(k)$ is the sequence for which the filter is matched, and $m'(k)$ is the sequence actually being processed.

To obtain a value for D which is independent of the magnitudes of the message and filter vectors m and m' , these must be normalised by scaling such that their lengths both become unity. This involves dividing each

by quantities $m(0)$ and $m(0')$ respectively, given by:-

$$m(0) = \sum_{i=1}^N m(i)$$

and

$$m'(0) = \sum_{i=1}^N m'(i)$$

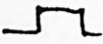
When D is computed with the scaled quantities, the filter mismatch can be examined as a function of shape only. It is effectively proportional to the cosine of the angle between the vectors M and M' in signal space (fig.4.8)

In the absence of a known and definite waveform for defect signals, it seems reasonable to choose as match waveform that having minimum mutual degradation of enhancement with all other waveforms. Table 4.4 shows mutual degradations between 5 waveforms which might be considered typical of defects, namely rectangular pulse, symmetric triangle, ramp, semicircle and one cycle of sinusoid. It was computed by substituting sequences of numbers representing the various waveforms in eqn.4.2, which were normalised before further computation. The sequences were each 21 samples long.

TABLE 4.1 DEGRADATION IN MATCHED FILTER GAIN

COMPARISON OF MATCH FOR DIFFERENT FILTERS

(the $\{i,j\}$ th entry is the match $\{D\}$ between waveforms i and j)

Filter Shape						Av. MATCH
						
	1	0.782	0.767	0.863	0	0.682
	0.782	1	0.600	0.950	0	0.664
	0.767	0.600	1	0.662	0.120	0.632
	0.863	0.950	0.662	1	0	0.695
	0	0	0.120	0	1	0.224

An average match can be computed by summing the row entries in table 4.1, and selecting the waveshape for which the sum is MAXIMUM, under the assumption that the various waveshapes are equally likely. The numbers

obtained show the semicircular pulse to give best performance, but this is only slightly better than the symmetric triangle and the rectangular pulse, and since the match waveforms suggested are somewhat arbitrary anyway, the difference cannot be considered significant. However, the rectangular pulse is much easier to compute than the symmetric triangle and semicircular pulses (since all multiplications in the convolution are by unity), and must thus be preferred in practical application.

The 'single cycle of sinewave' is orthogonal (appendix 'B') to the symmetric match waveforms (1), (2) and (4), and therefore generates a zero output from these filters at coincidence. Only for the asymmetric filter (3) is a finite output obtained. It is possible that some defect signals (for which the amplitude goes both above and below zero within the window of observation) are suppressed rather than enhanced using the symmetric filters used thus far in the simulation, and it might prove profitable to try some asymmetric forms.

The properties of the matched filter in its signal space representation are illustrated in figure 4.8. For

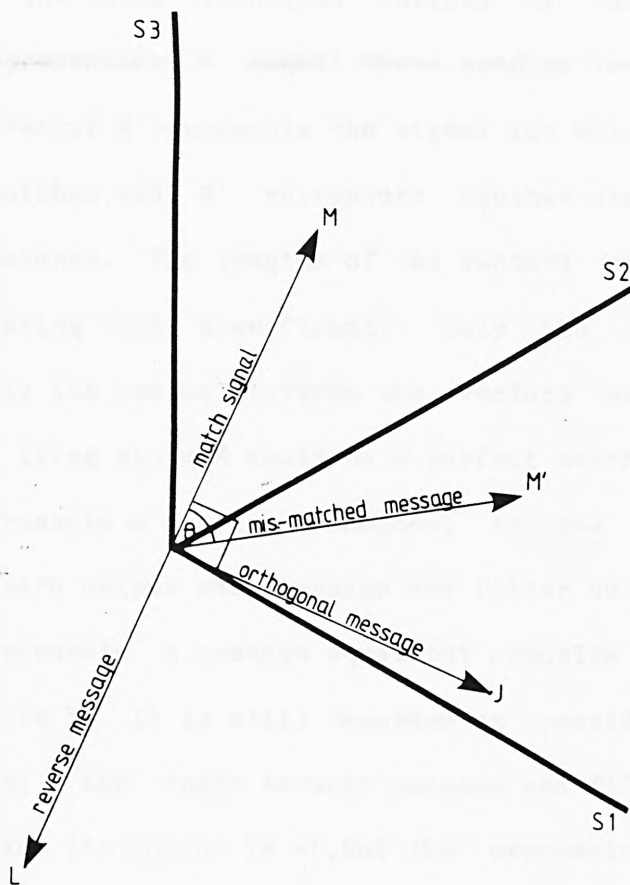


FIGURE 4.8 MATCHED FILTER DESCRIBED IN SIGNAL SPACE

clarity, only the three dimensions defined by the axes s_1, s_2, s_3 (representing a signal three samples long) are shown. The vector M represents the signal for which the filter is matched, and M' represents another signal, in general mismatched. The lengths of the vectors are for matched filtering not significant. Only the angle θ (more properly its cosine) between the vectors matters. Thus, any M' lying along M would be a perfect match. The vector J represents a message orthogonal to the filter M , which has zero output when message and filter coincide. Vector L represents a message equal but opposite to the filter waveform M . It is still regarded as providing a perfect match; the angle between message and filter is 180 degrees and its cosine is -1 , but the expression for contrast, Z , (eqn. 4.9) incorporates a squaring and is hence insensitive to sign. The proper interpretation is that M and L are perfectly matched, provided a negative going excursion is sensed in the decision stage.

If the waveshapes of the filters in a Matched Filter Bank are chosen to form an orthogonal set (App. 'B'), then each distinct signal waveshape will generate a unique vector of output amplitudes at coincidence.

The signals encountered in surface defect detection are generally two-dimensional, thus a two-dimensional matched filter should give better enhancement than a one-dimensional filter. The theory and signal space interpretation of such filters is a simple extension of that for one dimension (an explicit description is provided in (Norton-Wayne,1980a)),in which each sample requires two 'order' indicies for its specification,instead of one. The more samples used to define the match waveform,the more difficult it becomes in practice to provide a match to a given mean error.

To demonstrate by simulation the effectiveness of matched filtering for contrast enhancement, the defects seams ,sticker wrench and point wrench were examined. The first two could not be sensed even by lowering the detection threshold. Point wrench,though sensible in this way,required a high false alarm rate. All three kinds of defect have a distinct and characteristic shape and size in two dimensions. Thus,use of two-dimensional matched filtering seemed appropriate for these defects.

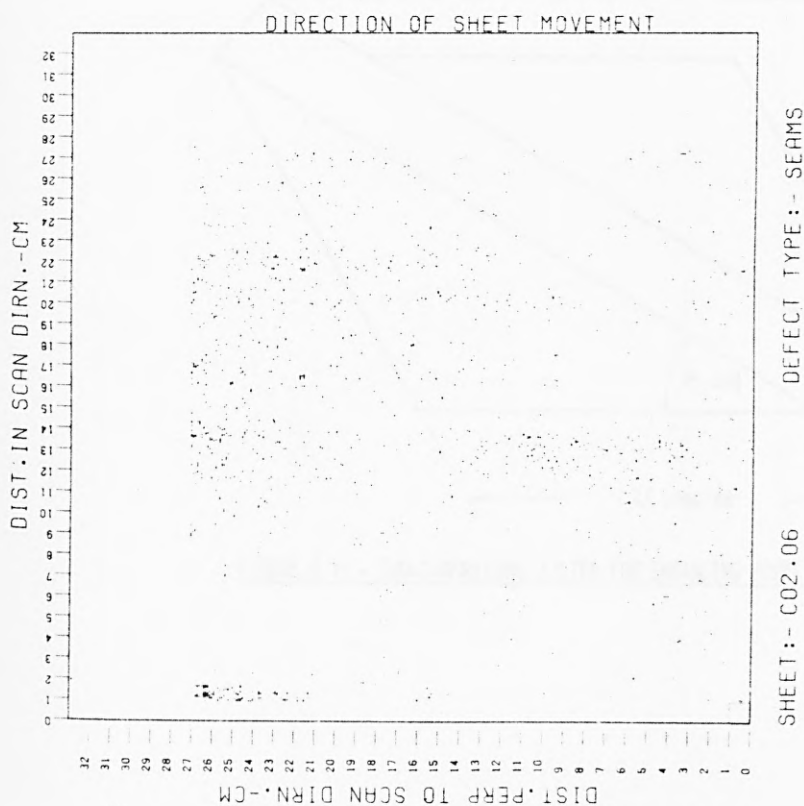
Seams take the form of lines which are low contrast absorbers,about 1mm wide, extending in the direction of

sheet movement. The standard SIRA detector when used alone cannot sense seams (fig.4.9). A two-dimensional filter comprising a rectangle 1cm. long by 1mm. wide was used (fig.4.10), producing the result shown in figure 4.12. Using a filter triangular in the x (scan) direction and rectangular in the y-direction (shown dotted in figure 4.10), produced no perceptable improvement in contrast, confirming the prediction that the enhancement obtainable by matched filtering is largely unaltered by small changes in the match waveform.

The enhancement produced by the two-dimensional filtering was in fact so effective that seams too faint to be perceived by the human eye were made detectable, confirming the prediction made in chapter 1 section 6, that visual inspection by machine should ultimately prove superior to human inspection, even when the human operative performs under ideal conditions of low speed, zero fatigue and perfect competence.

The approach was further used successfully to enhance point wrench, the filter being a bar 1cm. wide along the diagonal of a 1cm. square (figure 4.11) using the original insensitive threshold setting.

COLD ROLLED STEEL STRIP



COLD ROLLED STEEL STRIP

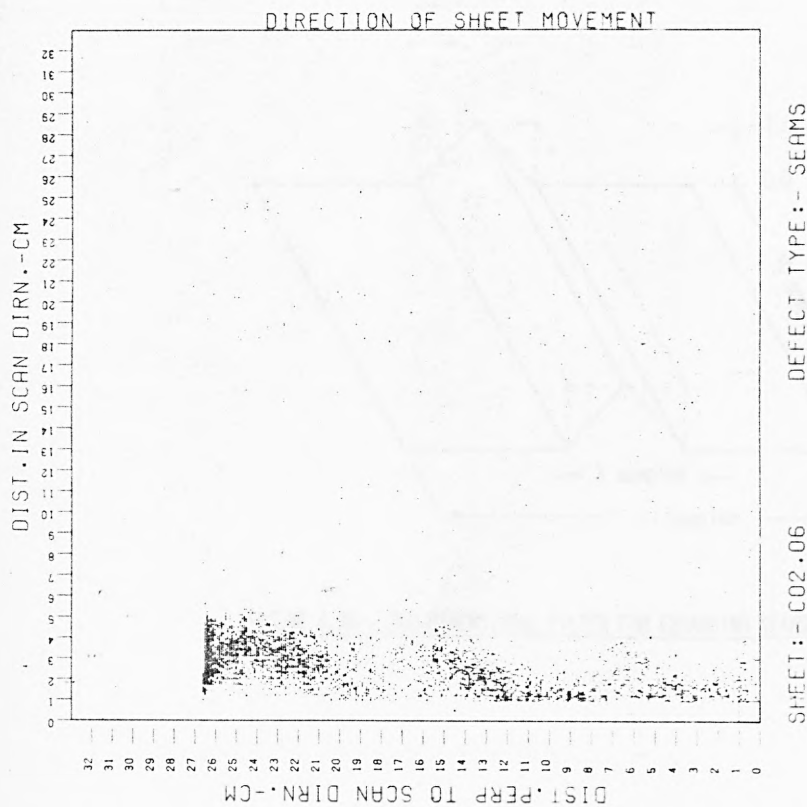


FIGURE 4.9 SEAMS UNDETECTED USING SIRA DETECTOR

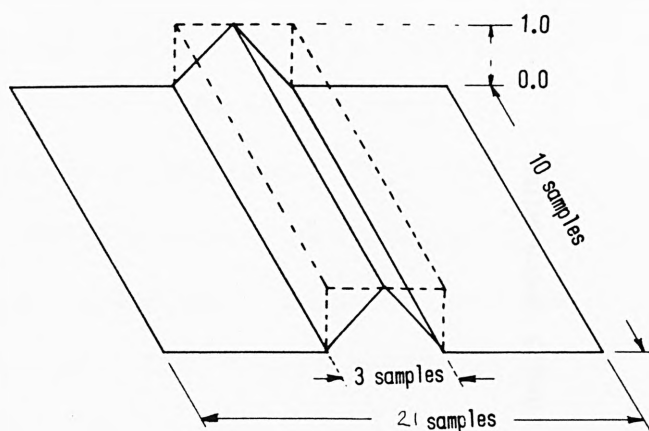


FIGURE 4.10 - TWO-DIMENSIONAL FILTER FOR ENHANCING SEAMS

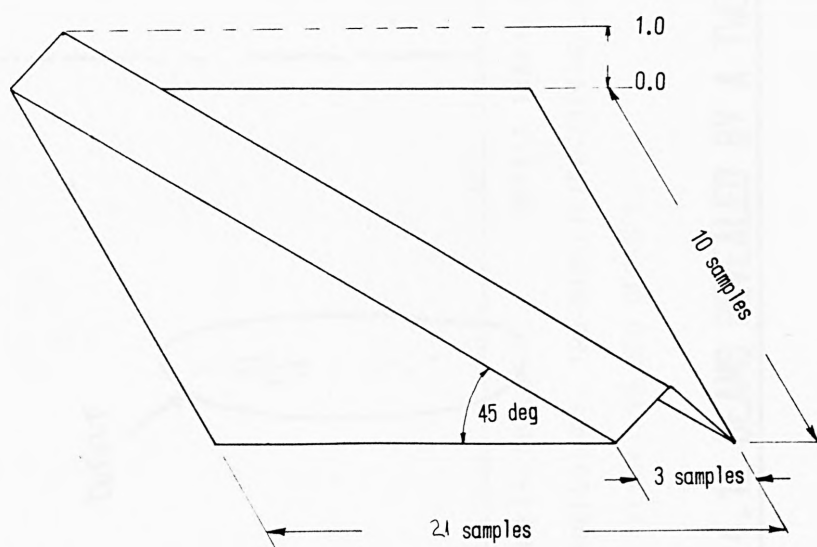


FIGURE 4.11 - TWO-DIMENSIONAL FILTER FOR ENHANCING POINT WRENCH

COLD ROLLED STEEL STRIP

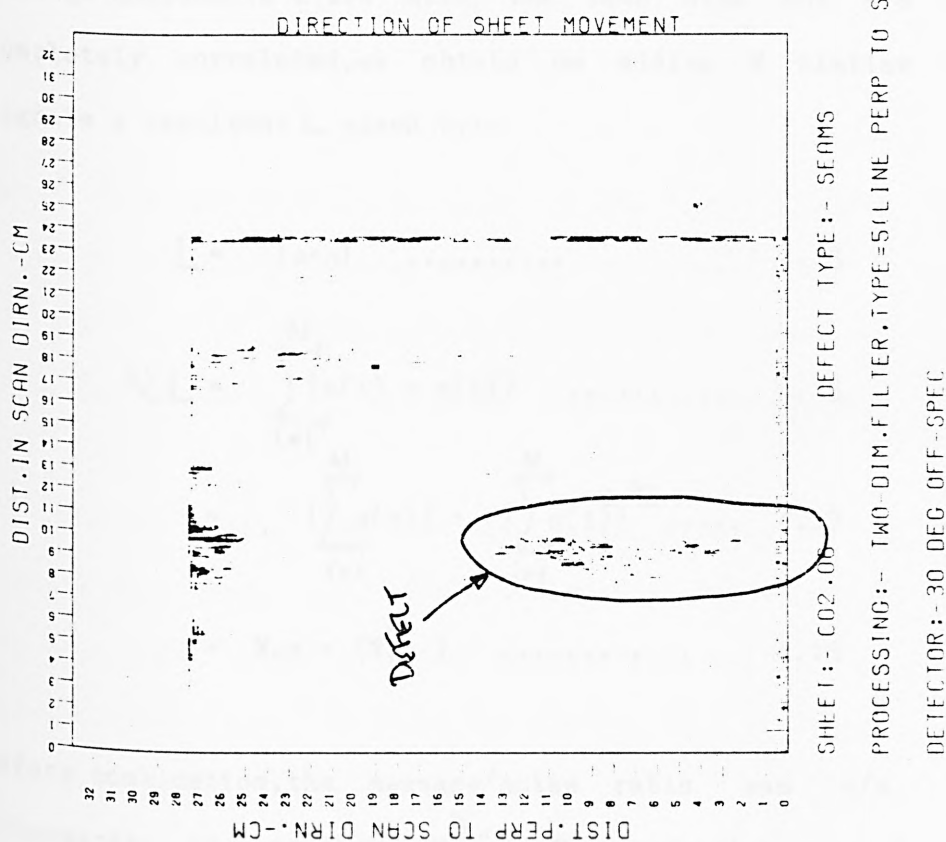


FIGURE 4.12 SEAMS REVEALED BY A TWO-DIMENSIONAL MATCHED FILTER

4.4 Contrast Enhancement Using Channel Combination

Another approach for improving message contrast is to form a linearly weighted sum of the signal components and to treat this as a scalar signal. The theory of this form of enhancement is discussed below. In the analysis, the noise components, $n(i)$, are assumed Gaussian, and of zero mean. For the simplest case in which the noise components of the signals being combined have the same amplitude and are completely uncorrelated, and the message components m are also the same size but are completely correlated, we obtain on adding N similar signals a resultant L given by:-

$$L = \{m+n\} \dots\dots\dots 4.15$$

$$N.L = \sum_{i=1}^N \{m(i) + n(i)\} \dots\dots\dots 4.16$$

$$= \left\{ \sum_{i=1}^N m(i) \right\} + \left\{ \sum_{i=1}^N n(i)^2 \right\}^{1/2} \dots\dots 4.17$$

$$= N.m + (N.n)^{1/2} \dots\dots\dots 4.18$$

Before combination, the message/noise ratio was m/n , afterwards it becomes $m/(N.n)^{1/2}$. Thus, combining the N

similar signals improves the message/noise power ratio by a factor N , and the amplitude ratio by $(N)^{1/2}$

In practice, the noise components are almost invariably mutually correlated, at least partially, and the signals have different amplitudes. In this case, the variance of the signal resulting from the linear summation of N mutually partially correlated signals is given by (Hoel, 1965):-

$$S_N^2 = \sum_{i=1}^N a(i)^2 \cdot S(i)^2 + 2 \sum_{i < j}^N \sum_{j=1}^N a(i) \cdot a(j) \cdot r_{ij} \cdot S(i) \cdot S(j) \quad \dots\dots 4.19$$

Here, $a(i)$ is the relative amplitude of the i th. signal, $r(i,j)$ is the coefficient of convolution between the i th. and j th. signals, and $S(i), S(j)$ are the standard deviations of the i th. and j th. signals, respectively.

For the simple case when the signals have equal standard deviation i.e., $S(1) = S(2) = S(3) = \dots = S$, have unit amplitude (i.e. $a(1) = a(2) = a(3) = \dots = 1$), and equal mutual correlations ($r(1,2) = r(1,3) = r(2,3) = \dots = r$), the expression for $S(N)$ reduces to:-

$$S^2(N) = N.S^2 + N(N-1)S^2r \quad \dots \quad 4.20$$

$$S^2(N) = N.S^2 \{1 + (N-1)r\} \quad \dots \quad 4.21$$

IF the contrast before combination was m/S , it will increase after combination to:-

$$N.m/N^{1/2} \{1 + (N-1)r\}^{1/2} . S \quad \dots \quad 4.22$$

As expected, this expression increases to the upper bound $N.m/S^{1/2}$ when $r=0$, and decreases to m/S (no improvement) when the noise components are completely correlated and $r=1$.

The question next arises, given a set of signals for which the correlations between the noise components are known, can a set of weights be chosen which will maximise the gain in contrast consequent upon combination, and, if so, then what are the weights.

The problem is illustrated in figure 4.13. This is a scatter diagram showing the resultant (vector) signal formed from two mutually partially correlated scalar signals $s(1)$ and $s(2)$. The noise component

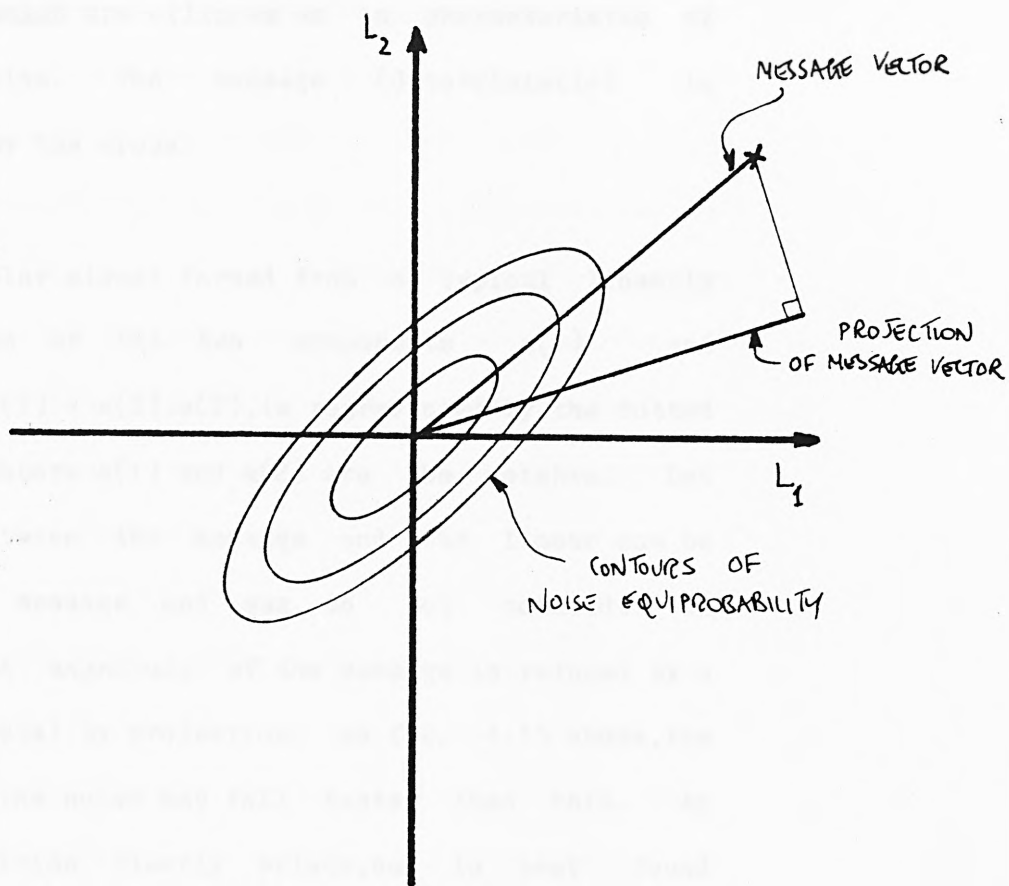


FIGURE 4.13 CONTRAST MAXIMISATION USING WEIGHTED
LINEAR COMBINATION OF SIGNAL COMPONENTS

(stochastic) is shown by the shaded contours of equal probability, which are ellipses as is characteristic of Gaussian noise. The message (deterministic) is represented by the cross.

The scalar signal formed from a typical linearly weighted sum of the two components $s(1)$ and $s(2)$, $S = a(1).s(1) + a(2).s(2)$, is represented by the dotted line. The numbers $a(1)$ and $a(2)$ are the weights. Let the angle between the message and the linear sum be θ . When message and sum do not coincide in direction, the magnitude of the message is reduced by a factor $\cos(\theta)$ by projection. As fig. 4.13 shows, the magnitude of the noise may fall faster than this. An optimum position clearly exists, but is best found computationally, or by trial.

The related question as to whether better performance could have been obtained by taking decisions separately for each signal component, then combining the outcomes logically, is dealt with in chapter 6.

One particular kind of defect considered in the

simulation analysis is particularly challenging in that application of none of the methods described prior to section 4.4 enabled the defect even to be sensed. This defect is sticker wrench, caused by the laps of coil sticking together during annealing. The surface is damaged when the laps are dragged apart during uncoiling, but the physical nature of the modification to the surface, and its effect on the scanner signal, are uncertain and appear to vary considerably for different occurrences of the defect.

The correlation properties of the noise have been measured for this problem (chap.2, section 4), but the direction of the message vector is unknown, and very difficult to determine. To enable this method of contrast enhancement to be used in the simulation investigation, the following procedure has been used. First, the best possible guess as to the direction of the message vector is made, guided by available extraneous information. The correct solution is then found iteratively by a series of trials, in which the vector of combination weights is varied along its three axes, and the position giving optimal performance hence determined.

This approach was applied in simulation to sticker wrench. It was believed initially that this deflects the incident light, causing the specular signal to fall, and both off-specular signals to rise. The nature of this signal is shown as vector (1) in figure 4.14 which assumes for simplicity that the signal is deterministic.

The linear combination optimal for increasing contrast is of the form:-

$$C(\text{opt}) = a(1).s(1) + a(2).s(2) + a(3).s(3) \dots$$

... 4.23

in which the coefficients $a(1)$, $a(2)$ and $a(3)$ are to be determined. Since we are concerned only with the direction of the defect message in signal space (otherwise the combination of components will produce a gain or attenuation), we should to be rigorous have $a(1)^2 + a(2)^2 + a(3)^2 = 1$. However, this variation is taken care of in the adjustment of the threshold. The hypothesis originated by workers at Hooghovens (De Jonge, 1979) that $a(1)$ should be negative, and $a(2)$ and $a(3)$ positive accords with their photograph of sticker wrench (fig. 4.19) but does not supply magnitudes for the

FIGURE 4.14 DETECTION OF STICKER WRENCH ANALYSIS
IN SIGNAL SPACE

coefficients.

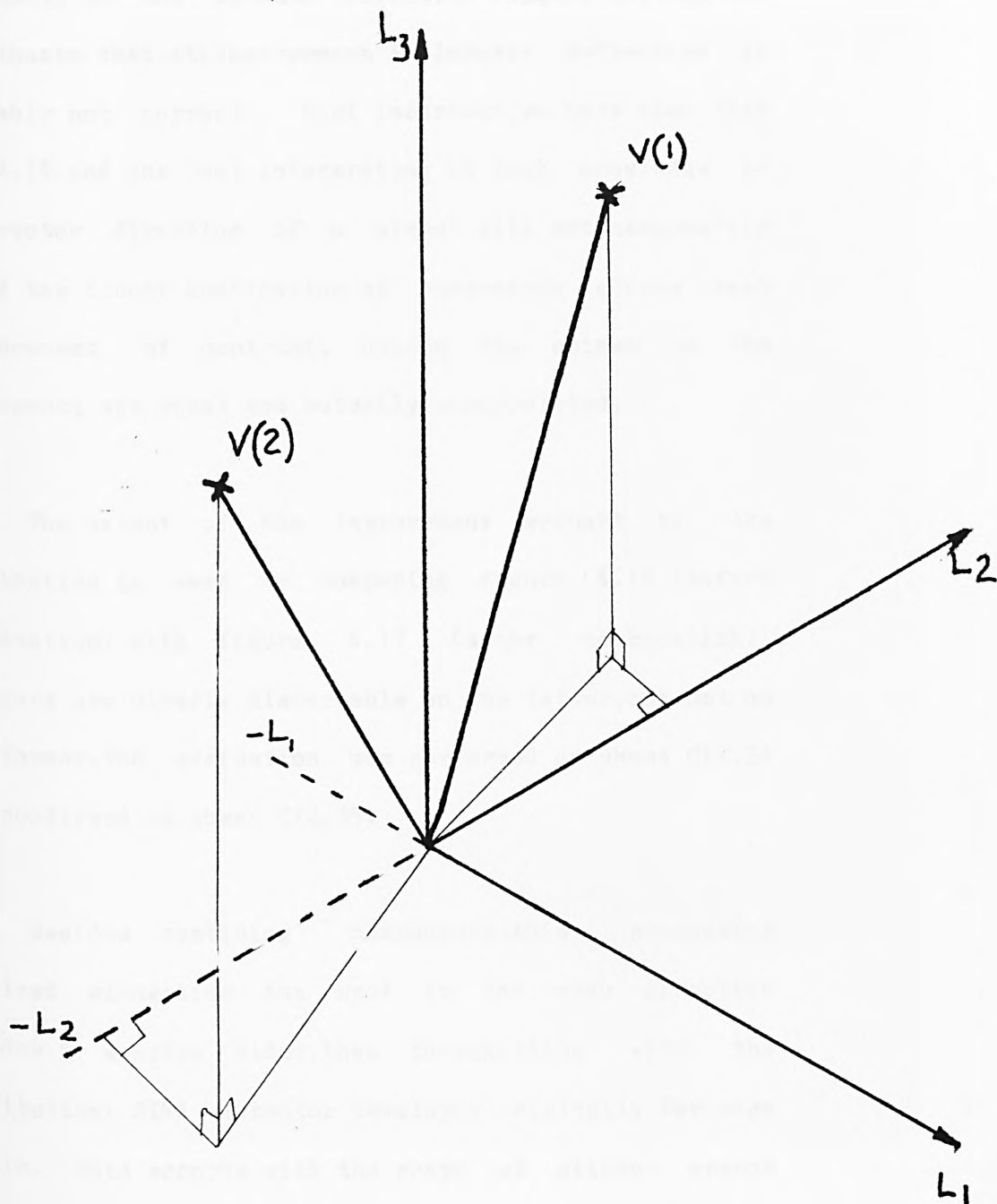
Several combinations were examined, and the optimal values for the weights were found empirically to be:-

$$a(1)=+.80 \quad a(2)=-.48 \quad a(3)=+.60$$

To establish that the values shown were indeed optimal, the chosen vector was 'rocked', (i.e., small increments were applied individually in each direction to the weights), and performance was seen not to be improved. The location of the optimum was not critical. Variation by + or - 0.1 in each coefficient produced little change in contrast.

Thus, the true position of the signal vector appears to be (2) in figure 4.14. The signal rises on both specular and 30deg.off-specular channels, and falls on the 7deg.off-specular channel. This does not accord with the Hooghovens prediction; the discrepancy is probably because if, as Hooghovens suggest, sticker wrench is a deflecting defect, the sensor must lie JUST OUTSIDE the mainlobe if it is to exploit this deflection. But visual examination of stickers often shows an INCREASE in

FIGURE 4.14 DETECTION OF STICKER WRENCH ANALYSED
IN SIGNAL SPACE



intensity in the specular direction (fig.4.15), thus the hypothesis that sticker wrench is largely deflecting is probably not correct. Most important, we have seen from fig.4.13 and the text interpreting it that knowledge of the vector direction of a signal will not necessarily yield the linear combination of components giving best enhancement of contrast, unless the noises in the components are equal and mutually uncorrelated.

The extent of the improvement wrought by the combination is seen by comparing figure 4.16 (before combination) with figure 4.17 (after combination). Stickers are clearly discernable on the latter, but not on the former. The evaluation was performed on sheet C12.24 and confirmed on sheet C12.35.

Besides combining components, this processing involved elongating the spot in the scan direction (window 5 samples wide), then thresholding with the longitudinal SIRA detector developed originally for edge strain. This accords with the shape of sticker wrench marks, as shown in figure 4.15. A block diagram of the processing sequence used is shown in fig.4.18.

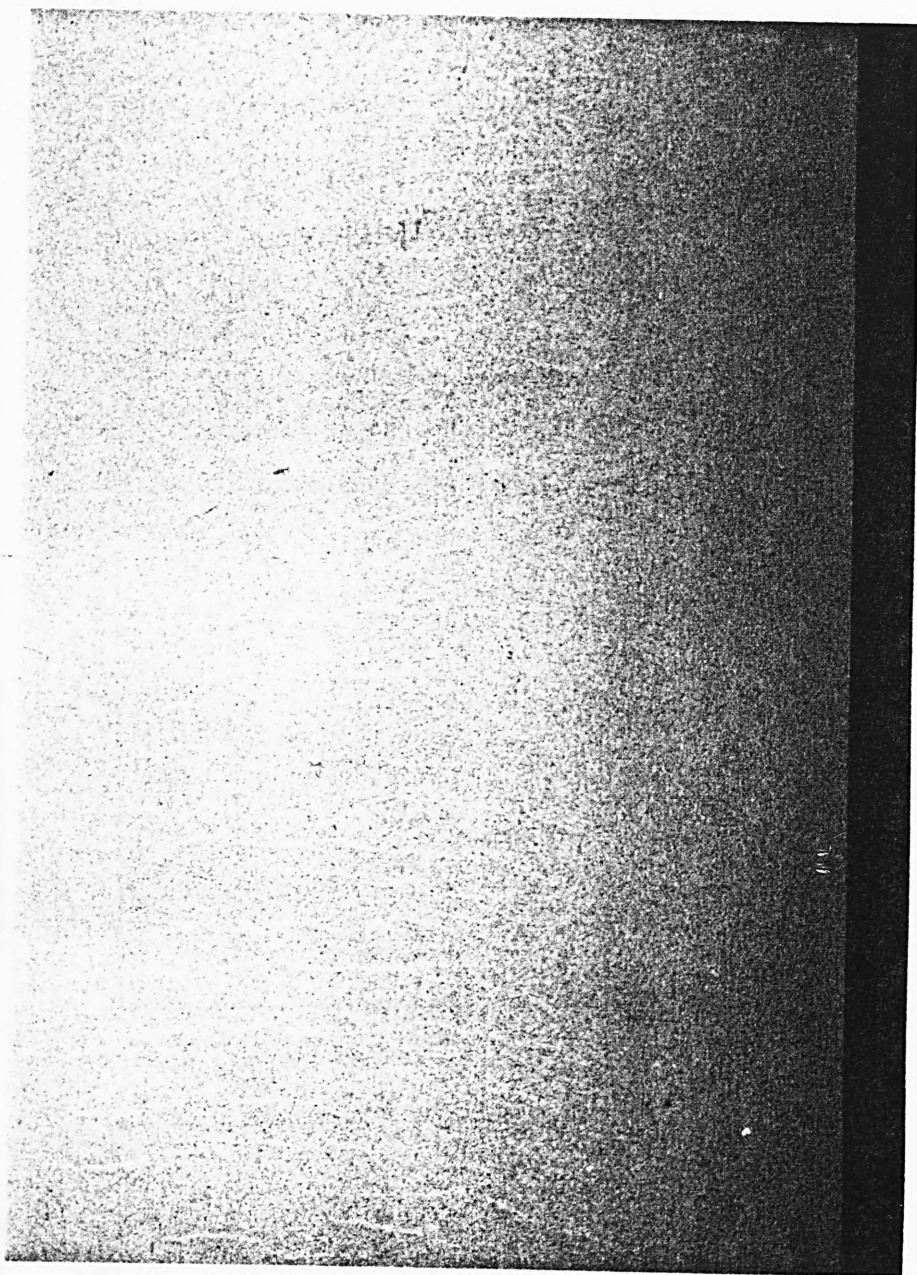
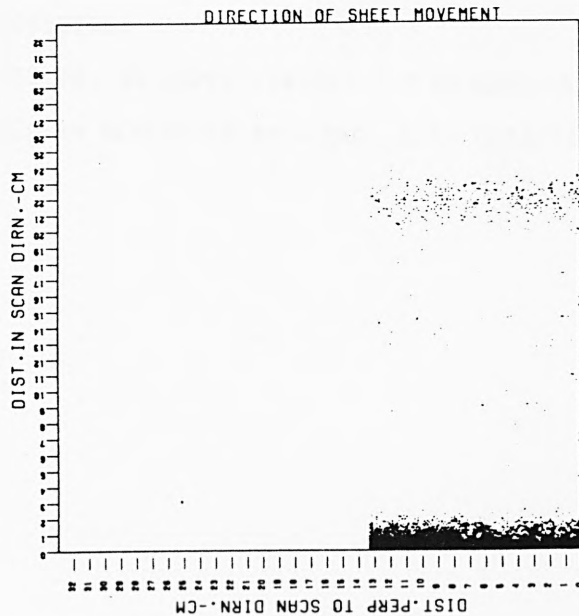


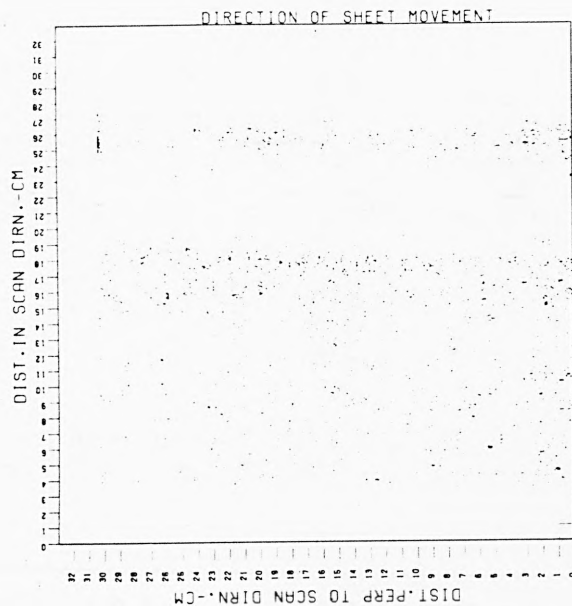
FIGURE 4.15 STICKER WRENCH - NORMAL CONTRAST
(source - British Steel Corporation)

COLD ROLLED STEEL STRIP



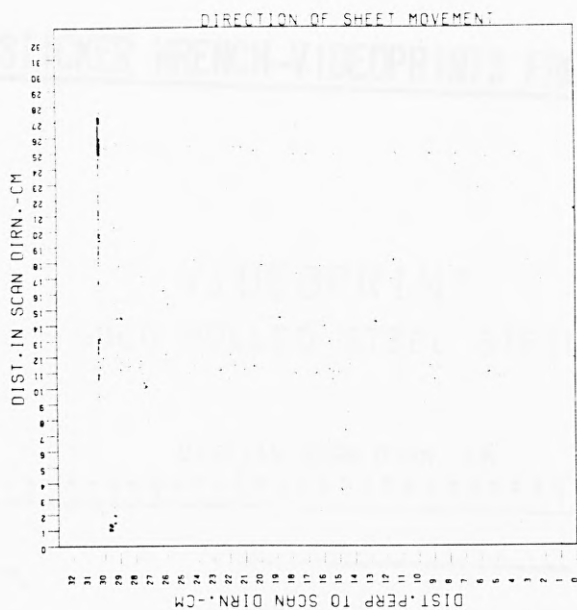
SHEET:- C12.24 DEFECT TYPE:- STICKERWARE
 PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=0.92
 DETECTOR:-SPECTULAR

COLD ROLLED STEEL STRIP



SHEET:- C12.24 DEFECT TYPE:- STICKERWARE
 PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.05
 DETECTOR:-5 DEG OFF-SPEC

COLD ROLLED STEEL STRIP



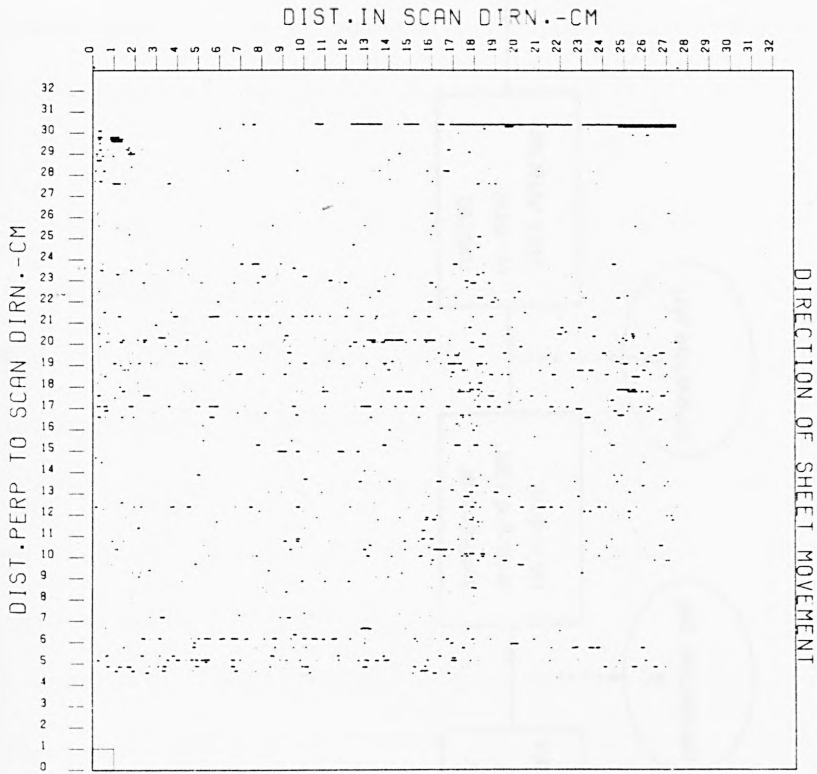
SHEET:- C12.24 DEFECT TYPE:- STICKERWARE
 PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.05
 DETECTOR:-30 DEG OFF-SPEC

FIGURE 4.16 STICKER WRENCH-VIDEOPRINTS USING STANDARD DETECTOR

(no trace of the defect is visible)

FIGURE 4.17 STICKER WRENCH-VIDEOPRINTS FROM OPTIMALLY PROCESSED
SIGNAL

VIDEOPRINT
COLD ROLLED STEEL STRIP



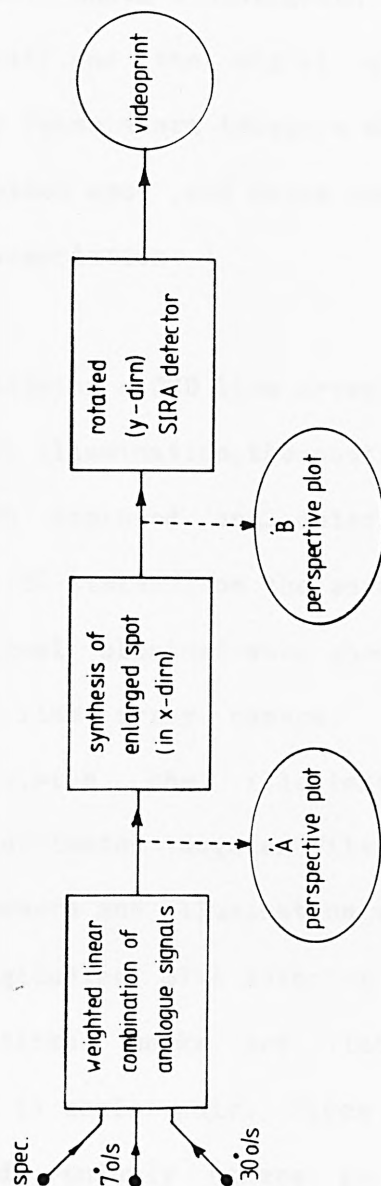


FIG 4.18 PROCESSING FOR DETECTION OF STICKER WRENCH

Despite this progress, successful detection of sticker wrench is not claimed. The defect varies too much in its properties (fig.4.19 shows a photograph of a sample giving reversed contrast), and the signal after thresholding contains too many false alarm triggers which are CLUSTERED due to the elongated spot, and hence cannot easily be removed by trigger association.

If, however, a camera containing a CCD line array is used as sensor, with incoherent illumination, the contrast of the sticker marks is much improved and detection becomes much easier. Figure 4.20 {taken from the work of Yaxley (1979)} shows the signal obtained when sheet C 12.24 was examined using a CCD line array camera. The sheet was viewed vertically, with the illumination incident obliquely from an uncollimated tungsten filament lamp. The exact position of camera and illumination were found to be uncritical. A longitudinal SIRA detector was used to generate triggers. Sticker marks are visible clearly, although the signal is scalar only. Since the noise triggers are distributed randomly (there is no spatial filtering) the trigger association to be described in chapter 5 should have no difficulty in disposing of them.

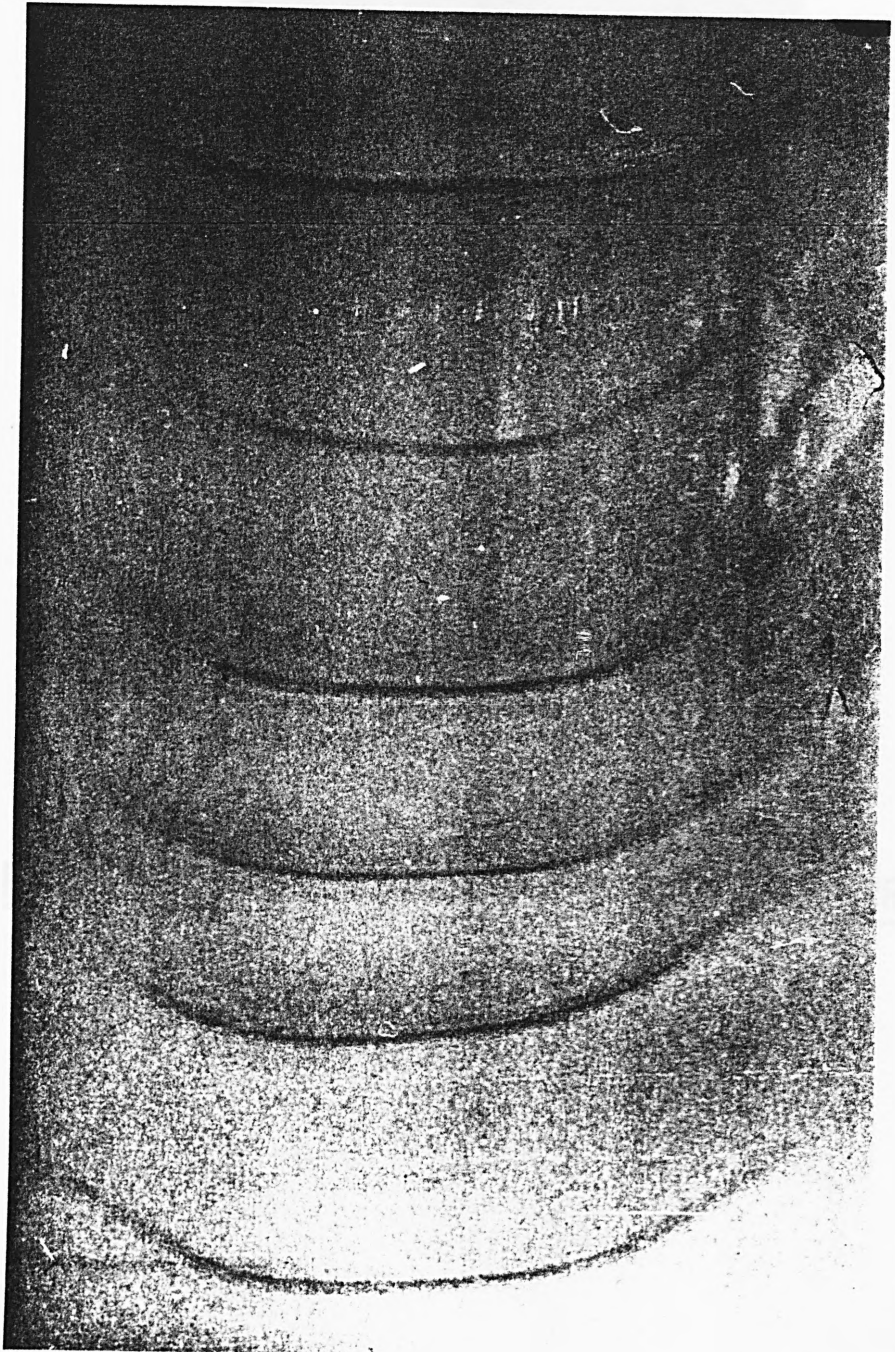


FIGURE 4.19 STICKER WRENCH – REVERSED CONTRAST
(source – Hooghovens)

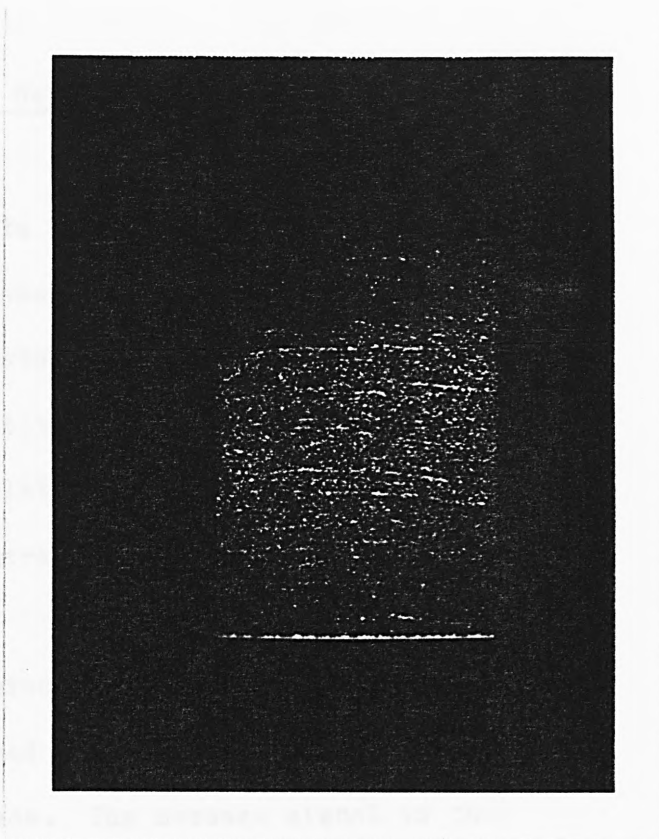


FIGURE 4.20 STICKER WRENCH - SENSED BY CCD ARRAY

(sheet C12.24; from Yaxley, 1980)

4.5 Detection of Non-local Defects

Virtually all defects encountered in automated surface inspection are essentially LOCAL, in that the message energy is concentrated within a few samples of signal, which are close spatially. For these, the 'window of observation' must be relatively small ('matched' to the defect) to obtain good contrast.

However, sometimes defects are encountered in which the message energy is spread (often rather thinly) over a substantial area of surface. The message signal is then GLOBAL, and a somewhat different approach must be used for improving contrast. The method appropriate for enhancement of non-local defects is generally specific to the type of defect, and the provision of a general methodology for their enhancement is thus not possible. Instead, we shall consider one specific kind of non-local defect in detail. This is chatter marks, one of the difficult defects encountered during the simulation investigation on cold rolled steel strip.

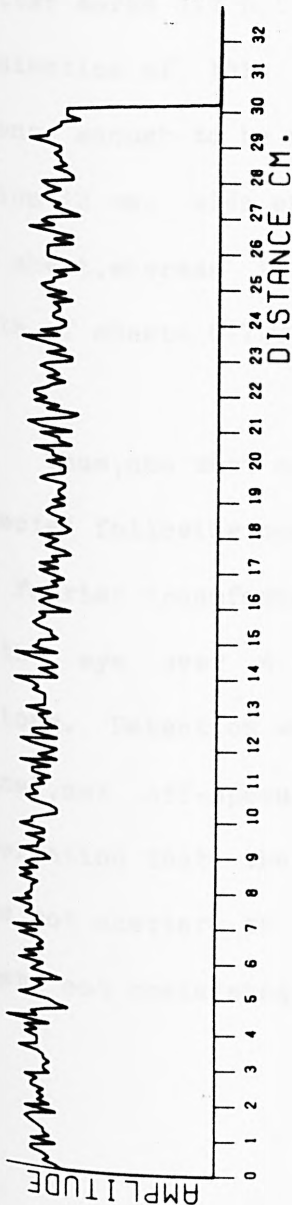
Chatter marks comprise bands alternately light and

dark, running along the scan direction. The contrast is poor and the intensity variation is not abrupt but is almost sinusoidal. However, the bands tend to be uniformly spaced. They have just sufficient contrast to be discernable to the human eye, and appear to absorb rather than to scatter or deflect.

Thus, the first stage in improving contrast is to average the sample amplitude within each scan. Since each scan contains 700 samples which are mutually independent (chap.2, sect.4), the contrast is thereby improved by a factor $(700)^{1/2}$, i.e. 26.5. The averaging produces a one-dimensional signal running down the sheet in the rolling direction. Figure 4.21(a) shows this signal plotted for sheet c17.04; despite the considerable increase in contrast due to the sample averaging, the bands are not discernable.

The periodicity is however almost certainly present, though still obscured by noise. The next step is thus to compute the power spectrum of this signal, which is the squared modulus of its Discrete Fourier Transform. This was performed using an FFT algorithm provided as a library subroutine. The theory behind this processing is

AVERAGED SCAN AMPLITUDE



POWER SPECTRUM

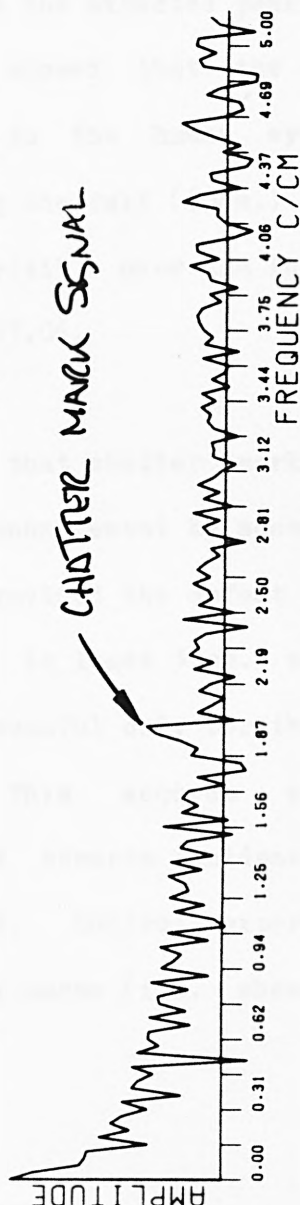


FIGURE 4.21 CHATTER MARKS DETECTED
USING SPECTRAL ANALYSIS

SHEET:-- C17.04

DEFECT TYPE:-- CHATTERMKS

PROCESSING:-- NONE

DETECTOR:--SPECTULAR

explained in appendix 'B'. The power spectrum is shown plotted (on a logarithmic scale) in figure 4.21(b). A peak in the signal power is clearly visible at a spatial frequency about 2 cycles/cm., which corresponds nicely to the measured wavelength of the chatter marks which is about 0.5cm. The initial measurement on sheet C17.04 was confirmed by a second experiment on sheet C17.06, but a further experiment on sheet C17.02 which also contains chatter marks did not produce the expected peak. Visual examination of this sheet showed that the defect was intense enough to be visible to the human eye over a region 12 cm. wide extending the full (30cm.) length of the sheet, whereas it was visible over the full (30cm.) width of sheets C17.04 and C17.06.

Thus, one must conclude that chatter marks can be detected following contrast enhancement by scan averaging and Fourier transformation, provided the defect is visible to the eye over a region at least 30cm. wide and 30 cm. long. Detection was successful only for the specular signal, not off-specular. This accords with the observation that the defect absorbs incident light, but does not scatter or deflect. Control experiments on sheets not containing chatter marks (i.e. sheets C12.24,

AVERAGED SCAN AMPLITUDE

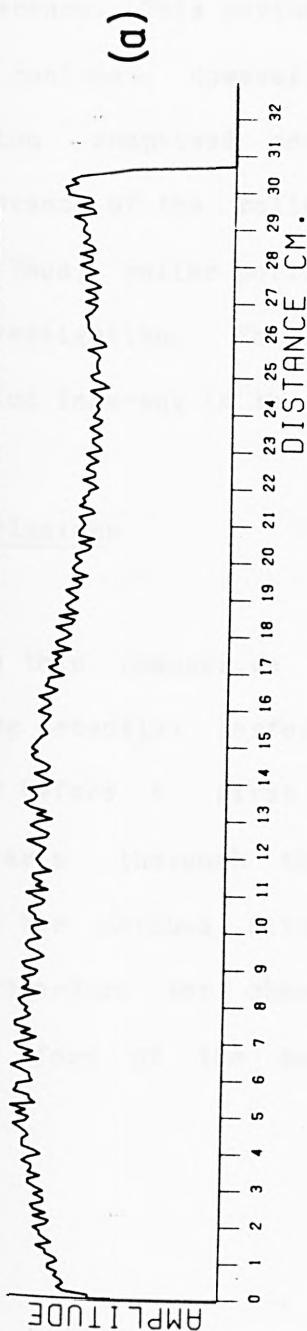
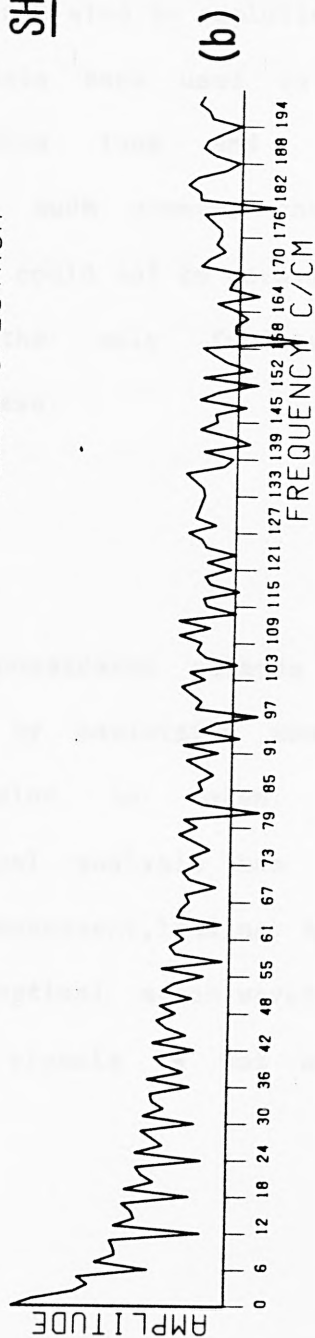


FIGURE 4.22 SPECTRAL ANALYSIS OF
SHEET WITHOUT CHATTER MARKS

POWER SPECTRUM



SHEET:- C99.10

DEFECT TYPE:- PRIMESHEET

PROCESSING:- NONE

DETECTOR:- SPECULAR

C13.10 ,C23.25, and C99.10 {fig.4.22}) gave a spectrum without significant isolated peaks.

Another kind of non-local defect is introduced during rolling by imperfections on the surfaces of the rollers. These generate marks on the rolled surface which recur with a periodicity equal to the roller circumference. This periodicity can also be exploited to enhance contrast. However, the data base used in the simulation comprised sheets 30cm long and the circumference of the rollers is much greater than 1 metre. Thus, roller periodicity could not be covered in this investigation. This is the only fundamental limitation inherent in the data base.

4.6 Conclusions

In this chapter, we have considered methods for improving detection performance by maximising message contrast before a first decision is taken. In particular, a thorough theoretical analysis has been provided for matched filter enhancement, leading to a sound procedure for choosing optimal match waveforms when the form of the message signals is not known

explicitly. The use of a linear weighted combination of the components of the vector signal has been investigated, and an experimental procedure demonstrated which has been shown to give a worthwhile increase in defect contrast. A procedure has been described for implementing the enhancement (i.e., for determining optimal values for parameters, given sample data for the specific problem). Operating on summed signal components has however certain disadvantages, to be discussed in chapter 6, section 2.

Simulation analyses have been presented which demonstrate the effectiveness of both matched filtering and component combination in increasing the detectability of appropriate defects. Finally, enhancement for non-local defects has been considered, and a methodology has been proposed, analysed theoretically and demonstrated using computational simulation, for one specific kind of defect since a general solution is for this problem believed to be impossible.

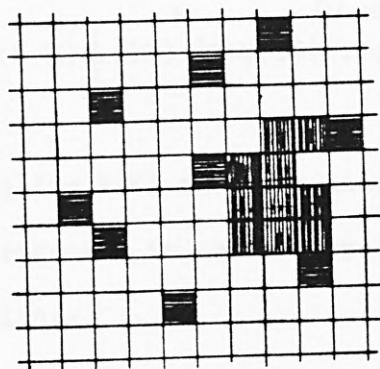
CHAP.5 ELIMINATION OF FALSE ALARM TRIGGERS USING ASSOCIATION

5.1 INTRODUCTION

It is commonly found in surface defect detection problems that the contrast of defect messages with respect to accompanying noise is very poor. Thus, although use of a low detection threshold proceeded perhaps by contrast enhancement can enable many defects having poor contrast to be sensed, the number of false alarm triggers generated is then so high that detection cannot be claimed. We now discuss some non-linear filters for two-dimensional binary signals, which eliminate false alarm triggers but retain triggers arising from defects. They exploit the property that triggers due to false alarms are distributed uniformly and at random over the surface, whereas triggers due to defects occur in clusters which, though differing considerably in shape with defect type, are relatively closely packed. The role of trigger association in an overall detection system is seen from the canonic form shown in figure 3.2.

The filters operate on a common principle, illustrated by figure 5.1. This shows the characteristic distributions of triggers due to noise (horizontal hatching) and due to defect messages (vertical hatching). It is assumed that the noise is white (i.e., that joint probabilities of all orders except the first are zero), and, further, have zero mean and constant variance. This has the consequence that the probability $p(0)$ that the signal within a given window of observation exceeds a particular threshold and generates a trigger pulse is equal for all windows, and is independent of the state of the signal within all other windows. For each cell representing a window of observation, the probability $p(0)$ is small (in practice, not greater than about 1 in 1000), so the probability of noise triggers occupying adjacent cells is then very small indeed, i.e. less than 3 in one million.

The probability of large clusters of triggered cells occurring by chance is correspondingly minute. For a selection of N cells, the probability $p(K)$ that K ($K < N$) of the cells simultaneously contain triggers due to noise is given by the binomial distribution:-



cells triggered by noise - 
 cells triggered by defects - 

FIGURE 5.1 PRINCIPLE OF NOISE ELIMINATION
 USING TRIGGER ASSOCIATION

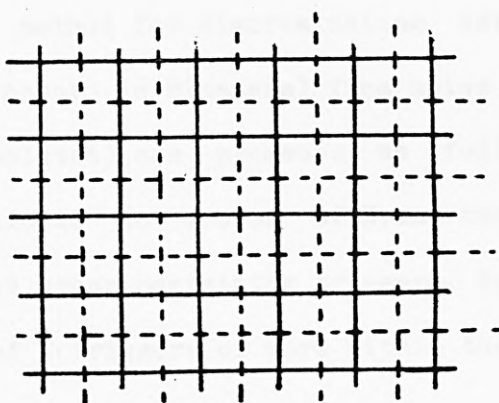


FIGURE 5.2 STAGGERED ASSOCIATION REGIONS

$$p(K) = \{N!/K!(N-K)!\} \cdot p(0)^K \cdot \{1-p(0)\}^{(N-K)} \dots\dots\dots 5.1$$

Further, the probability $P(K)$ that the selection contain K OR MORE noise triggers is given by the cumulative binomial distribution:-

$$P(K) = \sum_{L=K}^N \{N!/L!(N-L)!\} \cdot p(0)^L \cdot \{1-p(0)\}^{(N-L)} \dots\dots\dots 5.2$$

Because of the independence assumption, $p(K)$ and $P(K)$ depend only on $p(0)$, K and N , and are independent of the relative positions in the surface of the various cells. For reasonable values of $p(0)$, K and N , $P(K)$ becomes very small compared with $p(0)$.

To use the method for discriminating defects (for which triggers occur in clusters) from noise (for which triggers are isolated), one proceeds as follows. The cells are considered in groups of N , and the number of cells within each group containing triggers is counted. The occurrence of K triggers or more within the group due to noise alone is considered an event so rare than any such occurrence must be due to the presence of a defect. The number of cells N in the selection is chosen to be large enough that $P(K)$ is vanishingly small when noise

alone is present, with K selected such that the overwhelming majority of defects generate at least K triggers within the selection. The ratio $P(K)/p(0)$ is independent of the spatial distribution within the selection, but triggers due to defects tend to be closely packed. Thus without exact knowledge of the trigger cluster shapes due to defects, it is best to use selections of cells which are spatially compact. Design of a trigger association method then involves determination of an appropriate distribution and size N for the selection, of a value for the threshold parameter K , and of procedures for counting the triggers within the selection which can be implemented efficiently as processing algorithms for use in on-line instrumentation.

In this chapter, we describe three methods using the same general principle which can be used to reject noise triggers but retain message triggers. The methods are alternatives, in that they perform essentially the same function, but differ in the computation required, in sensitivity to message cluster shape, and in producing different degrees of distortion in message clusters as a penalty for rejecting the noise.

Parameters U and W describing the noise rejection and message distortion characteristics respectively for the alternative filters are introduced, and mathematical formulae are derived giving U and W for selected examples. Some computed numerical values for U and W are then presented so that quantitative comparisons can be drawn. Upper and lower bounds for the extent of the improvement in detectability for marginal defects are calculated, assuming extreme forms for the probability density function (i.e., Gaussian and negative exponential) of message and noise. Finally, the effectiveness of the methods in a particular practical application is demonstrated, and their performance compared, by simulation.

5.2 Association in a Box - Binary Matched Filter

Possibly the simplest way of eliminating isolated random triggers is as follows. We divide the surface into rectangular regions each I cells by J , and reject all triggers within a region unless the number present reaches some threshold K . If the 'per cell' probability of a trigger being generated due to noise alone is $p(0)$, then the probability $P(K)$ of K or more triggers

being present within the region due to noise alone is given by the cumulative binomial distribution:-

$$P(k) = \sum_{L=k}^N \{N!/L!(N-L)!\} \{p(0)\}^L \{1-p(0)\}^{N-L} \dots\dots\dots 5.3$$

in which N is the area of the region measured in cells, i.e. the product I*J

If the number of cells N in the region is large, and p(0) and the threshold K are small, the Poisson distribution is an adequate approximation to the binomial distribution. It is preferable computationally being simpler, (it is a function of one parameter instead of two.). Thus we have for equation 5.3 the alternative simpler expression:-

$$P(k) = \sum_{r=k}^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} \dots\dots\dots 5.4$$

Further economy in computation may be obtained by calculating Q(K) (= 1 - P(K)), which is given by:-

$$Q(k) = \sum_{r=0}^{k-1} \frac{\lambda^r e^{-\lambda}}{r!} \dots\dots\dots 5.5$$

In these equations, λ is the 'expected' number of triggers in the cluster, given by:-

$$\lambda = N \cdot p(0) \quad \dots\dots\dots 5.6$$

The binomial distribution has however been used for the computation of numerical values tabulated later in this chapter, since some of the combinations of N and p(0) considered do not satisfy the conditions under which the Poisson distribution is a good approximation to the binomial, and it would have been uneconomical to produce two sets of programs.

For typical values of 10 for each of I, J and K, P is about 10^{-15} . This assumes that the individual samples are independent. Any mutual dependence will reduce this figure.

Assuming that message triggers occur in clusters (possibly loosely packed) that contain at least 10, then we have a powerful method for eliminating noise triggers selectively. But we have also assumed that the message clusters lie wholly within one region. Particularly when

clusters form straight lines, the probability of their intersecting a region boundary and being lost is significant. Although this can be avoided to some extent by using a second set of association regions staggered with respect to the first,(fig.5.2) some distortions are unavoidable. Specification of the efficiency of the process requires two parameters. The first,the PROCESSING GAIN U,is defined by:-

$$U = \frac{\text{probability of trigger generation due to noise after association}}{\text{probability of trigger generation due to noise before association}}$$

on a 'per region' basis. For the numerical example just quoted,U is then given by:-

$$U = P(K)/N.p(O) \qquad \dots\dots\dots 5.7$$

Next,we have the DISTORTION FACTOR,W,defined as:-

$$W = \text{Number of triggers in a message cluster lost due}$$

to processing

Clearly, U is always less than unity, and must for good performance be as small as possible. W is a positive integer which should again be small, though negative values for W (implying filling in of message triggers lost due to noise) sometimes occur (section 5.4), and may be quite valuable. The parameters are generally in conflict; a system having a high value for U (favourable) is likely also to have a high value for W , which is unfavourable. The comparison of alternative methods for trigger association is further complicated in that W (often) and U (more rarely) may vary according to the shape of the message trigger clusters.

Some numerical values of the processing gain U consequent upon associating triggers within a region of total area N cells with threshold K are given in table 5.1 below:-

TABLE 5.1 PROCESSING GAIN, SINGLE LEVEL CELL ASSOCIATION

N	$-\log U (K=10)$	$-\log U (K$
-----	------------------	--------------

=20)

20	22.66	57.51
40	19.31	46.70
60	17.55	42.39
80	16.34	39.60
100	15.43	37.50
120	14.69	35.88.
140	14.07	34.50

$W=10$ for $K=10$, and $W=20$ for $K=20$.

This table was computed numerically, by inserting appropriate quantities for N and K into equation 5.1. A listing of the program used (F099) is appended to the thesis.

The performance of a trigger association system of the kind just described may in general be improved by breaking the process into two consecutive stages. In the first stage, the surface is divided into small 'elementary' regions, and association applied to each of these. In the second stage, these elementary regions are themselves considered as representing cells, which can contain triggers only if the first threshold has been

satisfied. They are considered in groups as 'super' regions, upon which the threshold test is repeated, in general with different parameters.

That this modification does indeed produce an increase in processing gain can be seen from the following example. Consider the region divided into a 3 by 8 cell grid containing a total of 64 cells shown in figure 5.3, and consider a detection to be successful if 4 cells or more contain triggers. (case 1). Then, suppose the same grid to be sub-divided into four 'super' cells, each containing 16 sub-cells. Let it be required that at least two sub-cells be filled in each supercell to trigger that cell, and at least two super-cells be filled to initiate a detection for the region (case 2).

Though this might at first sight seem to indicate merely that at least four cells must be triggered for detection, the two cases are clearly different. Whereas all combinations of four triggered cells will initiate a detection in case (1), some such combinations will not generate a detection in case (2). This occurs, for example, when one triggered sub-cell falls within each supercell. Thus, the two-stage trigger association

two-stage cell association.
 in a 2,2 association sequence,
 Δ are rejected, but,
 $\times$ are retained.

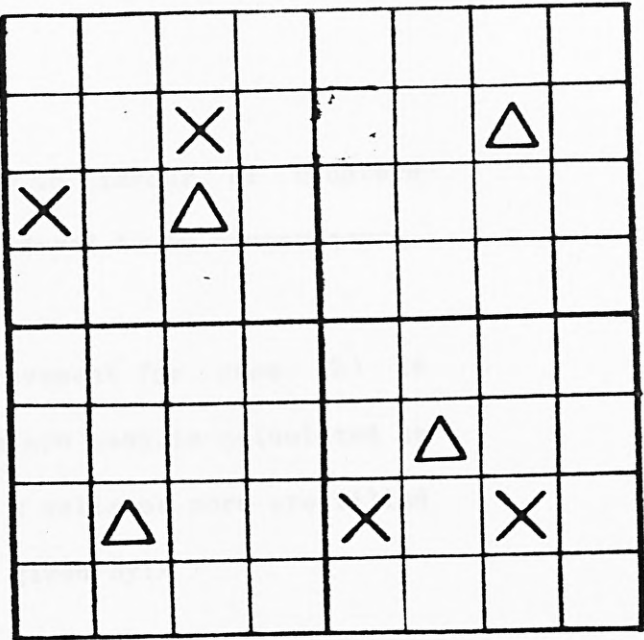


FIGURE 5.3 PRINCIPLE OF TWO-STAGE ASSOCIATION

trigger at sample i, j is retained
 only if there is also a trigger at
 at least one of the four positions
 $(i, j+1), (i+1, j-1), (i+1, j), (i+1, j+1)$

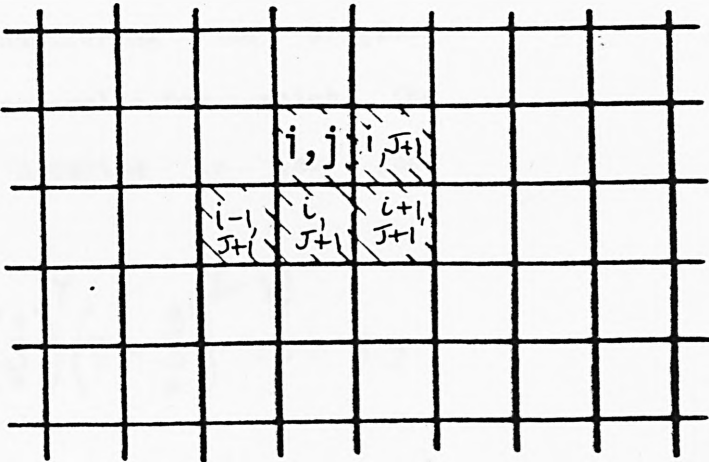


FIGURE 5.4 PRINCIPLE OF 'NEAREST NEIGHBOR'
(ADJACENCY) FILTER

discriminates even more strongly in favour of clusters which are compact, which is considered to be advantageous.

The magnitude of the improvement for case (2) in which two consecutive divisions are used is calculated as follows. The probability that x cells or more are filled in the first stage decision is given by:-

$$\sum_x^a = \sum_{\beta=x}^a \{a!/\beta!(a-\beta)!\} p(0)^\beta \{1-p(0)\}^{(a-\beta)} \dots\dots 5.8$$

here, x is the first stage threshold, and a is the number of sub-cells in each super-cell. The probability R that more than y 'super' cells are then filled in a second decision is obtained by considering the original super-cells themselves as sub-cells, for which the probability of a trigger occurring is, as just calculated, S , i.e.:-

$$R_{y,x}^{b,a} = \sum_{\gamma=y}^b \{b!/\gamma!(b-\gamma)!\} \left(\sum_x^a\right)^\gamma \left(1 - \sum_x^a\right)^{(b-\gamma)} \dots\dots 5.9$$

Here, b is the number of cells within the final decision region.

The process can be extended infinitely by iteration, e.g. three stages could be used instead of two, in which case we would have:-

$$\frac{V_{z,y,x}}{Y_{b,a}} = \sum_{\delta=z}^d \{d!/ \delta!(d-\delta)!\} \cdot (R_{y,x}^{b,a})^{\delta} \cdot (1-R_{y,x}^{b,a})^{d-\delta}$$

Here, d is the number of 'super super' cells in the third level decision cell. It is difficult to provide simple analytic expressions for the processing gains $U2 = P/R, U3 = P/V$, however some numerical values are provided in table 5.2 below, for a two-stage system.

TABLE 5.2 PROCESSING GAIN, TWO-LEVEL CELL ASSOCIAT

ION

N	SEQ.OF THRESHOLDS	-LOG U	W(MIN)
100	1 in 10, 10 in 10	19.65	
100	5 in 10, 2 in 10	22.73	10
100	2 in 10, 10 in 10	42.70	20

The value for W quoted is the minimum; the actual value will depend on the cluster shape. Table 5.2 was computed by substituting numbers into equations 5.8 and 5.9, using program F099, which is appended.

Operation of these 'association within a box' methods does not require that the boxes be rectangular. In fact, the approach works even when clusters are used whose members are not contiguous. The clusters must 'pack' the array (i.e., each cell must belong to a cluster) but overlapping is permitted. The processing gain U remains a function only of the number N of cells associated within the cluster, and of the threshold K , for given 'per cell' noise trigger probability $p(0)$. $-U$ is evidently maximum when the shape of the association cell is the same as that of the cluster of cells expected for the message. In this sense, the association cell shape can be considered 'matched' to the message, and we have a BINARY MATCHED FILTER. Alternatively, matching may be regarded as selecting the association shape having given N which minimises the distortion parameter W . The

device can be implemented as a binary mask which is convolved with the binary image. The mask may for simplicity be regarded as a rectangular array of dimension Y cells by X, with the cells at the '1' level having the shape of the pattern it is desired to detect. Binary convolution is regarded as computation of the sum S given by:-

$$S = \sum_{i=0}^{Y-1} \sum_{j=0}^{X-1} \alpha(k, l) \odot \beta(k+i, l+j) \dots\dots\dots 5.11$$

where $\odot$ denotes the logical 'AND', $\alpha(k, l)$ is the (k, l) th image cell and $\beta(k+i, l+j)$ is the (i, j) th mask cell measured from the bottom left of the convolution window.

When, and only when, the sum S exceeds a threshold K, the cells at the '1' level are retained in the filtered signal. Otherwise, they are set to zero. This approach is more powerful than that using fixed cells since the splitting of clusters by cell boundaries is avoided, but it is more expensive computationally. The value U in terms of N and K is again given by equation 5.7

5.3 Association Using Nearest Neighbour Cells

Although the cell association methods are easy to program they require that blocks of signal I cells by J be held in store, generally over the whole width of the signal. Further, the message loss occurring when a defect cluster contains just less than K cells (due for example to the cell boundary being crossed) is catastrophic. Thus, we now introduce a simpler method which requires that only one row (or column) of binary data be retained in store, continuously updated as new scans are acquired.

The surface being inspected is scanned in a raster, with successive scans in the x -direction. Then, for the cell whose x and y coordinates are J and I respectively, called the centre cell, we retain a trigger in this cell only if at least one of its nearest neighbours not yet considered as centre cell [i.e., the cells having co-ordinates $(I, J+1)$, $(I+1, J+1)$, $(I+1, J)$ or $(I+1, J-1)$] also contains a trigger (see figure 5.4) This arrangement is also called the adjacency (AJ) filter.

Let the 'per cell' probability of trigger generation be $p(0)$. Then, assuming that the probabilities of triggers being generated for each cell are mutually independent when noise only is being considered, the

probability P that an original noise trigger will be retained is given by:-

$$P = \sum_{K=1}^4 \{4!/K! \cdot (4-K)!\} \cdot p(0)^K (1-p(0))^{(4-K)} \dots$$

.. 5.12

For $p(0)$ small, this is only trivially greater than $4p(0)^2$. The processing gain U is thus very close to $1/4p(0)$. For a typical $p(0)$ of 0.001, U is then only .025. The method would thus seem to be markedly inferior to the 'association in a box' methods just described, a presumption which is confirmed by the experimental results provided in section 5.7. However, the distortion W for the method is small and moreover completely predictable. For a connected cluster containing N triggers, only one is removed so that W is 1, irrespective of the shape of the cluster or of its size. All clusters containing two or more triggers are retained, though they are reduced in size slightly. All isolated single triggers are rejected.

Better noise rejection may be obtained by applying this method repeatedly over the same data. Following the Nth iteration, all clusters containing N triggers or

fewer have been eliminated, the distortion factor W is equal to N , and the processing gain becomes U^{**N} .

Further, the method may be used to analyse the distribution of cluster sizes (which should with noise only be Poisson, provided $p(0)$ is very small), and so test the hypothesis of independence between cells. If the operation is applied repeatedly to the same binary data, and the number of triggers remaining after the $N+1$ th iteration is subtracted from that remaining after the N th, the difference will be the number of clusters containing exactly N triggers which were present in the original data.

The processing gain and distortion for the nearest neighbour (adjacency) detector are given in table 5.3:-

TABLE 5.3 CHARACTERISTICS OF THE ADJACENCY DETECTOR

NUMBER OF PASSES	$-\text{LOG}(\text{base } 10) U$	W
1	2.35	1
2	4.71	2

3	7.02	3
4	9.42	4
5	11.78	5
.	.	.
10	23.56	10

Table 5.3 was computed using the approximate relation $U = 1/4p(0)$, with $p(0) = 10^{*-3.1}$

5.4 Association Using a Bank of 'Add,Dump and Threshold'(ADT) Filters

A more sophisticated form of cluster association has been investigated by Lacroix et al (1979). This involves providing a binary store for each row (or column) of the data, to which it is submitted column by column (or, row by row) see figure 5.5 . Suppose we have a store for each row (termed a 'band'). When the n th column is examined, the stores for the bands 1,2,3 etc where the cell contains a trigger are incremented by one, whilst those for the remaining bands, which do not contain triggers, are decremented by one. The n th column of a new binary signal is then created, in which a trigger is placed in the cells for those bands for which the

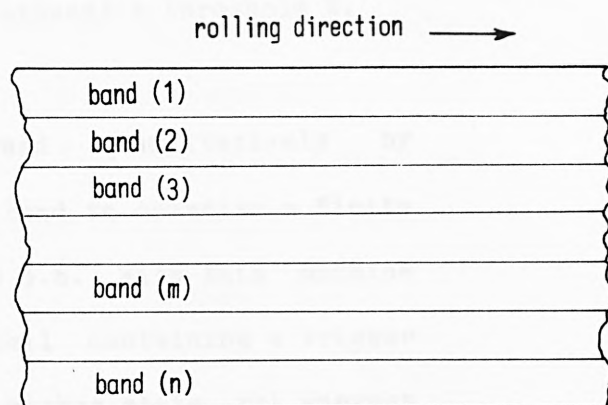


FIGURE 5.5 SHEET DIVIDED INTO STRIPS FOR 'ADD,DUMP AND THRESHOLD' (ADT) FILTER

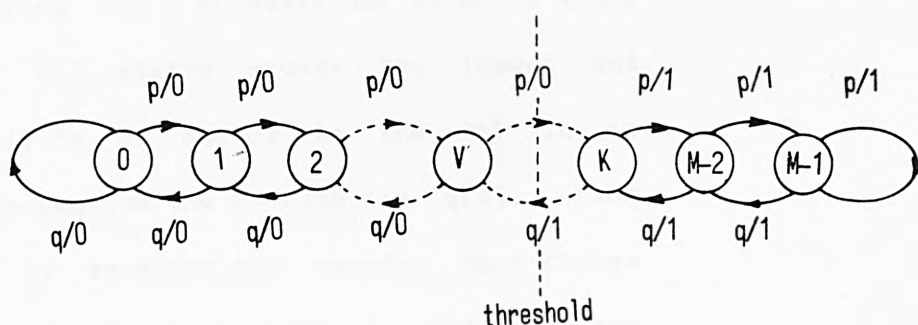


FIGURE 5.6 'ADT' FILTER AS A FINITE STATE MACHINE

current store content equals or exceeds a threshold K .

This approach is analysed quantitatively by assuming the store for each band to comprise a finite state machine, as shown in figure 5.6. With this machine in a state r , receipt of a cell containing a trigger causes it to ascend to the next higher state $r+1$, whereas a cell without a trigger causes it to fall to the next lower state $r-1$. The machine for each column is independent of those for all other columns and has M states, corresponding to the range of counts from 0 to $M-1$. Assume as usual that the probability of trigger generation per cell is $p(0)$ and that cells are mutually independent. The complementary probability of no trigger, $q(0)$, is equal to $1-p(0)$. Let the probability of the machine occupying the r th state due to noise alone be $S(r)$. Then, for all states except the lowest and highest, the probability of ascent to the $r+1$ th is $p(0)$, and that of descent to the $(r-1)$ th is $q(0)$. Each time a new cell is received, the machine must change state, unless it is already in the highest state or the lowest. If in the lowest state, it remains in that state if a triggerless cell is received; if in the highest it stays put despite receiving a trigger.

In the analysis which follows, the parameters $p(0)$ and $q(0)$ are abbreviated to p and q respectively for simplicity. The behaviour of the machine may be described by its state transition matrix, shown below:-

$$\begin{array}{c}
 \text{NEXT STATE} \\
 \left(\begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ M-1 \end{array} \right) = \left(\begin{array}{cccccc} q & p & 0 & 0 & 0 & \dots \\ q & 0 & p & 0 & 0 & \dots \\ 0 & q & 0 & p & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right) \times \left(\begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ M-1 \end{array} \right)
 \end{array}$$

STATE TRANSITION MATRIX

This specifies the probability that the system when in the state k will move on examining a new cell to any other state k' .

Let the probability of being in the (k) th state be

$P(k)$. The condition of statistical equilibrium (Feller, 1968) requires, for all states, that the probability of ascending from the (k) th state to the $(k + 1)$ th be equal to that of descending from the $(k + 1)$ th to the (k) th. That is, :-

$$p.P(k) = q.P(k + 1) \quad \dots\dots\dots 5.13$$

Thus we have :-

$$p.P(0) = q.P(1)$$

$$p.P(1) = q.P(2)$$

$$p.P(2) = q.P(3)$$

..... etc.

It follows from the above that :-

$$P(1) = P(0).p/q$$

$$P(2) = P(0).p^{**2}/q^{**2}$$

$$P(3) = P(0).p^{**3}/q^{**3}$$

..... etc.

The machine must always occupy one of the N possible

states, thus we must have:-

$$\{P(0) + P(1) + P(2) + \dots + P(N-1)\} = 1 \quad \dots\dots$$

..... 5.14

This gives for $P(0)$ the expression:-

$$P(0) = 1 / \{1 + p/q + p^2/q^2 + p^3/q^3 + \dots + p^{(N-1)}/q^{(N-1)}\}$$

The probability $P(k)$ that the (k) th state is occupied is then given by:-

$$P(k) = \{p^k/q^k\} / \sum_{l=0}^{N-1} \{p^l/q^l\}$$

However, when p is small, $q \sim 1$ and the only significant term in the denominator of the above is $p^0/q^0 = 1$. Thus (returning to our original notation) we can use the approximation:-

$$P(k) = p(0)^{**k} \quad \dots\dots\dots 5.15$$

which holds even for large N.

The probability $S(K)$ that the count will reach or exceed K is then given by:-

$$S(K) = p(0)^{**K} \quad \text{.....} \quad 5.16$$

since $p^{**}(K+1)$ is insignificant compared with p^{**K} . The processing gain U then becomes:-

$$U = p(0)^{**} (K-1) \quad \text{.....} \quad 5.17$$

For typical $p(0)$ of 0.001 and $k=4$, we have $U=10^{**}-9$, so the method is highly effective for rejecting noise triggers.

When a message cluster appears, $p(0)$ will rise from something negligibly small to something very near unity, and the machine will ascend to the next highest state with virtually every input sample. Since it will almost invariably start from state 0, K message cells (per band) must be received before the threshold is exceeded and message pulses appear in the output image. Thus, for

a message J columns wide, J.K triggers will be lost, giving for W the value:-

$$W = J.K \dots\dots\dots 5.18$$

The outstanding characteristic of this approach is its directional sensitivity. Defects which form lines along a column will be detected with distortion $W=K$ and $U = p^{**}(K - 1)$, whereas defects forming lines in the perpendicular direction will be missed entirely ($W= \text{inf.}$) unless they are at least K cells wide. The approach has an important strength in that gaps produced in a line defect by noise are often filled by the filter (W becomes negative), thus restoring a defect shape corrupted by missed detections. Some typical performance values (computed by substituting for p and K in equations 5.17 and 5.18) are provided in table 5.4:-

TABLE 5.4 PERFORMANCE OF 'ADD,DUMP & THRESHOLD'

FILTER

THRESHOLD, K	-LOG(base 10)	W(in filt.dirn)	W(pmp)
2	2.95	1	inf

5	11.83	4	inf
10	26.62	9	inf
20	53.25	19	inf

Lacroix and his collaborators (1979) improved the performance of the ADT filter by logically ORing the triggers within bands K cells wide (with K typically 6), and applying the signal from each band to an ADT filter. This raises the effective value of p to p', given by:-

$$p' = \sum_{n=1}^K \{K!/n! \cdot (K-n)!\} [p(0)]^n [(1-p(0))]^{K-n} \dots\dots\dots 5.19$$

The directional dependence of W, and the complexity of the hardware, are also much reduced.

5.5 Comparison of Methods Using Theory

Table 5.5 shows values of U for the binary filters

discussed, computed for the same realistic values of W and $p(0)$ so that a comparison may be drawn. Although the 'add, dump and threshold' (ADT) filter seems at first sight to be superior, it is effective only when the defect clusters are oriented strongly in one preferred direction. The ADT filter rejects virtually everything in the perpendicular direction. The adjacency (AJ) detector also looks attractive, except that it requires repeated examination of the data. Further, it always distorts the cluster. Cell association, on the other hand, does not distort clusters containing the threshold number of triggers or more. It can be implemented with only one pass over the data. However, simultaneous storage of several consecutive scans is required; the AJ filter needs storage of only one scan per pass desired.

TABLE 5.5 COMPARISON OF ASSOCIATION METHODS

(THEORETICAL PREDICTION)

Prob. per cell of trigger generation due to noise is 1

0**-3

METHOD

- LOG(base 10) U

W

cell assn

N=100,K=10	15.04	10
cell assn		
2stg.	22.73	10(min)
(5 in 10,2 in 10)		
AJ (10 passes)	23.5	10
ADT (K=10)	26.62	9 (in fitt. dim.)
		inf. (in prp. dim.)

5.6 Analysis of Potential Improvement in Detection Performance

Having established that trigger association can improve the performance of a defect detection system by making it more sensitive at a given false alarm rate, we now compute the extent by which the message/noise contrast of the defect has effectively been increased.

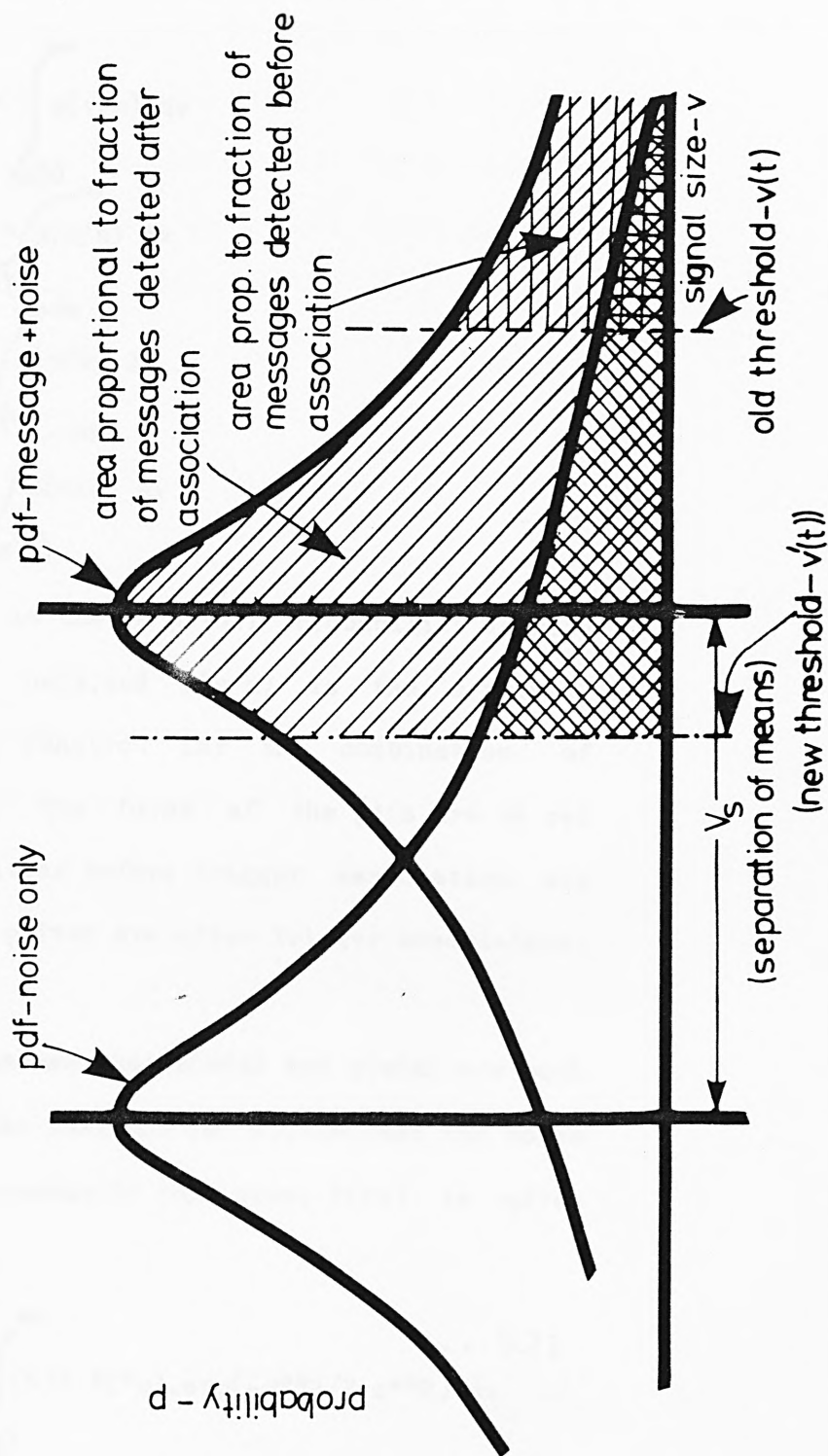
Suppose that the false alarm probability for the whole system is required to be $P(fa)$. If a simple

detection system were used which did not include trigger association, the detection threshold would (according to the Neyman-Pearson criterion) have to be set to some value $v(t)$, to achieve $P(fa)$. Since values for $P(fa)$ of the order of 10^{-10} are typically required, $v(t)$ must be set far away from the mean for the noise, and the detectability of low contrast defects is consequently poor. With trigger association which can reduce the false alarm probability from an initial high value $P'(fa)$ to the required lower value $P(fa)$ without losing message triggers, the first detection threshold can be lowered to a new value $v'(t)$ which gives much improved detection for low contrast defects.

The means by which improvement is obtained is illustrated in figure 5.7, to which we now refer. With a first threshold $v(t)$, the probability of defect detection is $P(d)$, which is low, but rises to a higher value $P'(d)$ when the threshold is lowered to $v'(t)$. The magnitude of the improvement is specified by the parameter X , defined by:-

$$X = P'(d)/P(d) \quad \dots\dots\dots 5.20$$

FIG. 5-7 Performance Improvement Following Trigger Association



To determine the magnitude of X , we use the relationships below (derived from eqs. 3.2 to 3.5, sec. 3.3):-

$$P(fa) = \int_{v(t)}^{\infty} p(v/n) dv$$

$$P'(fa) = \int_{v'(t)}^{\infty} p(v/n) dv$$

$$P(d) = \int_{v(t)}^{\infty} p(v/m) dv$$

$$P'(d) = \int_{v'(t)}^{\infty} p(v/m) dv$$

In the above, $p(v/n)$ is the 'a priori' probability density function for noise only, and $p(v/m)$ is the 'a priori' probability density function for the combination of message and noise. The forms of the pdfs are as yet unspecified. Quantities before trigger association are unprimed, those with primes are after trigger association.

Initially, we assume that $p(v/n)$ and $p(v/m)$ are both Gaussian with the same variance (s^{**2}), i.e. that the noise is additive to the message. In this case, $P(fa)$ is given by:-

$$P(fa) = \int_{v(t)}^{\infty} (1/2 \cdot \pi \cdot s) \cdot \exp(-v^{**2}/2 \cdot s^{**2}) \cdot dv \quad \dots 5.21$$

To obtain an explicit value for $P(fa)$, we note that $v(t)$ must be well away from the mean (certainly more than 3 standard deviations). In this case (Schwartz and Shaw, 1975) we can approximate the right hand side of equation 5.21 by the first term of its series expansion, i.e.:-

$$P(fa) = s \cdot \sqrt{\pi} \cdot v(t) \cdot \exp(-v(t)^2 / 2s^2)$$

.....5.22

Some values of this function for typical values of $v(t)$, expressed in terms of the standard deviation s , are given in table 5.6 below, and are shown graphically in fig. 5.8

TABLE 5.6 PROBABILITY OF FALSE ALARM WITH THRESHOLD

FAR FROM MEAN

GAUSSIAN AND NEGATIVE EXPONENTIAL DISTRIBUTIONS

NS

$v(t)$	$P(fa)$	$P(fa)$
(standard deviations)	(Gaussian)	(negative exponential)

4.2	$1.16 \cdot 10^{-5}$	$1.50 \cdot 10^{-2}$
4.9	$3.86 \cdot 10^{-7}$	$7.45 \cdot 10^{-3}$
5.6	$7.94 \cdot 10^{-9}$	$3.70 \cdot 10^{-3}$
6.3	$1.01 \cdot 10^{-11}$	$1.84 \cdot 10^{-3}$
7.0	$7.84 \cdot 10^{-13}$	$9.10 \cdot 10^{-4}$
7.7	$3.74 \cdot 10^{-15}$	$4.5 \cdot 10^{-4}$
8.4	$1.09 \cdot 10^{-17}$	$2.25 \cdot 10^{-4}$

We are here concerned with the detectability of low contrast defects, e.g. those for which $P(d)$ is, say, 5%. Reference to figure 5.8 shows that the separation of the means of the two distributions, $v(s)$, will then be 4.26s. Suppose now that the trigger association can reduce $P(d)$ by a factor (the coefficient U already defined) 10^{-8} , and that the $P(fa)$ required for the system overall

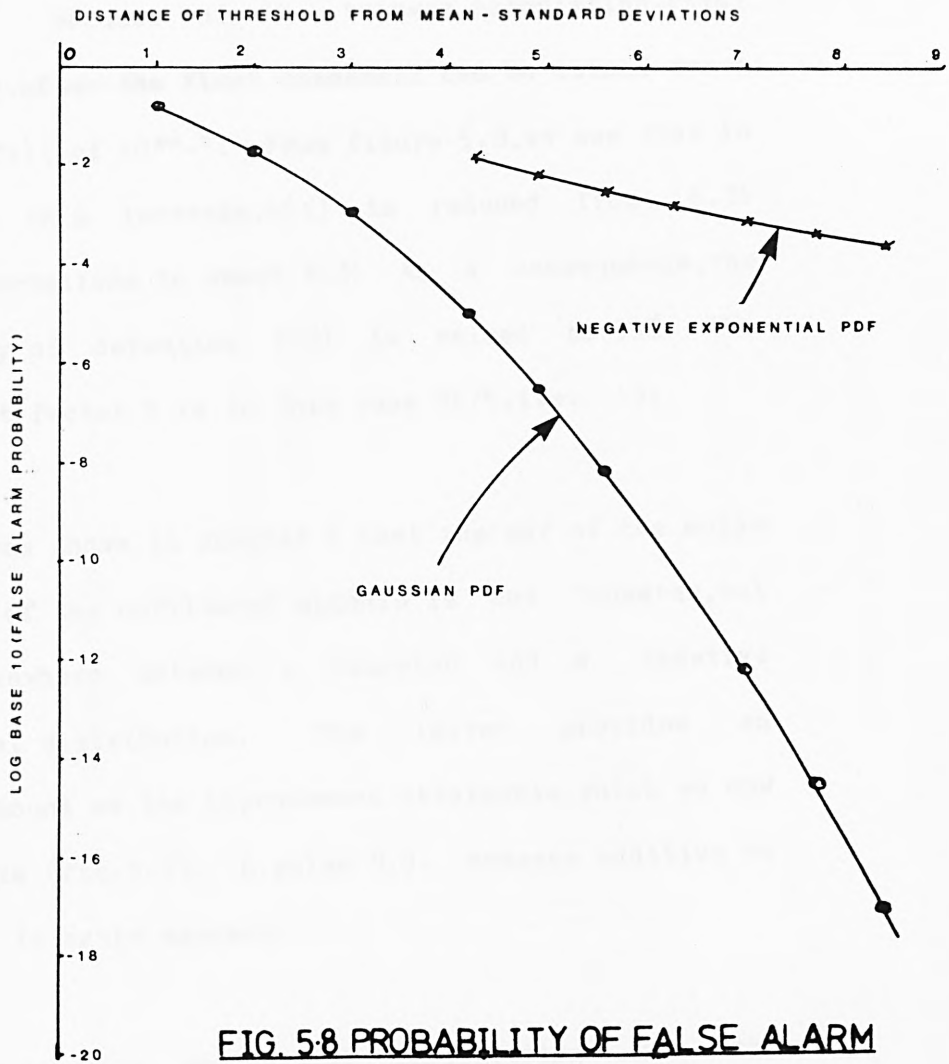


FIG. 5.8 PROBABILITY OF FALSE ALARM
VS. THRESHOLD LEVEL

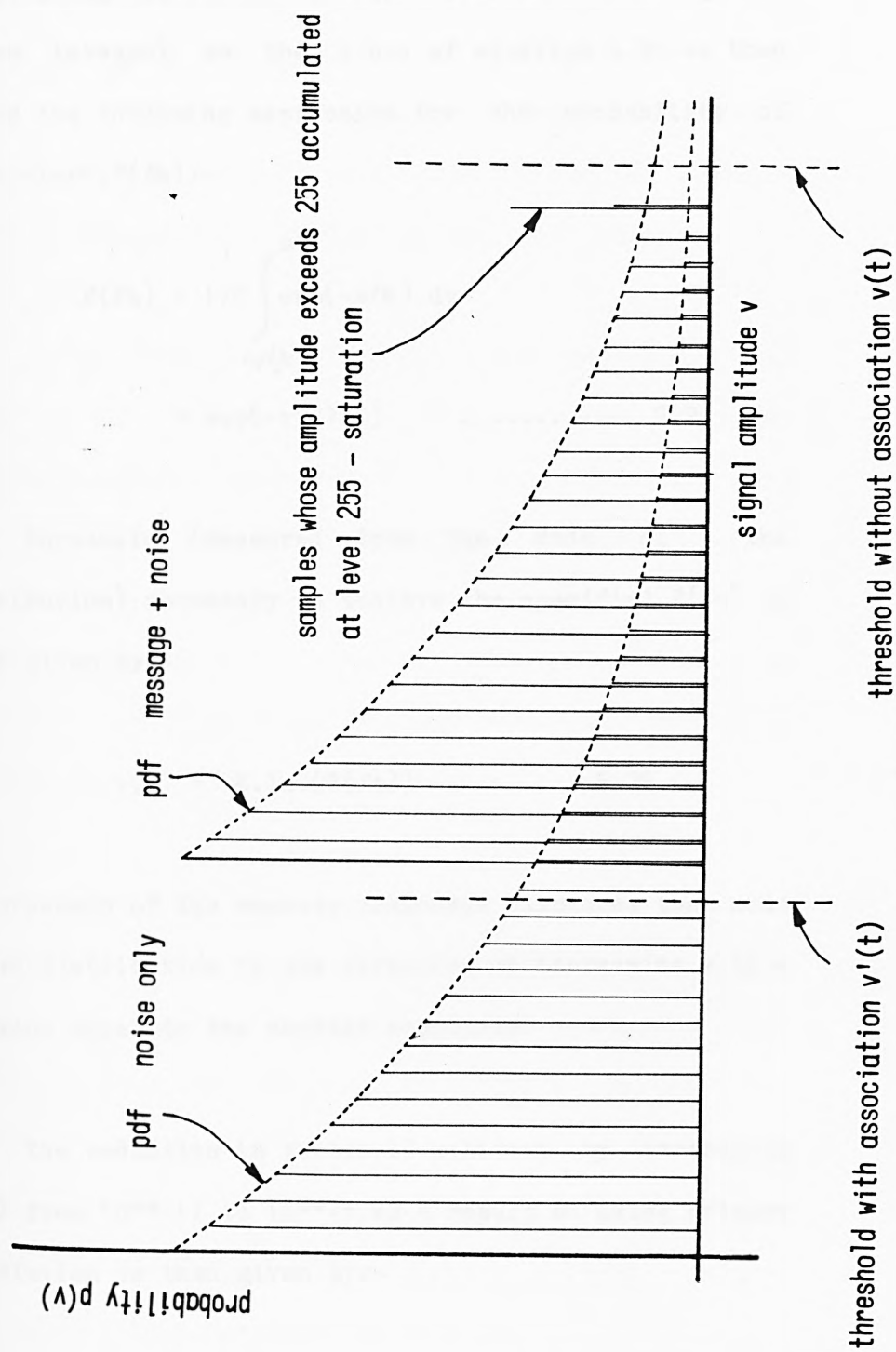
is 10^{-11} . Because of the trigger association, $P(f_a)$ immediately after the first threshold can be raised to a level $\{P'(f_a)\}$ of 10^{-3} . From figure 5.8, we see that in obtaining this increase, $v(t)$ is reduced from 6.36 standard deviations to about 3.3. As a consequence, the probability of detection $P(d)$ is raised to 95%. The improvement factor X is in this case $95/5$, i.e. 19.

It was shown in chapter 2 that the pdf of the noise component of the unfiltered signals is not Gaussian, but lies somewhere between a Gaussian and a negative exponential distribution. The latter provides an opposite bound on the improvement obtainable, which we now investigate (fig. 5.9). A pulse D.C. message additive to the noise is again assumed.

The negative exponential distribution has the general form:-

$$f(v) = (1/K) \cdot \exp(-v/K) \quad \dots\dots\dots 5.23$$

where K is a single parameter specifying the characteristics of the distribution. For the signals used in the simulation investigation, K is typically 8.



**FIGURE 5.9 PERFORMANCE IMPROVEMENT USING TRIGGER ASSOCIATION
NEGATIVE EXPONENTIAL NOISE, DISCRETE SIGNAL LEVELS, LIMITED
DYNAMIC RANGE**

Substituting the r.h.s. of eqn. 5.23 for the argument of the integral on the r.h.s of equation 5.21, we then obtain the following expression for the probability of false alarm, $P(fa)$:-

$$P(fa) = \frac{1}{K} \int_{v(t)}^{\infty} \exp(-v/K) dv$$

$$= \exp(-v(t)/K) \quad \dots\dots\dots 5.24$$

The threshold (measured from the mode of the distribution) necessary to achieve the specified $P(fa)$ is hence given by:-

$$v(t) = -K \cdot \ln.(P(fa)) \quad \dots\dots\dots 5.25$$

The presence of the message component displaces the mode of the distribution in the direction of increasing v , by a distance equal to the message amplitude.

The reduction in threshold achieved by increasing $P(fa)$ from 10^{-11} to 10^{-3} as a result of using trigger association is then given by:-

$$K \{ \ln.(10^{-11}) - \ln.(10^{-3}) \} \sim 147 \text{ for } K=3 \quad \dots\dots 5.26$$

The fall in threshold measured in standard deviations is thus about three times as great when the noise obeys a negative exponential distribution as for a Gaussian.

As is seen from figure 5.9, the probability of detection now becomes 100%. This provides the upper bound on available improvement. The actual distribution (chapter 2 section 3) lies somewhere between these two extremes, and is probably closer to the Gaussian than to negative exponential. Thus, we conclude that an improvement, X , of at least 20 times should be obtainable for marginal defects if trigger association is used.

Further, for a system whose signals are quantised uniformly to 256 levels with a mean of 128 (so that negative and positive excursions of the signal can both be sensed), then the minimum sensitivity for which the threshold might be set is about 10^{-7} , due to limited dynamic range. A setting giving a false alarm probability of 10^{-11} (as required without trigger association) just could not be obtained.

There seems to be little point in performing the computation for distributions which are closer to the true theoretical distribution, e.g. that provided by Stansberg (1981) since the distributions actually measured (e.g. fig. 2.9) do not conform exactly with theory due for example to variation of surface properties, and in any case vary widely from sample to sample.

5.7 Comparison of Methods Using Simulation

The comparative effectiveness of the various methods when applied to real data depends on several factors whose values are impossible to specify exactly, such as the extent to which triggers due to noise really are mutually independent, and the sizes and shapes of the trigger clusters arising from the various kinds of defect. For the cold rolled strip data used, samples in the scan (x) direction are found to be uncorrelated, but there is some correlation in the y -direction (chap. 2, section 4). However, the analysis provided in chapter 2.4 suggests that triggers not having the same scan index (x) should be completely independent. Only samples having the same x co-ordinate should show any

correlation, and even this should fall off rapidly with separation. Further, as may be seen from figure 6.9, the parameters of the noise are highly non-stationary, causing $p(0)$ to vary, and this effect is probably most significant. Consequently it is necessary to resort to simulation for meaningful comparison of the methods, and to be aware that the conclusions drawn from such simulations may in their detail apply only to the specific problem considered. The improvement obtained in the simulations can be expected thus to be somewhat worse than theory predicts.

Figures 5.10 to 5.16 illustrate the comparative effectiveness of the methods when applied in simulation to the detection of low-contrast defects on cold rolled strip. Figure 4.7 shows triggers obtained using a SIRA detector whose threshold is set so close to the noise that a high density of false alarms is obtained. The shape of the trigger cluster resulting from the defect (in this case, pinch marks) is well maintained, and the defect appears within the appropriate boxes.

Figure 5.11 shows the result of processing the data shown in figure 4.7 through a region association, with

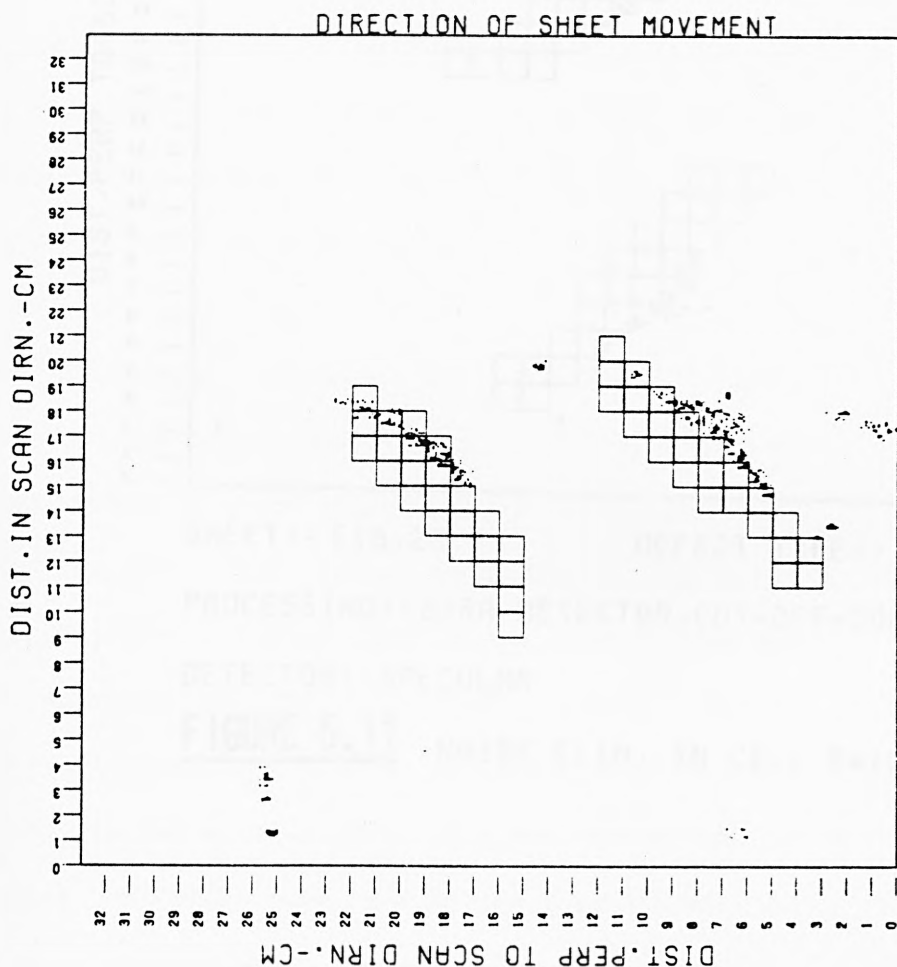
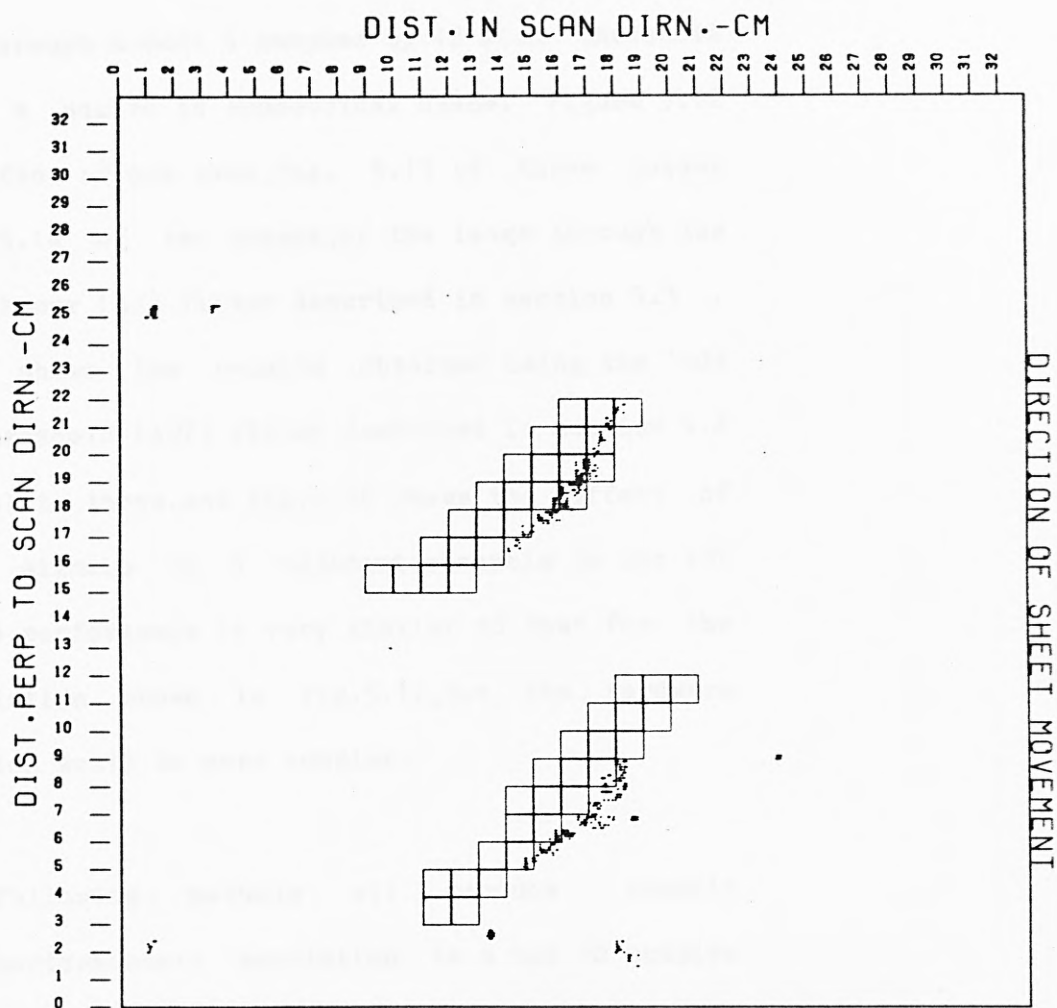


FIGURE 5.10 TRIGGER ASSOCIATION.CELL 10 SAMPLES BY 10 SAMPLES.THRESH=10.



SHEET:- E16.26

DEFECT TYPE:- PINCH MARK

PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.12

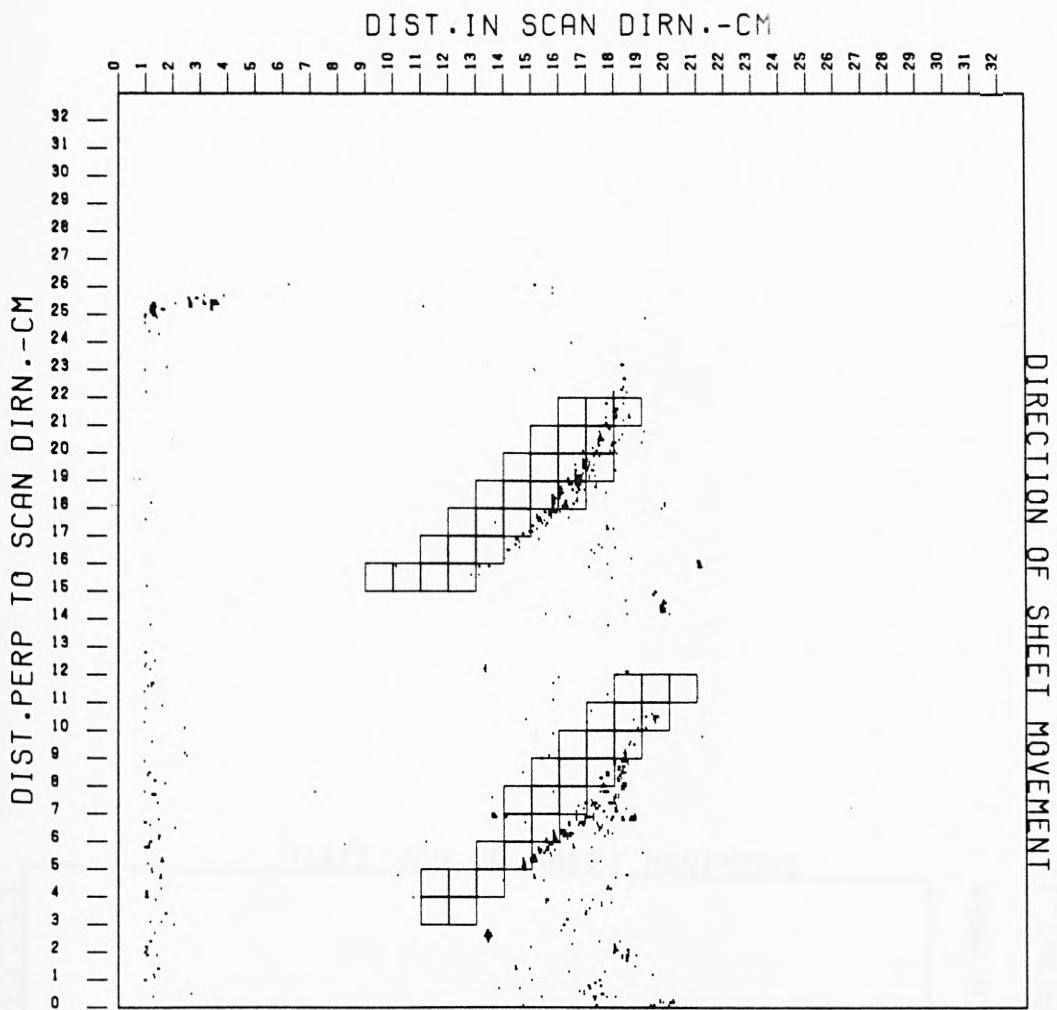
DETECTOR:-SPECULAR

FIGURE 5.11

NOISE ELIM. IN CELL 5*12,THRESH=10

square regions 10 cells by 10 and a threshold of 10 and fig. 5.11 through a cell 5 samples by 12 with threshold 10, which is a square in geometrical space. Figure 5.12 shows the effect of one pass, fig. 5.13 of three passes and figure 5.14 of ten passes, of the image through the nearest neighbour (AJ) filter described in section 5.3. Figure 5.15 shows the results obtained using the 'add, dump and threshold' (ADT) filter described in section 5.4, with K equal to three, and Fig. 5.16 shows the effect of 'ORing' the signals in 6 adjacent channels in the ADT filter. The performance is very similar to that for the cell association shown in fig. 5.11, but the hardware implementation would be more complex.

The following methods all produce roughly equivalent performance:- association in a box 10 samples by 10 with threshold 10, 10 passes of the adjacency (AJ) filter, and the ADT filter applied to groups of 6 adjacent channels. There is little to choose between them, and they are superior to the remaining methods (except possibly association in a box in stages, which was not investigated in the simulation). Thus, the choice as to which method should be used in a hardware implementation for on-line operation must depend on other



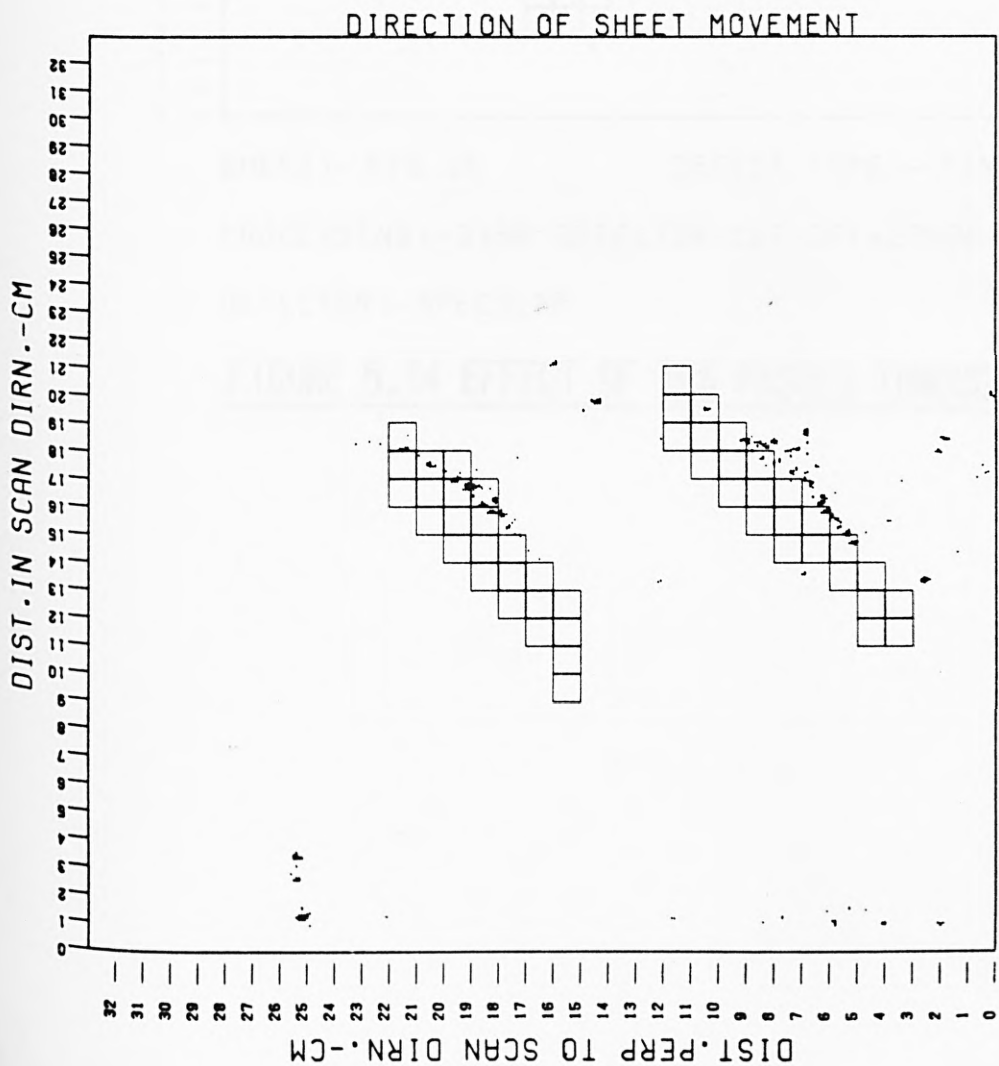
SHEET:- E16.26

DEFECT TYPE:- PINCH MARK

PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.12

DETECTOR:-SPECULAR

FIGURE 5.12 EFFECT OF ONE PASS THROUGH 'AJ' FILTER

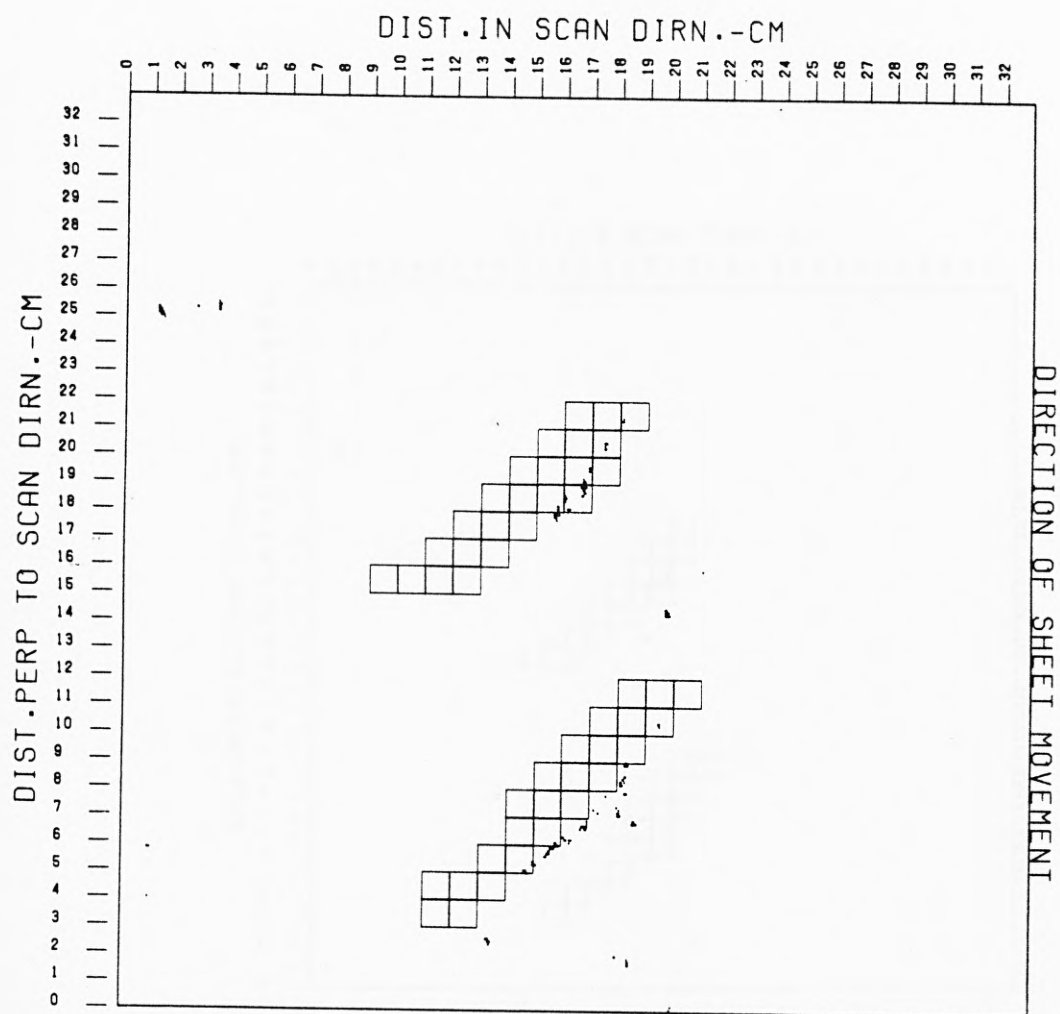


SHEET:-- E16.26 DEFECT TYPE:-- PINCH MARK

PROCESSING:--SIRA DETECTOR,CUT-OFF=20000,ATN=1.12

DETECTOR:--SPECULAR

FIGURE 5.13 EFFECT OF THREE PASSES THROUGH 'AJ' FILTER



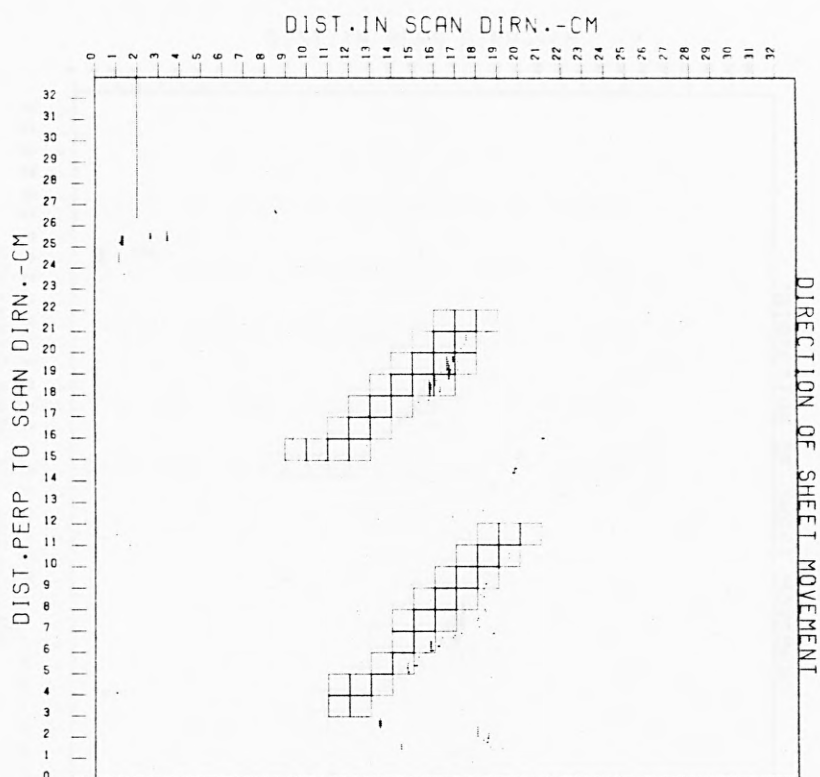
SHEET:- E16.26

DEFECT TYPE:- PINCH MARK

PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.12

DETECTOR:-SPECULAR

FIGURE 5.14 EFFECT OF TEN PASSES THROUGH 'AJ' FILTER



SHEET:- E16.26

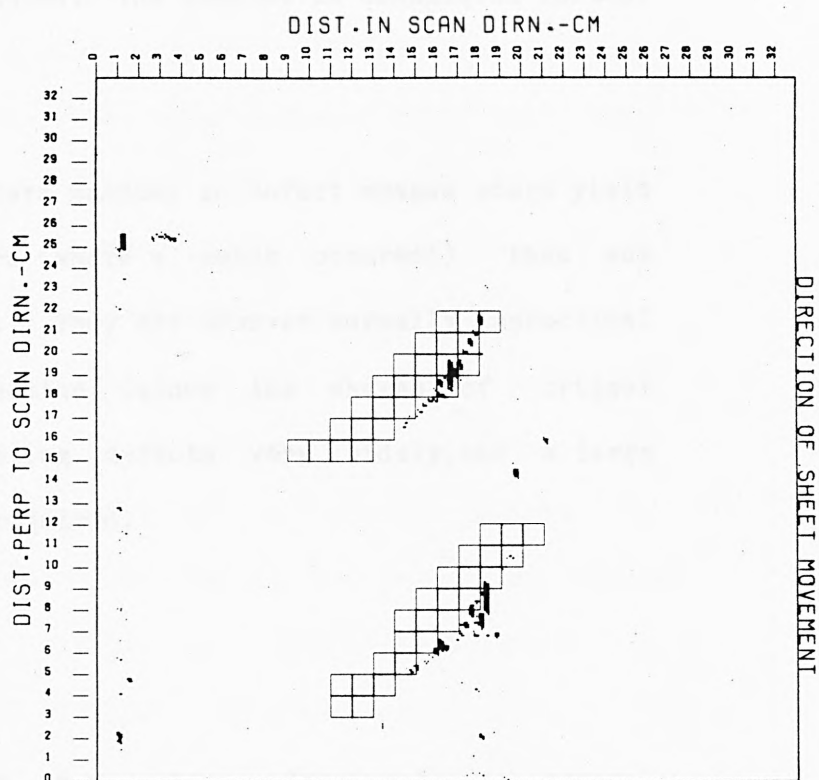
DEFECT TYPE:- PINCH MARK

PROCESSING:-SIRA DETECTOR.CUT-OFF=20000.ATN=1.12

DETECTOR:-SPECULAR

TRIGGER ASSOCIATION USING FTRIGS WITH K=3

FIGURE 5.15 EFFECT OF 'ADT' FILTER, ONE SAMPLE PER BAND



SHEET:- E16.26

DEFECT TYPE:- PINCH MARK

PROCESSING:-SIRA DETECTOR,CUT-OFF=20000,ATN=1.10

DETECTOR:-SPECULAR

&FRENCH&FILTER.THRESH.=4.6 LINES PER FILTER.

FIGURE 5.16 EFFECT OF 'ADT' FILTER,SIX SAMPLES PER BAND

considerations, such as cost, and compatibility with other operations required in the system. This latter may allow hardware to be shared. The problem is considered further in section 6.5.

Binary filters matched to defect shapes would yield better performance (where a match occurred!) than any method simulated. They are however normally impractical in surface inspection, since the shapes of trigger clusters representing defects vary widely, and a large number would be required.

5.8 Conclusions

In chapter 5, we have developed a general methodology for separating trigger clusters containing useful messages from isolated triggers constituting random noise in the binary signals remaining after the first stage of decision. We have further shown that the probability of detection for marginal defects is thereby increased very considerably. For example, for a pulse defect in additive Gaussian noise, a marginal detection probability of 5% is increased to a safe 95% by use of trigger association, without increasing the probability

of false alarm.

Three filters using this methodology have been described and compared. We have then used computational experiment to show that the methods work well when applied to real data, even when the assumptions made in deriving the theory (notably, that the triggers due to noise are mutually independent) are not strictly true. Inclusion of trigger association in a defect detection system is seen thus to be very worthwhile. There is on grounds of performance little to choose between the methods, and the final choice must be made using consideration of simplicity of implementation in hardware. This is considered in section 6.4.

CHAP.6 COMPOSITE DETECTION SYSTEM

6.1 Introduction

The analysis provided in chapters 3 to 5 developed a processing methodology for defect detection in automated inspection, and demonstrated its effectiveness using computer simulation, for a typical inspection problem.

However, the analyses hitherto presented have considered the methods to operate in isolation and assumed them to be independent of one another, and it is now necessary to consider them in combination. The methods may be combined in many different ways, and it is necessary to select the best possibility.

The methods of signal processing sometimes interact. For example, matched filtering to increase defect contrast increases the effective size of the window of observation, reducing the effectiveness of subsequent trigger association. Doubt also exists as to whether when practical as well as theoretical considerations are taken into account, it is better to

combine the components of the vector signal as soon as possible and treat the combination thereafter as a single scalar, or whether the components should be processed independently, and a final decision reached by logical combination of decisions made separately for the individual channels. These two viewpoints are extremes, since the components may in principle be combined at any point. Determination of the best position requires consideration of the theory of signal processing, of the cost of the hardware, and of the properties of the defect and noise signals.

The system must also accommodate variation in the properties of the surface noise, an aspect we have hitherto ignored.

Finally, to be of practical use, the selected processing scheme must be capable of being implemented in hardware, to operate at high speed, in a manufacturing environment, at a cost which is less than the savings achieved by incorporating the inspection.

The purpose of this chapter is hence to integrate the results of analyses provided in chapters 2 to 5

,and,by way of example,to propose a complete system for the specific task of inspecting the surface of cold rolled steel strip. Thus,section 6.2 discusses the alternative configurations and their trade offs from a theoretical viewpoint, and section 6.3 proposes an optimal configuration in the light of this discussion. Section 6.4 considers the hitherto-ignored problem of non-stationarity in the noise signal,and 6.5 investigates the hardware implementation and suggests a preferred form.

6.2 Trade-Offs

The first task is to consider a detection scheme based on the canonic form proposed in chapter 3,and establish an optimal configuration. The analysis has general applicability. It is not specific to the problem of inspecting cold-rolled steel strip.

In figures 6.1 to 6.4,a sequence of forms which the system might take for processing a vector signal comprising three mutually partially correlated scalar components (designated (1),(2) and (3)) is shown. As the sequence develops,the point at which the components are

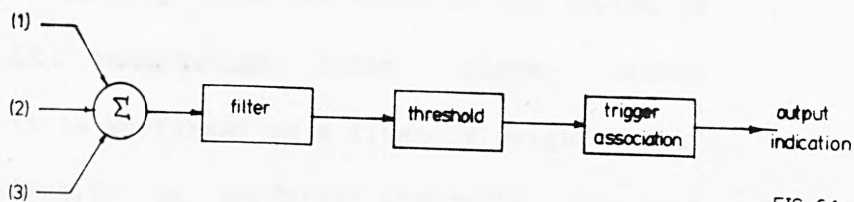


FIG. 6.1

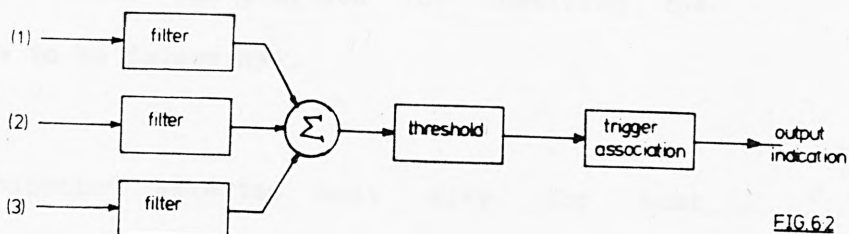


FIG. 6.2

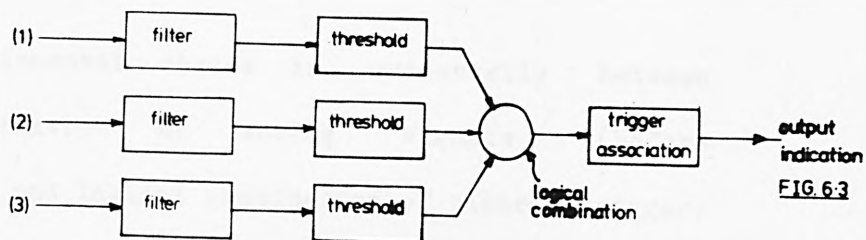


FIG. 6.3

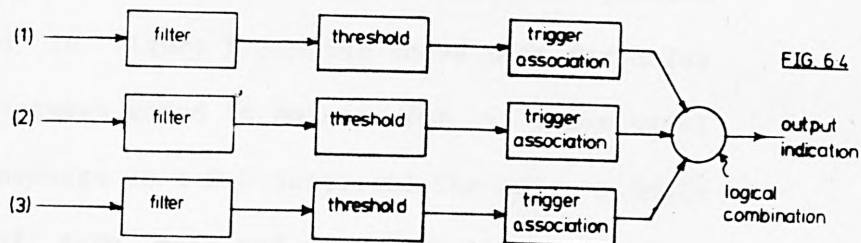


FIG. 6.4

combined moves steadily from the front of the system to the rear. If combination takes place before thresholding, it is performed as a linearly weighted sum. After thresholding, it is performed logically. The best form ('AND', 'OR' etc.) and position for combining the components are to be determined.

The combination selected must give the best probability of detection for ALL of the defect types which will be encountered, whilst maintaining the specified false alarm rate.

The fundamental choice is evidently between linear combination of analog signals (before thresholding), and logical combination of binary (trigger) signals after thresholding. The essence of the problem is illustrated in figure 6.5, which shows pdfs for noise only, and for message added to noise, for a two-channel system. The message is a D.C. pulse, and the noise signals are Gaussian, of zero mean, and are partially correlated. Both the message and the noise components in the two channels are equal.

To detect the message we can (a) impose thresholds

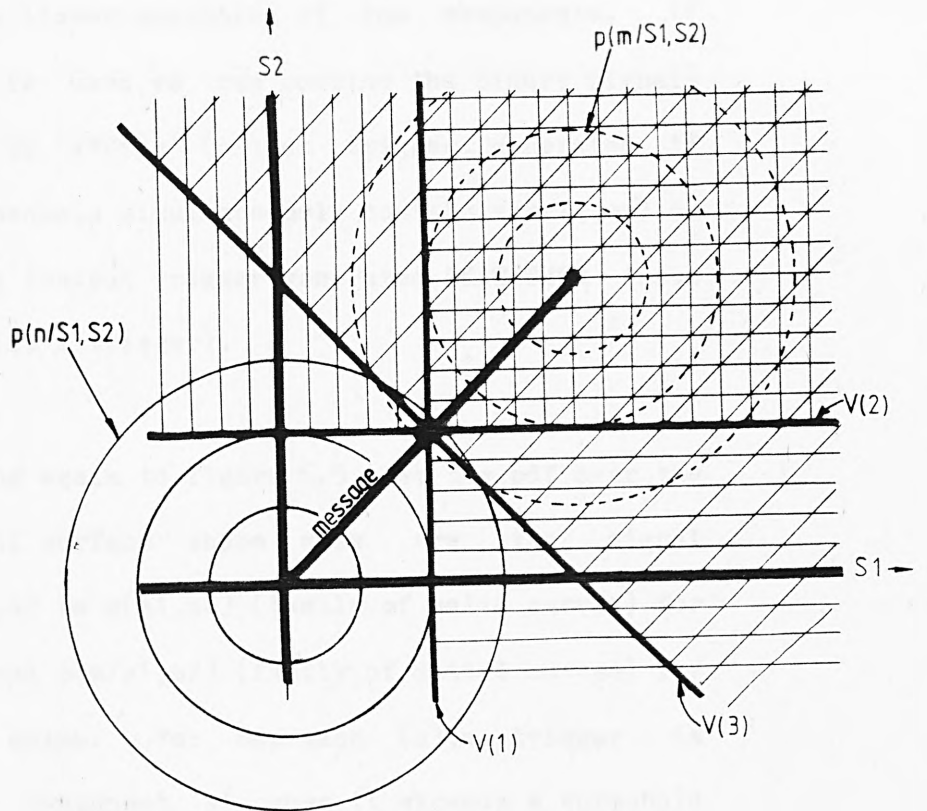


FIG. 65 REGIONS OF SIGNAL SPACE CONTRIBUTING NOISE TRIGGERS, FOR THREE THRESHOLDING SCHEMES

on each component individually, or (b) impose a single threshold on a linear summation of the components. If approach (a) is used, we can combine the binary signals logically (i) by 'ANDing' (output trigger generated if BOTH input channels simultaneously contain a trigger), or (ii) by 'ORing' (output trigger generated if EITHER input channel contains a trigger).

Referring again to figure 6.5, let the pdf over the two-dimensional surface whose axes are the signal components s_1, s_2 be $p(s_1, s_2)$ (family of solid curves) for noise only, and $p(m/s_1, s_2)$ (family of dotted curves) for message plus noise. For approach (a), a trigger is generated for component s_1 when it exceeds a threshold $v(1)$, and for s_2 when it exceeds a threshold $v(2)$. For case (b), the linear combination of s_1 and s_2 (i.e., $s_3 = w_1 s_1 + w_2 s_2$, where w_1 and w_2 are constants) must exceed a threshold $v(3)$. We shall for simplicity assume that $v(1) = v(2)$, that $w_1 = w_2$, and that $v(3)$ equals the square root of $(v(1)^2 + v(2)^2)$ and, further, that the signal components s_1 and s_2 are mutually uncorrelated. The pdfs shown on figure 6.5 are hence circular though under less stringent assumptions they would be elliptical. The vector which is the combination of signal components $s_1 +$

s2 is denoted 'message' on figure 6.5.

The false alarm probabilities for each thresholding are obtained by substituting the appropriate pdfs in eqn.(3.4), giving for case (a):- Components ANDed together:-

$$P1(fa) = \int_{\tau(1)}^{\infty} \{p(n/s1).ds1\} \cdot \int_{\tau(2)}^{\infty} \{p(n/s2).ds2\} \quad \dots 6.1$$

Components ORed together:-

$$P2(fa) = \left\{ \int_{\tau(1)}^{\infty} p(n/s1).ds1 \right\} + \left\{ \int_{\tau(2)}^{\infty} p(n/s2).ds2 \right\} \quad \dots 6.2$$

The false alarm probability $P3(fa)$ for case (b) is given by:-

$$P3(fa) = \int_{\tau(3)}^{\infty} p(n/s3).ds3 \quad \dots 6.3$$

The regions of the signal space within which noise alone will generate a false alarm are indicated for the three cases in figure 6.5. Signals exceeding the threshold τ_2 lie in the vertically hatched region in figure

6.5, signals exceeding v_1 lie in the region hatched horizontally, and signals exceeding v_3 lie in the region hatched diagonally. Signals falling in the horizontally and the vertically hatched regions generate triggers for an 'ORed' system, but for an 'ANDed' system, triggers are generated only by signals in the region hatched both horizontally AND vertically.

Since $P(f_a)$ always increases as the region of space over which it is obtained by integration increases, it is evident from fig. 6.5 that $P(f_a)$ will be largest for case (a)(ii) and smallest for (a)(i), with (b) falling in between, without a significant change in the probability of defect detection between the three cases. Though the probability of detection also decreases in the order (a)(ii), (b), (a)(i), the magnitude of the decrease is not intolerable, at least given the idealised message properties shown in figure 6.5. Thus, configuration (a)(i) would seem to be the best choice. This is confirmed by the quantitative analysis which follows.

In this, we substitute numerical values for the various quantities, and assume (in accord with the Neyman-Pearson criterion) a false alarm probability

immediately before trigger association of 10^{-3} ,
 (i.e., $P_1(fa) = P_2(fa) = P_3(fa) = 10^{-3}$. Assuming that the
 false alarm probabilities for the channels are equal
 before combination and that the noise is completely
 uncorrelated between them, we then have:-

For case (a)(i), $10^{-3} = P(fa) + P(fa)$, i.e., $P(fa) = 5 \times 10^{-4}$
 6.4

For case (a)(ii), $10^{-3} = P(fa) \cdot P(fa)$,
 i.e., $P(fa) = 3.3 \times 10^{-2}$ 6.5

For case (b), with the weighting coefficients w_1 and w_2
 equal, and with the distributions in the two channels
 mutually independent and of the same amplitude, then $P(fa)$
 for the linear combination is the same as that for the
 individual components, i.e., $P(b) = 10^{-3}$. This result
 follows immediately from the circular symmetry of the
 joint pdf for the two signals considered
 simultaneously, see fig. 6.5.

The next problem to be analysed is, should the
 signals in the various channels be combined before
 trigger association or after. This is easily resolved.

Let the probability of false alarm in each of the two channels be P , and be mutually independent between the channels. Trigger association applied separately to each channel will reduce this to $U.P$. Thus, we have five possible expressions for the resultant false alarm probability $P(fa)$, namely:-

(i) Apply trigger association to each channel independently, then combine the two outputs by 'OR'ing them together. This gives:-

$$P(fa) = U.P + U.P \quad \dots \quad 6.6$$

(ii) 'OR' together two channels, then apply trigger association, giving:-

$$P(fa) = U.(P+P) \quad \dots \quad 6.7$$

(iii) Apply trigger association separately to each channel, then combine the channels by 'AND'ing, giving:-

$$P(fa) = (U^{**2}).(P^{**2}) \quad \dots \quad 6.8$$

(iv) 'AND' channels together, then apply trigger association:-

$$P(fa) = U.(P^{**2}) \quad \dots \quad 6.9$$

(v) Apply trigger association to the analogue sum (case 'b'):-

$$P(fa) = U.P \quad \dots \quad 6.10$$

Recalling that the objective in attaining an optimal

design under the Neyman-Pearson criterion is to minimise $P(md)$ for given $P(fa)$, cases (i) and (ii) are seen to be equivalent, and the approaches may then be arranged in ascending order of preference as follows:-

(i and ii){worst}, (v), (iv), (iii){best}

However, if the MESSAGE signals in the channels being combined are not well correlated, combining triggers by 'ANDing' will cause an enormous increase in $P(md)$, and will thus be unacceptable.

We conclude then that if the components are to be combined by 'ORing', approach (ii) should be used since it is more economical in hardware. If on the other hand the channels can be 'ANDed' together, it is preferable to apply the trigger association before combination, since this yields a significant improvement in performance.

Extension of the analysis to three (or more) dimensions increases the sizes of the differences between the various alternatives, but retains the order of preference.

6.3 Recommended Configuration

Unfortunately, the signals met with in practice differ considerably from the ideal form suggested in figure 6.5. The noise amplitudes in the various channels can be made equal by adjustment of channel gains, but the message components are generally quite unequal, are not (Dainty, 1971) additive with the noise, and also vary according to the kind of defect. Figure 6.6 shows more typical forms which the message might take. Message (Q) is effectively indistinguishable from noise in channel s_1 considered individually (evidenced by the projected view) and should ideally be sensed in channel 2 alone. Message (W), though correlated between the channels, generates a negative going indication in both s_1 and s_2 . The system must detect messages of both types.

There may well be no reason to assume that any particular direction of signal space is more likely to contain message signals than any other, when a substantial ensemble of different defects must be considered simultaneously. The system must detect messages occurring in all directions with equal facility.

Superficial analysis then suggests that the channels should be 'OR'ed together, using a double

threshold in each channel so that triggers are generated by both positive and negative excursions from the mean. Association then takes place after combination, since this yields the simplest hardware. However, a better solution is to use the non-linear detector whose decision surface is indicated by the 'xxxxxxx' chain on figure 6.6. The noise is assumed Gaussian (reasonable if it has been filtered, see chap2, sect3) and is partially correlated between the channels. Thus, the decision surface for detection, in the complete absence of 'a priori' information concerning the defect signals should be an ellipse, positioned on a contour of equal joint probability. Any signal vector falling within the ellipse is regarded as being noise, any falling outside as a defect message.

The coefficients a, b, θ specifying the ellipse which yield the required false alarm probability $P(fa)$ given the noise variances s_1, s_2 in the two channels and their mutual correlation $r(1,2)$ are extremely difficult to calculate in closed form, except for the simple case in which the noise magnitudes are equal in the two channels, and are uncorrelated. This case we now examine; it is illustrated in figure 6.7.

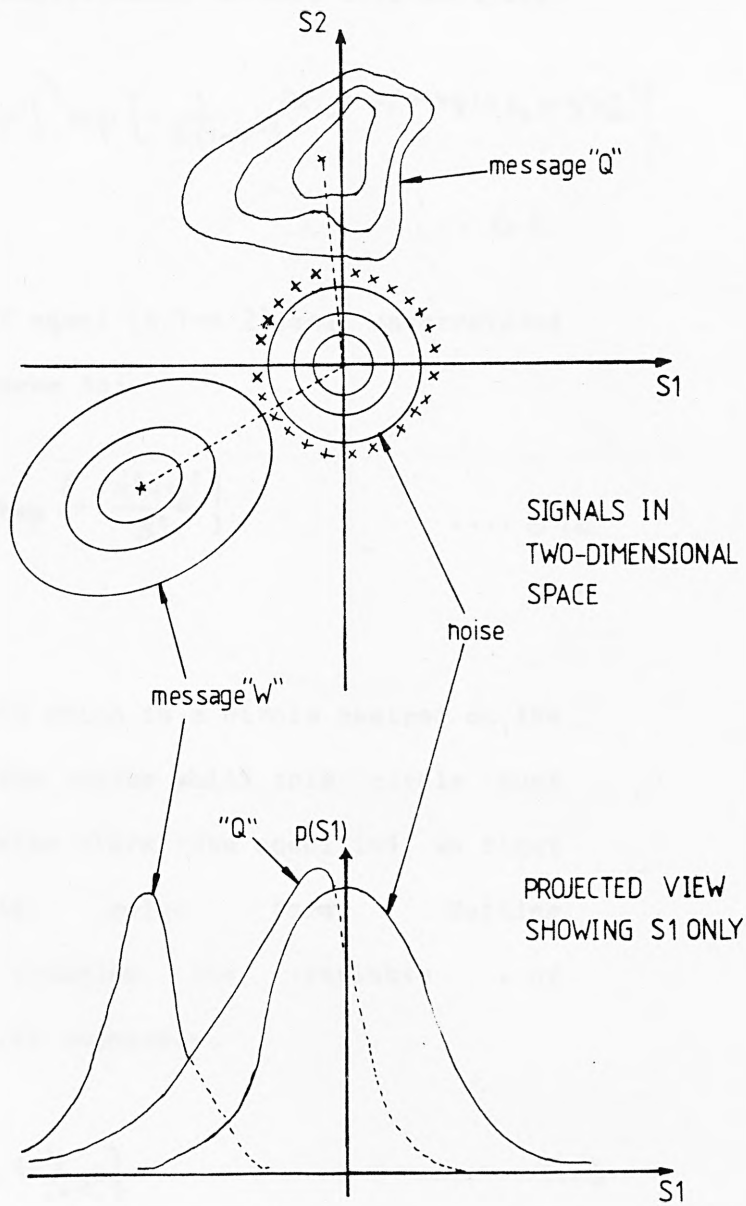


FIG. 6.6 SOME MORE REALISTIC MESSAGE SIGNALS

The pdf of the noise, assumed to have zero mean, is:-

$$p(x, y) = \frac{1}{2\pi S_1 S_2} \{1 - r^2\}^{\frac{1}{2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left(\frac{x^2}{S_1^2} - 2rxy/S_1 S_2 + \frac{y^2}{S_2^2} \right) \right\}$$

.... 6.11

Under our assumption of equal ($S_1 = S_2$), and uncorrelated ($r=0$) signals, this reduces to:-

$$p(x, y) = \frac{1}{2\pi S^2} \exp \left\{ -\frac{x^2 + y^2}{S^2} \right\}$$

.... 6.12

We require a threshold which is a circle centred on the origin. To calculate the radius which this circle must have to achieve the false alarm rate specified, we first reduce eqn. 6.12 to polar form. Setting $(x^2 + y^2) = V^2$, and changing the variable of integration, equation 6.12 becomes:-

$$p(V) = \frac{V}{S^2} \exp \left\{ -\frac{V^2}{2S^2} \right\}$$

.... 6.13

This is a Rayleigh distribution (app'B') and is circularly symmetric. It is a function of V only, not the angle ϕ . The false alarm probability we are

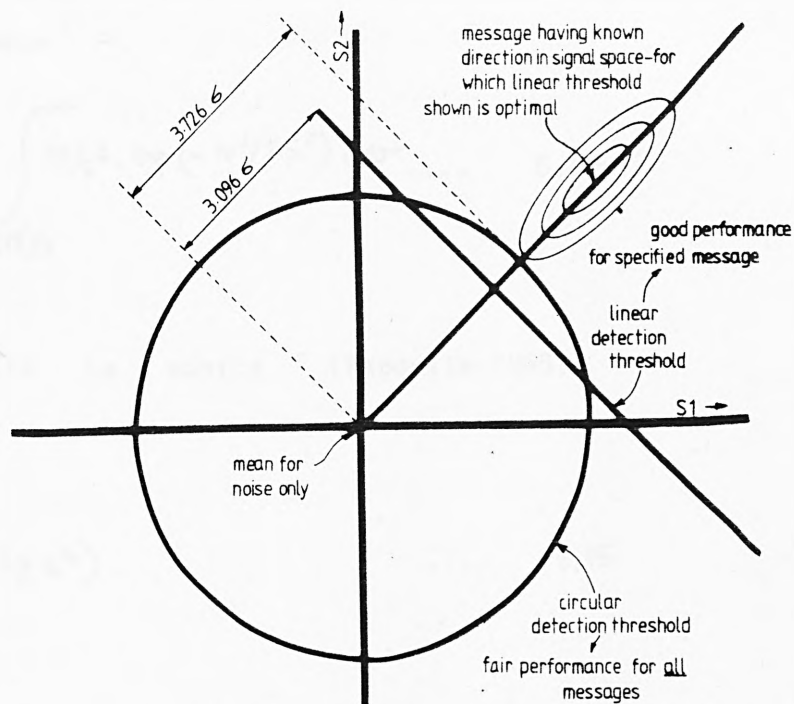


FIG 67 DEGRADATION IN DETECTION PERFORMANCE WHEN ALL FORMS OF MESSAGE MUST BE DETECTED WITH EQUAL EFFICIENCY

seeking, P' , is the probability that V exceeds a threshold $V(t)$, as indicated below:-

$$P' = \int_{v(t)}^{\infty} p(v) dv = \int_{v(t)}^{\infty} v/s^2 \cdot \exp(-v^2/2s^2) dv \dots 6.14$$

Fortunately, eqn. 6.14 is easily (Papoulis, 1965) integrated to yield:-

$$P' = \exp(-v^2(t)/2s^2) \dots 6.15$$

Not unexpectedly, this threshold is larger (reducing the detectability for marginal defects) than that for the simple (type (b)) detector, in which the threshold is set on the sum of the two signals. The distinction is evident immediately from figure 6.7, in which the area outside the circle is greater than that to the right of the linear threshold. However, by making this slight sacrifice in probability of defect detection for one preferred direction, we have made the detector sensitive equally to defects whose message vectors point in all directions.

To compute an estimate of the degradation in detectability thus caused to a marginal defect, we first re-arrange equation 6.15 to be explicit in $V(t)$, and obtain:-

$$V(t) = \sqrt{2} S \left\{ \lg_e (1/P') \right\}^{1/2} \dots 6.16$$

Substituting for P' the typical value 10^{-3} then yields:-

$$V(t) = 3.726s$$

This compares with the value for $V(t)$ of 3.096 (extracted from standard tables of the cumulative Gaussian distribution) obtaining when the two signals are combined linearly. The effect of the shift of threshold by about 0.6 standard deviations is to reduce the probability of detection for an 'additive' defect in Gaussian noise from 95% to about 50%.

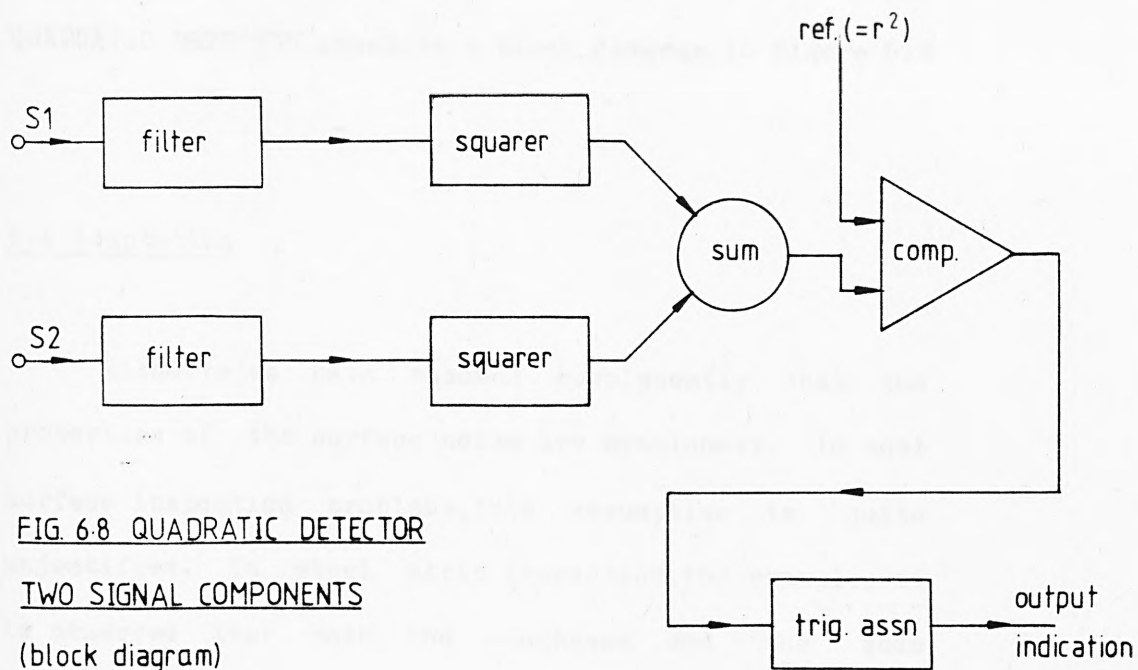


FIG. 6-8 QUADRATIC DETECTOR

TWO SIGNAL COMPONENTS

(block diagram)

This detector may be implemented as the WEIGHTED QUADRATIC DETECTOR shown as a block diagram in figure 6.8

6.4 Adaptation

Hitherto we have assumed complacently that the properties of the surface noise are stationary. In most surface inspection problems, this assumption is quite unjustified. In steel strip inspection, for example, it is observed that both the roughness and the mean reflectivity of the surface vary significantly even within a particular coil. This causes the mean and variance of the noise signal, the form of its pdf and cross-correlations between channels to vary. Some idea as to the extent of this variation, both within a given coil and between coils, may be obtained from data gathered during a series of on-line trials, performed by the British Steel Corporation at Port Talbot (Williams, 1973). The parameter recorded was trigger count measured during successive fixed length (10 second) time intervals, with a fixed detection threshold. The two curves plotted in figure 6.9 are records 1 and 9 from 13 collected during the trials, and are those showing the

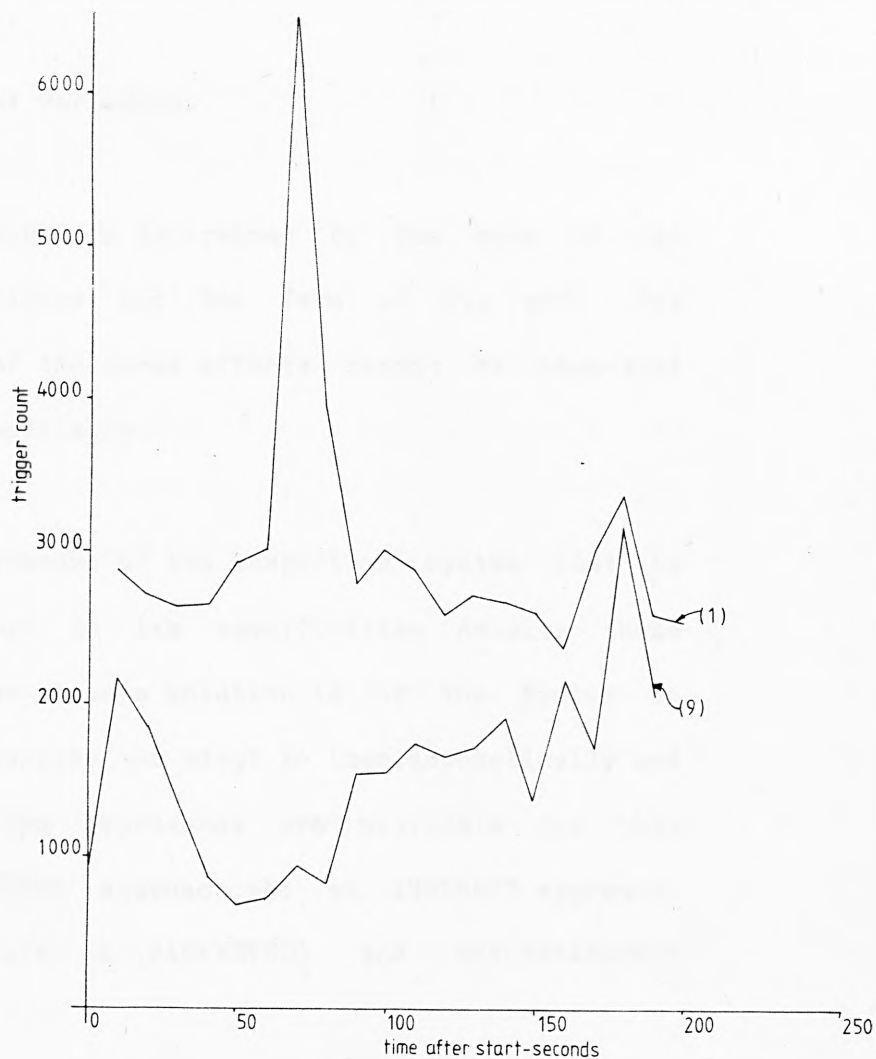


FIG 6.9 VARIATION IN TRIGGER COUNT MEASURED IN ON-LINE TRIALS

maximum range of variation.

Trigger count is determined by the mean of the signal, its variance and the form of its pdf. The contributions of the three effects cannot be separated from the data available.

The performance of the inspection system must be made to conform to its specification despite these variations. The obvious solution is for the system to sense the variations and adapt to them, automatically and in real time. Two approaches are available for this adaptation, a DIRECT approach, and an INDIRECT approach. The latter exists in PARAMETRIC and NON-PARAMETRIC versions.

In the direct approach, the number of triggers within a given area of surface (e.g. within five scans) is counted, and the threshold raised or lowered to eliminate the difference between the number of counts expected and the number obtained. This involves feedback, i.e., an error must exist to generate a correction signal. Further, the quantity measured has a variation due to sampling error which is large compared to its mean, thus

its estimation is unreliable. For a typical false alarm probability $P(fa)$ of 10^{-3} per sample, observation over 5 scans would yield an expected count of 5 triggers, with variance 5, since the count would obey a Poisson distribution. Finally, the occurrence of defects will raise the threshold, rendering the system locally less sensitive. The adaptive system might mistake low contrast defect signals for increased noise amplitude, if they occurred at high density, and might then raise the threshold instead of indicating detections.

The direct approach is however absolute in that the parameter immediately relevant in the detection process, i.e. the density of false alarm triggers, is controlled directly.

In the indirect approach, the probability density function (pdf) for the noise is used to set the detection threshold to give the false alarm probability desired. In the non-parametric variation, the pdf is first estimated, by constructing a histogram (by counting the number of samples having levels 1, 2, 3 ... 254, 255, 256) over a particular region of sheet. By making this large enough to include say 25,000 samples, (an area 10cm. by

10 cm. would suffice with the parameter values used in the simulation), the number of samples within each bin can be made large enough to make the estimate reliable statistically.

The threshold to achieve a false alarm probability $P(fa)$ is then determined by summing the contents of the histogram bins downwards from 256, until a total count S is obtained such that the ratio S/N , with N the total number of samples in the histogram, equals $P(fa)$. The threshold is then set to be that of the lowest bin whose contents must be included in the sum S . That is, if bins containing samples at levels 256, 255, 254, ... 231, 230 have to be included, the threshold will be 230. A sample at level 230 will generate a trigger, but a sample at level 229 will not.

In the parametric variation, a form is assumed for the pdf, and parameters describing this are then estimated. The noise when filtered is found to be a good fit to a Gaussian distribution, which is described completely by two parameters, mean and variance (app' B'). The arithmetic required for the parametric and non-parametric variations is roughly similar, but the

non-parametric approach makes no assumptions regarding the forms of the distributions, and must therefore be preferred.

Ultimately, a Kalman filtering approach could be used, to make the system optimally adaptive to changes in covariance for a multicomponent signal, as well as to mean and variance for individual components.

It is possible to achieve perfect adaptation using the non-parametric version of the indirect approach. This involves storing the analog data from say K consecutive scans, computing an optimal threshold for this data, and then applying this to the data for which it was computed. In view of the rapidly reducing cost of semiconductor memory, this possibility might well prove practicable.

The preferred solution in terms of simplicity and economy might however be to dispense with the adaptation. It is found from on-line measurements (Williams, 1978), fig. 6.9, that false alarm probability varies over a range 6:1 with a fixed threshold setting, at a maximum rate about 1% per metre of strip. Assuming

that the pdf of the noise is Gaussian and that the detection threshold is set to give the specified $P(fa)$ as a maximum during the variations, the separation between threshold and mean is 2.3 standard deviations. The 6:1 variation in false alarm probability then corresponds to an increase in the separation between threshold and mean of only 0.6 standard deviations. This will cause $P(cd)$ for a defect of fair contrast to vary between 95% and 85%, which is trivial. For a typical low contrast defect, $P(cd)$ will vary between 5% and 2%. Thus, dispensing with adaptation may well prove acceptable, in that the improvement in performance adaptation would provide might well not justify the extra cost entailed.

6.5 Hardware Implementation

In this section, we first discuss the general problem of implementing instrumentation for the automation of visual inspection in hardware which is capable of operation in real time. Then, a system is proposed for the specific task of detecting defects on the surface of cold rolled steel strip.

Having selected a signal processing methodology which is in principle capable of effective automated surface inspection, it is now necessary to establish that it can be implemented in hardware. Instrumentation for on-line surface inspection must be able to process data in each of several parallel channels at more than 30Mbits/sec., and must be tolerant of the harsh operating environment usual in manufacturing plants, i.e., of dirt, vibration, humidity, temperature variation, susceptibility to tampering, and so on. The hardware implementation must further be economic, i.e., must cost less (when amortised over a reasonable period) than the saving resulting from its incorporation into the manufacturing process.

Because of these exacting requirements, automated visual inspection lies at the upper limit of the capability of conventional signal processing hardware. Consequently, we first discuss some new techniques for high speed signal processing which are potentially applicable to automated visual inspection. These are evaluated in competition with conventional methods, which would involve special purpose processing architectures. It is already clear that a standard processor, such as a

PDP11 minicomputer could when used alone not do the job even with software written optimally in assembler since it would not be fast enough. However, a specially designed front end might well be used, to reduce the data rate before the general purpose computer becomes involved.

Processing could using conventional methods be either analog or digital, or could be some combination of these. Further, it might be desirable to perform some of the processing using state-of-the-art techniques such as coherent optics, fibre optics, surface acoustic wave (SAW) or charge coupled (CCD) devices. In discussing these techniques, we shall consider compatibility, cost, data throughput rate, availability, robustness, and range of operations available. For the latter, it is necessary to take into account the operations shown in table 6.1, which are required in the processing, and must be implemented by the hardware.

TABLE 6.1 PROCESSING OPERATIONS FOR AUTOMATED VISUAL INSPECTION

- (a) Linear amplification
- (b) squaring
- (c) linear combination of analogue signals
- (d) thresholding, involving comparison of an analogue signal with a reference level.
- (e) two-dimensional filtering (fixed filter response)
- (f) storage of several scans of signal, before and after binarisation.
- (g) convolution of the binary signal with binary functions
- (h) combination of binary signals logically
- (i) estimation (in real time) of statistical

properties such as the p.d.f of signals.

The processing required in automated visual inspection must operate with a bandwidth of roughly 30 Mhz in each of perhaps 3 channels, with 8 bit accuracy. Any analog operation may be implemented digitally and vice versa, though not necessarily with equal facility. The processing method which, whilst able to perform all the processing required, is least expensive and is practicable in a manufacturing environment, must be selected.

The rate, C at which a signal processing channel can handle information is governed by Shannon's second law:-

$$C = BW \log_2(1 + M/N) \quad \text{..... bits/second}$$

Here, BW is the channel bandwidth in hz., and M/N is the message/noise power ratio. White, Gaussian noise uncorrelated with the message is assumed. The maximum quantity of information which may be contained in a message of duration T seconds is thus $T.C$ bits. The maximum value for M/N may be restricted by poor dynamic range, as well as by self noise introduced by the

processing. The product T.BW (referred to as 'time-bandwidth product') is thus a useful parameter for comparing processing devices, but is not an absolute criterion. One can however be confident that good processing devices will be able to store long segments of signal, at high bandwidth, with large dynamic range and a minimum of self-noise.

Filtering is very frequently required in the processing, and it is well to remember that ANY response of practical significance may be obtained using a transversal filter, for which good design techniques (e.g. that of Parks and McLellan, 1972) are available.

Power consumption is also significant, since power dissipated by processing hardware has to be removed to prevent overheating.

We shall consider first the analog alternative, for which some highly efficient new methodologies have recently become available. These are optical processing, surface acoustic waves (SAW), and charge coupled devices (CCD).

COHERENT OPTICAL PROCESSING is based on a fundamental property of lenses, that the complex signal formed by a lens in its image plane is the Fourier transform of the complex distribution of light over the aperture in the object plane. In the transform plane, filtering is accomplished by blocking out, attenuating or phase delaying particular spatial frequencies. Correlation is implemented by placing in the transform plane a screen whose transmission is that of the conjugate of the complex (i.e., phase and amplitude) Fourier transform of the signal with which correlation is required. This exploits the theorem (mentioned in appendix 'B') that Fourier transformation converts correlations to multiplications, and vice-versa.

A second lens then performs an inverse Fourier transformation, restoring the signal to the spatial domain.

Such systems have time-bandwidth products of typically $3 \times 10^{**5}$, with dynamic range and self noise determined by the devices used to transform the signals to and from optical form. Acousto-optic modulators (Casasent, 1978) are commonly used to obtain optical

signals from electronic signals. In automated inspection applications, the original signal is optical anyway, and it may be possible to dispense with this first transducer (and with the conventional scanner!). The system developed by the Axel Johnson company and reported by Sjolin as long ago as 1972 (Sjolin, 1972) used the Fourier transforming property to increase the contrast of defect messages before transducing them into electronic signals. Unfortunately, at the current state of commercial availability, coherent optical processing cannot perform most of the operations mentioned in table 6.1, and a hybrid system would be necessary. In fact, the only operation for which coherent optics would offer attractive advantages is filtering. Coherent optical processors normally require a laboratory environment to function properly, which is free of dust and vibration. The advantages they offer in a few operations (notably two-dimensional filtering) would probably not justify their inclusion when only detection (rather than identification) of defects is required.

Circuit optical techniques (in which signals are conveyed along fibres of transparent material), were developed originally for transmission of telephone

signals, but have been further developed for use in state-of-the-art radar. In automated visual inspection, commercially available fibre optics could profitably be used for transporting signals within a plant, say from scanner to processor, since they are immune to the high level of electrical interference which is endemic in manufacturing plants, and have a high data rate. Recently, however, additional processing operations have become available in compatible form (Diffard and Chang, 1979). These include tapped, weighted delay lines, which can be used to produce transversal filters. An opto-electronic A/D converter has been produced, which gives 6-8 bits of resolution at 500Mwords/sec. Time-bandwidth products in excess of 10^{**6} are available, which is better by two orders of magnitude than SAW devices which offer strongest competition. Thus, a fibre-optic front end comprising transversal filters followed by opto-electronic conversion to a digital electronic signal may become competitive when the necessary components become available commercially at an acceptable price! Fibre optic processing must be considered an option for use in the future, when the technology has developed further. Unlike coherent optical systems, fibre optics can easily be packaged to

render them tolerant of harsh environments.

SAW devices comprise strips of piezoelectric material, typically quartz. An interdigitated electrode at one end, deposited as a metal film, excites an acoustic wave, which propagates along the surface at a speed typically 10^{-5} that of light. At the remote end, it excites another interdigitated electrode to generate an electrical output signal. Structures deposited or etched on the surface in the path of propagation modify the signal 'en route', by (for example) rejecting certain frequencies. Alternatively, the surface of propagation may be regarded as constituting a delay line, and outputs from intermediate electrodes distributed at intervals along the line, summed to implement a transversal filter.

SAW devices typically have delays of 0.5 microseconds (Gautier and Tournois, 1981), bandwidths from 5×10^4 Hz. to 0.4 times the centre frequency, and centre frequencies of 10^7 to 10^9 Hz., giving time-bandwidth products from 40 to 16,000. SAW devices thus make good analogue delay lines, filters, convolvers and so on. Signals in other bands can be processed using SAW devices by mixing the signal with a sinewave, and

filtering out an appropriate sideband.

CCD devices comprise in essence a sequence of storage sites diffused into a silicon chip. Charge is transferred from one site to the next using clocking pulses. The quantity of charge in each site codes the amplitude of the signal. CCD devices thus use a sampled analogue signal (appendix 'B').

In some chips, the charge at each of perhaps 64 sites may be monitored directly via a non-destructive output connection. The chips may thus be used to produce transversal filters. For other chips, connections are possible only to the input and output ends of the chain. These have many more sites, perhaps several thousand, and are useful for example, as delay lines. Clock frequencies from 2kHz to 20 Mhz are possible. The lower limit is set by charge leaking away from the storage sites, the upper limit by a degradation in transfer efficiency (which is typically 0.99999 in the latest devices) as the clock frequency increases. Time bandwidth products may be as great at 10^{**3} per chip, but are usually much smaller.

CCD devices are being developed (Eden and

Dahimy,1979) using gallium arsenide instead of silicon. These are difficult to manufacture, and are hence not yet available commercially. They promise clocking at rates in excess of 1 gigahertz (10^{**9} Hz.)

Operations such as summation of signals, multiplication by a constant and amplification are readily performed by analogue operational amplifiers implemented as analogue integrated circuits. However, systems based on such devices are normally hard wired, and cannot easily be programmed using software.

The alternative approach is to process digitally, by sampling the signal and quantising it to perhaps 256 levels (8 bits) at the camera output. Use of an adaptive linear amplifier immediately before digitisation is highly desirable, to make full use of the digital system's limited dynamic range, in the face of gross variations in signal due e.g. to change in overall surface reflectivity. Thereafter, all processing may be digital.

Digital hardware which processes data at up to about 40 M words/second is readily available commercially as high speed TTL. Emitter-coupled logic (ECL) can

improve this by perhaps half an order of magnitude. The initial stage in a digital processor is an A/D converter, which takes as input the analog (or sampled) signal generated by the scanner, and generates as output a sequence of digital words. In inspection systems eight bit words generally provide sufficient information. To include all information present in a signal covering a passband extending from D.C. to 30 Mhz, without introducing error due to aliasing, requires samples spaced according to the Nyquist criterion (appendix.'B'). Given that the boundary of the passband at 30Mhz will not drop sharply to zero, samples separated by 10^{-8} second are required. Converters operating on the 'flash' principle are commercially available which meet this requirement. However, this rate is two-and-a-half times faster than that of fast TTL.

The principal disadvantage of digital implementation is the high package count required, which leads to high power dissipation, many interconnections (which degrade reliability), and difficulty in isolating components from interference due to transients. The cost of the chips, even of memory sufficient to store several consecutive scans, each containing 2 thousand samples

digitised to 8 bits, is however very reasonable. All operations listed in table 6.1 are readily implemented digitally. The 'square' matched filter recommended in section 4.6 is implemented merely by summing N consecutive samples, for example.

Another possibility would be to use a two-dimensional array processor such as a CLIP [Duff et al., 1973]. This is a programmable digital device which gives a speed improvement between 10^6 and 10^4 times that for a fully serial processor, depending on the operation. However, CLIP is currently expensive (about \$70,000), has a resolution only 96 points by 96, and is not yet available commercially. Although it processes in parallel, it must be loaded serially. It is thus apparently uncompetitive for a system solely to detect defects, but might be appropriate for the more complex instrumentation required for defect identification.

In chapter 1, we remarked that the signal processing required for defect detection in automated visual inspection is closely analogous to that required for target detection in radar. A recently published

comparison of digital, CCD and SAW techniques for radar, by Margoun and Mchugh (1979) is thus of interest. They conclude that unless the highest processing rates are indispensable, then SAW devices (which are available commercially) are best avoided, because the mixing operations required to match operating frequencies introduce an excessive complication. They recommend digital techniques using LSI as current front runners, but promised developments in CCD devices (notably, decrease in COST) are likely to render them competitive, for operations for which CCD devices are applicable. CCD devices are potentially superior in size, weight, and power consumption. They operate on a sampled, analog signal (appendix 'B'), which involves no loss in information provided the Nyquist criterion is fulfilled. The signals used in digital systems require on the other hand quantisation of samples to fit a finite wordlength, which introduces noise, limits dynamic range, and may cause errors to appear in computations. CCD devices used as processors are of course immediately compatible with CCD devices used in cameras, another reason for preferring them.

Digital systems have a definite advantage in being

programmable, using software or firmware. Thus, the structure of a system may be changed very easily. Most important, the values of operating parameters (e.g., the response of a transversal filter) may easily be changed. However, special purpose architecture is probably essential to achieve the high operating speeds required, and programmability is really vital only during research and development, until an acceptable processing scheme has been selected. Otherwise, programmability may only introduce the opportunity for degrading system performance by tampering.

In figure 6.10, a system proposed for the specific task of detecting visible defects on cold rolled steel strip is shown in block diagram form.

The strip is scanned by 3 CCD linescan cameras (1728 or 2048 elements per line), synchronised so that a vector signal is obtained. One camera views at the specular angle, the other two at large (~ 30 deg.) angles at either side of the specular. Non-coherent illumination is used, obtained preferably from a fluorescent tube driven by high frequency A.C., with its emission energy concentrated around 0.5 micron

wavelength. A system of cylindrical lenses is used to concentrate the illumination along a narrow strip perpendicular to the direction of movement of the inspected material. The CCD scanner generates its signal as a sequence of pulses of analog signal, which is immediately compatible with processing using a sampled or digital approach. If purely analog processing is desired, a sample/hold is necessary. At the highest linespeeds, the samples would probably have to be elongated in the y-direction, with an aspect ratio of about 3:1. This is likely to enhance rather than degrade the contrast of many defects, since these tend to be elongated in the y-direction.

The majority of the defects should be readily detected by the weighted quadratic detector shown in figure 6.8, which hence comprises the initial processing. The squared signals in each channel are summed and compared with a threshold to generate trigger pulses. This gives the useful advantage that both positive and negative signal excursions in all channels can be detected with one threshold test. An option to vary the gain in each channel independently gives the detector the facility to operate ellipsoidally rather than merely

spherically, so that the system can be adjusted to incorporate more detailed information regarding the distribution in signal space of message vectors, which may become available following extensive experimentation on a particular production line.

On the specular channel only, a two-dimensional analogue store is provided, which holds the signals from the latest K scans. The minimum K is about 10 but a value of 20 may give better performance. A two-dimensional longitudinal filter (as described in section 4.3) operates on the data in this store, as a matched filter to increase the contrast of longitudinal defects such as seams. The filter is implemented by tapping each element of a region of the stored area L samples wide by K samples long, and forming a linear weighted sum of the tap outputs. This forms a two-dimensional transversal filter. The output of this filter is compared with a threshold, and the trigger pulses obtained combined with those from the main (weighted quadratic) detector by 'ORing'.

The longitudinal detector described in section 4.2 is implemented by forming a linearly weighted sum of m th

signal samples in the first $(K - 1)$ scans stored, which is compared with the m th sample in the K th scan to provide the trigger pulse. Since the whole of the signal is convolved with this detector, one filter and comparator suffice to apply the longitudinal test over the whole width of the strip. This is appropriate for detecting sticker wrench, but gating may be added to define regions close enough to the edge of the sheet to contain edge strain. The trigger pulses provided by this detector are again combined with the others by 'ORing'.

Trigger association is applied to the triggers after combination to eliminate false alarms using a binary store, which holds the triggers from the last N scans. These are convolved with a window N samples in the y (sheet movement) direction, by M samples in the x (scan) direction, and the number of triggers present is summed to implement the 'association in a box' method suggested in chapter 5. Simulations described in that chapter suggest that values of 10 will suffice for both N and M .

The output of the trigger association, within intervals corresponding to the inspection strips shown in

figure 1.1, constitutes the inspection data provided as output indication.

The detector for chatter marks (the only global defect considered) is not included in figure 6.10. It is possibly best implemented by software using an auxiliary microprocessor. This would sum the sample levels in N consecutive scans, and Fourier transform the N-sample long vector thus obtained. At linespeeds close to the maximum envisaged of 25 metres/sec., inspection would have to be on a sampling basis. 100% coverage of the strip surface would not be possible, because of the time which a microprocessor would take to perform the computation. A hardwired processor giving 100% coverage would be possible, but unnecessarily expensive.

If required, the unprocessed outputs of the three channels could be combined (as suggested in figure 3.2) to yield an indication as to whether the defect absorbed, deflected or scattered, as an aid to subsequent identification of the defect type.

The two two-dimensional filters suggested in figure 6.10 share a common organisation, although one filter is

analog and the other binary. This is shown in figure 6.11. Each scan is circulated into a store which is effectively a delay line, under the command of a clocking pulse. The line holding each scan holds the same number of samples as are gathered by the CCD cameras during each scan. The first K elements are provided with taps. Having passed through the n th delay line, the signal is routed into the $n+1$ th and so on, and data emerging from the N th (i.e., last) line is discarded. Thus, the untapped part of the 5th stage delay line in figure 6.11 could be omitted if there were no following stage. This architecture enables the signal to be convolved with an arbitrary filter of dimension L samples by N , to allow filtering (analog or binary) to be applied to the whole signal, without 'edge effects' such as those described in section 5.2 which necessitate overlapping boundaries.

It remains merely to suggest whether analog or digital hardware should be used to implement this particular system. We have already noted that digital hardware could perform all processing, but that a fast A/D converter would be needed in each channel, and an array of D/A converters might be needed in a real-time two-dimensional analog filter. Fast digital squarers are

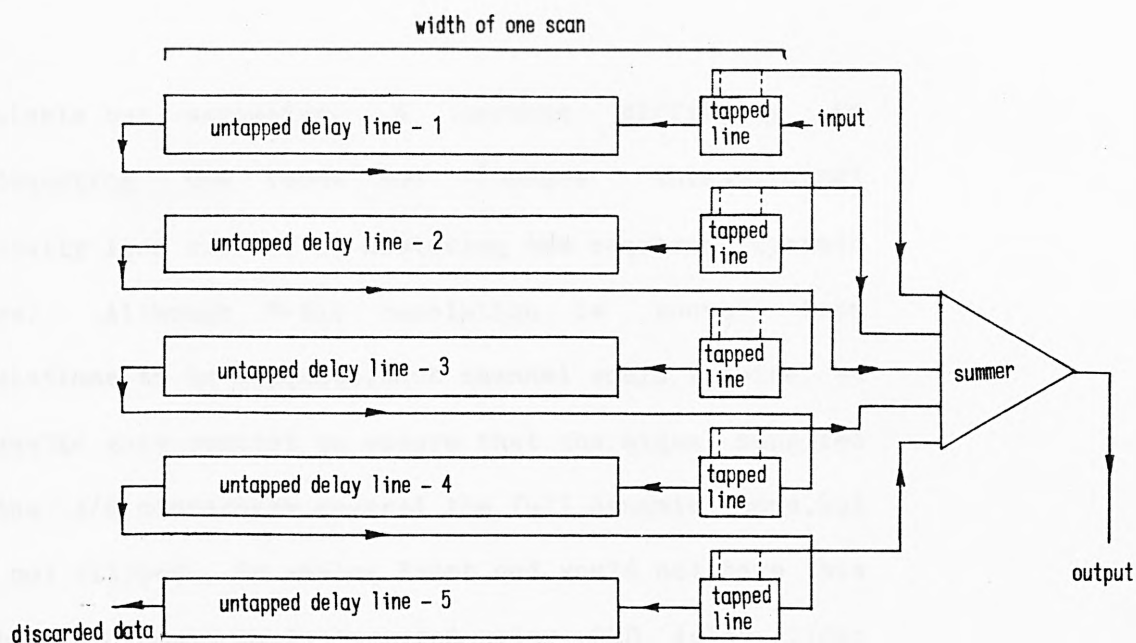


FIGURE 6.11 TWO-DIMENSIONAL DELAY LINE FILTER

(analogue and binary forms have the same organisation)

available but expensive. A serious difficulty in implementing the front end (before thresholding) digitally lies however in achieving the required dynamic range. Although 8-bit resolution is known from simulations to be adequate, each channel would require an automatic gain control to ensure that the signal supplied to the A/D converters covered the full dynamic range, but was not clipped. An analog front end would not have this problem, and could be implemented using CCD delay lines for the filters. SAW filters would provide a more complex alternative if CCD proved not to be fast enough. Thus, it is recommended that a sampled analogue front end processor be used.

The processing required following thresholding can be implemented easily using binary logic, either TTL (fast but high-current) or CMOS (low current and less susceptible to interference, but a little slower). There is no competing alternative which merits consideration. A programmable digital system (based presumably on a standard minicomputer) would be needed, to co-ordinate the process overall, and organise the output data for presentation to a human operative, and for record keeping.

6.6 Conclusions

In this chapter, we have discussed the best form which the complete system for automated defect detection, formed by combination of individual methods discussed, and implemented in the best possible way in hardware, should take.

This depends on properties of both noise and message components of the vector signal. The properties of the noise are understood quite well (chap.2), but are highly variable. The chief difficulty imposed by the noise is in fact the need to accommodate variation. However, the only definite information concerning the message signals generated by defects (discussed also in chapter 2) is that they are almost invariably unknown. The inspection system must thus be configured to respond to all significant forms for the defect, which generally leads to a slight loss in overall performance. In this circumstance, given a vector signal, a configuration in which thresholding is applied to the sum of squares of the signal components is recommended.

It has been shown that if (as is usual) both message and noise signals are largely uncorrelated between channels, then the best method of combining the components is to sum the analogue signals, then to threshold this sum, and apply association to the resulting triggers. This gives both best performance and simplest hardware.

If however more definite information is available concerning signals characteristic of a few kinds of important defect, then special purpose channels may be included to accomodate these. This procedure is recommended for sticker wrench, for example.

A suggested system for the particular case of cold rolled steel inspection is shown in figure 6.10

It is shown that although it is not difficult to make the system adaptive to variation in noise level, it might well be preferable to ignore the variation, since the improvement in performance obtained by adaptation might well not justify its cost.

Finally, the implementation of the proposed system

in hardware has been discussed. The emergence of new technologies has provided a wealth of new forms of processing, but most of these have not been developed far enough commercially to be applicable with benefit. Nevertheless, implementation in hardware should cause no difficulty, despite the high linespeeds required. A hybrid system is recommended, with a 'front end' using analogue processing by special purpose hardware, probably CCD devices. This would be hard wired. Its function is to extract from the data gathered by the scanner information useful in the inspection process, resulting in a substantial compression of data. It is followed by processing using binary logic, for example TTL or CMOS, to combine triggers generated by the various channels and eliminate false alarms. The final stage would comprise a programmable general purpose microprocessor, for system co-ordination and result analysis. The instrumentation suggested would probably cost less than \$100,000 per installation in 'batch' quantities, could be developed quickly, and would be viable economically.

CHAP.7 CONCLUSIONS

7.1 Introduction

In this chapter, the results obtained in the previous chapters are summarised and combined to provide an overall conclusion. This is in two parts. The first (section 7.2) provides general conclusions reached concerning the theory and methodology of signal processing as applicable to automated visual inspection, particularly when applied to the detection of defects visible on the surface of material produced as a continuous sheet or strip. The second part (section 7.3) recommends an optimal system for detecting defects on a specific material, cold rolled steel strip, which has been given special attention in the thesis as a sample material for validating the various processing concepts which have been investigated. The problems involved in hardware implementation are discussed with particular reference to this material. Finally, in section 7.3, the shortcomings of the contemporary state-of-the-art in the theory and analysis for automated visual inspection are reviewed. A direction is then suggested for future work.

7.2 Conclusions Regarding Signal Processing Methodology

We have shown that concepts and methodologies for detecting message signals accompanied by noise developed originally chiefly for use in target detection in radar and sonar may be applied very effectively in automated visual inspection. Their use is particularly valuable in the detection of surface defects of low contrast. Modifications and extensions have however been necessary, particularly since the message signals in inspection problems are of unknown form, may vary widely for a particular task, and even within a particular defect class. The signals encountered in surface inspection are, further, vectors having typically at least three components which may be partially correlated, and describe a two-dimensional phenomenon. The accompanying noise is generally non-Gaussian unless filtered.

Thus, the detection process has been reduced to a canonic form, in which it is broken into a sequence of successive stages, namely contrast enhancement, thresholding and trigger association. These may to a first approximation be analysed independently.

Methodologies have been developed to perform the processing required by each stage in the canonic model, and have been validated by experiment performed by computer simulation. Procedures for selecting the parameters which must be specified at each stage have been developed, to yield optimal operation, even if the forms of signals and other essential data are not available explicitly. Most important, an original and quantitative theory has been developed for stage 3 in the canonic detection process, in which false alarm triggers due to noise are eliminated selectively, but triggers due to defects are retained. The result of these labours is thus a sound, general, quantitative and comprehensible theory for detection applicable particularly to automated visual inspection. We have also shown how the theory may be applied to the design of instrumentation systems for particular cases.

Two principal methods have been proposed for contrast enhancement, namely matched filtering and weighted linear combination of signals. The absence of reliable and accurate descriptions of the waveforms characteristic of surface defects makes design of optimal

matched filters hazardous. A procedure for determining an optimal filter when the signal waveform is not known explicitly has been suggested, and investigated theoretically. For surface defect detection, a rectangular match waveform has been shown to be about as efficient for contrast enhancement as the triangular waveforms used initially as the MFB detector, but is much easier to compute. Simulation in which a two-dimensional matched filter was used to enhance seams has confirmed this prediction.

The linear weights giving optimal contrast improvement by linear combination are also impossible to compute unless the message waveform is known explicitly as well as the correlation properties of the noise. They may however be determined experimentally, and this approach has been demonstrated experimentally to give worthwhile contrast enhancement for sticker wrench.

The Neyman-Pearson has been shown to be the best approach for setting the decision threshold which generates triggers indicative of defects from the analogue signal. Alternative approaches which promise superior performance cannot be used, once again, because of

lack of information concerning the properties of the signals generated by defects.

Although adaptation of the detection threshold to accomodate variation in surface noise would certainly improve the performance of a detection system, the degradation resulting if it is omitted is not particularly large. Thus, a non-adaptive system may well prove acceptable, in which a small loss in performance is traded for a substantial gain in cost.

The improvement in performance produced by using trigger association has been shown to be dramatic. A reduction in false alarm trigger probability of ten orders of magnitude is readily obtained, without significant message loss. It is surprising therefore that trigger association approaches are rarely used for discriminating message from noise in image analysis. This is probably because, although earlier investigations (notably Lacroix, 1979) had indicated its value generally, no quantitative analysis had been performed, and the scale of its effectiveness appears not to have been realised. It was shown in chapter 5 that the improvement obtained from trigger association is largely independent

of the scheme used to implement the association, for a given data storage capacity. Values for alternatives have been computed and are tabulated in the chapter. The alternative chosen should then be that which is easiest to provide in hardware.

Early in the investigation, the detection performance demonstrated during simulation was very poor for certain defects. More powerful processing methods were consequently applied, and improved the performance. Eventually, however, all information available in the original signals had been exploited for a particular type of defect (sticker wrench), without acceptable performance being achieved. In this case, further improvement could be obtained only by improving the performance of the scanner, either by modifying its operating parameters, or by adding another component to the vector signal, or by use of an alternative type having improved sensitivity. The first two possibilities would require re-scanning of the data base, the second additionally the installation of one or more additional photomultiplier sensors. Thus, in this investigation, only the third possibility could be explored. An analysis provided in chapter 2, initiated by experiments performed

by the SIRA Institute, indicated that scanners based on CCD arrays should provide better defect contrast, because of the absence of speckle noise. Experimental work by Edy (1977), and more notably Yaxley (1979) demonstrated the improved effective sensitivity of CCD sensors conclusively. These devices are new, having first appeared only in 1971, and are being developed rapidly. They will almost certainly dominate as sensors in visual inspection instrumentation, as manufacturing problems such as low yields and fixed pattern noise are overcome, and prices fall.

It has also been demonstrated that for defects whose message location in component space is known, a worthwhile increase in defect contrast is available by forming a linear weighted sum of the components of the vector signal and treating this as a scalar. However, there is as yet no analytic method for finding the optimal weights, and design by trial has been necessary.

Implementation in hardware of the processing suggested should however present no problem, even at the highest linespeeds. It should become even faster and

cheaper as hardware development proceeds, and new processing methodologies (CCD, SAW, optical fibres etc.) become available in commercial hardware.

7.3 Conclusions Regarding Systems Aspects and Hardware Implementation

These conclusions are presented with reference particularly to the inspection of cold-rolled steel strip. Firstly, we have established that the inspection performance required for industrial viability should be achievable. This specifies detection of 95% of defect occurrences, with no more than 0.2% of good material wasted due to false alarms, as detailed in section 1.2. Although only 35 of 52 classes of defect known to occur on cold rolled steel strip were actually processed in the simulations, the 17 not experimented upon are sufficiently similar to those which were, that they should succumb to the same methods of processing. The results obtained suggest that incoherent illumination should be used, and initial conclusions regarding the detectability of sticker wrench will have to be confirmed using a much wider range of samples.

It is further established that the processing speeds required in the inspection should also present no insurmountable difficulties, though, as discussed in section 2.2, controversy remains as to whether CCD array line scanners are sufficiently sensitive for a linespeed of 25 metres/ second. A calculation provided in section 2.2 suggested that despite doubts expressed by the SIRA Institute CCD devices should be adequate, but this also remains to be confirmed by experiment.

There is no doubt at all concerning the ability of commercially available hardware to operate fast enough, for all the processing operations which are required.

Regarding the system configuration, particularly for cold rolled steel strip, it seems that filtering of the analog signal for contrast enhancement (stage '1' in the canonic form) before thresholding is justified only for certain kinds of 'difficult' defect. For these it would however be worth making the filter two-dimensional. The shape of the filter has been shown not to be critical, given the wide variation observed in the properties of defect message signals.

7.4 Suggestions for Further Work on Automated Visual Inspection

Tasks in automated visual inspection form a hierarchy of increasing difficulty. Detection of surface defects against a background of noise is required in the inspection of many materials, including sheet metals such as stainless steel and tinplate, magnetic recording surfaces on tape and disc, photographic paper and film, rubber and plastic, printed materials, textiles, and so on. Satisfactory systems have been developed for inspecting some of these, and the methodology developed within the thesis should be readily applicable to the remainder. When the material surface is patterned, or when the material structure introduces a quasi-random noise which is larger than most defects, then more powerful processing will be required. This applies particularly to the inspection of textiles, which is known to constitute a significant problem which is still largely unsolved.

The effectiveness of surface defect detection

systems is often limited by the presence on the surface of marks whose contrast is as great as that of typical defects, but which do not degrade the utility of the material. There is, further, often a requirement for defects to be identified by type, so that material may be directed to the most appropriate user, or so that corrective action may be applied. Though there has been some progress in automating the identification by type of visible defects (Norton-Wayne and Hill, 1974), (Logan and Macleod, 1974), (Logan, 1974), (Popovici, 1976), (Hill, 1977), (Saridis and Brandin, 1979), (Chittineni, 1982) this problem remains essentially unsolved, and merits further investigation. Although adequate methods (such as feature space pattern recognition) are available for the signal processing required, processing in simulation of vast quantities of stored data is necessary. Identification is thus an order of magnitude more difficult than detection. The most successful work thus far reported on surface defect identification is that by Chittineni (1980), on magnetic recording materials and abrasives.

Automating the visual inspection of discrete objects is generally more difficult than inspecting strip

products. Additional degrees of freedom such as position and orientation are introduced into the sample, which the inspection system must tolerate. Further, the values of individual items are generally much greater (between 2 and 3 orders of magnitude), than for strip material having the same area, and the efficiency required of the inspection is correspondingly higher. On the other hand, the decision implied by a detection process may for discrete objects be more specific (i.e., to reject or accept INDIVIDUAL ITEMS) and it is possible to extract particular items from a stream for more thorough examination, if desired. Thus, the cost per unit throughput acceptable for the system is much greater. Also, the Neyman-Pearson criterion recommended for the decision stage in strip product inspection is then no longer optimal.

For the automated detection of defects in printed circuit boards, for example, (Norton-Wayne and West, 1992), an inverse Neyman-Pearson criterion is appropriate, in which the detection threshold is set to perceive virtually all (say, 99%) of significant defects. A second stage in the inspection can then be used to eliminate the large number of false alarm indications

generated. This may (in early stages of the development of instrumentation) be visual. The instrumentation guides the human eye to the region of the putative defect, eliminating the boredom, fatigue and consequent inefficiency which would result if the eye had to scan unaided. Thus, sequential analysis (Wald, 1947) may be applied to improve overall efficiency.

When steel and other strip products are inspected, it is not generally possible to isolate individual defect indications or particular regions of material, and quality designation has to be made for a complete coil. Defects present within a coil are counted, and provided the count is less than some threshold, the coil is designated as being satisfactory. The false alarm rate must thus be kept low. There is thus a fundamental difference in the operational configuration between inspection for discrete (e.g. PC boards) and strip material. The former may be repaired immediately and, if not repairable, individual items can be discarded immediately. Defects in strip materials are on the other hand often not eliminated from the manufacturing process until the final product has been assembled.

Despite the appearance of CCD cameras, there is still scope for application of laser scanners, and for improvement in their design. Laser scanners have very high resolution and sensitivity, and are easily configured to yield a vector output signal. Further, as was pointed out in chapter 2, defect contrast should be much improved by scanning with radiation whose wavelength is commensurate with the size of the defects, since the speckle introduced by the surface roughness would be correspondingly smaller, and there would be a 'resonance' effect in the interaction between radiation and defect. A carbon dioxide laser operating at 10.6 micron wavelength could be used; the necessary sensors have been developed for military purposes. By using even longer wavelengths, e.g. in the sub-millimeter region at 100 microns, the rough surface could be made into a weak phase reflector, so that information describing the statistical properties of the surface contour could also be extracted. Further, innocuous marks caused by dirt would be transparent to the long wave radiation, which would be affected only by the structure of the metal beneath. The technology for handling electromagnetic radiation is however poorly developed in this region, and hardware would be expensive.

There is some scope also for the exploitation of new signal processing methods which have become available. Although O Bray's early work suggested (O Bray, 1973) that time series analysis was not a profitable approach in surface defect detection, the vast development in this area that has occurred subsequently may render his conclusions invalid. The new time series methods are, most particularly, inherently multi-dimensional whereas O Bray's approach was one dimensional.

Automated visual inspection thus has a bright future. The factor which dominates the rate of progress in the application of machine vision in industrial automation is the rate of progress in imitating the operation of the human eye-brain system by signal processing machinery, particularly when semantic concepts are involved. Other important factors include the cost of mounting suitable investigations (excessive because of the simulation effort required), and the introduction of concepts and methodologies unfamiliar to and untrusted by the overwhelming majority of practicing engineers.

The recent realisation by the Establishment in this

nation (and others), that automation is essential to continuing economic health, and that this automation requires the cost-effective automation of vision, and the development of signal processing whose capability approaches human intelligence, should facilitate progress greatly.

APPENDIX 'A' - DATA BASE and PROGRAM PACKAGE

This describes the procedure for validating data recordings supplied by the SIRA Institute, and adding them to a central data base for experimentation using computer simulation. The program package used for implementing the various processing methods is then explained.

The data base was configured to provide signals characteristic of those obtained from the surface of cold rolled steel strip when defects are present, in a form suitable for analysis on a digital computer.

Appendix A.1 Data Base

The data was gathered by scanning sheets of steel containing defects of the 37 types contained in the target performance specification. A few sheets containing other types of defect were also included, together with some sheets which were defect free. The sheets had been gathered by BSC from the rolling mills at Port Talbot and Llanwern, and were scanned at the SIRA Institute using automated instrumentation and stored on industry standard magnetic

tape. This is the form in which the data arrived at The City University.

The sheets were each about 30 cm.square to facilitate handling, and to ensure that the defects were specified adequately without an excess of redundant data. It had been intended originally to gather at least 10 sheets containing defects from each class, however the samples were obtained from only two sources over a restricted period of about two years. Several defect types occur only sporadically (e.g. chatter marks), and for some of these, it turned out to be impossible to gather the number of samples originally intended. To ensure that the data included a good statistical spread of properties, each sheet was extracted from a different coil, and coils were chosen which had been rolled on the widest possible selection of lines and at widely separated times. Further, the sheets were chosen to be as flat as possible, and to include only one kind of defect.

Data was supplied to TCU on 9-track industry standard (1.25 cm. wide) magnetic tapes, at 315 bits/cm. The scanner used was a SIRA type 1500 laser scanner, as illustrated in figure 2.3. It contained three

sensors, one accepting light at the specular angle (15 degrees from normal), a second at 7 (nominally 10) degrees from specular, a third at 30 degrees from specular. During data acquisition, one scan was recorded from each sensor in turn, then the sheet was advanced by one scan width (i.e., 1 millimetre) and a further series of scans taken. Each scan could include up to 1024 samples of signal amplitude, each sample quantised to 256 levels (3 bits). The samples were spaced by 0.4 mm in the scan direction. In addition to these scans, the data for each sheet included trigger signals (both positive and negative) generated by a hardware SIRA detector which was part of the data gathering system. This was not adjusted to follow variation in surface roughness from sheet to sheet, thus its output provides a rather pessimistic impression of defect detectability. It was included chiefly to facilitate validation of the data. Information such as the serial number of the sheet, the date on which it was gathered and the number of scans was also included.

Documentation was provided with the tapes to facilitate validation and interpretation of the data. This comprised:-

REC'D 2/2/2011

S I R A

CO-ORDINATE POSITION OF DEFECT ON SHEET STEEL SAMPLES

DEFECT: ROLL MARKS, BEUSE

OTHER MARKS :- $\left\{ \begin{array}{l} 1 : \text{scratch} \\ 2 : \text{roll marks} \\ 3 : \text{unpainted mark} \end{array} \right.$

SAMPLE NO: 16C-11 SEVERITY MED

X=DEFECT

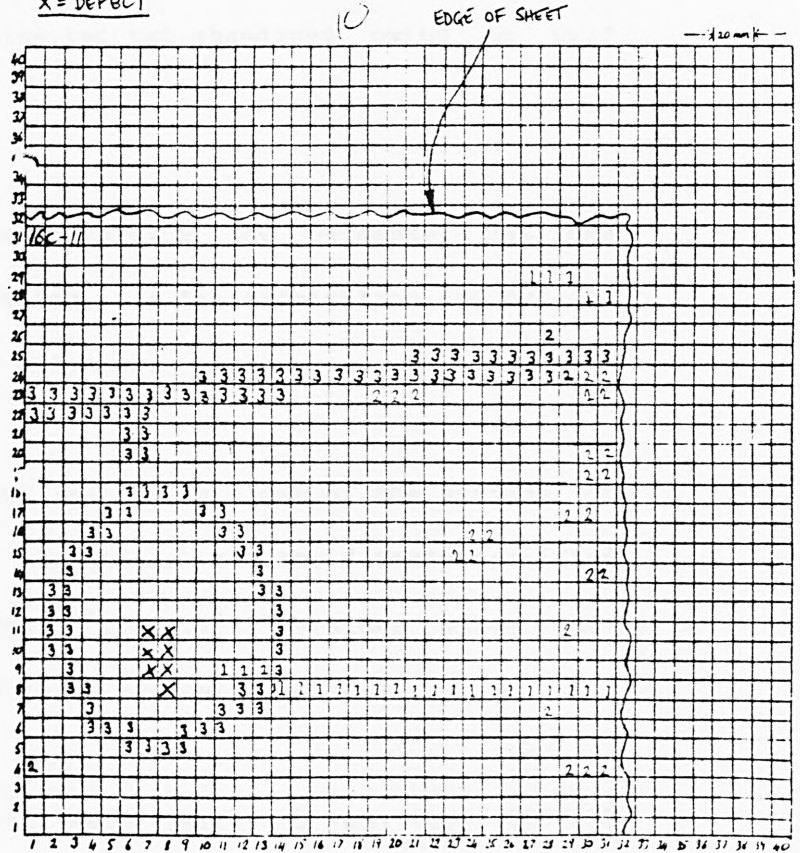


FIGURE A.1 LOG SHEET SHOWING DEFECT LOCATIONS
INDICATED ON SURFACE MAP

(a) The teletypewriter record produced when the sheet was scanned, which contained the sheet identification number, the number of scans on each sheet, the number of sheets on each magnetic tape (normally there were 7), warning of the presence of sheets whose scanning was started but abandoned owing to poor sheet quality, etc.

(b) A log sheet recording the parameter settings used during the scanning, e.g. the attenuation and cut-off levels for the SIRA detector.

(c) A log sheet containing a map of the surface of the sheet, with the locations of the defects specified within 1 cm. squares. (fig.A.1). Markings other than the defect are seen to be present sometimes.

The simulations were performed on the CDC7600 machine at the University of London Computer Centre (ULCC). This was the fastest computer available for the project, and had some useful peripherals such as a microfilm plotter. It was unfortunately operable in batch mode only, with programs submitted on punched cards.

Addition of a SIRA generated data tape to the data base being accumulated at ULCC required the following series of operations. First, the data was translated from 9 track tape to 7 track. Then, the data was re-formatted to incorporate some additional information, and to obtain more compact storage. This included:-

(a) Separating the analog data from the three sensors, and storing the data from each sensor in a single two-dimensional array 330 words by 100. Seven samples of amplitude are packaged into each 60 bit word, thus the maximum sheet size was 700 samples (28cm.) in the scan (x) direction, and 33cm in the perpendicular (y) direction. This enables the data from one sensor to be held in RAM store for a whole sheet.

(b) The trigger data from the SIRA detectors was re-formatted and stored in a one-dimensional array containing the co-ordinates of up to 4000 triggers. Negative going triggers were stored for the specular scan, and positive going triggers for the two off-specular scans.

(c) The co-ordinates of the 1 cm.squares within

which defect was present and also the name of the defect (e.g. 'lamination') was punched onto cards and added to the record. If the defect occupied more than 100 squares, it was considered to cover the whole sheet, in which case square (1,1) only was marked, to indicate that the defect locations had not been omitted accidentally.

This was performed using program F063.

Some errors are inevitable during data acquisition, such as sheets mis-numbered, or gathered upside-down. These were detected by a validation stage, which was followed by editing to correct the errors.

Validation comprised comparing videoprints of triggers generated by the SIRA detector with the defect maps, and examination of the perspective plots. Videoprints and perspective plots comprise data plotted on microfilm to facilitate visual examination. They are explained later in the appendix. Properties such as the presence of triggers arising from the identification number stamped at the top left-hand corner of the sheet, and whether trigger clusters evidently due to

defects fitted into the defect boxes were utilised in validation.

It became apparant during validation that about 20% of sheets were bent sufficiently to corrupt the data, that often marks other than the 'official' defect were present, sometimes at a higher contrast than the defect, and that rust appeared very quickly, even when the sheet was protected by a coating of palm oil. Because of rust, only two out of ten sheets deemed to be defect free were of use for setting thresholds.

Mislabelling was easily corrected, and upside-down sheets were inverted by the computer. However, it was considered possible that the nature of the signal obtained from certain critical defects (such as sticker wrench) depends on the relation between the direction in which the beam was incident during scanning, and the rolling direction, and some such sheets were consequently re-scanned. Data from bent sheet cannot be corrected in the computer. The solution to this problem (which was not implemented during the project) is to hold the sheets flat during scanning using a magnetic chuck.

All records were stored in duplicate on different magnetic tapes, to guard against loss of data due to deterioration of the tapes and accidental erasure.

The composition of the data base is shown in table A.1

Appendix A.2 Program Package

Three types of computer program were used for the simulation investigation- programs for the storage, editing and re-formatting of data, programs for signal processing, and programs for the display of data and the outcome of detection experiments in pictorial and graphical form, normally by plotting on microfilm. The programming package was configured as a family of mutually compatible sub-routines, such that a wide variety of processing schemes could be simulated by setting up an appropriate combination, co-ordinated by a relatively simple master program. To ensure economy in computer usage, and to facilitate the selection and location of particular records, the data base was configured such that the data from each sheet could be selected by being specified on a punched card, and could be held in RAM

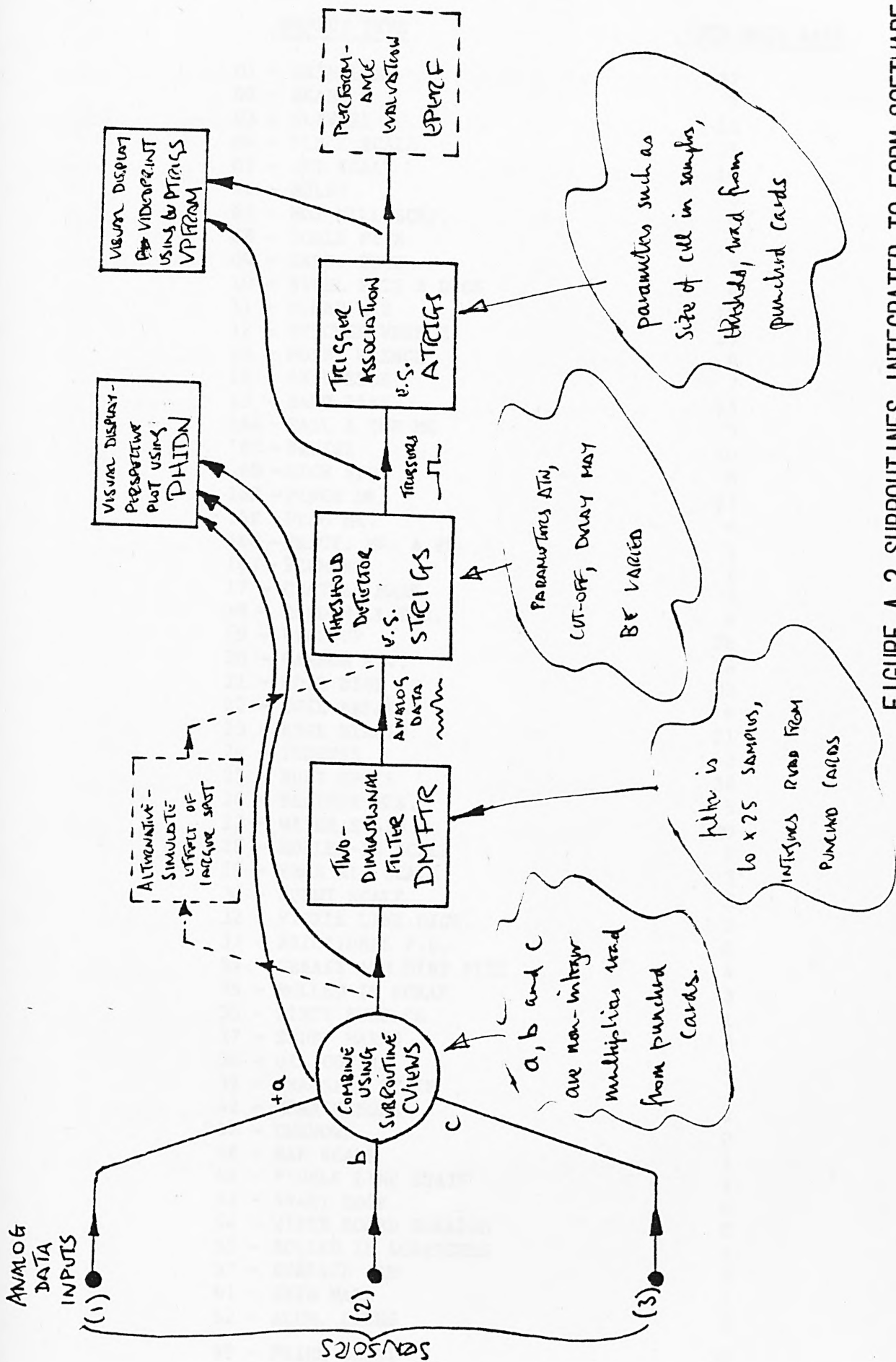


FIGURE A.2 SUBROUTINES INTEGRATED TO FORM SOFTWARE PACKAGE

TABLE A1 - STATE OF DATA BASE

<u>DEFECT TYPE</u>	<u>TCU DATA BASE</u>
01 - SKIN LAMS	37
02 - SEAMS	7
03 - SLIVERS	13
04 - FLECK SCALE	35
05 - JET SCALE	10
06 - HOLES	3
07 - HOT MILL SCRP.	9
08 - SCALE PITS	3
09 - SERR. EDGE	3
10 - STEEL PITS & DIGS	4
11 - SCRAP MKS	15
12 - STICKER WRENCH	25
13 - POINT WRENCH	9
14 - OXID EDGE	7
15 - SAND PITS	15
16A - TAIL & CNR MK	3
16C - BRUISE	10
16D - HIGH SPOT	8
16E - PINCH MK.	17
16F - DRAG MK.	4
16G - FRACT. MK. & FK.	1
16J - FEWMARK	2
17 - CHATTER MARK	3
18 - COLD MILL SCR.	4
19 - PICK-UP	26
20 - CARBON P.V.	14
21 - COIL DIGS	32
22 - COIL BREAK	9
23 - EDGE STRAIN	21
24 - INDENTS	2
25 - RUST SPOTS	38
26 - FEATHER MKS.	5
27 - WATER STAIN	20
28 - ROLLED-IN SCALE	1
29 - BRKR R1. SCALE	3
30 - BURNT SCALE	3
32 - PICKLE LINE DIGS.	0
33 - FRICTIONAL P.U.	0
34 - GREASE AND DIRT PITS	4
35 - ROLLED-IN SCRAP	3
36 - DIRTY SURFACE	1
37 - SCUFF MARKS	8
38 - OIL CONTAM.	2
39 - TRANSFER STAIN	3
42 - HERRINGBONE	2
44 - UNKNOWN	0
46 - BAR SCALE	1
49 - PICKLE LINE STAIN	1
53 - SNAKY EDGE	0
54 - WIPER BOARD SCRATCH	0
55 - ROLLED IN SCRATCHES	1
57 - SURFACE JAM	1
61 - SKID MARK	1
62 - ALUM. LINES	0
99 - PRIME SHEET	10

store throughout the processing, together with essential annotation including the location of the defect as specified by the 1cm.squares. The package was further configured to be able to handle signals from any number of sensors with equal facility.

A common filestructure was adopted to ensure that the data base and all processing subroutines were mutually compatible. This was:-

ID, ITYPE, IA, NDEFS, LCDEFS, ICO, IPM

in which:-

ID contains the sheet number in A10 format (e.g.21.09, with 21 indicating the type of defect, and 09 indicating the serial), packed into a single word.

IA contains the analog data for the sheet, in an integer array 330 words by 100, with 7 samples each of 3 bits packed into a 60 bit word.

NDEFS contains the number of 1cm. squares within which the defect lies, in I2 format in a single word.

LCDEFS is a one-dimensional array containing 100 words, with each word containing the location of one square containing defect markings. Each square is specified by its X and Y coordinates in a 1cm. grid. LCDEFS also contains the serial number of the defect. For example, the square specified by 2312704 contains markings from defect number 231, and has X co-ordinate 27 and Y co-ordinate 04.

ICO is a one dimensional array of 4000 words, each containing the location (in samples, measured from the bottom left hand corner of the sheet) of one trigger sample. For unprocessed data these contained triggers generated using the unoptimised SIRA detector when the data was gathered, which were replaced during processing by triggers generated by simulated detectors (without, of course, destroying the original data base).

The parameter IPM indicates the sensor, as follows:-

IPM=1 specular

IPM=2 10-degree off-specular

IPM=3 30-degree off-specular

A.2.1 Programs for the Creation, Verification, Editing and

Protection of the Data Base.

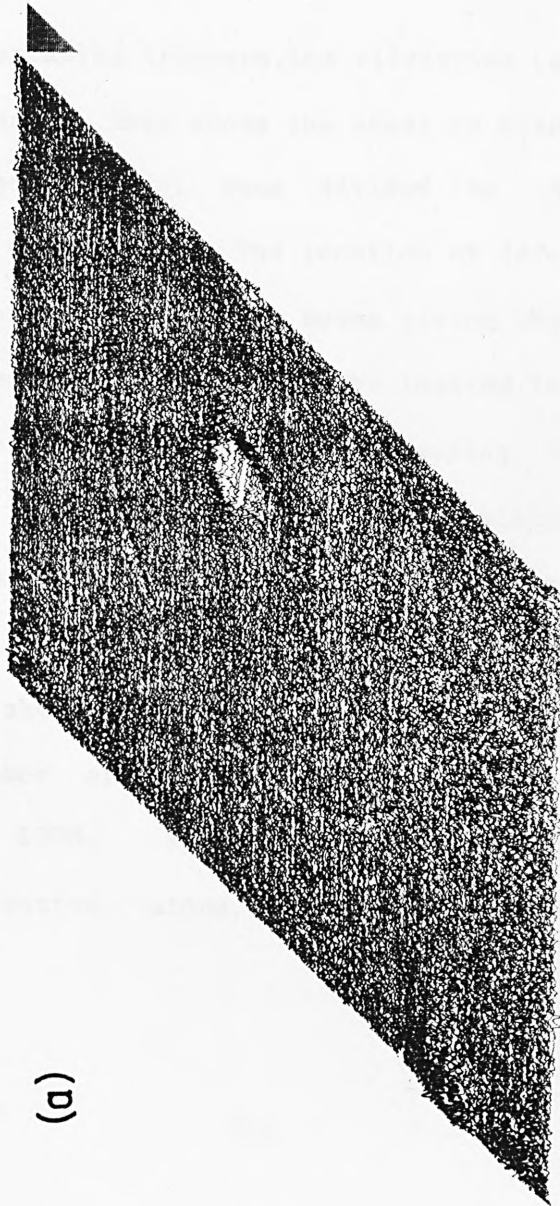
Data is displayed graphically by plotting it on microfilm using a Calcomp 1670 plotter. This has a resolution of 4096 points by 4096, but plots monochrome only, and is intended for plotting line drawings rather than grey scale images. Two methods were developed for displaying analog data, namely:-

(i) Perspective plots, in which the analog data from each scan is plotted above and a little to the right of that from the previous scan. Figure A.3(a) is an example of a perspective plot. 'Hidden Lines' are normally eliminated in the plot, to improve clarity.

Imperfections in data due to poor sample quality (sheet bent or grossly contaminated), or to mistakes during scanning (sheet upside down, saturation) are readily apparent from perspective plots. Signals from defects having reasonable contrast stand out as pulses above or below the noise signal due to surface roughness. Some correlation between the noise in successive scans is generally visible, compatible with fig. 2.12. However, it is difficult to locate markings on the sheet precisely

PERSPECTIVE PLOT
COLD ROLLED STEEL STRIP

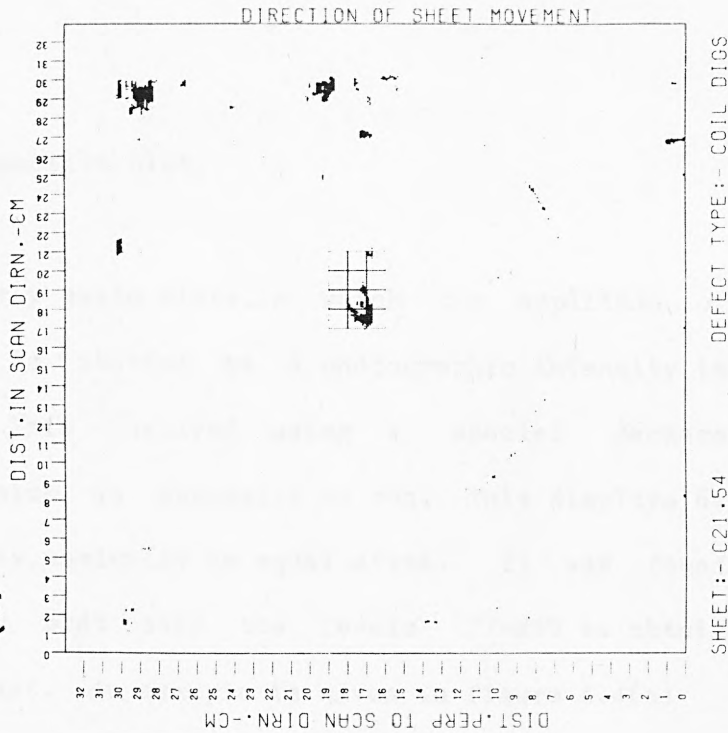
(a)



C21.54 DEFECT TYPE:- COIL DIGS
PROCESSING:-NONE
DETECTOR:-5 DEG OFF-SPEC

VIDEOPRINT
COLD ROLLED STEEL STRIP

(b)



SHEET:- C21.54 DEFECT TYPE:- COIL DIGS
PROCESSING:-NONE
DETECTOR:-5 DEG OFF-SPEC

FIGURE A.3 PERSPECTIVE PLOT AND VIDEOPRINT COMPARED

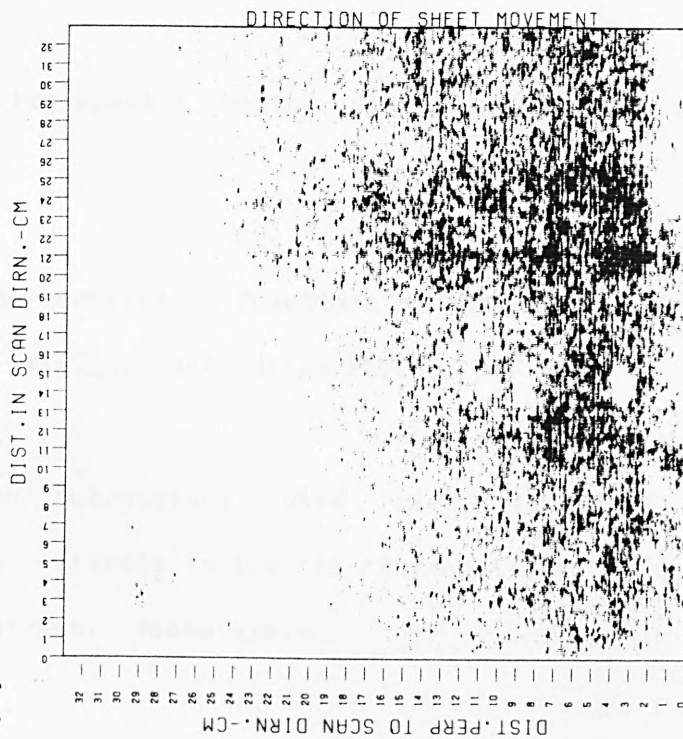
using a perspective plot.

(ii) Grey scale plots, in which the amplitude of each sample is plotted as a photographic intensity, in plan view. This involved using a special package ('Picpac') which is expensive to run. This displays 64 levels of grey, nominally in equal steps. It was found preferable to plot only the levels 127-255 to obtain usable contrast. An example is given in figure A.4(a)

For displaying triggers, the videoprint (example, fig A.3 (b)) is used. This shows the sheet in plan view, with horizontal and vertical axes divided to show true distance in centimetres. The location of defects on the original sheet is indicated by boxes giving the outlines of 1 cm. squares supplied in the logging information. Triggers generated by the signal processing subroutines (supplied in the standard array ICO) are plotted as dots in their appropriate location. Thus, the pattern of triggers may be compared with the defect distribution on the original sheet very easily. For economy in memory usage, the number of triggers the package can handle is limited to 4000. When this number have been generated, detection stops, and the videoprint is

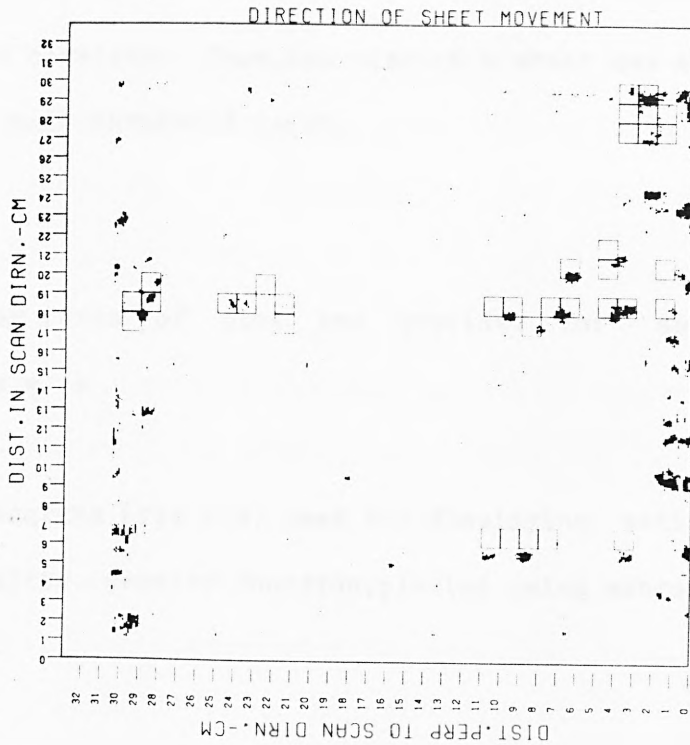
VIDEOPRINT
COLD ROLLED STEEL STRIP

(a)



SHEET:- C21.35 DEFECT TYPE:- COIL DIGS
PROCESSING:- NONE
DETECTOR:- SPECULAR

(b)



SHEET:- C21.35 DEFECT TYPE:- COIL DIGS
PROCESSING:- NONE
DETECTOR:- SPECULAR

FIGURE A.4 GREY SCALE PLOT AND VIDEOPRINT COMPARED

considered complete. Thus, the size of a sheet may appear to change with threshold level.

Other forms of plot are provided for special purposes, e.g.:-

Histograms (fig 2.9) used for displaying estimates of probability density function, plotted using subroutine HSTPLOT.

Scatter diagrams (fig 2.13) used for displaying correlation relationships between pairs of components of vector signals. Subroutine PSCAT is used.

Fourier spectra together with the original signal (fig 4.21)

Cross-correlation functions of successive pairs of waveforms (fig 2.14) subroutine PLTCON.

Three subroutines have been developed for processing signals prior to thresholding, to enhance defect contrast. These are:-

(a) CVIEWS, which generates a scalar signal which is a linear weighted sum of the three component signals obtained from the three sensors. The weights are supplied via punched cards.

(b) SMDATA, which synthesises the signal which would be obtained by scanning with a spot N times the length of the scanner spot used, and M times its width. This is achieved by convolving N consecutive samples together for M consecutive scans, and summing the results.

(c) DMFTR, which convolves the signal with a two-dimensional array which is N samples along the scan (x) direction, and M samples in the y -direction. The maximum value of N is 25, that for M is 10. Since the samples are separated by 0.4 mm in the scan (x) direction and 1 mm in the y direction, the maximum size filter is approximately square. The values of the filter elements are supplied from punched cards.

Four subroutines have been produced for thresholding the signal, which is the fundamental decision stage in the detection. All involve comparing an analog signal with a reference signal, such that when signal

crosses reference, in either the negative direction or the positive direction, a trigger is generated for the particular sample or samples. The output is thus binary. The subroutines are:-

(a) STRIGS, a simulated SIRA detector, which uses a non-recursive low-pass filter to generate the reference signal. The delay between filtered and unfiltered signals, the reference signal cut-off frequency and attenuation, and the direction of the threshold (positive going or negative going) were specified on punched cards.

(b) MFB, which simulates the one-dimensional matched filter bank with up to six shapes possible. The shapes of the filters, and the thresholds for each shape, are supplied on punched cards. The software had been developed by Hill (1977) for use in his tinplate study, with minor modifications to provide compatibility.

(3) TTRIGS, which implements a simulated SIRA detector turned through 90 degrees to operate on signals composed of corresponding samples from consecutive scans, i.e. in the y-direction, in contrast to strigs which operates along the x-direction.

The following subroutines were prepared for association of triggers after thresholding, to distinguish randomly distributed noise triggers from clustered message triggers :-

(a) ATRIGS, which rejects triggers within a cell N samples by M, unless at least I are present within the cell. The parameters N, M and I were supplied on punched cards. ATRIGS implements the 'association in a box' filter described in section 5.2.

(b) PROXT, which rejects triggers for which no adjacent sample also contains a trigger. This implements the 'AJ' filter discussed in section 5.3

(c) FTRIGS, which rejects triggers at sample number K unless a store for sample K contains a count of L or higher. When the input for sample K is at level '1', the store is incremented, otherwise it is decremented. This implements the 'ADT' filter described in section 5.4. It is also possible to 'OR' together the signals in N adjacent scans. The parameters L and N are supplied on punched cards.

The following subroutines were written to measure properties of the surface noise signal, and to plot them where relevant on microfilm.

(i) A subroutine for measuring histograms of sample levels, to obtain an estimate of the probability density function(pdf).

(ii) SPCTAN computed the power spectra for individual scans, and the average power spectrum for all scans for a particular view of a sheet. The FFT algorithm (available as a library subroutine) was used to compute a complex spectrum as a first step.

(iii) LCOR computes the cross-correlation between scans (at a specified separation) from a particular sheet, by direct multiplication of corresponding samples from which the mean has been removed, then summing and normalising for variance.

(iv) F094 (not configured as a subroutine) analyses cross-correlations between data from different sensors, i.e. between the components of a vector signal. The covariance matrix is found for the signals from the

three sensors, and its eigenvectors and eigenvalues are computed.

In addition, subroutines were provided (a) for automatic assessment of methods for defect detection (EPERF), by counting the number of defect triggers falling within defect squares, without actually plotting the videoprint, (b) for detection of non-local but periodic defects (DCHTS) using the FFT algorithm to compute a Fourier power spectrum, which was scaled logarithmically for display (fig 4.21), and (c) for implementing the analysis and synthesis of correlated waveforms described in chapter 2 section 5. n

APPENDIX 'B' - NATURE AND PROPERTIES OF SIGNALS

In this appendix some concepts and methodologies used in the body of the thesis to describe signals and analyse them are defined and explained.

A SIGNAL is a physical observable which conveys information. To be analysed theoretically signals must be represented mathematically by functions. Signals generally appear naturally in ANALOG form, for which the mathematical representation is a function $f(x)$, where x is a continuous independent variable, e.g. space or time. Generally, the properties of signals are independent of the nature of the variable x , thus concepts developed for time variant signals apply equally to spatial signals. For analog signals, $f(x)$ may take an infinite range of values within an interval (the DYNAMIC RANGE of the signal) and is defined for all values of x . It must be finite and single valued, though it may possibly be COMPLEX, in which the function $f(x)$ has both real and imaginary components.

A VECTOR signal is composed of two or more scalar (component) signals $f(1,x)$, $f(2,x)$, $f(3,x)$...etc., in which

each component possesses individually the properties required for a scalar signal. The components may be related to one another, i.e., knowledge of $f(1,x)$ may provide information concerning $f(2,x)$.

For processing by computer, signals must be represented in an alternative, DISCRETE form. This comprises a sequence of numbers, $g(1), g(2), g(3), \dots, g(k), \dots$, where k is an integer indicating the order of $g(k)$ within the sequence. To enable processing to be implemented economically in hardware for high speed operation, it is further necessary to constrain each $g(k)$ to occupy only specified discrete values within the dynamic range, e.g. the integers 0 to 256. This yields a DIGITAL representation.

Digital signals are obtained from analog signals by SAMPLING, followed by QUANTISATION. Sampling involves multiplying the analog signal by a sequence of unit impulses, which converts the analog signal into a sequence of numbers (the values of $f(x)$ at the impulses). However, the numbers may still occupy an infinity of values within the dynamic range. The NYQUIST CRITERION states that provided the impulses are separated (on

average) by a distance less than X , where $X=1/2W$, and W is the highest frequency at which energy is present in the signal, then the analog signal may be recovered uniquely from the sequence of samples. That is, all information present in the analog signal is retained in its discrete representation. This is strictly true only if the signal segment is of infinite length.

To produce the digital representation, each sample of the discrete signal is rounded off to the closest of a finite number of discrete levels. This process, termed QUANTISATION, removes information irreversibly from the signal manifested by the addition of random noise which is white with a uniform distribution. The amount of noise added is inversely proportional to the number of levels.

Properties of signals, and relationships between them, can often be described very lucidly by using a SIGNAL SPACE (fig.4.8, B.1) representation. In this, the samples describing a discrete signal are considered to define the axes of a Euclidean space. Each distinguishable signal, i.e. each different waveshape, is in its sampled representation, specified by a distinct

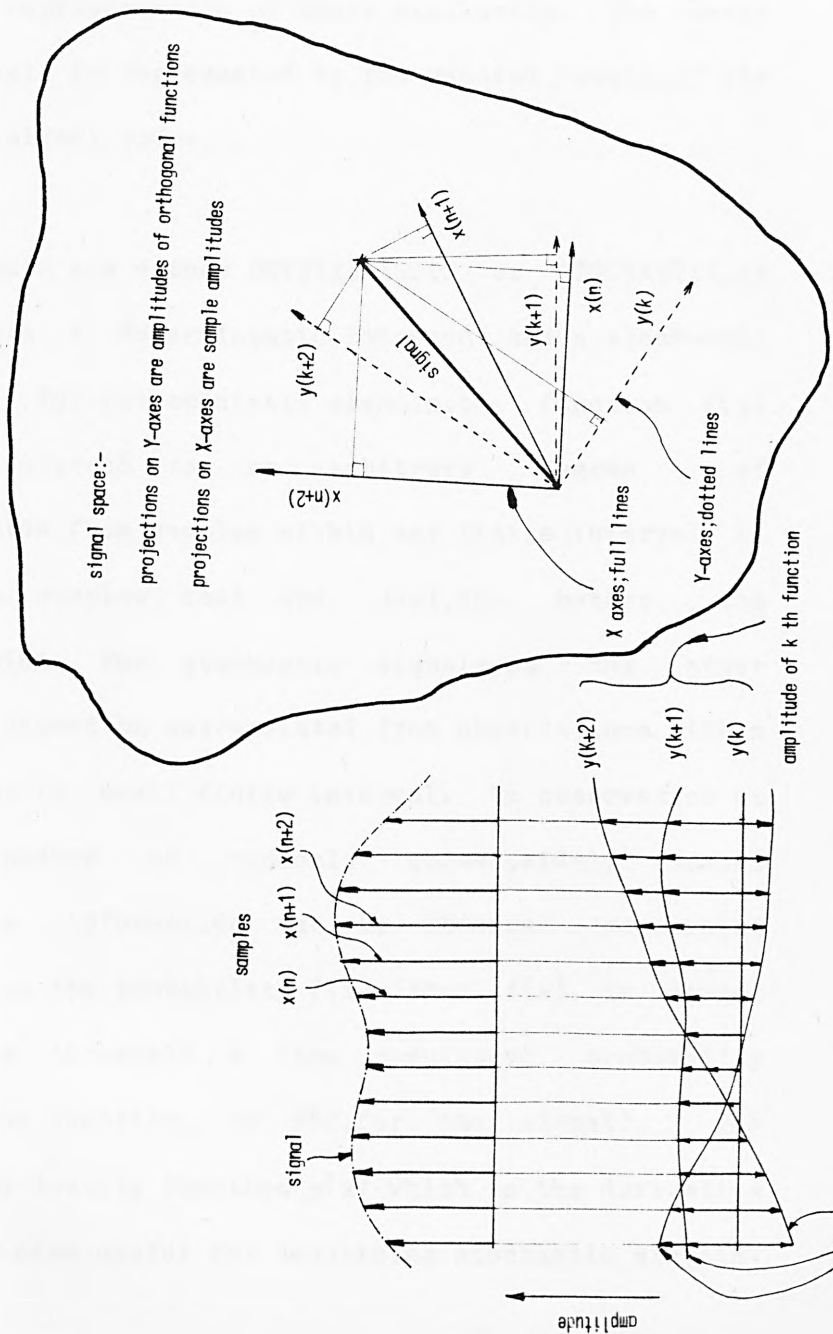


FIGURE B.1 ORTHOGONAL REPRESENTATION DESCRIBED IN SIGNAL SPACE

point within the space. The Euclidean distance between different signals in the space provides a useful (inverse) representation of their similarity. The energy of a signal is represented by the squared length of its vector in signal space.

Signals are either DETERMINISTIC or STOCHASTIC, or contain both a deterministic component and a stochastic component. For deterministic signals, the function $f(x)$ may be inferred to an arbitrary degree of approximation, from samples within any finite interval X . The more samples that are used, the better the approximation. For stochastic signals, on the other hand, $f(x)$ cannot be extrapolated from observations within an arbitrarily small finite interval. By observation at a large number of randomly chosen, widely spaced points, some information can be obtained concerning $f(x)$, such as the probability $P(x)$ that $f(x)$ is larger than some threshold x (the cumulative probability distribution function, or PDF, for the signal). The probability density function $p(x)$ which is the derivative of $P(x)$ is also useful for describing stochastic signals.

A stochastic processes can alternatively be

regarded as an ordered sequence of random variables. The properties of the most general kind of ensemble require for their specification statement of the joint probabilities of all orders. If however the ensemble is ordered into a sequence, specification is generally much simpler. When the joint probability density between the k th and l th members of the sequence is a function only of their separation ($k-l=m$), the process is said to be stationary. Specification of the joint pdfs of all orders is in this case provided by the AUTOCORRELATION FUNCTION, $A(m)$, for the signal. This is the expectation of the joint p.d.f. of signal samples at a separation m . $A(m)$ is 1 for $m=0$ for a stochastic signal, and decreases (not monotonically) with m .

The form of stochastic signal met most frequently is the Gaussian signal, for which the probability density function has the form: _

$$p(x) = 1/\{\sqrt{2\pi}\cdot s\} \cdot \exp([x-x(0)]^2 / 2s^2)$$

The parameters $x(0)$ {the mean} and s {the standard deviation} define the distribution. Gaussian stochastic signals are completely specified by their p.d.f. and

autocorrelation functions.

Gaussian signals are interesting because they are generated naturally by many physical processes and are tractable mathematically. The CENTRAL LIMIT THEOREM states that the p.d.f. of the distribution resulting from the combination of a large number of uncorrelated random variables having zero mean converges towards the Gaussian form, irrespective of the pdfs of the variables individually.

The signal resulting from the joint occurrence of two uncorrelated Gaussian signals having variances $s(x)^{**2}$ and $s(y)^{**2}$, and coefficient of correlation r , has the form:-

$$p(x,y) = \frac{\{1-r^2\}}{2\pi s(x).s(y)} \exp \left\{ -\frac{1}{2[1-r^2]} \left(\frac{x^2}{s(x)^2} - \frac{2xy}{s(x).s(y)} + \frac{y^2}{s(y)^2} \right) \right\}$$

If the signals have further the same variance s^{**2} and are uncorrelated, the above equation then reduces to:-

$$f(x,y) = \frac{1}{2\pi} \exp \{ -(x^{**2} + y^{**2}) / 2s^{**2} \} / 2s^{**2}$$

On transforming this to polar co-ordinates r, θ , and incorporating the property that the distribution is now a function of r only, we obtain the RAYLEIGH distribution:-

$$f(r) = r \cdot \exp\{-(r^2/2s^2)\}/s^2$$

The mean (expectation) of the Rayleigh distribution is (Papoulis, 1965) $s/\sqrt{2}$. The cumulative distribution function $F(b)$ that the signal has amplitude greater than some threshold b is then given by:-

$$F(b) = \int_b^{\infty} f(r) \cdot dr = \exp(-b^2/2s^2)$$

The GAMMA distribution has the form:-

$$f(x) = C^{b+1}/G(b+1) \cdot \exp(-cx) \quad \{x > 0\}$$

where b and c are arbitrary constants, and G is the gamma function. If b is integral, $G(b+1) = b!$. The mean (expectation) of the Gamma distribution, $E(x)$, is:- $(b+1)/c$

The NEGATIVE EXPONENTIAL distribution (a special case of the gamma distribution, when $b=0$ and $c=1/g$), has

the form:-

$$f(x) = \{1/g\} \cdot \exp(-x/g)$$

where $x > 0$, and g is a parameter. The mean of the negative exponential distribution is g .

In analysing stochastic signals, it is frequently necessary to infer the properties of a POPULATION of signals, by analysis of finite segments of signal which must be maintained as short as possible for economy in computation. The estimates thus obtained are said to be UNBIASED if the estimate is equal to the true value for the population, is ASYMPTOTICALLY UNBIASED if any bias vanishes as the quantity of sample data used increases, and is CONSISTANT if the mean square error of the estimate converges to zero as the quantity of sample data analysed increases. The estimates used in this work {chap 2, sect.4} possess the latter two properties.

The components of stochastic vector signals may be related to one another mutually. A useful measure of this relationship for signals of zero mean is provided by the CROSSCOVARIANCE, $C(x,y)$ between two signals x,y .

This is defined as the expectation of the product $f(x).f(y)$. The CORRELATION between $f(x),f(y)$ is defined as $C(x,y)/(s(x).s(y))$, where $s(x)$ and $s(y)$ are the standard deviations of the individual signals. For Gaussian signals, components which are uncorrelated are also mutually independent, i.e., the expectation of the probability of the joint event $f(x),f(y)$ is the product of the probabilities of the individual events, $P(x).P(y)$.

For a scalar signal, the AUTOCOVARANCE, $C(x,x)$, describes the expectation of the product $f(x).f(x-k)$, where $f(x-k)$ is k samples distant from the reference sample $f(x)$. The autocovariance C is related to the autocorrelation A by:-

$$A = C/s^{**2}$$

where s^{**2} is the signal variance.

For multidimensional vector signals, the mutual relationships between the components are described by the covariance matrix, $[C]$, for which the (i,j) th entry is the covariance between the i th and j th components. The entry on the main diagonal (i,i) is the variance of the i th

component. $[C]$ is symmetric and positive semidefinite, consequently its eigenvalues are always non-negative. By dividing the (i,j) th entry in $[C]$ by the product of the standard deviations of the i th and j th signals, the correlation matrix, $[A]$, is obtained. This matrix is also symmetric and positive semidefinite, and the entries on the main diagonal are all unity.

The properties of a signal may often be clarified by orthogonal transformation. The signal is regarded as comprising a sequence containing N samples, of which the (k) th is designated $f(k)$. Following transformation, the (k) th component is expressed as a linearly weighted sum of orthogonal signals $g(l,k)$, where k is a sample index and l identifies the orthogonal signal within the family, i.e.:-

$$f(k) = \sum_{l=1}^N a(l) \cdot g(l,k)$$

The basis functions $g(l,k)$ have the property that:-

$$\sum_{m=1}^N g(l,m) \cdot g(u,m) = 0 \quad \text{, unless } l=u, \text{ in which case the sum is non-zero.}$$

If ,in addition,the sum is unity when $l=u$,the functions are orthonormal.

Using the definition of orthogonality,it is seen that the coefficients $a(l)$ can be computed from the expression:-

$$a(l) = \sum_{k=1}^N f(k).g(l,k)$$

The term $g(l,k)$ is the KERNEL of the orthogonal representation. The $a(l)$ are scalar numbers,which may be complex if the $g(i,x)$ are complex,even if $f(x)$ is real. For a vector signal containing N components, N orthonormal basis functions are required. The transformation may be visualised as multiplication of the signal regarded as an N element vector $[F]$ by an N -square matrix $[G]$,to yield a transformed signal $[A]$ which is also a column vector. From the definition of orthonormality just given, the matrix $[G]$ has the property:-

$$[G].[G'] = [I]$$

Here, $[G']$ is the transpose of $[G]$. This implies that the matrix representing an orthogonal transformation has its inverse equal to its transpose.

The significance of orthonormal transformation may be appreciated by reference to a signal space.(fig B.1). The signal $(x(k), \text{where } k \text{ is an index})$ comprises a sequence of samples numbers representing amplitude. The separation between members of the sequence may equally represent space, time etc. in which axes representing the original (unsampled) representation are shown as solid lines, whilst those indicating the orthogonal functions are shown dotted. The effect of the orthonormal transformation is to rotate the axes of the space about the origin. The signal remains unaltered. Its components in the new representation are its projections on the rotated axes. Not only is information within the signal preserved, as for any nonsingular transformation, but the relationships between different signals in the space, as represented by relative lengths and the angles between the signals, are also preserved.

Orthogonal functions are in general GLOBAL, in contrast to signal samples which are LOCAL. The way in which the same signal can be described by two alternative representations, one local, the other global, is explained in fig.(B.1)

Although many orthogonal representations are available, one particular form, the Discrete Fourier Transform (DFT), has been used exclusively in this work. In the DFT, the orthogonal functions are the discrete sequences $\exp(2 \times \pi i \times nm/N)$, where n is the sample index, m is the function index, N is the number of samples in the signal. N is also the number of functions in the representation. These are sampled sinewaves designated by frequency which is the number of zero crossings per unit length function when the d.c. component is removed. Both symmetric (cosine) and antisymmetric (sine) forms are included. The DFT may be computed speedily using the Fast Fourier Transform (FFT) algorithm which is normally provided as a standard library subroutine. It reduces the number of complex multiplications to transform an N -element sequence from N squared to $N \log N$.

Several important theorems relate the spatial and spectral representations of a signal. PARSEVAL'S theorem states that the energy of a signal (the squared length of its vector representation in signal space) is unchanged on orthonormal transformation (fig.B1). The WIENER-KHINCHIN theorem states that the autocovariance function of a signal is the Fourier transform of its

power spectral density, i.e.:-

$$C(x,x) = \text{DFT}\{a(k).a'(k)\}$$

where $a(k), a'(k)$ are the real and imaginary parts respectively of the transformed signal. The CONVOLUTION theorem states that time domain convolution (a vector multiplication) corresponds in the frequency domain to a scalar point-by-point multiplication,

This theorem assumes implicitly that the signal sequence is repetitive outside its interval of definition.

DECOMPOSITION AND SYNTHESIS OF STOCHASTIC VECTOR SIGNALS Consider an ensemble of N scalar Gaussian stochastic signals having zero mean, of which the k th has zero lag autocovariance $c(k,k)$, and the zero-lag cross-covariance between the i th and j th members of the ensemble is $c(i,j)$. This ensemble may be expressed as a linearly weighted sum of N uncorrelated Gaussian signals. The idea is illustrated by the signal flow graph shown in figure 2.16.

This property may be used to uncouple signals which are not independent, to increase message contrast and to disclose relationships which have physical significance but are not perceptable in the original signal.

The form of this decomposition (i.e., the auto- and cross-covariances of the N uncorrelated signals and the weighting coefficients) is generally not unique but the Karhunen-Loeve decomposition is preferred, and will be described.

The vector of correlated signals is denoted R , with its k th member $r(k)$. The vector of uncorrelated signals is S , with its l th member $s(l)$. The fraction of $s(l)$ which must be used in the synthesis (or decomposition) of $r(k)$ is $m(l,k)$. Thus, each $r(k)$ is formed from the following summation:-

$$r(k) = \sum_{l=1}^N m(l,k) s(l)$$

The coefficients $m(l,k)$ may be considered to form a square matrix M , thus the process whereby the vector R is generated from the vector S may be expressed by the matrix equation:-

$$[R] = [M].[S]$$

The K/L theorem states that $[M]$, $[S]$ and the covariance matrix for $[R]$ which is $[C]$ are related by the following equation:-

$$[A] = [M'].[C].[M] \quad \text{with } [M'].[M] = [I]$$

where $[A]$ is a diagonal matrix whose (k,k) th entry is the variance of the k th uncorrelated signal, $s(k)$, and $[M]$ is orthogonal. Since $[C]$ is real and symmetric, the elements of $[A]$ are never negative.

This decomposition has the following advantages:-

(i) The elements of $[S]$ are provided in descending order of amplitude $\{s(k) > s(k+1)\}$, and $s(k)$ decreases in the most rapid way possible as k increases. Thus, if $[R]$ is considered to be composed of only the first J elements of $[S]$, the best mean square approximation is obtained for given J .

(ii) The equation for the decomposition is solved by finding eigenvectors and eigenvalues for a symmetric

matrix ,i.e.by using an algorithm which is normally provided as a standard library subroutine on most computer systems.

The approach can be used alternatively to synthesise a vector signal $[R]$ having a given $[C]$, using a random number generator to create each element of $[S]$, for use in simulation experiments.

REFERENCES AALDERINK, B.J. (1976) and M.W.C. DEJONGE

'Automatic Surface Inspection of Cold-Rolled Sheet at the Temper Mill' Journees Internationales de Siderurgie, Bruxelles-Duesseldorf, Art 6.2.1

AKUTSU, S (1976) et al. 'Automatic Surface-Flaw Inspection of Cold Strip' Journees Internationales de Siderurgie, Bruxelles-Duesseldorf 1976 Art. 6.2.2

ANTONSSON, D. (1979), P. DANIELSSON, B. KRUSE and T. OHLSSON 'A System for Automated Visual Inspection of Printed Wiring Boards' Internskrift LITH-ISY-0264 (Univ. of Linköping, Sweden) .

ANON (1981) 'Electronic 35mm. camera' Sensor Review, vol.1 no.1, Jan 1981 p41

AYRES, P. (1962) 'Matrices' Schaum, New York, p112

BARKER, A.J. (1976a), R.A. BROOK and R.N. WEST 'Preliminary Report on the Use of Multiple Detectors in Automatic Surface Inspection Systems' Prel.Rept. on SIRA project 6/4742/02 6th July 1976

BARKER,A.J.(1976b),R.A.BROOK and R.N.WEST 'Interim Report on the Use of Multiple Detectors in Automatic Surface Inspection Systems-Specification of the Optimal Data Logging Configuration for Scatter Pattern Classification' rept on SIRA project 6/4742/02 29th October 1976

BARKER,A.J. , and R.A.BROOK (1978) 'A Design Study of an Automatic System for On-Line detection and Classification of Surface Defects on Cold Rolled Steel Strip' Optica Acta vol25 #12 pp1189-1196

BARTLETT,M.S.(1955) ' An Introduction to Stochastic Processes' ,Camb. Univ. Press

BECKMANN,P. and A.SPIZZICHINO (1963) 'The Scattering of Electromagnetic Waves from Rough Surfaces' Pergamon

BEYNON,J.D.E. and D.R.LAMB (1980) 'Charge Coupled Devices and their Applications' McGraw-Hill

BLACKMAN,R.B.(1958) and J.W.TUKEY 'The Measurement of Power Spectra',Dover,New York,1958.

BRITISH STEEL CORPORATION 'Surface Defects Present on

Steel Sheet and Strip' B.S.C. booklet 1283/2.73/3M

BROOK, R.A. (1971a) 'Automatic Inspection of Flat Surfaces' Steel Times Annual Review 1971 pp91-98

BROOK, R.A. (1971b) 'An Experimental Automatic Surface Inspection System' Metron, vol 3, p219-223

BROOK, R.A. (1976), J.F. CLARIDGE and C.F. WATTS 'Automatic Inspection of Flat Strip in the Steel Industry' Proc. VII Imeko Congress, London, May 1976 paper AML/154

BROOK, R.A., D.J. PURLL, G.H. JONES and D.O. LEWIS (1977) 'Practical Experience of Image Processing in On-Line Industrial Inspection Applications' SPIE vol. 130 pp84-97

BROOK, R.A. (1979), C. WATTS, L. NORTON-WAYNE and W.J. HILL 'Designing an Automatic System for Surface Inspection of Cold Rolled Strip' Iron and Steel International, Oct 1979 pp289-290

BOWERS, D (1970) 'The Automatic Inspection of Aluminium Sheet' DSA Internal Report, The City University.

BUSSGANG, J.J. (1955) and D.MIDDLETON 'Optimum Sequential Detection of Signals in Noise' IRE Trans. Inf. Theory, vol. IT-1, pp. 5-18, Dec. 1955.

CASASENT, D. (1978) 'Optical Signal Processing' in vol. 23 of series 'Topics in Applied Physics', Springer, Berlin.

CHANDLEY, P.J. (1979) and H.M. ESCAMMILA 'Speckle from a Rough Surface when the Illuminated Region Contains Few Correlation Areas: the Effect of Changing the Surface Height Variance' Opt. Comm vol 29 no. 2 pp 151-154

CHITTINENI, C.B. (1980) 'Signal Classification for Automatic Industrial Inspection' IEEE Workshop on Digital Signal and Waveform and Analysis, Miami Dec 5, 1980.

CHITTINENI, C.B. (1982) 'Signal Classification for Automatic Industrial Inspection' IEE Proc. vol. 129 pt. E no. 3 May 1982 pp. 101-106

CLARKE, G.M. (1976) and J. BEDFORD 'Application of the Minicomputer to High-Speed, On-line Optical Inspection Systems' Proc. IMEKO VII pp AQC/119 pp 1-9

CLARKE,G.M. and J.BEDFORD (1977) 'Laser Scanning Systems for Image Acquisition' SPIE vol.130 pp63-74

DAINTY,J.C. (1971) 'Detection of Images Immersed in Speckle Noise' Optica Acta ,vol.18 no.5 pp327-339

DAINTY,C.(1975)[ed.] 'Laser Speckle and Related Phenomena',Springer,Berlin 1975.

DIFFARD,G.M. and C.T.CHANG (1979) 'Applications of Fiber and Integrated Optics to Radar Signal Processing' Proc.SPIE vol.130 pp164-170

DONNELLY,R.K. (1971) The Automatic Inspection of Steel Sheet' M.Sc. Thesis 1971 The City University.

DUDA,R.O. (1973) and P.E.HART 'Pattern Classification and Scene Analysis', J.Wiley,New York.

DUFF,M.J.B.(1973),D.M.WATSON,T.J.FOUNTAIN and G.K.SHAW 'A Cellular Logic Array for Image processing' Patt. Rec.,vol.5 pp229-247

EDEN,R.C. and I.DEYHIMY (1979) 'GeAs Integrated Circuits

and Charge Coupled Devices (CCD'S) for High Speed Signal Processing' Proc.SPIE vol.180 pp182-189

EDY,G. (1977) 'Automatic Surface Inspection Using Microprocessors' M.Sc. Thesis,Department of Systems Science,The City University.

FELLER,W (1968) 'An Introduction to Probability Theory and its Applications' vol.(1),3rd edition,John Wiley.

FUJII,H,J.UOZOMI and T.ASAKURA (1976) 'Computer Simulation Study of Image Speckle Patterns with Relation to Object Surface Profile' J.Opt.Soc.Am. vol.66 no.11 pp1222-1236

FUJII,H and T.ASAKURA (1977) 'Roughness Measurements of Metal Surfaces Using Laser Speckle' J.Opt.Soc.Am. vol.67 no.9 pp1171-1176

FUJII,H and J.W.Y.LIT (1973) 'Surface Roughness Measurement Using Dichromatic Speckle Pattern: An Experimental Study' Applied Optics,vol17 no.17 pp2690-2694

GAUTIER, H. and P. TOURNOIS (1981) 'Very Fast Signal Processing as a Result of the Coupling of Surface Acoustic Wave and Digital Technologies' IEEE Trans. Microwave Theory and Techniques, MTT-29, no. 5, pp 404-409

GOODMAN, J. W. (1975) 'Statistical Properties of Laser Speckle Patterns' in 'Laser Speckle and Related Phenomena' ed. C. Dainty, Springer pp. 9-77

GREIS, P. and W. HOERSTER (1977) 'Comparison of Presently Known Measuring Instruments for Surface Defects (as at December 1976)' VDEH Report 4.024 (sup. to BFI reports 482 and 592)

GRUENHOFER, H. G. et al. (1967) 'Surface Defects on Cold Rolled Strip and Sheet' Verlag Stahleisen M. B. H. Duesseldorf

HARIHARAN, P. (1977) 'Statistics of Speckle Patterns Produced by a Rough Surface' Optica Acta, vol. 24 no. 9 pp 979-987

HANCOCK, J. C. and P. J. WINTZ (1966) 'Signal Detection

Theory' McGraw-Hill

HECHT,D.C. (1979) 'Acousto-Optic Signal Processing
Device Performance' Proc.SPIE vol.180 pp150-159

HILL,W.J. (1972) 'Automatic Inspection - A Progress
Report Covering the Period April 1972 to September 1972'
ISC Rept. DSA/WJH/32

HILL,W.J. (1977) 'Defect Recognition in Automated
Surface Inspection' Ph.D Thesis The City
University,London

HOEL,P.G. (1971) 'Introduction to Mathematical
Statistics' 4th ed.,Wiley,p.195

DE JONGE ,M.W.C.(1973) 'Statistical Considerations
Concerning the Detection of Linear Type Surface Defects'
Hooghovens Rept ARCH LAB 35434

DE JONGE,M.W.C. (1979) 'Study and Development of a
Scanner Device for Detection of Surface Defects on Sheet
Steel from a Skin-Pass Rolling Mill' ECSC Final
Report,res.no.E6/E36,conv.no. 6210-60/6/601

IWAI, J (1981), N. TAKAI and T. ASAKURA 'The Autocorrelation Function of the Speckle Intensity Fluctuation Integrated Spatially by a Detecting Aperture of Finite Size', Optica Acta, vol. 28 no. 10 pp. 1425-1437

IEEE PROCEEDINGS vol. 59 no. 5 May 1980

JONES, A. R. (1977) 'Investigation of Polar Scattering Patterns Occuring when Laser Light is Incident on Typical Mild Steel Defects' Brit. Steel Corp. Welsh Lab., Tech. Note no. B/2338

LACROIX, M (1979), J. PINARD and F. PONS 'Improvement of the Surface Quality of Hot Rolled Strip by Automatic Testing on the Strip Rolling Mill' *Final rept. E.C.S.C. Cnr. 6210 GA/3/305*

LOCKE, D. H. (1975) 'Lithographic Plate Inspection Using Fibre Optics' Proc. SPIE vol. 60 pp 39-44

LOGAN, I (1974) and J. E. S. MACLEOD 'An Application of Pattern Recognition to the Automatic Inspection of Steel Strip Surfaces' Proc. 2nd Intl. Jt. Conf. on Pattern Recognition, Copenhagen 1974 pp 286-290

LOGAN, I (1974) Ph.D. Thesis, Univ. of Glasgow.

MARGOUN, J and P.G. McHugh (1979) 'Radar Signal Processing-Digital, Charge-Coupled ' Proc.SPIE vol.180 pp230-241

MCDONOUGH, R.N. (1971) 'A Canonical Form of the Likelihood Detector for Gaussian Random Vectors' J.A.S.A. Vol.49 no.2(pt.1) pp402-406

MCKECHNIE, T.S. (1975) 'Speckle Reduction' in 'Laser Speckle and Related Phenomena' ed.C.Dainty, Springer pp123-204

MILANA, E and F.RASELLO (1981) 'An Optical Method for On-Line Evaluation of Machined Surface Finishing' Optica Acta, vol.28 no.1 pp111-123

MORAAL, J. (1971) 'Effect of two Different Methods of Training on the Results of a Visual Inspection Task' ECSC rept. 2442/72c RCE

MORAAL, J (1974) 'Visual Inspection of Sheet Steel-An

Ergonomic Study in the Dutch Steel Industry', Le Travail Humain, vol.37 no.1 1974 pp35-52

MURISON, A., P.M. GRANT and R.M. De LA RUE (1979)
'Applications of Optical Techniques to Signal Processing'
in 'Case Studies in Advanced Signal Processing' IEE
pub.180, pp34-45

MUNDY, J.L. (1980) 'Visual Inspection of Metal Surfaces'
Proc. National Computer Conf., 1979 pp227-231

NORDQUIST, K.G. (1974) and A.L. MILLGARD 'Automatic
Surface Inspection and Classification' The Iron and Steel
Engineer, June 1974 pp67-70

NORTON-WAYNE, L. (1972) 'Automatic Inspection of
Surfaces' City University Research Report LNW/DH/30 May
1972

NORTON-WAYNE, L. (1974) and W.J. HILL 'The Automatic
Classification of Defects on Moving Surfaces' Proc. 2nd
Intl. Joint. Conf. on Pattern Recognition, Copenhagen 1974
pp496-473

NORTON-WAYNE, L (1977), W.J.HILL and R.A.BROOK 'Automated Visual Inspection of Moving Steel Surfaces' Brit.Jnl of NDT, vol.19, no.5 pp242-248

NORTON-WAYNE, L and W.J.HILL (1979) 'The City University Contribution to the Final ECSC Report' City University report DSS/LNW-WJH/183 March 1979.

NORTON-WAYNE, L., G.WEST and C.KOULOPOULOS (1980) 'Progress in the Inspection of Binary Patterns' TCU research report DSS/LNW-GW-CK/.....

NORTON-WAYNE, L. (1980), W.J.HILL and C.F.WATTS 'The Automatic Visual Inspection of Steel Strip' Proc.5th Intl.Conf. Automated Inspection and Product Control Stuttgart, June 1980

NORTON-WAYNE, L. (1980a) 'Signal Theory' Notes to accompany M.Sc. course, Dept.of Systems Science, The City University.

NORTON-WAYNE, L. (1980b) 'On the Removal of Random Noise from Binary Images' -a comparison Study' Proc.5th Intl.Intl.Conf. on PR (IEEE pub 80CH1499-3) 1980 pp

1130-1133.

NORTON-WAYNE, L., (1982) and G.A.W.WEST 'Instrumentation for the Automated Visual Inspection of PC Boards' Proc.5th European Conf. on Electrotechnics, (EUROCON), Copenhagen, June 14th to 18th 1982. pp 832-836

OBAY, C.D. (1973a) M.Sc.Thesis 'The Detection of Surface Defects by Automatic Surface Inspection', The City University

OBAY, C (1973b) 'Detection of Surface Defects by Adaptive Forecasting Techniques' Instrument Systems Centre Report CDO/DH/60, 1973

OBAY, C.D. (1982) Ph.D.Thesis (in Preparation)

PARKS, T.W. and J.H.MCLELLAN (1972) IEEE Trans.CT-19 189-194

PAPOLIS, A.(1965) 'Probability, Random Variables and Stochastic Processes' McGraw-Hill (p.195)

POPOVICI, V. (1970) 'Application of Syntactic Pattern

Recognition to Defect Classification' Ph.D. Thesis, The City University.

RAISBECK, G. (1963) 'Information Theory' MIT Press

REYNOLDS, P.M. (1966) 'The Surface Inspection of Metal Sheet and Strip by Optical Methods' Metallurgical Reviews, vol.11 pp.89-96

ROOT, W (1970) 'An Introduction to the Theory of the Detection of Signals in Noise' Proc IEEE vol.59 no.5 May 1970 pp610-623

SARIDIS, G.N. (1976) and D.M.BRANDIN 'An Automatic Inspection System for Flat Rolled Steel' Purdue Univ. Laboratory of Applied Industrial Control, PLAIC Rept.78

SARIDIS, G.N. (1979) and D.M.BRANDIN 'An Automatic Inspection System for Flat Rolled Steel' Automatica, vol.15 pp505 - 520 1979

SAWATARI, T. (1972) 'Surface Flaw Detection Using Oblique Angle Illumination' Applied Optics, Vol.11 no.6, Jan '72 pp.13437-1344

SCHUSTER, A. (1898) 'On the Investigation of Hidden Periodicities, with Application to a Supposed 26-Day Period of Meteorological Phenomena' Terrestrial Magnetism, Vol. 3 pp13-41, March 1898.

SJOLIN, U (1972) 'Optical Inspection of Metal Surfaces' Laser Focus, July '72 pp 67-70

SKOLNIK, M. J. (1962) 'Introduction to Radar Systems', McGraw-Hill

STANSBERG, G. T. (1931) 'On the First Order Probability Density Function of Integrated Laser Speckle' Optica Acta vol. 28 no. 7 pp917-932

STAPELY, B. E. (1965) 'Automatic Inspection of Metal Strip for Visual Defects' Iron and Steel Engineer, Nov. 1975 pp99-108

SCHWARTZ, M. and L. SHAW (1975) 'Signal Processing-Discrete Spectral Analysis, Detection and Estimation' McGraw-Hill

SYKES,G. and C.BURNS (1963) 'Closed Circuit Television for the Inspection of Fast Moving Surfaces' Jnl.IEE vol.9 July 1963 pp296-299

THOMAS,W.O. (1961) 'Quality Control in the Production of Steel Sheet for Deep Drawing and Pressing' Sheet Metal Industries,Sept1961 pp659-"674.

VAN DAELE(1977),A.OOSTERLINK and H.VAN DEN BERGHE 'Television Scanners' Proc.SPIE vol 130 pp75-82

WATTS.C. (1977a) 'Advancement of Surface Inspection Instrumentation for Cold Rolled Steel Strip',ECSC Proj. no.7210.GA/3/304 Tech Rept.(1), Jan 1977

WATTS,C.(1977b) ECSC Proj no. 7210.GA/3/304 Tech Rept.(2) July 1977

WATTS,C.(1978a) ECSC Proj no. 7210.GA/3/304 Tech Rept.(3) Jan 1978

WATTS,C.(1978b) ECSC Proj no. 7210.GA/3/304 Tech Rept.(4) July 1978

WALD, A. (1947) 'Sequential Analysis' Dover, New York.

WHALEN, A.D. (1971) 'Detection of Signals in Noise'
Academic Press

WEBBER, K.R. (1971) 'The Application of Stochastic
Filters to Automatic Surface Inspection' M.Sc. Thesis. The
City University 1971.

WECKLER, G.R. (1975) 'Image Detection for Industrial
Gauging and Inspection' Proc. SPIE vol. 60 pp14-19

WIENER, N. (1930) 'Generalised Harmonic Analysis' Acta
Mathematica, vol. 55, pp117-259

WILLIAMS, G. (1978) 'Trials with SIRA 550 Scanner on no. 1
Temper Mill, Port Talbot' results up to 1/12/78 BSC
Research Lab, Port Talbot, Internal Report.

YAXLEY, D (1979) 'Improved Methods for the Detection of
Surface Defects on Sheet Steel' M.Sc. thesis, The City
University, 1979.