



THE UNIVERSITY *of* LIVERPOOL

A System for the Digitisation and Enhancement of Industrial Radiographic Film Images

Thesis submitted in accordance with the
requirements of the University of Liverpool
for the degree of Doctor of Philosophy

by

Tony Colclough

Department of Electrical Engineering and Electronics

August 1995

Acknowledgements

I would like to express thanks to my supervisor Dr J. S. Smith for his advice and support throughout this project. I would also like to acknowledge the help and advice supplied by the staff of the mechanical workshop of the Department of Electrical Engineering and Electronics during the design and construction of the mechanical scanning apparatus. Finally, I would like to thank the EPSRC and Mr. A. Paton and Mr. S. Fisher from BNFL who sponsored this research.

Abstract

Film radiography is probably the most sensitive method of non-destructive testing for internal defects within components. A typical industrial installation may require the radiography of many thousand components, each of which will require several views to build up a full picture of its internal structure. The storage of such a large number of radiographic film plates is a major undertaking. The aim of this research was to develop a system for the digitisation and processing of industrial radiographic film images. Converting images to digital form significantly simplifies their storage, archiving, reproduction and communication.

The film digitisation system used a linear CCD array photosensor for image acquisition. The film image was moved across the sensor using a high precision two axis scanning mechanism. Image illumination was provided by a scanning laser spot, projected through the film image. The apparatus produced images with a maximum spatial sampling resolution of 14 μm and a digital fidelity of 12 bits/pixel. The modulation transfer function of the image acquisition system was measured to be 0.6 at a spatial frequency of 10 cycles/mm, ensuring accurate contrast reproduction of very small features. The maximum pixel data rate of the system was 625 kHz.

The image acquisition and processing system was constructed around a network of transputer processors. This arrangement provided concurrent software control of the different hardware subsystems used during image acquisition. Software was also developed to allow interactive enhancement of the digitised images produced by the system. Enhancement algorithms included noise reduction, image sharpening, contrast enhancement and edge detection. Lossless compression of image data was used to reduce the size of data files without compromising image quality.

Contents

1.Introduction	1
1.1 Review of Existing Film Digitisation Technology	2
1.2 Objectives	3
2. Review of Theory and Standards Concerning Radiographic Film Digitisation	5
2.1 Introduction	5
2.2 Radiographic Image Formation	6
2.3 Geometric Unsharpness	7
2.4 Film Characteristics and Inherent Unsharpness	9
2.5 Sensitivity Limitations	13
2.5.1 Contrast Sensitivity	14
2.5.2 Crack Sensitivity	15
2.6 Image Quality Indicators	16
2.7 Radiograph Interpretation	18
2.8 Performance of the Human Eye	19
2.9 Real Time Radiography	23
2.9.1 Fluoroscopy and Image Intensifier RTR Systems	24
2.9.2 Solid State RTR Systems	26
2.10 Conclusions and Specifications of a Film Digitisation System	27
2.11 Summary	30
3. Review of Solid State Imaging Technology	34
3.1 Introduction	34
3.2 Photodiodes and Dark Current	35
3.3 Photodiode Array Sensors	36
3.3.1 The CCD Shift Register	37
3.3.2 Improvements to the Basic CCD Structure	38
3.3.3. A Practical CCD Image Sensor	40
3.4 Measures of CCD Performance	42

3.4.1 Dark Signal Output and Dark Signal Nonuniformity	42
3.4.2 Linearity and Gain Nonuniformity	43
3.4.3 Saturation and Blooming	45
3.4.4 Spectral Sensitivity Characteristics	46
3.4.5 Image Lag	48
3.4.6 Modulation Transfer Function	48
3.4.7 Charge Transfer Inefficiency	49
3.5 Commercially Available Solid State Image Sensors	50
3.6 Summary	51
4. The Transputer Processor	53
4.1 Introduction	53
4.2 Parallel Processing	53
4.3 The Transputer	54
4.3.1 Transputer Architecture	54
4.3.2 Link Communications	56
4.3.3 Link Services and Interconnection	57
4.4 Interfacing Transputers to External Hardware	58
4.4.1 The Inmos IMS C011	59
4.5 Summary	60
5. A Radiographic Film Digitisation System	62
5.1 Introduction	62
5.2 System Outline	62
5.3 Transputer System	64
5.4 Texas Instruments TC104 CCD Sensor	65
5.4.1 Clocking Requirements	66
5.4.2 Drive Hardware	68
5.4.3 Output Signal Processing	74
5.4.4 Circuit Layout	79
5.5 Sony ILX503 CCD Sensor	81
5.5.1 Clocking Requirements	83
5.5.2 Drive Hardware	84

5.5.3 Output Signal Processing	86
5.6 Data Acquisition System	89
5.6.1 Analogue to Digital Conversion	89
5.6.2 Interface Circuit	92
5.7 Image Display	94
5.7.1 MosaiQ Frame Store	94
5.7.2 Colour Monitor	95
5.8 I/O Control	96
5.8.1 Input Card	96
5.8.2 Output Card	96
5.9 Motor Subsystem	96
5.9.1 Vertical Motion	97
5.9.2 Horizontal Motion	99
5.9.3 Motor Control Logic	100
5.10 Laser Scanner	100
5.10.1 Laser Source	102
5.10.2 Galvanometer Scanners	104
5.10.3 AE1000 Scan Controllers	104
5.10.4 Dual Channel DAC	105
5.10.5 Exposure Time Synchronisation Logic	106
5.10.6 Differential Link Adapters	107
5.11 Mechanical Apparatus	108
5.12 System Construction and Layout	111
5.12.1 Main System Rack	111
5.12.2 Sensor Hardware	112
5.12.3 Laser Scanning Rack	113
5.12.4 Host PC	114
5.13 Summary	114
6 Sensor System Performance	117
6.1 Introduction	117
6.2 Dark Signal and DSNU	117
6.3 Sensor System Noise	120

6.4 Measuring the Resolving Power of an Imaging System.....	125
6.4.1 The Point Spread and Line Spread Functions	127
6.4.2 Measurement of Point and Line Spread Functions.....	129
6.4.3 The Modulation Transfer Function.....	135
6.4.4 Resolving Power of Radiographic Films.....	141
6.5 Test Pattern Images	143
6.6 Comparison of TC104 and ILX503 Sensors	144
6.7 Summary.....	147
7. Software Development	150
7.1 Introduction	150
7.2 Programming Parallel Systems.....	150
7.3 Overview of the 3L C Language and Associated Tools	152
7.3.1 The 3L C Language	153
7.3.2 Configuration Utility	156
7.4 Configuring the Processor Network	157
7.4.1 Hardware Configuration	157
7.4.2 Software Configuration	159
7.4.3 System Services.....	161
7.5 Host Link Server.....	163
7.6 Menu Driven Software Front End	164
7.7 Summary.....	167
8. Image Capture Performance and Software.....	168
8.1 Introduction	168
8.2 Image Format.....	168
8.3 Image Capture.....	171
8.3.1 Sensor Saturation.....	173
8.3.2 Laser Scanning Algorithm.....	174
8.3.3 Sensor Calibration	174
8.4 Image Display	179
8.5 Image Archiving	180
8.6 Digitisation of Industrial Radiographs	182

8.7 Summary.....	186
9. Image Enhancement.....	188
9.1 Introduction	188
9.2 Image Processing Fundamentals.....	188
9.2.1 A Digital Image Model.....	188
9.2.2 Image Enhancement Overview.....	189
9.2.3 Pixel Mask Operators	190
9.3 Image Enhancement Techniques	191
9.3.1 Noise Reduction	191
9.3.2 Image Sharpening	192
9.3.3 Contrast Enhancement.....	193
9.3.4 Edge Detection	195
9.4 Software Implementation	198
9.4.1 Parallel Image Processing.....	198
9.4.2 Pixel Buffer	200
9.5 Results	201
9.5.1 Image Smoothing and Sharpening.....	201
9.5.2 Histogram Equalisation	202
9.5.3 Linear Contrast Enhancement	207
9.5.4 Edge Enhancement	210
9.6 Summary.....	213
10 Image Data Compression.....	215
10.1 Introduction	215
10.2 Lossless Compression	216
10.2.1 Entropy Coding.....	216
10.2.2 Data Pre-processing and Entropy Reduction.....	221
10.3 Lossy Compression Methods.....	225
10.4 Tests on Compression Methods.....	226
10.5 Compression of Radiographic Images.....	229
10.6 Summary.....	230

11 Conclusions and Suggestions for Further Work.....	232
11.1 Conclusions	232
11.2 Further Work	233
Appendix A : Circuit Diagrams	A-I

1. Introduction

Welded joints are often a critical area for component failure. Tiny defects within a weld may cause failure of the joint. A poor weld may exhibit such faults as gas pores, cracking and slag inclusion [1]. Stress concentrations occur around such defects which may lead to progressive crack propagation and eventual failure. Modern fabrication procedures often call for 100% inspection of welds to identify defects. Simple surface inspection of a welded joint may not reveal potentially serious internal faults, and hence some form of volumetric testing is required. One of the commonest, and most sensitive, forms of sub-surface weld assessment involves the inspection of radiographic images of welds [2].

A large industrial installation, such as a chemical plant or power station, may require the archival of many thousands of radiographic film plates. Standard industrial practice dictates that these film images should be retained for the life time of the plant, which may typically be decades. These film images are liable to degrade with time, unless they are stored in a controlled, light sealed atmosphere. Clearly, considerable savings in physical storage space may be made by converting these film images into digital form.

Assessing the quality of a welded joint from a radiograph is a complex and highly skilled task. The implications of an error may be expensive scrap and re-work costs or, more seriously, component failure. Hence, in order to ensure accuracy of image assessment, it is essential to maintain the definition of the original image during any digitisation process. Any loss of image quality would compromise the system performance.

In addition to considerations of storage and archiving, several other advantages may be gained from converting a film image to digital form. Inspection procedures may benefit from the application of modern digital techniques for data transmission, reproduction, processing and interpretation. A digitised image may be enhanced using computer image processing techniques in order to more easily identify faults. Areas of an image may easily be magnified and features made clearer. Digitised images also allow for computer aided assessment of the welds. Such a system may eventually be able to

completely supersede the requirement for inspection of radiographs by a skilled human operator.

1.1 Review of Existing Film Digitisation Technology

Very little technical literature currently exists concerning the problem of radiograph digitisation, although a good summary of current technology can be found in a report by the USNRC¹ [3]. An early attempt at film digitisation was a system developed by BIX Engineering for BNFL during the 1980's [4]. This system used a high resolution video camera to view film mounted on a servo controlled scanning table. Although this system was put into service, the images it produced were generally unacceptable for high integrity NDT² applications.

A more recent product is the Elumisys Lumiscan film digitiser [5]. This system uses an integrating cylinder light collector to detect a scanning beam laser spot. The Lumiscan digitiser provides a pixel spacing of 100 μm over an area of 14 inches by 17 inches and achieves a maximum pixel data rate of 150 kHz. This system provides no capabilities for image enhancement and processing and is very expensive, costing tens of thousands of pounds.

The film digitisation system developed during this project used a linear array photosensor, scanned across the film plate to build up a two dimensional image. This system of image acquisition closely resembles recently developed technology used in document scanners [6] [7]. An example of this type of system is the Kodak RFS 2035 film scanner [8], designed for desk top publishing applications. This system digitises standard 35 mm colour film negatives at a resolution of 12 bits/colour using a CCD area array sensor consisting of 2048 pixels by 3072 pixels, with a 12 μm pixel spacing. The maximum pixel data rate achieved by the Kodak RFS 2035 was 50 kHz.

The linear CCD array sensors used during film digitisation are readily available from several manufacturers either as single chip analogue sensors or in the form of self

¹ U. S. Nuclear Regulatory Commission

² Non Destructive Testing

contained digital cameras. An example is the Cronin S Series of Linescan Cameras [9], which incorporate a 1024 element, wide aperture linear CCD array consisting of 25 μm by 2.5 mm pixel elements. This system digitises images at a resolution of 12 bits/pixel and retails at £2100. However, no off-the-shelf linear CCD camera currently on the market offered the required resolution or the flexibility of operation required for this project. Instead, it was decided to build a custom made linear CCD image acquisition system.

1.2 Objectives

The principal objective of this research was to produce high quality digital images from radiographic film for use in high integrity NDT applications. The main requirement of such a system was that it should be capable of reproducing all the features visible in the original film image, without introducing any spurious data during the digitisation process.

Before embarking on the design of the image digitisation apparatus it was first necessary to gain some measures of the sensitivity of modern radiographic film imaging techniques and relate this information to the factors affecting the quality of digitised images. By carrying out a detailed literature review of these topics it was possible to produce a specification for a system capable of producing digitised images of the desired quality.

A prototype image acquisition system was then be constructed. This system used a linear array photosensor for direct contact scanning of the film image. Film illumination was by a scanning laser spot mechanism. Detailed measurements of the imaging performance of this apparatus were then made in order to assess whether an acceptable image quality was achieved.

Methods of digital image enhancement were then investigated in order to increase the subjective quality of the image and assist the operator in the task of radiograph interpretation. Finally, it was proposed to develop methods of lossless data compression in order to reduce the volume of digital data to be stored, without compromising the integrity of the image.

References

- 1 W. Kenyon, "Basic Welding and Fabrication", Pitman, 1979.
- 2 R. Halmshaw, "Non Destructive Testing", Arnold, 1987.
- 3 "Review and Evaluation of Technology, Equipment, Codes and Standards for Digitisation of Industrial Radiographic Film", U. S. Nuclear Regulatory Commission, Washington D. C. May 1992.
- 4 "Radiographic Interpretation System Technical Report", BIX Engineering Ltd.
- 5 "Lumiscan Technical Notes", Elumisys Ltd, 1993.
- 6 G. Nagy, "Optical Scanning Digitisers", Computer **16** 5, 1983, pp 13-24.
- 7 H. Tseng, J. Ambrose & M. Fattahi, "Evolution of the Solid State Image Sensor", Journal of Image Science **29** 1, 1985, pp 253-259.
- 8 "Kodak RFS 2035 Film Scanner Technical Literature", Eastman Kodak Company, 1993.
- 9 Cronin Line Scan Cameras Product Outline, Cronin Electronics Ltd, 1992.

2. Review of Theory and Standards Concerning Radiographic Film Digitisation

2.1 Introduction

One of the most sensitive and widely used industrial methods of non-destructive testing for internal flaws in components is radiography. The basic principle of radiography consists of using ionising radiation (usually X-rays or gamma rays) to produce images of internal features of solid objects. This chapter gives a brief summary of some relevant theory concerning the theory and practice of modern industrial radiography. More detailed accounts can be found from many sources in the relevant literature [1] [2].

Before approaching the problem of producing digital images from radiographic film it is necessary to briefly discuss some basic characteristics of radiographic images. The amount of useful information contained in a radiographic film is affected by many factors, most notably radiographic technique, object size and composition, imaging geometry, film characteristics and film processing. The sources and relative contributions to overall image quality of these different factors will be discussed in this chapter in order to give some idea of the sensitivity of film radiography. In addition it should be noted that any radiographic film image must be assessed by a human viewer. Hence, the sensitivity of the human eye to small changes in detail and contrast must be considered when assessing the ability of an observer to resolve features within an image.

A review of several relevant industrial standards was undertaken in order to gain an indication of the sensitivity of modern industrial radiographic techniques. These standards are quoted extensively throughout this chapter. Relevant standards relating to industrial radiography of welded joints include BS2910, BS2600, BS5230, BS3915 and ASME V [3] [4] [5][6] [7]. More stringent standards are set for high integrity applications such as within the nuclear industry, these include BS5500, AESS6001 and NF0081 [8] [9] [10].

By analysis of the factors affecting flaw detection and by reference to the available standards an idea of the minimum detectable flaw size is derived. A short review of the

relevant literature concerning radiographic film digitisation is then given. Finally, the information discussed in this chapter is used as the basis for defining the required operating parameters of a radiograph digitisation system.

2.2 Radiographic Image Formation

The absorption of a thin monoenergetic radiation beam of intensity, I_0 , by an object of thickness, t , and linear absorption coefficient, μ , will result in only partial transmission of the beam. The intensity of the transmitted beam, I_t , is governed by equation (2.1).

$$I_t = I_0 \exp(-\mu t) \quad (2.1)$$

This principle is the basis of all radiographic imaging techniques and can be illustrated by considering the simple experimental geometry shown in Figure 2.1. If a point source of ionising radiation is placed some distance from a photographic film, and a dense test specimen is placed in contact with the film, the amount of radiation reaching the film will depend on the absorption characteristics of the test object.

The linear absorption coefficient of electromagnetic radiation by air is negligible when compared to that of solids. Hence if the specimen under inspection contains an air filled cavity the radiation will pass through this portion of the specimen with much less attenuation than radiation passing through the bulk material. The film directly beneath the cavity will be subject to a greater intensity of ionising radiation causing increased film blackening in this area.

If the film to source distance is much greater than the specimen thickness, the image will be an approximately true size, sectional view of the cavity taken at right angles to the beam direction. The image of the cavity will be highly dependent on its orientation within the specimen. If the cavity is a thin crack, running parallel to the beam direction, the image produced will be of a thin section through the crack. However, if the crack runs perpendicular to the beam the full length of the crack will be visible on the radiograph.

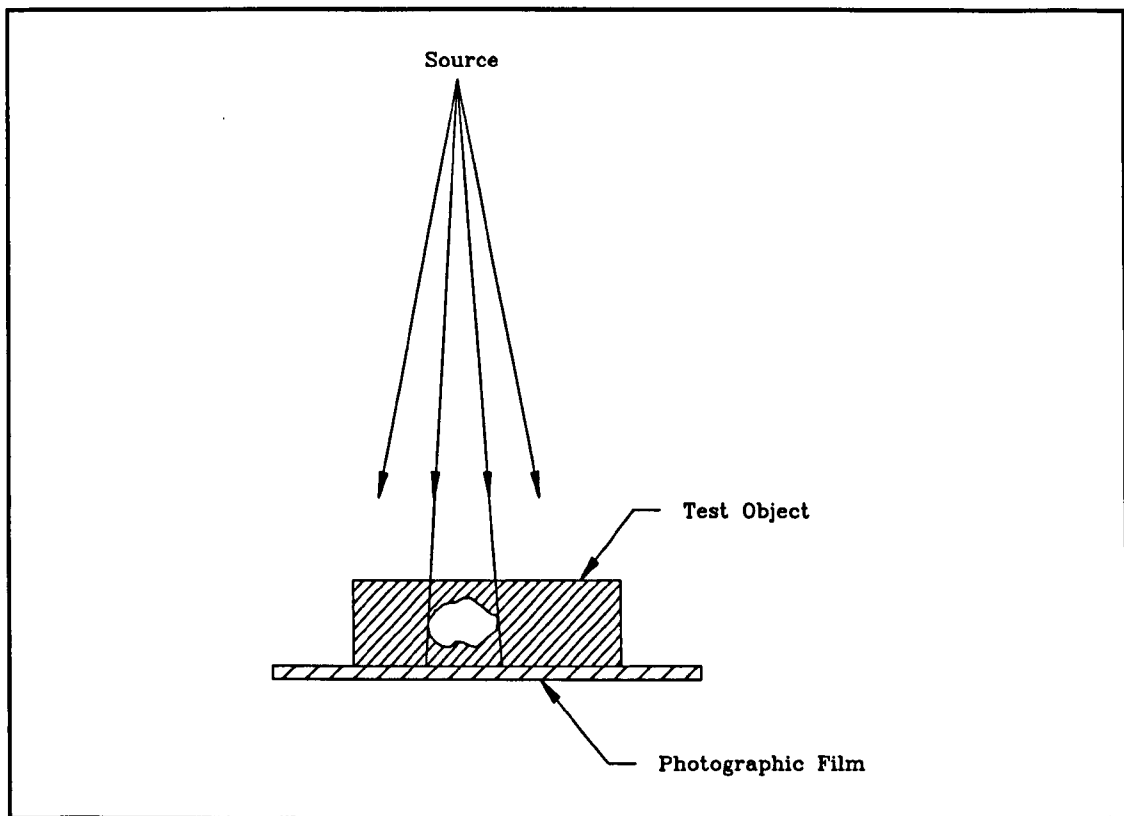


Figure 2.1 Radiographic Imaging of a Cavity

2.3 Geometric Unsharpness

Any practical source of ionising radiation will emit rays from a small focal spot rather than a single point. This finite source size can lead to some degree of blurring around the edge of features. This phenomenon, known as geometric unsharpness, is illustrated in Figure 2.2. In this example a stepped block, of maximum thickness, t , is being imaged using an extended source of width, ϕ , placed some distance, F , from the photographic film. Radiation beams from different positions within the source form separate images of the stepped block, leading to an overall blurring of the sharp edge. From simple geometry it can be seen that the maximum unsharpness, U_g , is given by:

$$U_g = \frac{\phi t}{F - t}$$

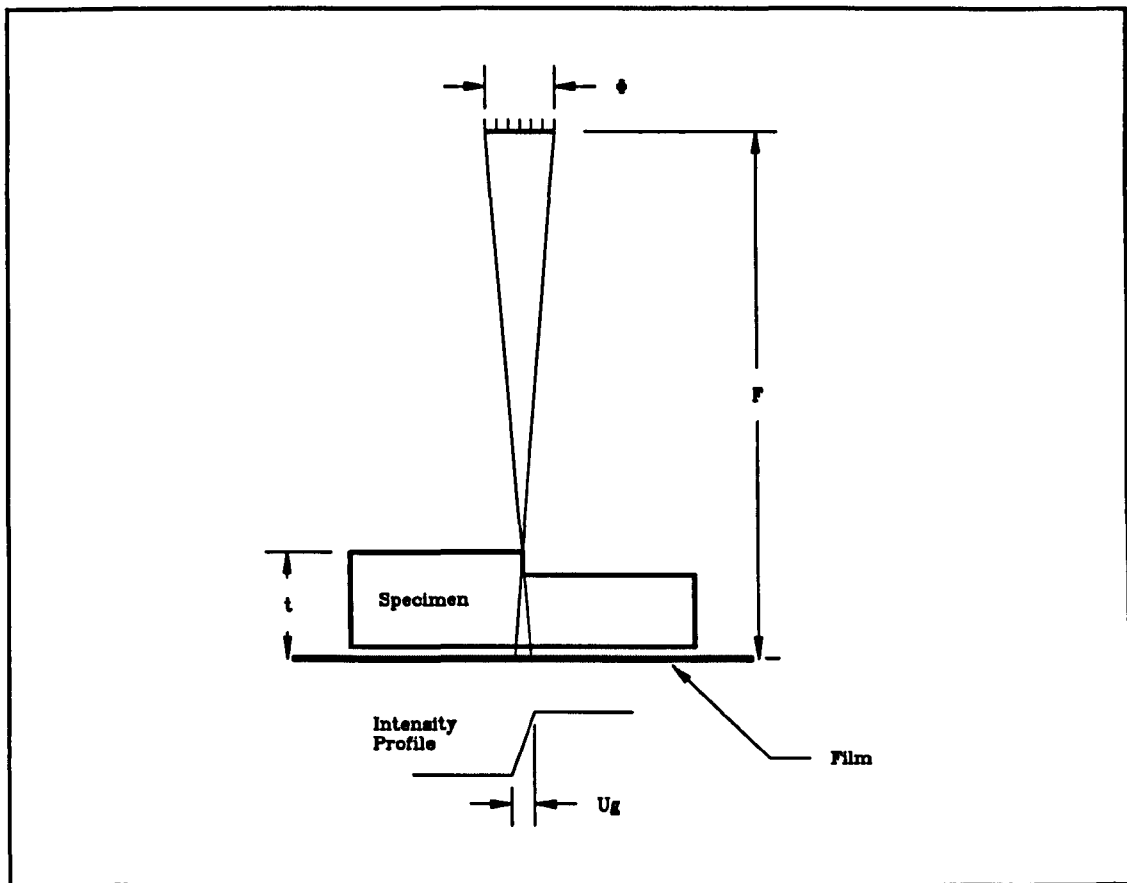


Figure 2.2 Geometric Unsharpness

Obviously, increasing film to focus distance, F , will reduce the effect of geometric unsharpness. However practical considerations limit the amount to which F may be increased. Since for all practical purposes X-rays and Gamma-rays cannot be focused, the source radiates out a diverging beam whose intensity decreases with distance from the source. This decrease in intensity is governed by the inverse square law, so that the radiant intensity from the source, I , at a distance, F , from the focal point, is given by the equation:

$$I \propto \frac{1}{F^2}$$

Increasing the film to focus distance, F , will rapidly decrease the radiation intensity reaching the film and necessitates a much longer exposure time. High power X-ray sets are expensive to run for long periods, similarly gamma sources are expensive and have a limited half life so that their use must be optimised. In addition it should be noted that there is a limit on focus size, ϕ , for a practical radiography set and that the

position of a test piece will often limit the achievable focus to film distance. These considerations place a practical limit on minimum achievable geometric unsharpness. Most industrial regulatory authorities specify some guidelines as to the maximum acceptable geometric unsharpness limits of practical radiographic techniques. A selection of these standards are summarised in Table 1.

Standard	Maximum Geometric Unsharpness, u_g	Conditions
ASME V	500 μ m	$t < 50$ mm
EN44	200 μ m	
NF0081	80 μ m	

Table 1 Industrial Standards for Geometric Unsharpness Limits

2.4 Film Characteristics and Inherent Unsharpness

Radiographic film consists of a central emulsion layer containing silver halide crystals suspended in gelatine; this emulsion layer is protected on either side by a flexible base material. The emulsion layer is generally much thicker than that used in conventional photographic films due to the fact that X-rays are much more energetic than visible light, and consequently are less easily absorbed.

Radiation absorbed by the emulsion causes ionisation leading to the deposition of silver grains. After photographic processing of the exposed film these silver grains appear as dark areas within the film emulsion. The degree of film blackening caused by image formation is usually described in terms of optical density. Optical density, D , is a measure of film opacity to visible light expressed as a ratio of incident light, I_0 , to transmitted light, I_t , passing through the film. Since the range of densities possible within a film image is vast, a logarithmic scale is used to express optical density, as shown in (2.2).

$$D = \log_{10} \left(\frac{I_0}{I_t} \right) \quad (2.2)$$

The density of any point on an exposed radiograph will always be greater than zero even if no direct radiation reached that point from the source. This effect is mainly due

to the density of the film base material. A certain amount of film darkening also occurs due to chemical processing of the image. These phenomena lead to the formation of a uniform base density level known as the fog level. This fog level tends to obscure image detail and should be minimised wherever possible. Table 2 summarises the maximum acceptable base fog levels specified by the relevant industrial standards.

Standard	Maximum Base Fog
BS2910	0.3
AESS6001	0.3

Table 2 Industrial Standards for Base Fog Limits

Much work has been carried out to assess the optimum optical density, at which low contrast image features are most readily discernible to the human eye [11]. These studies show general agreement in placing this optimum optical density between 2.5 and 2.8. At high exposure, there is a limiting density of between 10 and 15 due to complete ionisation of silver halide crystals. In order to ensure that all features on a radiograph are readily perceptible by the human eye there are standards set for the required density range of all features within the area of interest on a radiograph. These density limits are summarised in Table 3.

Standard	Minimum Density	Maximum Density
BS2910	2.0	3.0
AESS6001	2.0	3.5
ASME V	1.8	4

Table 3 Industrial Standards for Density Limits Within a Radiograph

As previously discussed absorption of X-rays or gamma rays within the radiographic film emulsion layer leads to ionisation of a silver halide atom and the consequent deposition of a silver crystal. An absorbed X-ray quantum may possess enough energy to release secondary electrons from the parent atom. These electrons may cause additional ionisation and further formation of silver grains. This process leads to the

formation of image grains in a small area around the point where the X-ray struck the emulsion. This blurring effect is known as the inherent or film unsharpness, U_f .

The area of inherent unsharpness depends on the number of secondary electrons produced during absorption and the distance they travel through the emulsion before being re-absorbed. High energy radiation will obviously cause more energetic collisions which in turn will produce more electrons with correspondingly high kinetic energies causing them to travel further through the film emulsion. The distance travelled by electrons before re-absorption has been shown experimentally to be more dependent on the ratio of silver to gelatine in the emulsion than on the actual silver halide grain size. In any case, inherent unsharpness is not greatly affected by film type, and is usually assumed to be dependent only on beam power. The relationship between beam power and inherent unsharpness has been investigated by many researchers [1], the results of which are summarised in Figure 2.3.

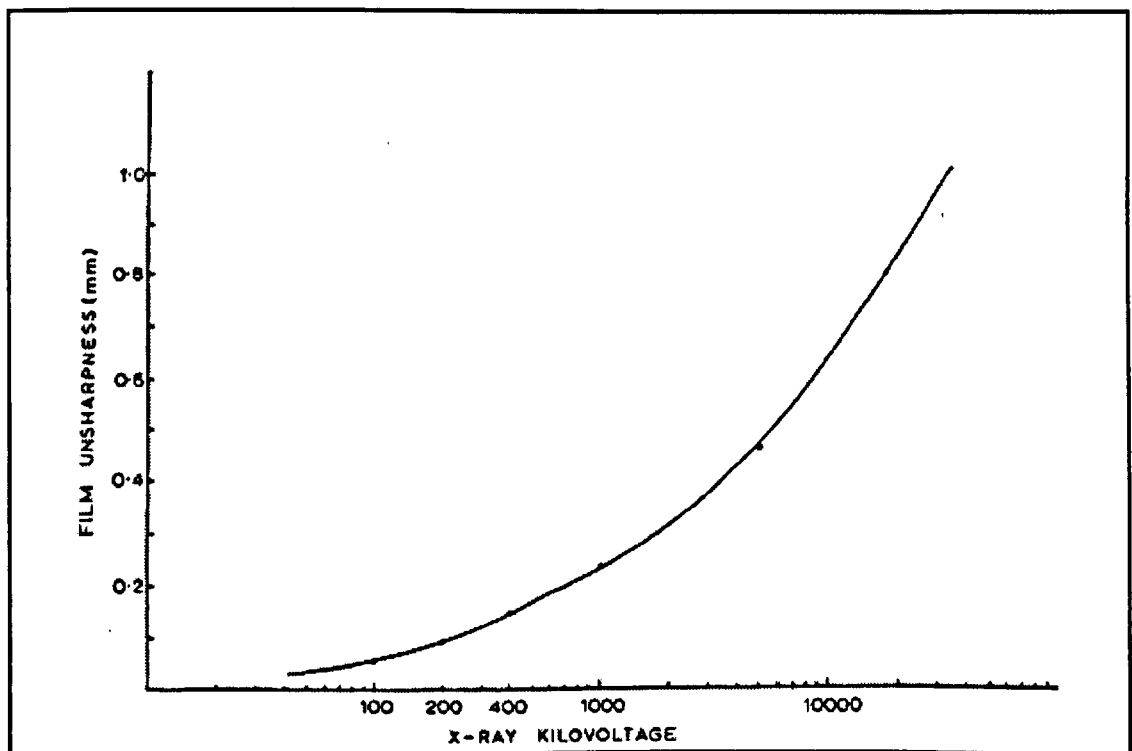


Figure 2.3 Experimentally Measured Film Unsharpness

The density of image formed on a film depends on the amount of ionising radiation reaching each point in the image during the exposure period. The amount of radiation

incident on the film is measured as the exposure, E , which is defined as the multiple of the incident radiant intensity, I , and the exposure time, t .

$$E = I t$$

The degree of film darkening produced by a given exposure level depends on the type of film used. The characteristic curve for a given film type is obtained by plotting film density as a function of radiant exposure, as illustrated in Figure 2.4. The gradient of this curve at a given film density level is often referred to as the film contrast or film gradient, G_D , and is a valuable measure of film sensitivity.

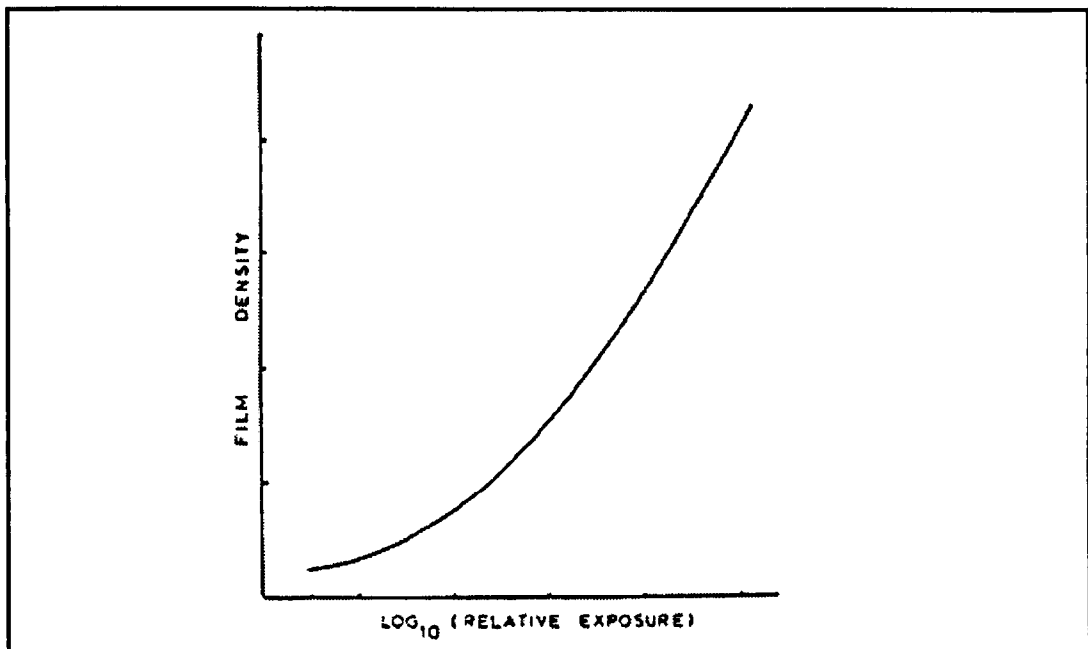


Figure 2.4 Film Density Characteristics

Film sensitivity may be increased by the use metallic intensifying screens, usually made of thin lead sheets. X-rays absorbed by these screens cause the secondary emission of electrons and low energy X-rays which in turn are more easily absorbed by the photographic film. The use of intensifying screens increases film sensitivity by a factor of between 2 and 5, but causes a slight increase in inherent unsharpness.

As previously discussed, radiographic film consists of silver halide grains suspended in a central emulsion layer. The high energy associated with ionising radiation dictates that grains must be reasonably large if a significant number of absorption events are to occur during the exposure period. Individual silver halide grains within an unexposed

radiographic film may vary in size between approximately 0.2 μm and 0.8 μm depending on the type of film used. Secondary ionisation effects within the film will cause many separate silver grains to be formed by a single absorbed X-ray quanta. Consequently, radiographic film images will have a pronounced grainy appearance compared to those produced by conventional photography.

The type of radiographic film used is highly dependent on factors such as required image quality, exposure time and radiographic technique. Films containing large silver halide grains are more likely to absorb incident radiation, and so require shorter exposure times to form an image. The sensitivity of a film to incident radiation is known as its speed, with fast films requiring short exposure times. In general reducing grain size increases image quality at the expense of film speed.

Many types of radiographic film are available from a variety of manufacturers and several classification systems exist to assess their performance. Unfortunately film classes are not generally consistent between any two manufacturers or standards. Recent standards such as ASTM E 746-87 have attempted to develop a film classification system based on measured sensitivity values, using standard test objects [12]. Ciorau has shown that under identical radiographic conditions the measured sensitivity of different films increases linearly as silver halide grain size is reduced [13].

2.5 Sensitivity Limitations

It is important when assessing a radiograph to have some idea of the sensitivity of the process used. For instance, it is futile to attempt to detect small cracks in a specimen if the radiographic process used is not sensitive enough to pick them up. Similarly, when considering the digitisation of radiographic film images it is necessary to ensure that all image features are reproduced without the inclusion of redundant or spurious data. The term radiographic sensitivity is necessarily vague since image quality is a somewhat subjective concept and is contributed to by several factors such as film characteristics, overall image density, inherent and geometric unsharpness and radiographic technique [14].

Radiographic sensitivity can be roughly divided into two categories, contrast sensitivity and resolving power. These two factors interact to set limits on overall image quality.

Contrast sensitivity refers to the ability of a radiograph to respond to small changes in specimen thickness by producing corresponding changes in film density. Resolving power is the ability of a radiographic process to accurately reproduce small features such as cracks within the test specimen.

This concept of contrast sensitivity and resolving power can be extended to the assessment of any imaging system. When considering the digitisation of radiographic films, contrast sensitivity depends on digitisation resolution, in other words the number of separate grey level values which may be reproduced from the analogue intensity variation of the image. Similarly, the resolving power of a digital imaging system is related to the spatial sampling resolution of the continuously varying image intensity.

2.5.1 Contrast Sensitivity

Consider the simple problem of imaging a stepped block of thickness, x , and step thickness Δx , as shown in Figure 2.5. If the block is reasonably large, any change in image contrast will be independent of feature size and dependent only on contrast sensitivity. Halmshaw has derived a theoretical expression for the minimum contrast sensitivity of a radiographic technique using this simple image model [15].

Factors affecting contrast sensitivity will include the characteristic film gradient, G_D , the linear absorption coefficient, μ , of the radiation source in the block material, and the ratio of scattered radiation, I_s , to direct radiation, I_D . Assuming a change in block thickness, Δx , causes a just discernible change in radiograph density, ΔD , the contrast sensitivity, S , of the technique is approximately :

$$S = \frac{\Delta x}{x} = \frac{2.3\Delta D}{\mu x G_D} \left(1 + \frac{I_s}{I_D} \right)$$

Typical values for ΔD range from 0.008 to 0.01 depending on viewing conditions and overall film density.

Experimental work has shown that practical results agree well with this equation. At first glance contrast sensitivity appears to be a useful measure of a radiograph's ability to record features, relating as it does parameters concerning film, exposure, specimen thickness and the eye's sensitivity to small density changes. Unfortunately contrast

sensitivity only measures a radiograph's ability to detect small changes in material thickness and does not measure ability to resolve small features. In practice we are not concerned with imaging stepped blocks but in identifying small features such as cracks. Although the contrast sensitivity of a radiograph may be a useful indication of the quality of radiographic technique used in its production, care must be taken in drawing conclusions about the ability to detect real image features.

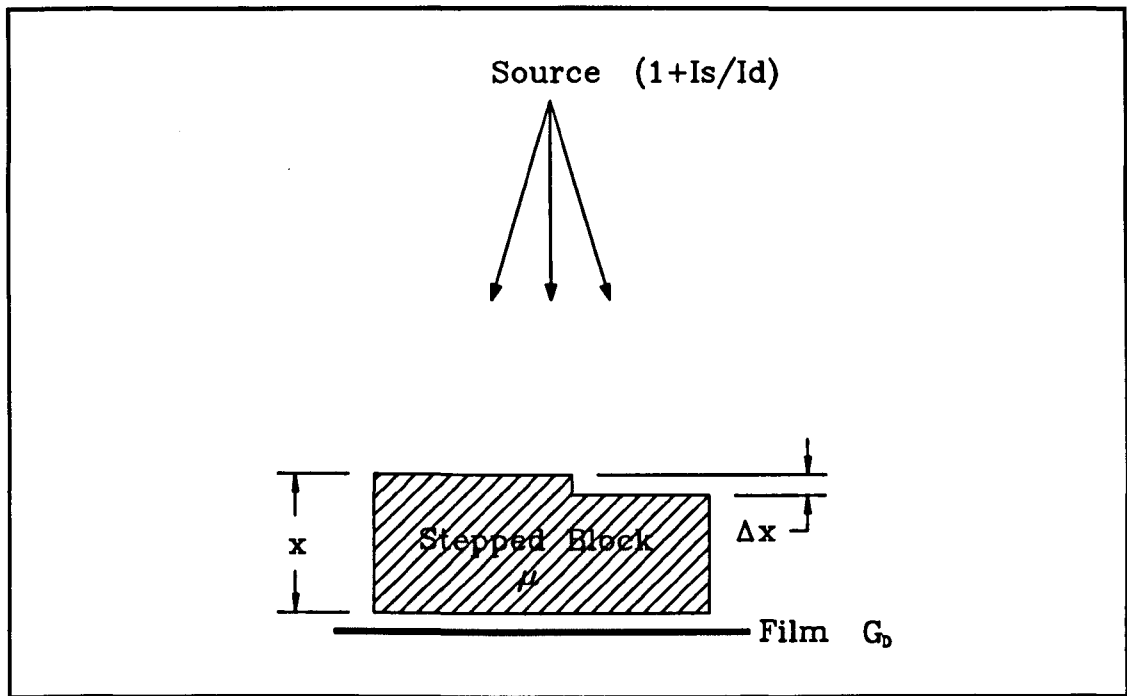


Figure 2.5 Contrast Sensitivity Model

2.5.2 Crack Sensitivity

A similar analysis may be carried out to assess the sensitivity of a given radiographic technique to small internal cracks within a test specimen. A commonly used formula for detection of a crack running at an angle, θ , to the radiation beam, with depth, d , and width, W , is given by :

$$d.W = \frac{2.3F\Delta D}{\mu G_D} \left(1 + \frac{I_s}{I_D}\right) (d\sin\theta + W\cos\theta + u_i) \quad (2.3)$$

F is a form factor with a value between 0.65 and 1.0. When considering the special case of a smooth crack ($F = 1$), of small dimensions ($w < 20 \mu\text{m}$), running parallel to the beam direction ($\theta = 0$), this formula reduces to :

$$d.W = \frac{0.02}{\mu G_D} \left(1 + \frac{I_s}{I_D} \right) u_t \quad (2.4)$$

Equation (2.4) illustrates that, when considering the detection of small cracks, the total unsharpness, u_t , becomes a major factor influencing resolving power. Assuming the two major contributors to the unsharpness of a radiographic image are geometric unsharpness, u_g , and inherent (film) unsharpness, u_f , the total unsharpness, u_t , of a radiograph is given by:

$$u_t = \sqrt{u_g^2 + u_f^2}$$

It has been shown that the use of test objects consisting of small surface notches to simulate cracks provide an accurate measure of attainable crack detection performance. The results obtained using this method show good correlation with work on specimens containing stress induced cracks which were subsequently destructively tested to assess the size and number of cracks present [16]. As a general guide, experimenters have found that smooth cracks imaged under ideal conditions with high fidelity equipment represent the limit of current radiographic imaging techniques if their dimensions are as summarised below.

- width = 5 to 20 μm
- depth < 5 % of object thickness
- length < 10 mm

These values agree roughly with guidelines set out in BS2910, concerning radiographic flaw sensitivity. However, it should be noted that these crack dimensions do not necessarily correspond to the same dimensions reproduced in the film image, since unsharpness effects will tend to widen the image of any crack.

2.6 Image Quality Indicators

In order to gain some quantitative information about the sensitivity of a radiograph to small features in a specimen it is standard practice to use a device called an image quality indicator (IQI) or penetrameter. An IQI is a test object of strictly controlled dimensions, manufactured of the same material as the specimen being radiographed,

which can be mounted on the specimen in order to estimate the minimum detectable feature size.

Many different types of IQI exist, all of which produce slightly different measures of radiographic sensitivity. The most common types of IQI are the wire IQI, the step/hole IQI, the plaque IQI and the duplex wire IQI. Most industrial standards include a requirement for all radiographs to include an IQI and quote minimum acceptable IQI sensitivities that must be attained. It should be noted that each type of IQI will give a different sensitivity value for the same radiograph and care must be taken to record which type of IQI was used as well as the sensitivity value achieved.

The detection of real flaws from a radiograph of a weld or casting, even under optimum radiographic conditions, cannot be derived directly from IQI sensitivities. The size, shape and orientation of the flaw all contribute to the final image. It is, however, reasonable to assume that a poor IQI sensitivity will correspond to a poor flaw sensitivity and an increase in IQI sensitivity is liable to produce some increase in flaw sensitivity.

The most simple form of IQI consists of a stepped block with increasingly small step sizes, the smallest step discernible on the processed radiograph is taken as a measure of radiographic sensitivity. This type of IQI corresponds to the case discussed above in the derivation of contrast sensitivity. As previously discussed contrast sensitivity is not a good measure of a radiograph's ability to resolve small flaws, hence stepped IQI's are no longer used as the sole measure of image quality.

Probably the most commonly used type of IQI in British industry is the wire type. This IQI consists of a set of parallel wires of varying thickness mounted in a flexible envelope. The wire diameters are dependent on the thickness of the specimen being radiographed. The mounting envelope should be optically transparent, with a low X-ray absorbency. Measured wire IQI sensitivity is usually quoted as the minimum visible wire diameter expressed as a percentage of weld metal thickness. Many standards include specifications for wire type IQI's, these include BS3971, DIN54109 and ASME V [17] [18] [8].

Different industrial standards tend to show close agreement in specifying minimum requirements for IQI sensitivity. An example of these specifications is given in Table 4 which shows the BS2910 standard for minimum acceptable IQI sensitivity to be achieved for single wall X-ray radiography using a wire type IQI. More stringent requirements are set by the Atomic Energy Authority Standard AESS 6001 which specifies IQI sensitivity should be less than 1% at all times, unless special circumstances apply.

Weld Thickness (mm)	Measured IQI Sensitivity (%)
2	2.5
6	1.6
12	1.6
25	1.3
50	1.0
100	0.8

Table 4 Minimum Acceptable IQI Sensitivities to BS 2910

A relatively new concept in radiographic sensitivity assessment is the use of separate IQIs to measure contrast sensitivity and resolving power. This type of IQI uses a stepped block to measure contrast sensitivity and a duplex wire IQI to measure resolving power [19][20]. Research has shown that this type of IQI exhibits a good correlation between measured IQI sensitivity and the ability to detect real cracks in a specimen [21].

When considering the problem of film digitisation it should be noted that IQI's offer a good target image to assess the imaging capabilities of a system. A minimum requirement of any such system should be that any IQI feature visible on a radiographic film should be reproduced in the digitised image.

2.7 Radiograph Interpretation

It should be noted that one of the most demanding aspects of radiographic NDT involves interpretation of the radiographic image. Many different types of defect may

be detected by radiography, each with a different characteristic appearance. Obviously some types of defect will be more difficult to detect than others.

Another consideration which should be borne in mind is that not all defects are a cause for component rejection. In general planar defects, such as cracks, are much more serious than spherical defects. This is because the sharp edges of planar defects produce high stress concentrations which may lead to progressive crack propagation, and component failure. As an example of the different types of defect which may be encountered and their relative seriousness, Table 5 summarises the acceptance criteria laid down by BS5500.

Defect Type	Permitted Maximum
Cracks and Tears	Not Permitted
Lack of Fusion	Not Permitted
Lack of Root Penetration	Not Permitted
Isolated Pores	$\varnothing < 3\text{mm}$ $\varnothing < \text{metal thickness}/4$
Localised Porosity	2 % by area
Wormholes	length $< 6\text{mm}$ width $< 1.5 \text{ mm}$
Slag Inclusions	length $< \text{material thickness}$ width $< \text{material thickness}/10$
Tungsten Inclusions	As Isolated Pores
Copper Inclusions	Not Permitted
Undercut	Depth $< 0.5 \text{ mm}$
Excess Penetration	Height $< 3 \text{ mm}$

Table 5 Weld Acceptance Criteria to BS5500

2.8 Performance of the Human Eye

Since the aim of the proposed film digitisation system is to aid in the interpretation of visual information by a human observer, it is necessary to consider the phenomenon of

human visual perception. Any assessment of the response of the eye relies on the subjective response of an observer to changes in visual stimuli. This makes quantitative assessment difficult and leads to a rather vague understanding of the human visual response, combining elements of both physiology and psychology. However, much experimental work has been carried out in this field which provides a useful guide to human visual response [22].

The human eye operates over a huge dynamic range of light intensities, somewhere in the order 10^{10} . Over this range the subjective response to changes in brightness is very approximately logarithmic. The eye is not capable of distinguishing the entire range of intensities at any one time, instead it adapts to some overall brightness level at which it can distinguish only a limited range of intensities. Intensities outside this limited range will be seen either as pure black or pure white, unless the overall brightness level changes enough to cause a change in brightness adaptation level. The contrast sensitivity of the eye rapidly diminishes when trying to distinguish objects not at the brightness adaptation level. In practice when the eye scans an image it will be continuously adapting to the local brightness level of each region within the image. As well as being affected by image brightness the adaptation level of the eye will depend on ambient lighting conditions.

Research indicates that under ideal viewing conditions, a qualified film interpreter can detect density variations as small as 0.008 OD [23]. This finding agrees well with classical work carried by Weber concerning the ability of the eye to discern small contrast changes over a range of brightness adaptation levels [24]. These experiments measured a subjects ability to distinguish a change in image intensity, ΔI , from a uniform background intensity, I . Weber found that at any darkness adaptation level, the eye has a dynamic range of approximately 2.2 Log units over which the ratio of ΔI to I remains virtually constant, at a value conventionally referred to as the Weber ratio.

$$\text{Weber Ratio} = \frac{\Delta I}{I} = 0.02 \quad (2.5)$$

Combining equation (2.2) and equation (2.5) gives an estimate of the minimum discernible density change, ΔD , within a radiographic film image.

$$\Delta D = \text{Log}\left(\frac{I + \Delta I}{I}\right) = \text{Log}(1 + 0.02) = 0.0086$$

The ability of the human visual system to resolve image features is dependent both on the contrast between areas of different intensity, and the size of these areas. If the spacing of these contrasting areas is regular, it may be characterised in terms of spatial frequency. When considering the response of the eye, the most meaningful measure of spatial frequency is the number of intensity cycles per degree subtended at the eye. The human visual response to changes in both contrast and image size has been investigated by measuring the ability to resolve sinusoidally varying intensity test patterns of varying spacings [25].

The findings of this research are summarised in Figure 2.6, which plots contrast sensitivity as a function of spatial frequency. It can be seen that the human visual response is more sensitive to certain spatial frequencies than others. In particular at spatial frequencies above 50 cycles per degree, virtually no contrast information can be discerned. This corresponds to the case of a 15 μm wide feature, viewed at a distance of about 100 mm, and represents the limit of the human eye's ability to resolve detail. In practice small details will not consist of sharply defined grating patterns, but irregular, unsharp edges, which will be more difficult to resolve.

The number of possible grey levels, G , displayed in a digital image depends on the digitisation fidelity expressed in bits per pixel, N .

$$G = 2^N$$

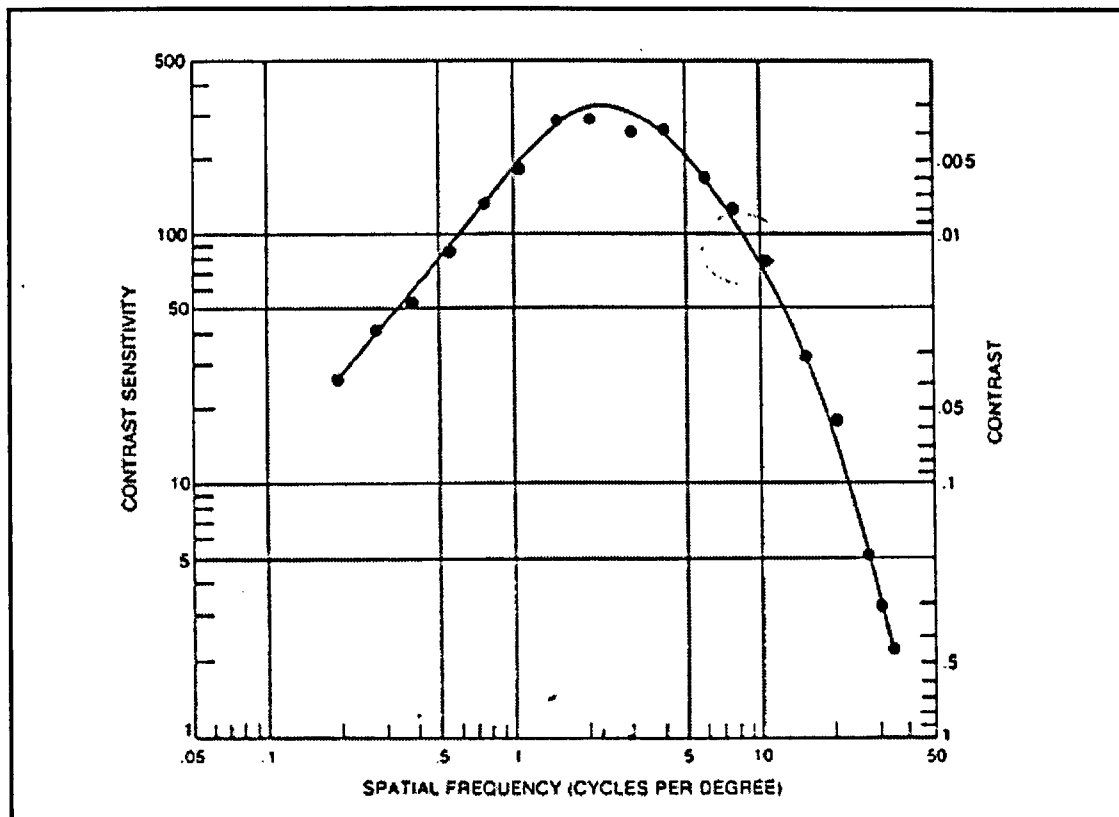


Figure 2.6 Contrast Sensitivity of the Human Visual System

Ideally the intensity difference between adjacent grey levels should be small enough that the digital image appears to the observer as a constantly varying light intensity function. As the displayed digital resolution is reduced, images begin to exhibit a loss of contrast and a degree of false contouring so that data capture at low digitisation fidelity would discard useful information available on the original radiograph. However, if digitisation fidelity is high the system may well be processing a large amount of redundant data. It should also be noted that high digitisation fidelity does not necessarily assure high data accuracy, since the signal being sampled may itself be noisy and inaccurate. Empirical evidence suggests that a digital display offering between 100 and 200 separate grey level quantisation levels gives a smooth variation of intensity so that the image appears continuous to the human eye [26].

Although digital image processing techniques can make visible more contrast information than is available in an unprocessed image, these techniques are limited by the resolution of the display system used. Most monochrome display adapters have a digital fidelity of 8bits/pixel and high resolution systems are available displaying 10

bits/pixel. In addition it should be noted that post-processing of high fidelity data may lead to the accentuation of spurious data, leading to false indications of image features.

Finally, it should be noted that the perception of visual data is highly dependent on viewing conditions and the experience of the observer. Radiographic film inspection involves placing the film image on an illuminated screen and observing the light transmitted through the radiograph. Industrial standards specify that the area of interest should be masked off and viewed in a darkened room, in order to eliminate the effects of extraneous light from the viewing screen and the surrounding environment. BS2600 states that a minimum luminance of 30 cd/m^2 should be transmitted through the radiograph.

2.9 Real Time Radiography

Although most industrial radiography still uses some type of photographic film system, this method has several disadvantages. A major drawback associated with film radiography is the time delay incurred due to the need for film processing before the image can be viewed. In particular this limitation prevents the development of interactive inspection systems, where the test specimen orientation can be adjusted to enhance defect visibility. In addition, when considering the large number of radiographic images required during the inspection of a large industrial installation, such as a power station, it is clear the production of film images will be expensive and lead to problems with image storage and reproduction.

Much work has been undertaken over the last few years to develop a system which produces acceptable radiographic images for immediate viewing and evaluation [27]. This type of technique is referred to as real time radiography (RTR), although systems may not strictly operate in real time, due to delays caused by hardware and software image processing. Despite many advances over the last decade no real time system can currently match the high quality imaging and sensitivity of film systems.

RTR systems frequently produce digital images, which may benefit from computer techniques such as image processing and automated image interpretation. Digital systems also benefit from the ease with which digital images may be stored, transmitted

and reproduced. Many of these techniques used by RTR systems for image manipulation and evaluation are equally applicable to film digitisation apparatus.

2.9.1 Fluoroscopy and Image Intensifier RTR Systems

The earliest work carried out on real time radiography involved a technique known as fluoroscopy, illustrated in Figure 2.7. This type of system uses a fluorescent screen which converts incident X-radiation into visible light, which is subsequently detected by a CCTV¹ camera. The output image from the camera is usually digitised and held in a frame store, where it can be manipulated and enhanced using digital image processing techniques, before being displayed on a television monitor. Since incident radiation causes the screen to fluoresce, bright areas within the image will correspond to areas of low absorption within the specimen. Hence, the contrast distribution will be the reverse of that observed in a film image. One advantage of fluoroscopic imaging systems is that the image may be enlarged simply by increasing the distance between object and screen.

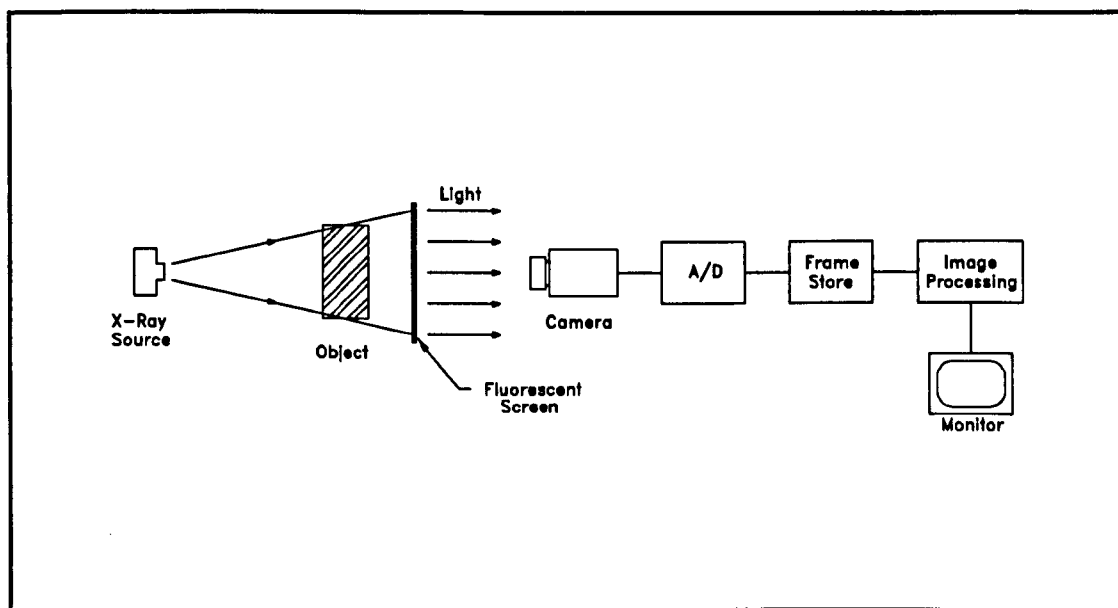


Figure 2.7 Fluoroscopic Real Time Radiography

Many different types of conversion screen have been developed mostly using a crystalline scintillator materials such as cadmium sulphide or caesium iodide. More

¹ Closed Circuit Television

recent research has investigated the use of two dimensional arrays of semiconductor photodiodes instead of fluorescent screens. However, both types of conversion process are characterised by an inefficient conversion of X-rays to visible light, resulting in very low intensity images.

These dark images typically exhibit much less contrast information than those produced by film systems. In order to optimise the light output of fluoroscopic systems it is necessary to increase the size of the scintillator crystals used in the conversion screen, resulting in images having a pronounced grainy appearance. Hence, the performance of fluoroscopic screen systems at resolving small variations in contrast and detail is significantly worse than that achieved by conventional film systems.

Since the main disadvantage of fluorescent screen systems is the low luminous intensity of images formed by the conversion screen, some method of amplifying the luminous output from such a system is required. One solution to this problem involves using an image intensifier tube which converts an X-ray image to light, then to electrons and finally back to light, which may then be viewed using a camera. A simple image intensifier RTR system is illustrated in Figure 2.8. The image produced by an image intensifier system is typically several hundred times brighter than that formed using a conventional fluorescent screen. This results in images exhibiting a significant increase in contrast and detail resolution. The major disadvantage of image intensifier systems is that the inspection area is limited by the diameter of the intensifier tube, which is usually less than 300 mm.

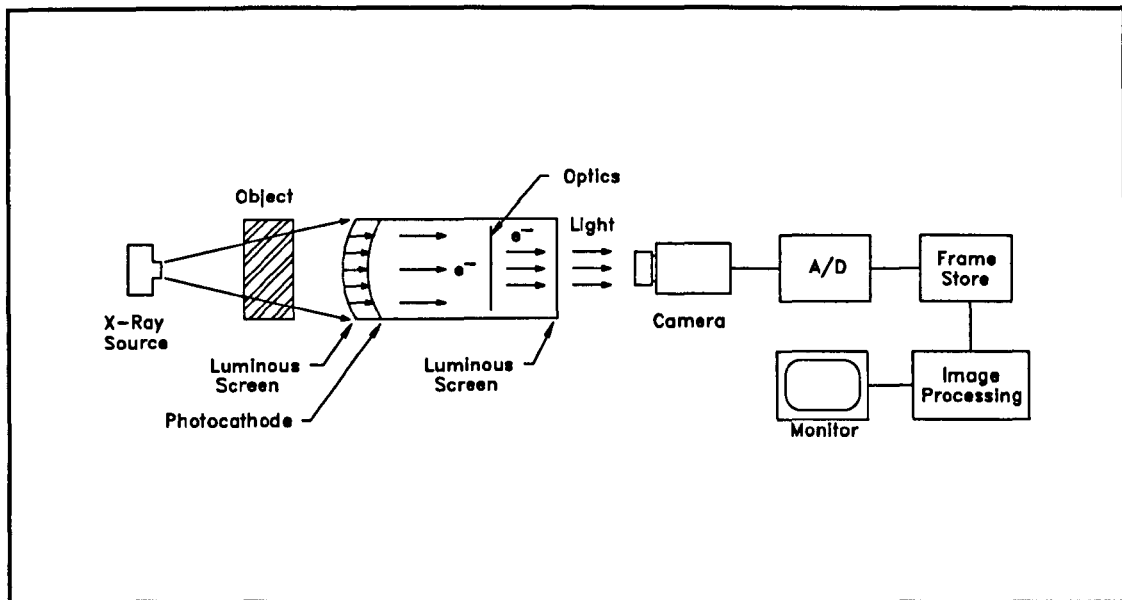


Figure 2.8 X-Ray Intensifier Real Time Radiography

2.9.2 Solid State RTR Systems

The performance of the most sensitive X-ray image intensifier systems is still significantly poorer than comparable film radiography techniques. Recent efforts to improve the quality of RTR images have involved the use of solid state imaging technology. These systems use an array of X-ray sensitive semiconductor elements to convert radiant energy directly to a sampled video signal. As with the other RTR systems discussed previously, this video signal may then be digitised, processed and displayed on a television monitor. Solid state X-ray imaging systems usually use a linear sensor array which is scanned behind the irradiated specimen to build up a two-dimensional image. This principal is illustrated in Figure 2.9.

Low power X-radiation, below about 5 keV, may be detected using a standard photodiode array sensor [28]. Detectors designed for high power X-rays usually incorporate an array of scintillator elements, which convert incident radiation into light, followed by a photodiode array. Modern semiconductor technology produces sensors with high sensitivity to X-radiation and high spatial sampling resolution. Consequently, solid state RTR systems can produce high quality radiographic images. Comparative tests between solid state and image intensifier systems have demonstrated a significant improvement in image quality when using solid state sensors [29]. However, film radiography still remains by far the most sensitive industrial radiographic technique.

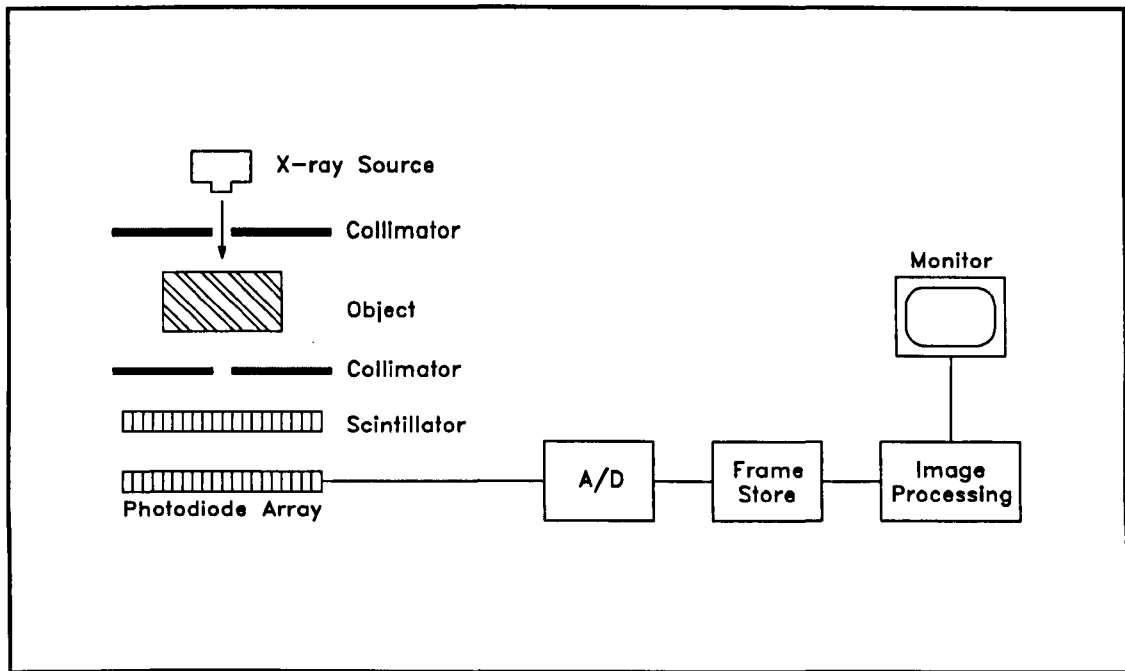


Figure 2.9 Solid State Real Time Radiography

2.10 Conclusions and Specifications for a Film Digitisation System

The preceding discussion has identified several factors influencing the ability of an observer to detect indications of internal flaws within a test object, using film radiography. These factors include flaw size and orientation, radiographic technique, the type and intensity of the radiation source used, viewing conditions, overall film density, film characteristics and the response of the human visual system. The effect of these factors will now be summarised with a view to specifying the required operating parameters for a radiographic film digitisation system.

One major factor limiting the sensitivity of radiographic film systems is the total unsharpness of the image. The overall effect of image unsharpness is to spread contrast data over an extended area. The two main contributors to the total unsharpness of a radiograph are geometric unsharpness and inherent unsharpness. It has been shown that even under ideal conditions, both sources of unsharpness are unavoidable. Although geometric unsharpness is highly dependent on specimen thickness and radiographic technique, for a relatively thin specimen, practical values will typically be between 20 μm and 100 μm . Inherent unsharpness depends on the intensity of the radiation source used and, to a lesser degree, on film characteristics and may vary between 50 μm and several millimetres.

The discussion of the sensitivity of radiographic film systems identified two separate components, contrast sensitivity and resolving power. A theoretical derivation of contrast sensitivity was proposed which has been demonstrated to agree well with values measured by IQI. This work indicates that thickness changes of approximately 1 % of total object thickness should be detectable, in the absence of spatial resolution limitations.

It was noted that reducing object size tended to reduce the amount of contrast information reproduced in the image. This effect is particularly relevant to the detection of small defects such as cracks. Both measured and theoretical data concerning crack detection suggest that small cracks with a width less than 20 μm represented the limit of modern high fidelity radiographic techniques. It was also noted that, when imaging such small features, unsharpness effects tend to predominate, limiting the size of the reproduced artefact.

The performance of the human visual system is also a major factor affecting feature detection in film radiography. When viewing an image, the human visual response operates within a very restricted region of its total dynamic range, with the range of intensities typically varying by a factor of less than 10^3 . The contrast sensitivity of the eye limits the minimum change in film density which may be detected, under ideal viewing conditions, to about 0.008. Feature size also affects the resolving power of the eye with a limiting spatial frequency of around 50 cycles per degree, corresponding to the detection of objects less than 15 μm wide viewed at close range.

The only current standard relating to the digitisation of radiographic film images is contained in the ASME V regulations (section 5, Article 2, Appendix III). Due to the scarcity of relevant research in this area the standard is necessarily vague and does not attempt to define acceptable operating parameters. However, the document does provide for the use of digital image processing techniques to enhance these digitised images and specifies that the evaluation procedure should include assessment of both modulation transfer function (MTF) and IQI sensitivity measurement.

The most pertinent British Standard is BS7009 which relates to the real time radiography of welded joints [30]. This document states that digitised radiographic

images should be subject to the same acceptance standards as film images, including the assessment of IQI sensitivity values. BS7009 also recommends a limiting resolution of not less than 5 cycles/mm as determined using a calibrated test pattern or modulation transfer methods.

Another relevant document is a review of current technology relating to radiographic film digitisation commissioned by the U. S. Nuclear Regulatory Commission (USNRC) [31]. This report notes the absence of international standards relating to the subject, and suggests operating parameters for digitisation of high quality radiographic film images for use in the nuclear industry. The main recommendations contained in the USNRC report are summarised in Table 6.

Parameter	Specification
Contrast Sensitivity	0.01 OD to 0.03 OD
Sampled Spatial Resolution	< 50 μm
Dynamic Range	1 OD to 4.5 OD
Modulation Transfer Function	≥ 0.33 @ 5 cycles/mm
Digitisation Fidelity	12 bits/pixel
Data Compression	Lossless

Table 6 Summary of USNRC Recommendations for Radiographic Film Digitisation

As a general observation it may be stated that the aim of any film digitisation system should be to retain all the detectable features from the original radiograph without the introduction of any spurious information due to system noise. The image displayed by the system should retain all discernible contrast information visible to the naked eye and may well, when allied to computer image processing software, provide more contrast information than that seen on the original radiograph.

The required performance of the system will vary greatly depending on the proposed industrial application. It is proposed here to summarise minimum standards for use in the nuclear industry, since this is probably the most onerous field of industrial

radiography. The same technology may then be applied to less critical fields with a corresponding reduction in digitisation resolution and sampling density.

In consequence of the preceding discussion it was felt that modern high resolution video cameras, having a display resolution of around 1000 pixels by 700 pixels, were inadequate for the task of digitising large radiographic films. In order to attain the required spatial sampling resolution it was decided to use solid state semiconductor image sensors, which can be manufactured in the form of very closely spaced linear or area photosensor arrays. Since the task of film digitisation involves the scanning of extended areas, linear array sensors were preferred, so that both spatial sampling resolution and scan area could easily be varied.

The two most significant factors affecting image quality were identified as spatial sampling resolution and digitisation fidelity. The two sensors evaluated during this project were both general purpose CCD photosensors, with spatial sampling resolution of between 10 μm and 14 μm . This resolution was deemed ample for even the most demanding imaging tasks, such as the crack detection. A digitisation fidelity of 12 bits/pixel, corresponding to 4096 separate grey scale levels, was chosen so that a large amount of contrast data was available for use by the image processing software. However, it should be noted that high digitisation fidelity does not necessarily guarantee good contrast sensitivity, since signal noise may obscure legitimate image data. Typical noise characteristics quoted by manufacturers of CCD sensors give signal to noise ratios of around 1000:1.

In the final analysis, the performance of a radiographic interpretation system can only be gauged by comparative tests with the performance of an experienced inspector. Therefore properly controlled tests must be designed and used as the basis of any justification of the system's success. Prior to this, it was proposed to undertake modulation transfer and IQI detection tests in order to measure the sensitivity of the film digitisation equipment.

2.11 Summary

This chapter contains a brief discussion of radiographic theory and the sensitivity limits of film radiography methods. Several factors affecting the minimum detectable feature

size have been considered, including contrast sensitivity, crack sensitivity and the performance of the human visual system. From this discussion a specification has been developed for the required operating parameters of a high integrity radiographic film digitisation system.

References

- 1 R. Halmshaw, "Industrial Radiology Techniques", Wykeham Publications, 1971.
- 2 Industrial Radiography Handbook, Agfa Gevaert Ltd, 1980.
- 3 BS2910, "Radiographic Examination of Fusion Welded Butt Joints in Steel Pipes", British Standards Institution, 1986.
- 4 BS2600, "Radiographic Examination of Fusion Welded Butt Joints in Steel Plates", British Standards Institution, 1983.
- 5 BS5230, "Measurement of Speed and Contrast of Industrial Radiographic Films", British Standards Institution, 1975.
- 6 BS3915 Steel Pressure Vessels for Nuclear Reactors", British Standards Institution, 1965.
- 7 ASME Boiler and Pressure Vessel Code, Section V, "Nondestructive Testing", American Society of Mechanical Engineers, 1992.
- 8 BS5500, "Unfired Fusion Welded Pressure Vessels", British Standards Institution, 1991.
- 9 Atomic Energy Standard Specification AESS6001, "Radiography of Welds in Plate, Sheet, Bar or Tube", U. K. Atomic Energy Authority, 1983.
- 10 NF0081, "B. N. F. L. Standard Procedures for Fabrication", 1988.
- 11 P. Ciorau, "The Influence of Density on the Detection of Linear Discontinuities", Proc 5th Eur Conf NDT, London, Vol. 3, 1987, pp 2017-2024.
- 12 ASTM E 746-87, "Method for Determining Relative Image Quality of Radiographic Film", Annual Book of ASTM Standards, vol. 03.03, 1988.
- 13 P. Ciorau, "A Contribution to the Testing and Classification of Industrial Radiographic Films", NDT & E International, 23 3, June 1990, pp 131-156.
- 14 M. Evrard and J. Dubresson, "Image Quality in Radiography", Non Destructive Testing, 4 6, December 1971, pp 396-400.

- 15 R. Halmshaw "Industrial Radiology Techniques", Wykeham Publications, 1971, pp 166-169.
- 16 P. Ciorau, "Critical Comments on the Use of X-rays and Ir-192 for Detecting Linear Discontinuities", NDT International, **23** 6, December 1990, pp 345-350.
- 17 BS3971, "Image Quality Indicators for Industrial Radiography", British Standards Institution, 1980.
- 18 DIN54109, "Nondestructive Testing - Image Quality of Radiographs of Metallic Materials", Deutsches Institut Fur Normung, 1978.
- 19 CERL IQI Specification, Parts A, B and C, Central Electricity Research Laboratories, 1973.
- 20 EN 462, Part 5, "Image Quality Indicators (Duplex Wire Type)", Comite Europeen de Normalisation, 1992.
- 21 M. J. Feaver, "Comparative Performance of BSI and CERL Image Quality Indicators in Detecting Natural Cracks in Steel", NDT International, **11** 1, 1977, pp 13-20.
- 22 S. S. Stevens, "Handbook of Experimental Psychology", Wiley Press, 1951.
- 23 ASTM E-7, "Visual Actuity Testing for Radiographic Examiners in Nondestructive Testing", 1991.
- 24 T. N. Cornsweet, "Visual Perception", Academic Press, 1970.
- 25 F. W. Campbell and L. Maffei, "Contrast and Spatial Frequency", Scientific American, Vol. 231, November 1974, pp 106-114.
- 26 R. C. Gonzalez and P. Wintz, "Digital Image Processing", Second Edition, Addison-Wesley, 1987, pp 21-29.
- 27 R. Halmshaw, "An Assessment of the Present Status of X-ray Television Fluoroscopy Real Time Radiography Methods", British Journal of NDT **30** 4, 1988, pp 257-258.
- 28 "Application of Reticon Photodiode Arrays as Electron and X-ray Detectors", E. G. & G. Reticon Application Note 101, 1990.
- 29 Y. M. Zhu, D. Babot and G. Peix, "A quantitative Comparison Between Linear X-ray Sensitive Array and Image Intensifier Systems", NDT International, Vol. 23, August 1990, pp 214-220.

- 30 BS7009, "Applications of Real Time Radiography to Weld Inspection", British Standards Institution, 1988.
- 31 "Review and Evaluation of Technology, Equipment, Codes and Standards for Digitisation of Industrial Radiographic Film", U. S. Nuclear Regulatory Commission, Washington D. C. May 1992.

3. Review of Solid State Imaging Technology

3.1 Introduction

Attempts have been made to digitise radiographic film images using a conventional high quality television camera for image acquisition [1]. However the high pixel sampling density required to produce acceptable results means that only small areas may be imaged in a single frame at the required resolution.

A major drawback associated with the use of a conventional camera and lens arrangement is that complex calibration procedures are required to relate dimensions in object space to the corresponding dimensions re-produced in the image. This calibration procedure must be repeated each time the camera is re-focused. In practice this makes direct measurement of image features cumbersome and prone to error.

The most simple method of obtaining high spatial resolutions is to use a small area photosensor in direct contact with the object [2]. In this way a direct relationship of each sampled point in the object to each pixel in the image is obtained. In addition the simplified optical arrangement removes the field distortion associated with lens systems.

Solid state image sensor technology offers the ideal means of achieving a direct contact scanning system, such as the one described above. A linear array photosensor was chosen for this project because it offered the flexibility to vary image dimensions and sampling interval by mechanical scanning, whilst offering the high speed image acquisition associated with modern solid state imaging technology.

Since their first practical implementation in the early 1970s solid state imaging devices have found applications in fields as diverse as optical character recognition (OCR), compact television cameras, facsimile machines, medical imaging, and document scanning [3]. Modern semiconductor fabrication methods allow an extremely high density of photosensitive elements to be deposited on a single chip. This has resulted in cheap, compact sensors that are simple to operate and offer very high resolution imaging. Both linear (up to 7000 elements) and area (up to 1024 x 1024 elements) array sensor are commercially available.

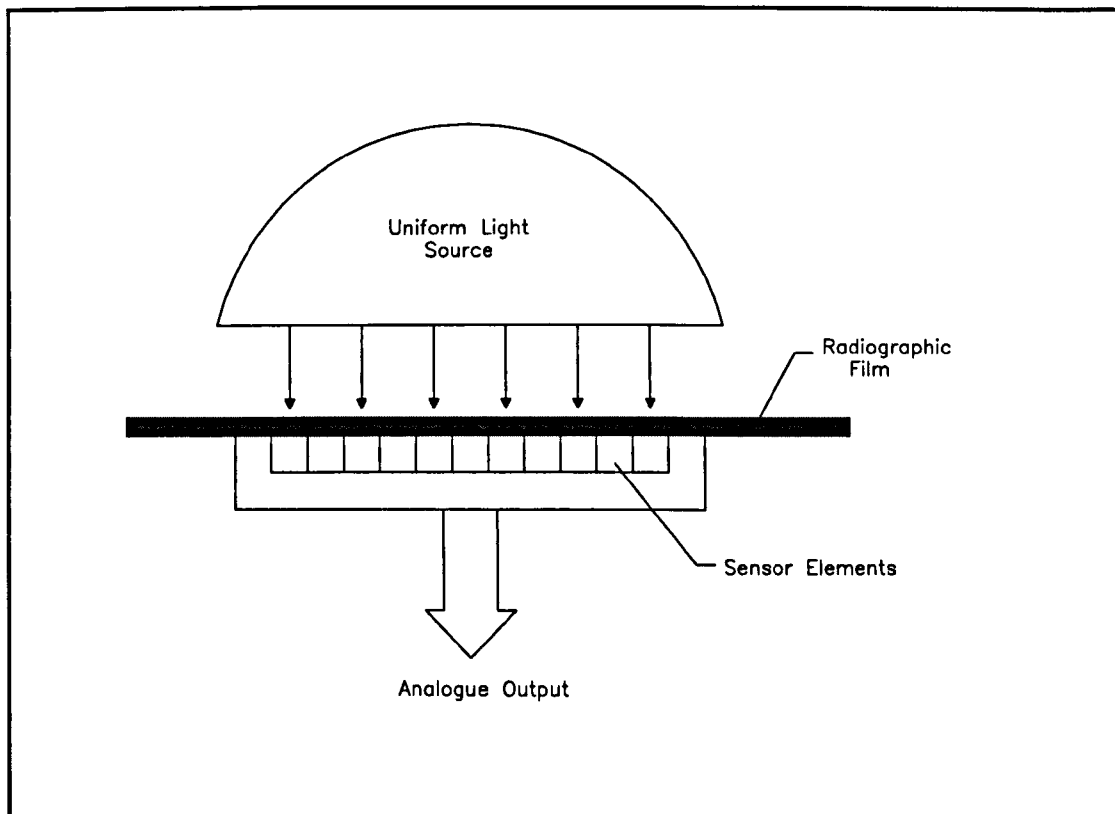


Figure 3.1 Direct Contact Image Scanning

Solid state image sensors may be sub-divided into three categories, self scanned diode arrays [4], charge injection diode (CID) arrays [5] and charge coupled devices (CCD) [6]. These three technologies have many similarities and hybrid sensors combining more than one principle are currently available.

Most recent advances in solid state imaging have involved CCD technology, resulting in sensors exhibiting high sensitivity, low noise characteristics and high sampling resolution. Consequently CCD sensors were chosen for this project due to their excellent imaging capabilities. A short review of CCD concepts will be given in this chapter in order to identify the relevant performance characteristics and possible sources of image degradation associated with their use.

3.2 Photodiodes and Dark Current

The basis of most solid state image sensors is the reverse biased p-n junction photodiode, operating in the integration and storage mode [7]. A brief discussion of this device will be necessary in order to define some important operating characteristics.

If a p-n junction is reverse biased to a preset potential, a quantity of charge will accumulate within the depletion layer electric field. If the junction is then open circuited the depletion layer capacitance will discharge. Under zero illumination the only current flowing across the p-n junction will be due to thermally generated minority charge flowing under the influence of the depletion layer electric field. This dark current is independent of applied reverse bias and dependent only on temperature.

Incident luminous energy causes the production of electron-hole pairs in the depletion layer by the same mechanism as thermally generated charge. The rate of photo-generated charge production, and hence the discharge current, is directly proportional to the instantaneous incident light intensity. An additional thermally generated dark current also flows, this current can be reduced by cooling the device.

If the photodiode is periodically re-charged, by applying a reverse bias, and then allowed to discharge, the amount of charge removed will be proportional to the integral of illumination intensity taken over the discharge time. The interval between successive re-charges of the photodiode is known as the integration time. This mode of operation increases the sensitivity of the sensor by integrating the incident illumination over the exposure period. In addition the integration time may be varied in order to maximise the output signal and increase dynamic range.

3.3 Photodiode Array Sensors

In order to construct a practical image sensor some method must be found to combine the signal from several photodiodes into a common video signal, consisting of a serial chain of pixels. An early approach to this problem was the self scanned photodiode array shown in Figure 3.2, where each photodiode is connected via a MOS switch to a common video line. The switches are then successively clocked, connecting each photodiode in turn to the video line. The requirement of each photodiode within the photodiode array to be connected to an associated MOS switch necessitates a reasonably complex array structure and limits the minimum size of each photosite.

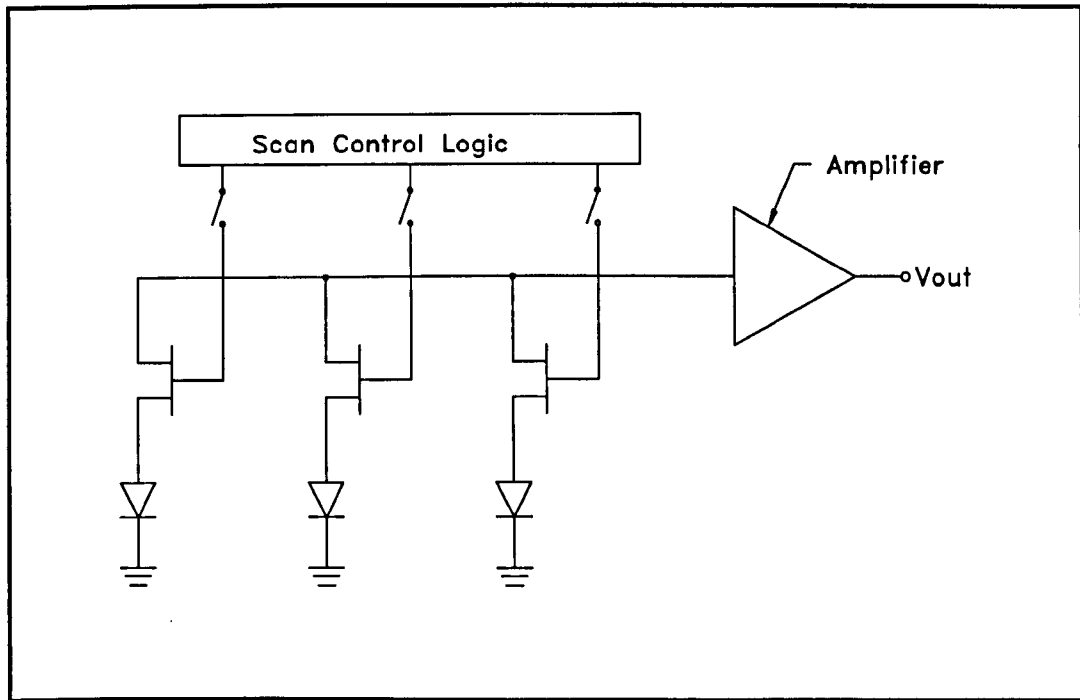


Figure 3.2 A Self Scanned Photodiode Array

3.3.1 The CCD Shift Register

The field of solid state image sensors was revolutionised by the advent of the charge coupled device (CCD) shift register. The charge coupled device operates by the manipulation of electrical fields within a semiconducting material to control the motion of individual "packets" of charge. Using a CCD shift register, charge packets from an array of photosensors may be clocked through the device and delivered to the output as a serial sequence of pixels.

Figure 3.3 illustrates the most simple implementation of a CCD shift register, first developed in 1969 by Boyle and Smith, using what is known as a three phase surface channel device [8]. The device consists of a series of metal electrodes isolated from a channel of p-type Silicon semiconductor by an insulating Silicon Dioxide layer.

Raising the potential of an electrode causes the formation of a "potential well", a localised area which is attractive to electrons, just under the semiconductor surface. Potential wells are illustrated in the diagram by potential energy profiles shown as dotted lines. By manipulation of electrode voltages the position of potential wells within the semiconductor can be controlled.

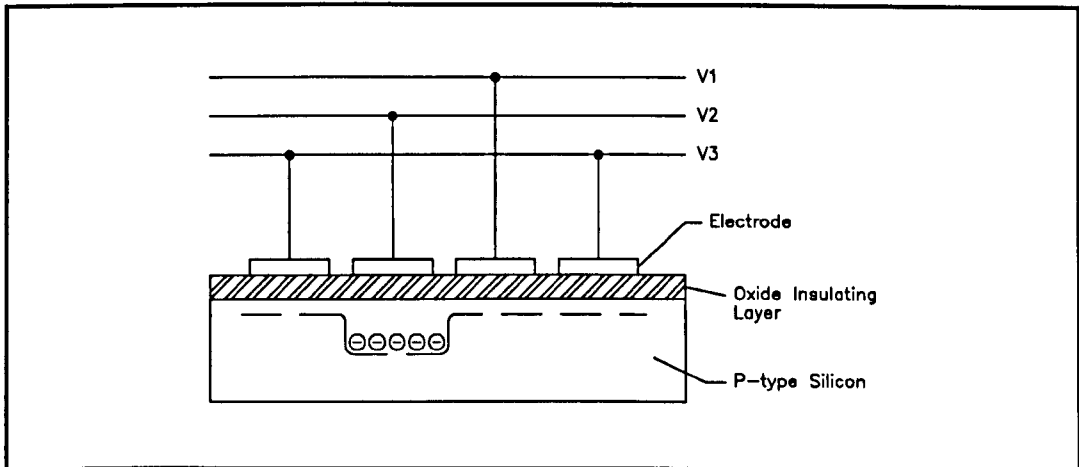


Figure 3.3 Charge Storage in a Three Phase CCD Element

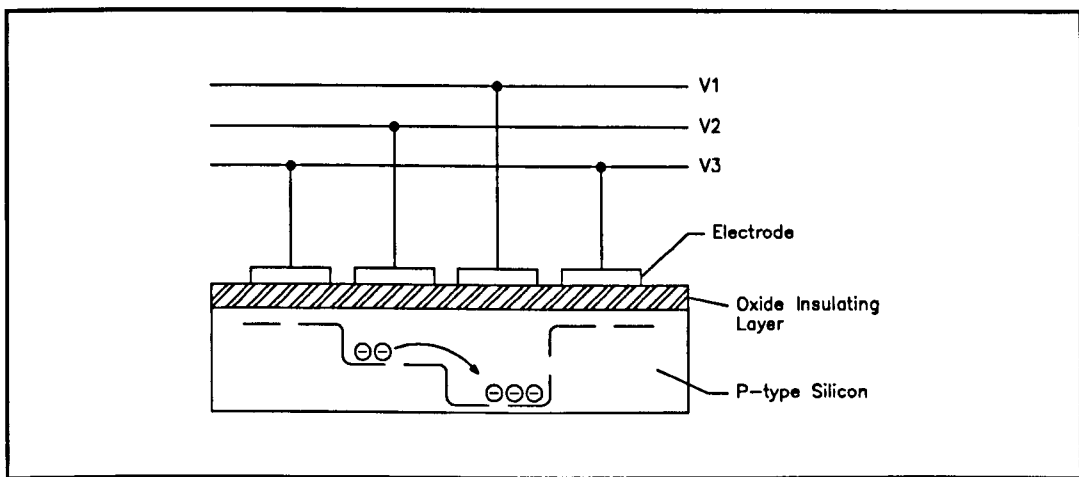


Figure 3.4 Charge Transfer Between Three Phase CCD Elements

Figure 3.3 shows a single electrode, V_2 , switched on so that an electron charge packet is constrained within the resulting potential well. In Figure 3.4, electrode V_3 is raised to a higher potential than V_2 creating a deeper potential well and causing the charge packet to move to the right.

In order to control the motion of charge packets within such a device the voltages at three adjacent electrodes must be controlled by clock signals operating 120° out of phase. Hence, this structure is referred to as a three phase CCD shift register.

3.3.2 Improvements to the Basic CCD Structure

The structure outlined above stores charge packets just below the surface of the silicon semiconductor channel, in close proximity to the interface with the insulating oxide

layer. The interaction of charge packets with this junction prevents the complete transfer of charge between shift register elements and introduces noise.

These problems can be alleviated using the so-called buried channel CCD structure, in which a layer of n-type silicon is deposited within the p-type substrate [9]. This has the effect that the potential minimum now occurs within the bulk semiconductor, away from the insulating oxide layer, as shown in Figure 3.5.

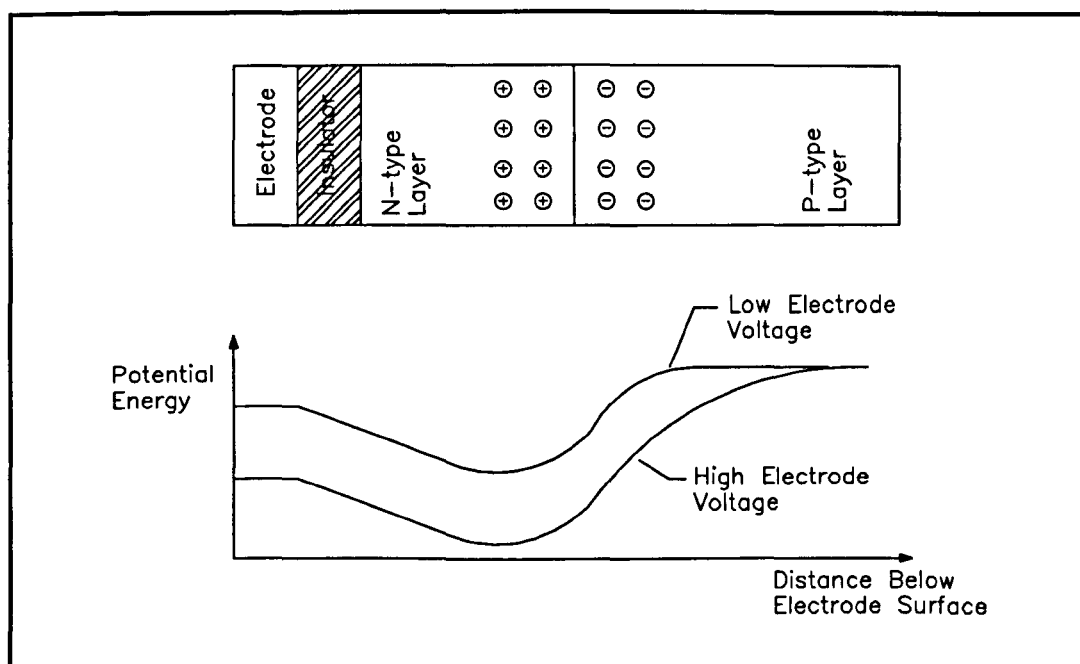


Figure 3.5 The Production of Potential Wells Within A Buried Channel CCD

The clocking requirements of a CCD register can be simplified by selectively changing the doping concentration of the n-type semiconductor just under a portion of each electrode as shown in Figure 3.6. This doping (shown in the diagram as + signs) creates a potential gradient within each element of the shift register, so that permanent potential wells and barriers are formed. This structure can achieve uni-directional motion of charge packets by clocking alternate electrodes using two clock signals 180° out of phase, and is known as a two phase CCD shift register.

A structure known as virtual phase has been developed in which each CCD element consists of a single surface electrode and an additional p-type electrode, biased at the substrate potential, deposited into the surface of the silicon [10]. Selective doping under each electrode produces a series of potential steps which can be manipulated by

a single clock signal to produce charge transfer, as illustrated in Figure 3.7. Virtual phase technology offers several advantages over conventional CCD registers including ease of fabrication, simplified clocking requirements, low dark current and high quantum efficiency.

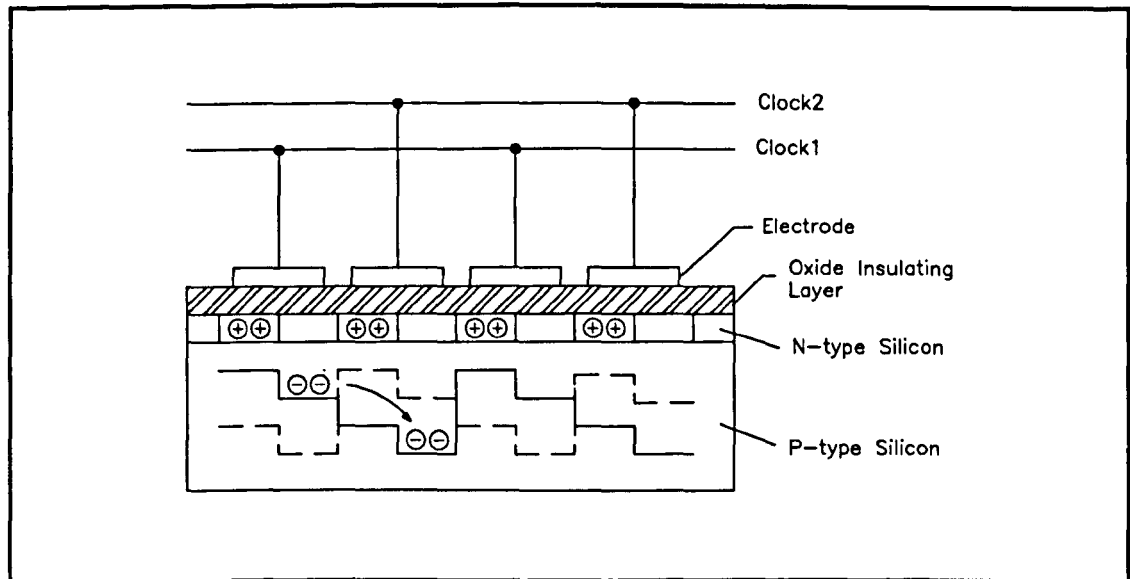


Figure 3.6 A Two Phase Buried Channel CCD Structure

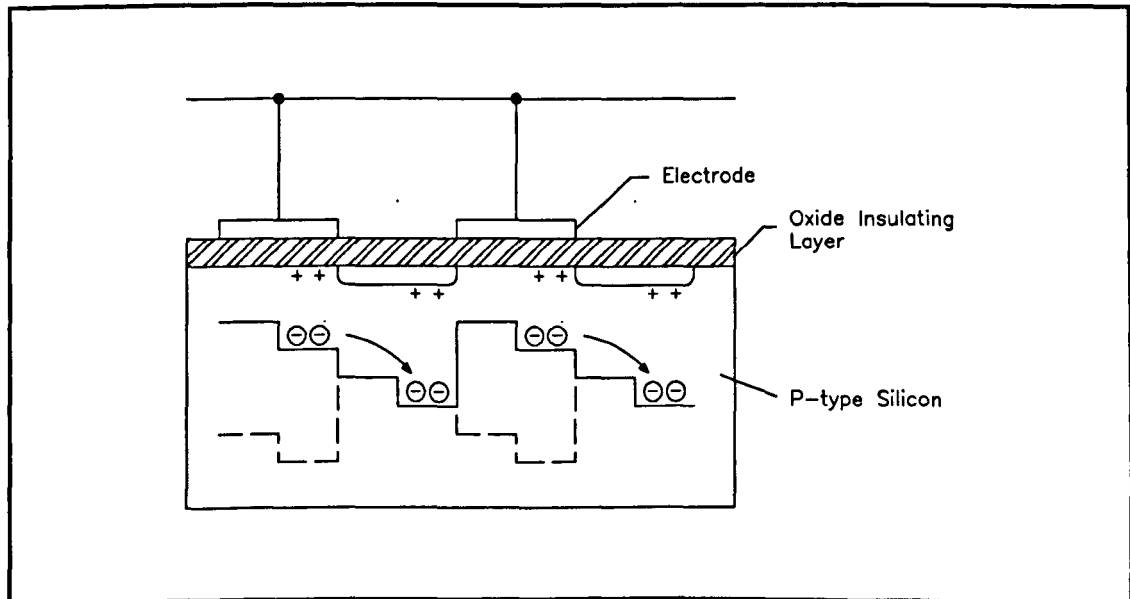


Figure 3.7 A Virtual Phase CCD Structure

3.3.3 A Practical CCD Image Sensor

A simple linear array CCD image sensor, such as the one shown in Figure 3.8, may be constructed from four components, a photosensor array, a transfer gate, a CCD shift register and a charge detection amplifier.

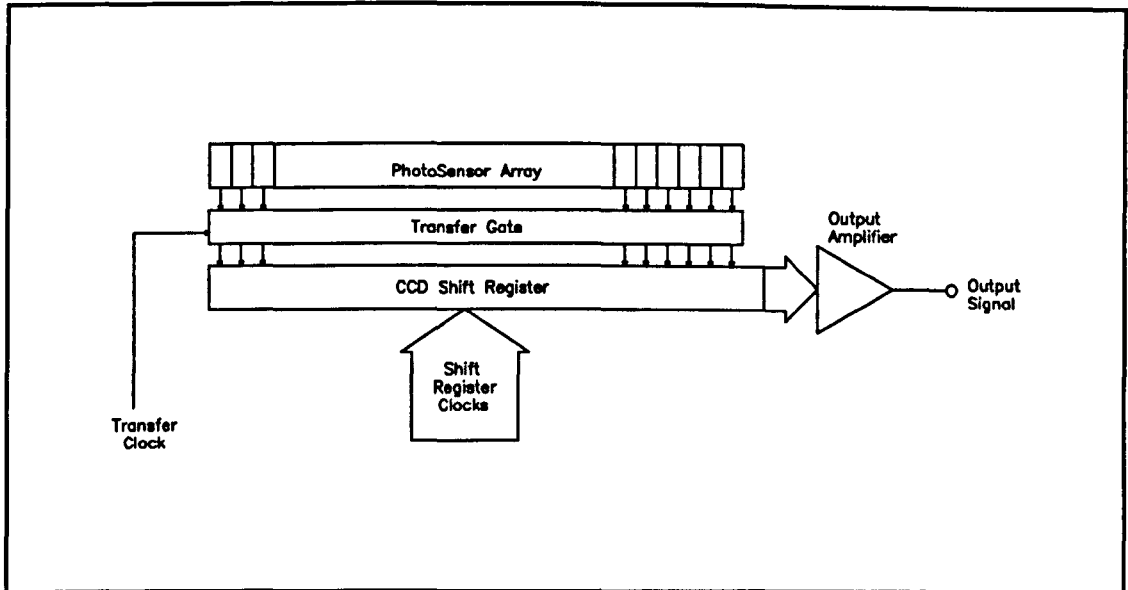


Figure 3.8 A Practical CCD Image Sensor Structure

Signal charge accumulates within the photosensors at a rate proportional to the incident light intensity during the period of light exposure, known as the integration time. As previously discussed a small amount of additional charge is generated due to thermal effects. At the end of the integration time the transfer gate is opened allowing accumulated charge to flow from the photosites into the CCD shift register elements. Signal charge is then clocked along the shift register as discrete charge packets which are delivered one at a time to the charge detection amplifier, which converts them into a voltage signal.

The Transfer Clock controls the operation of the transfer gate and hence determines the integration time during which signal charge is allowed to accumulate. The transfer clock period must be long enough to ensure that all signal charge packets are moved through the shift register to the output amplifier before the transfer gate is opened again to admit another set of charge packets into the shift register. The transfer clock period may be extended in order to increase signal charge build-up within photosites, although excessive amounts of charge will saturate the output amplifier. In addition it should be remembered that dark signal charge also increases with integration time, so the maximum integration time is determined by the tolerable level of dark signal.

3.4 Measures of CCD Performance

3.4.1 Dark Signal Output and Dark Signal Nonuniformity

An important feature of any CCD sensor is the dark level output signal voltage. As previously discussed in Section 3.2, dark level output is due to thermally generated charge within the sensor elements which accumulates during the integration period. Since, at constant temperature, the rate of thermally generated charge production is constant, dark signal voltage will be directly proportional to sensor integration time.

The build up of dark signal charge within a p-n junction photodiode is the result of a constant dark discharge current flowing during the integration time. It has been shown that the dependence of dark current, I_{DARK} , on ambient temperature, T , (in Kelvin) is given by the following equation [11].

$$I_{\text{DARK}} \propto \exp(kT) \quad (3.1)$$

(k = Boltzmann's constant = 1.381×10^{-23})

This exponential growth of dark signal with temperature illustrates the need to cool CCD sensors for optimum performance in low noise applications.

Dark signal voltage can be compensated for by the provision of an optically black sensor element within the pixel array. This sensor element is covered with an opaque layer so that only thermally generated charge contributes to the output signal. The dark signal voltage can then be subtracted from each illuminated pixel output to produce an "optically black clamped" signal.

Dark signal nonuniformity (DSNU) refers to the variation in dark signal voltage of individual pixels within a sensor array, as illustrated in Figure 3.9. DSNU is defined as the maximum variation in dark level output from the mean dark level voltage measured across all elements in the pixel array. In general DSNU increases with temperature and is a major contributor to overall output signal noise only in low light conditions.

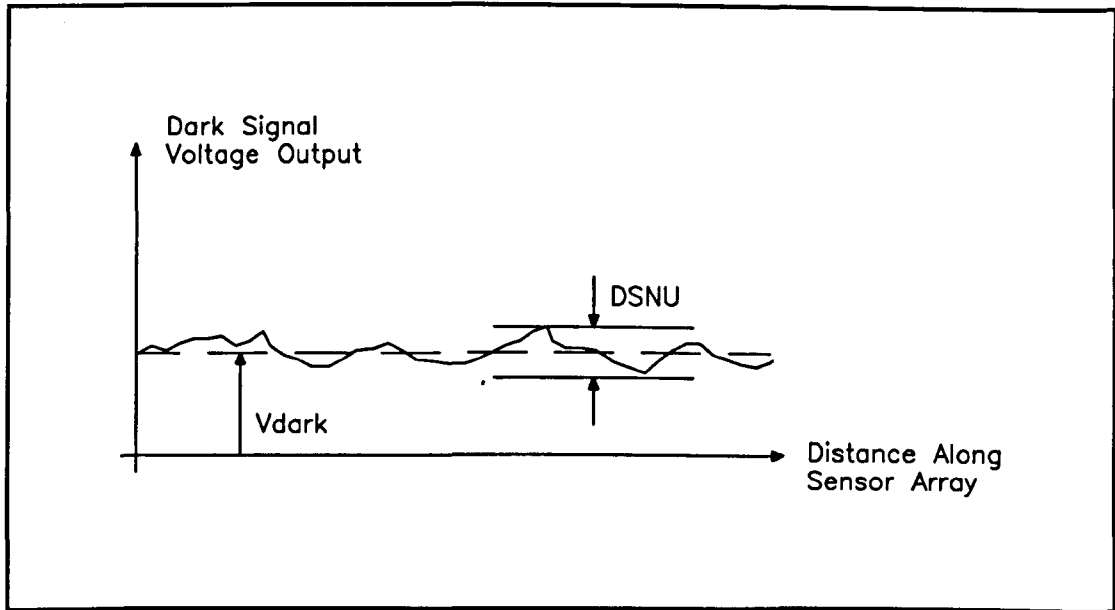


Figure 3.9 Dark Signal Nonuniformity

Dark signal nonuniformity is also known as fixed pattern noise because it produces a constant noise signal which overlays any output due to luminous input. This effect can be compensated for by the use of a calibration scan under dark conditions. However, this technique is rarely used since modern sensor structures limit DSNU to low levels, typically less than 1 % of saturation output voltage.

3.4.2 Linearity and Gain Nonuniformity

In an ideal CCD image sensor the output voltage signal will increase linearly with incident light exposure. An important performance criterion of CCD imagers is the degree of linearity exhibited as light exposure is varied. Good linearity is essential to ensure an accurate reproduction of the contrast variations present in an image. In addition, the dynamic range of the sensor is determined by the exposure range over which linearity is maintained.

Luminous exposure, E , is calculated as the integral of luminous power input, I , (usually specified in W/m^2) during the exposure time.

$$E = \int I \cdot dt \quad (3.2)$$

As previously discussed the output from a photodiode is a linear function of incident light exposure, so provided that all signal charge is transferred from the sensor element

into the CCD shift register, each sensor element will exhibit a high degree of linearity. However, each photodiode has an associated linearity constant which will not necessarily be the same for any two elements within an array.

Hence, each photosensitive element within a sensor array exhibits a slightly different characteristic signal transfer function. This concept is illustrated in Figure 3.10, where the output signals from three separate photosites are plotted as a function of luminous exposure. The voltage output of each element increases linearly from a dark signal, V_{dark} , at zero exposure to a maximum saturation output. The effect of this thermally generated fixed pattern noise is exaggerated in the diagram, and in most practical cases it is insignificant.

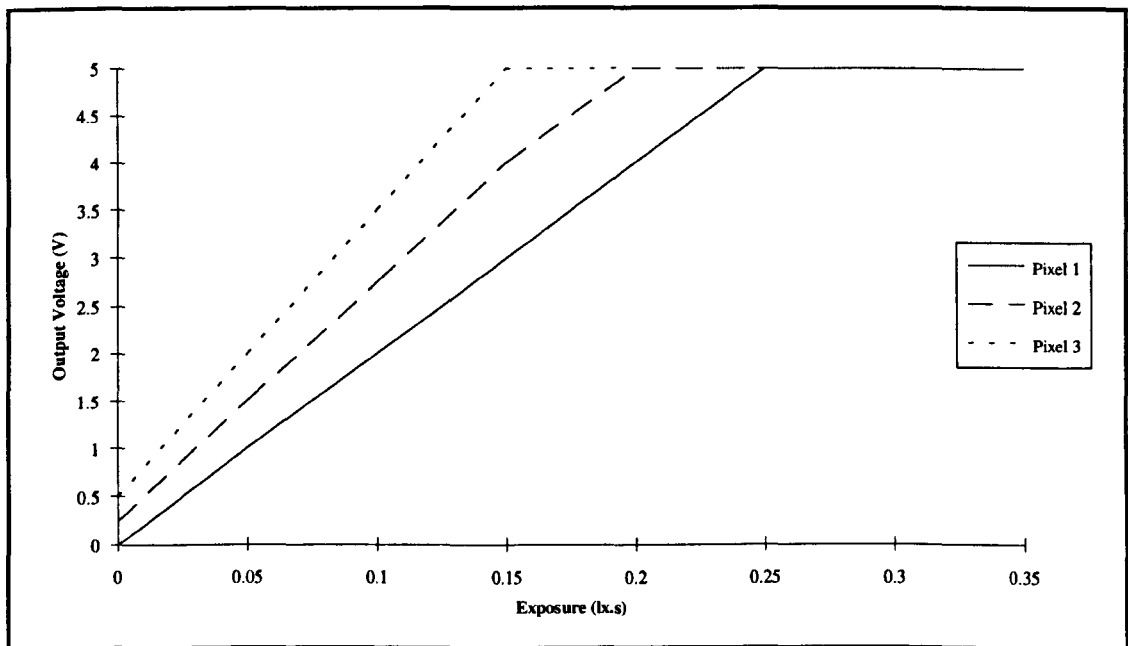


Figure 3.10 Linearity Characteristics of Three Separate Photosites

The gradient of the characteristic curve, effectively the signal gain of the image sensor, varies for each element in the sensor array. This variation is known as gain nonuniformity (GNU) and is a major contributor to signal noise.

Consider an element, n , within a sensor array with associated dark signal voltage, $V_{\text{dark}}(n)$ and signal gain, $K(n)$. The voltage output, $V_{\text{OUT}}(n)$, from this element would be :

$$V_{\text{OUT}}(n) = V_{\text{dark}}(n) + K(n).E \quad (3.3)$$

In practice the effects of dark signal nonuniformity are typically much less than 1% and may usually be ignored. Gain nonuniformity is usually measured under uniform illumination at 50% of saturation output and defined as the difference between maximum and minimum pixel outputs expressed as a percentage deviation from the mean output, V_{avg} .

$$GNU = \frac{V_{max} - V_{min}}{V_{avg}} \times 100\% \quad (3.4)$$

Commercially available solid state image sensors all exhibit some degree of gain nonuniformity, often as much as 10%. Tanaka [12] proposed a method of compensating for nonuniformity effects by first calibrating the imaging system under uniform illumination. This principle, along with details of its software implementation are discussed in Chapter 8 of this thesis.

3.4.3 Saturation and Blooming

As incident luminous exposure to a sensor element increases there is a corresponding rise in the amount of charge that must be stored within each CCD element. The dynamic range of any CCD image sensor is limited by the finite charge holding capacity of these shift register elements. The capacity of the potential wells in which charge is stored is dependent on CCD structure, and in any buried channel device is proportional to the doping concentration within the buried channel. The sensors used in this project had well capacities in the order of 10^3 electrons.

If a CCD sensor with full well electron capacity, N_{FULL} , and quantum efficiency, ξ , is illuminated by light of frequency, ν , the saturation exposure, E_{SAT} , necessary to cause signal charge to just fill the potential well is :

$$E_{SAT} = \frac{h\nu \cdot N_{FULL}}{\xi} \quad (3.5)$$

(h = Plank's constant = 6.626×10^{-34})

If the signal charge held with an element of a CCD shift register exceeds the capacity of the associated potential well, charge will overflow into adjacent storage elements.

This effect, known as blooming, results in a “smearing” of bright image data across several pixels.

Blooming may be prevented by ensuring signal charge never exceeds the capacity of the CCD storage elements. Most modern CCD image sensors achieve this requirement by ensuring the charge necessary to cause shift register overflow is greater than that which would cause saturation of the output amplifier. In this way amplifier saturation occurs before blooming in the shift register. This method of blooming elimination results in output characteristics shown in Figure 3.10, where output signal voltage increases linearly with incident exposure until the resultant signal charge level produces amplifier saturation. The saturation exposure level is defined as the luminous input where linearity is lost and the gradient of the characteristic curve begins to fall. Any further increase in light exposure will have no effect at the output, until signal charge begins to exceed well capacity, causing blooming into adjacent pixel elements.

Anti-blooming control may also be provided by novel sensor structures which include a charge overflow drain separated from the sensor element array by a variable potential barrier [13]. In this case any excess charge accumulated within a sensor element overflows the potential barrier and is swept into the drain, rather than being transferred to the CCD shift register. This feature prevents blooming but causes a “clipping” of high intensity image data

3.4.4 Spectral Sensitivity Characteristics

The sensitivity of a photodiode sensor is highly dependent on the wavelength of incident illumination. This spectral sensitivity dependence is illustrated by a plot of sensitivity against incident wavelength, shown in Figure 3.11 for the Sony ILX503 linear CCD array sensor. Spectral sensitivity, $K(\lambda)$, is defined as the output voltage of a sensor element per unit incident exposure to light of a given wavelength. It can be seen that the sensitivity of the ILX503 reaches a peak for wavelengths around 500nm, this corresponds to visible green light.

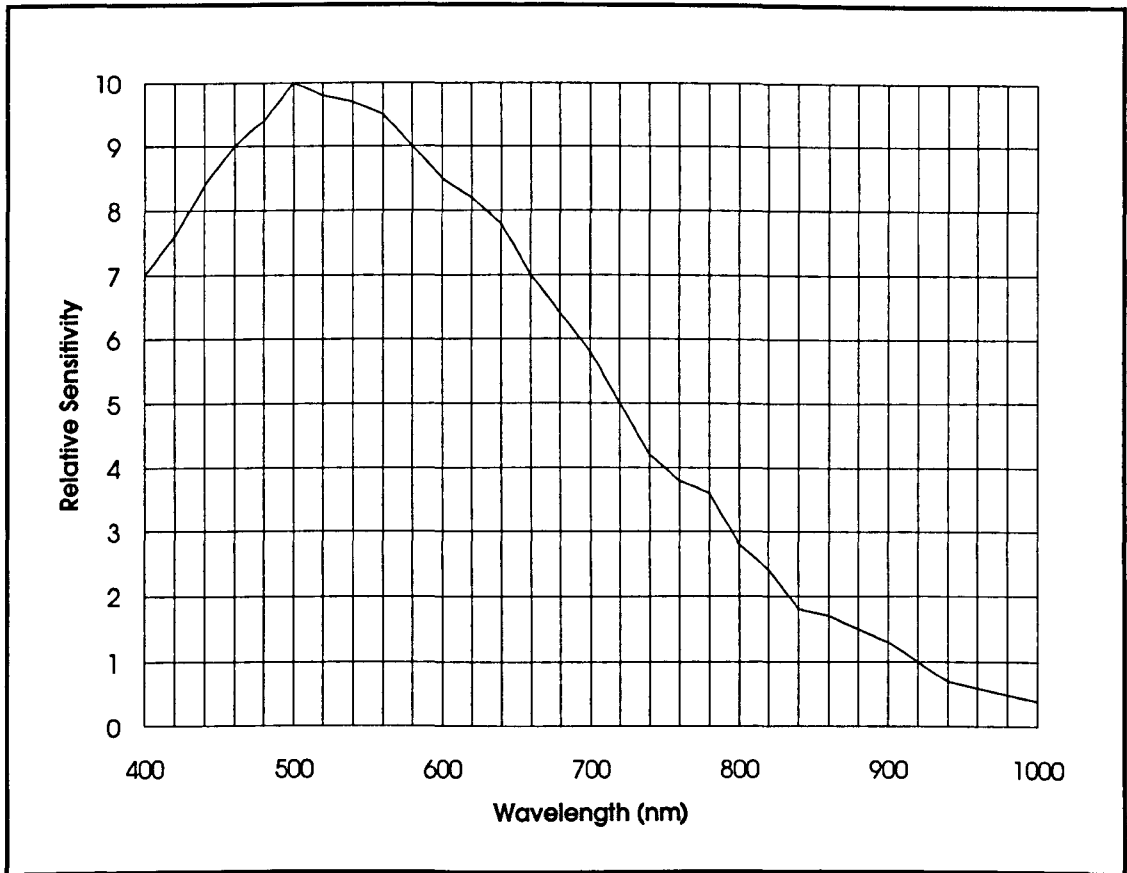


Figure 3.11 Spectral Sensitivity Characteristics of the Sony ILX503 Sensor

Neglecting thermally generated signal charge and nonuniformity effects, the voltage output, V_{OUT} , of a sensor when illuminated with monochromatic light of a wavelength λ , is simply the multiple of the incident luminous energy, E , during the exposure period and the spectral sensitivity, $K(\lambda)$, of the sensor at the specified wavelength.

$$V_{OUT} = K(\lambda).E \quad (3.6)$$

In the case of a light source consisting of a continuous spectrum, the output voltage is calculated as the integral of the individual voltage components at each wavelength over the specified range.

$$V_{OUT} = \int E(\lambda).K(\lambda).d\lambda \quad (3.7)$$

It is clear that the sensor output signal is highly dependent on the spectral distribution of incident illumination. Although output signal levels can be increased by extending the integration period, this will increase the thermally generated dark current noise and

decrease the speed of data acquisition. Hence, matching light source characteristics to device spectral sensitivity ensures optimum operation of the sensor.

3.4.5 Image Lag

At the end of each signal integration period the transfer gate is opened and charge accumulated in the sensor elements is moved into the CCD shift registers. It is a characteristic of CCD image sensors that not all of the signal charge is transferred into the shift registers. Excess charge remaining in the sensor element combines with signal charge associated with the subsequent scan to produce a phenomenon known as image lag. The effect of image lag is to produce images exhibiting a smearing of image data between the corresponding pixels from adjacent scan lines.

Image lag is highly dependent on sensor structure and recent attempts to eliminate it have focused on the provision of electric fields across the transfer gate to encourage complete charge transfer [14]. Image lag is also dependent on clocking conditions, since it is important to ensure sufficient time for complete charge transfer across the transfer gate. However, with the use of recently developed sensor structures combined with careful transfer gate clocking, image lag can be virtually eliminated.

3.4.6 Modulation Transfer Function

The ability of an imaging system to resolve contrast changes in an object tends to decrease with object size. This phenomenon can be quantified in terms of the modulation transfer function (MTF), which relates the ability of an imaging system to faithfully reproduce sinusoidally modulating intensity patterns of varying spatial frequencies. MTF analysis provides a generalised system transfer characteristic to assess the ability of a system to resolve both small changes in image contrast and feature size.

An upper limit on spatial frequencies reproduced by a solid state imaging system is set by the effective pixel size of photosensitive elements. This limit, known as the Nyquist frequency, is equal to half the spatial frequency of the sensor elements, since at least two sample points are needed to reproduce each intensity cycle. For a sensor with pixel element spacing, d , and corresponding spatial sampling frequency, f_{sensor} , the Nyquist frequency, f_n , is given by :

$$f_n = \frac{f_{\text{sensor}}}{2} = \frac{1}{2d} \quad (3.8)$$

Modulation Transfer Function analysis is a generalised technique which can be applied to any kind of imaging system. Much work has been carried out to assess the MTF characteristics of standard radiographic films [15]. Hence MTF analysis offers a basis for direct comparison of the imaging performance of radiographic imaging and solid state imaging techniques. The theoretical basis of MTF analysis and its application to solid state imaging performance is discussed at length later in this thesis.

3.4.7 Charge Transfer Inefficiency

As each signal charge packet is transferred from one storage element of a CCD register to the next, a small proportion of charge is left behind. This residual charge combines with the charge packet from the adjacent element causing a "smearing" of image data across pixel boundaries. This phenomenon is known as charge transfer inefficiency.

The tolerable level of charge transfer inefficiency is very low since charge is lost with each transfer along the CCD shift register. Consider an isolated signal charge packet being transferred between the first two elements of a CCD register with charge transfer efficiency, η_1 . If Q_1 is the total signal charge held in the first element, then the amount of charge transferred between elements, Q_2 , will be :

$$Q_2 = \eta_1 \cdot Q_1 \quad (3.9)$$

Total transfer efficiency, ϵ , is defined the multiple of individual transfer efficiencies at each element of a CCD shift register. For an n element shift register with individual element transfer efficiencies $\eta_1, \eta_2, \eta_3 \dots \eta_n$, the total transfer efficiency will be

$$\epsilon = \eta_1 \times \eta_2 \times \eta_3 \times \dots \times \eta_{n-1} \times \eta_n \quad (3.10)$$

The blurring effect caused by charge transfer inefficiency leads to a reduction in the theoretically attainable modulation transfer function. It has been shown that modulation transfer function, MTF, is related to total transfer efficiency and spatial frequency, f , by the following equation [16].

$$\text{MTF} = \left[-\varepsilon \left(1 - \cos \pi \frac{f}{f_n} \right) \right] \quad (3.11)$$

Since total transfer efficiency decreases with the number of elements in a shift register this sets an upper limit on the length of practical sensor arrays. Transfer efficiency is determined predominantly by shift register construction, in particular ensuring that a sufficient electric field exists between adjacent storage elements to effect complete charge transfer. Modern designs have achieved charge transfer efficiencies for individual CCD elements of over 99.999 % under ideal conditions.

High transfer efficiencies can only be achieved if sufficient time is provided for charge transfer between CCD storage elements. This requirement calls for careful control of shift register clock signals and sets an upper limit on attainable data rates of CCD devices.

3.5 Commercially Available Solid State Image Sensors

A survey of available solid linescan sensors was made in order to assess the attainable performance of current technology in this area. The range of products manufactured by EG & G Reticon [17], Sony [18], Texas Instruments [19], Thomson [20] and Fairchild Weston [21] were examined.

Time constraints only allowed for the evaluation of two of the available sensors, the Texas Instruments TC104 and the Sony ILX 503. The characteristics considered in the choice of sensors for evaluation were high spatial resolution, good signal to noise ratio, long scan length, high speed operation, high sensitivity, drive circuit complexity and price.

Any useful CCD imaging system requires a certain amount of additional hardware to generate the drive clock voltage waveforms required by the sensor and process the resultant output video signal into a useful form. Most manufacturers offer some kind of drive hardware to accompany their basic CCD sensors. This support hardware ranges from simple evaluation boards to video processing circuits and integrated digital linescan cameras [22]. In general these cameras are prohibitively expensive and do not offer the flexibility of operation required for this project.

It was decided, for the purposes of this project, to construct all the components necessary for an operational CCD imaging system, rather than using one of the off the shelf systems described above. This allowed more precise control over CCD sensor operating parameters and experimentation with different low noise techniques for improving image quality. A detailed discussion of work carried out to build a CCD imaging system is given in Chapter 5.

3.6 Summary

Previous chapters noted the need for very high resolution imaging techniques in order to preserve all the information visible in a radiographic film image. Hence, the selection of an image sensor for use during this project was of crucial importance. CCD sensors were chosen for this project due to their high pixel density, good stability with both time and temperature, linearity and low noise characteristics.

This chapter has given an overview of recent advances in solid state imaging technology. From this discussion it should be clear that the linear array CCD sensor was an ideal choice for high resolution imaging applications. The main operating characteristics of CCD image sensors were then summarised in order to indicate the potential sources of image degradation. These findings will be used later in this thesis in order to evaluate the relative merits of the two different sensors used during this research.

References

- 1 Automatic Radiograph Inspection System User Manual, BIX Systems Ltd.
- 2 G. Nagy, "Optical Scanning Digitisers", *Computer* **16** 5, 1983, pp 13-24.
- 3 H. Tseng, J. Ambrose & M. Fattahi, "Evolution of the Solid State Image Sensor", *Journal of Image Science* **29** 1, 1985, pp 253-259.
- 4 Y. Talmi & R. Simpson, "The Self Scanned Photodiode Array: A Multichannel Spectrometric Device", *Applied Optics* **19** 9, 1980, pp 1401-1414.
- 5 E. Arnold et al, "Charge Injection Imaging", *IEEE Trans. Electron Devices*, ED-18, 1971, pp 1003-1008.
- 6 G. F. Amelio, "Charge Coupled Devices", *Scientific American* **230** 2, Feb. 1974, pp 23-31.

- 7 R. H. Kingston, "Detection of Optical and Infrared Radiation", Springer-Verlag Ltd. 1978, pp 64-76.
- 8 W. S. Boyle & G. E. Smith, Bell System Tech. J. 49, 1970, p 587-593.
- 9 T. Yamada, H. Okano & N. Suzuki, "The Evaluation of Buried Channel Layer in BCCD's", IEEE Trans. Electron Devices, ED-25, May 1978, pp 544-546.
- 10 J. Hyncek, "Virtual Phase: A New Approach to Fabrication of Large Area CCD's", IEEE Trans Electron Devices, ED-28, 1981, pp483-489.
- 11 D. F. Barbe & S. B. Campana, "Imaging Arrays Using the Charge Coupled Concept", in Advances in Image Pickup and Display, Vol. 3, Editor B. Kazan, Academic Press, 1977, p 204.
- 12 S. Tanaka, "A Need for Nonuniformity Correction in Solid State Sensors", SPIE Proceedings, Vol. 350, 1982, pp 275-282.
- 13 T. Murphy & D. Debs, "Anti-Blooming and Integration Control", Fairchild Weston Application Note, 1987.
- 14 Sony Linear Sensor Application Notes, Sony Semiconductor, 1992.
- 15 R. H. Morgan et al. "The Frequency Response Characteristics of X-ray Films and Screens", Am. J. Roentgenol. 92, August 1964, pp 426-440.
- 16 A. F. Milton, "Charge Transfer Devices for Infrared Imaging", in Optical and Infrared Detectors, Editor R. J. Keyes, Springer-Verlag Ltd, 1977, p 205.
- 17 E. G. & G. Reticon Image Sensing Products Data Book, E. G. & G. Reticon Ltd, 1994.
- 18 Sony CCD Cameras & Peripherals Data Book, Sony Semiconductor, 1992.
- 19 Texas Instruments Optoelectronics and Image Sensors Data Book, Texas Instruments, 1992.
- 20 Thomson CCD Products Data Book, Thomson Composants Militaires et Spatiaux, 1993.
- 21 Fairchild Weston CCD Data Book, Fairchild Weston CCD Imaging Division, 1987.
- 22 Cronin Line Scan Cameras Product Outline, Cronin Electronic Ltd, 1992.

4. The Transputer Processor

4.1 Introduction

As discussed in detail in Chapter 5 of this thesis, the radiograph digitisation system performed several distinct functions including, signal acquisition, image display, laser light scanning, stepper motor control and image processing. During image acquisition, several of these functions were required to be performed concurrently. These separate subsystems may be viewed as a series of independently operating modules which require some degree of intercommunication.

The transputer network model of concurrently operating, intercommunicating processes is ideal for constructing a modular system such as the one outlined above [1]. Each distinct function may be performed as a separate process running in parallel with all the others. In addition, parallelism offers drastic speed increases during computationally intensive operations such as image processing [2], because work can be distributed between available processors. Many applications of transputers to such distributed processing systems can be found in literature [3].

Since the transputer processor acted as the basic building block from which this modular system was constructed, it will be necessary to briefly discuss the architecture of the transputer. Particular attention will be paid to the interfacing of transputer devices to construct parallel networks.

4.2 Parallel Processing

The original, and still pre-dominant computer architecture involves a single processing element executing a set of sequential instructions. Modern microprocessor technology has made vast improvements in the operation of these sequential processors. However, the demands being made on processor speed by a modern computer applications, such as real time image processing, often exceeds the capabilities of such sequential systems [4].

An obvious method of speeding up such computationally intensive operations is to divide the workload between several processing elements operating in parallel. Any such parallel system requires some mechanism for communication of data between

processing elements. The speed of these communications is often an important factor affecting the attainable speed of a parallel system.

The design of program code to be implemented in parallel represents a fundamental change in approach for the software engineer. The application must be broken down to identify operations that may be carried out concurrently without the need for shared data. These tasks must be divided between processing elements in such a way as to optimise the spread of processing workload. Poorly implemented parallel systems may result in several processing elements being idle while they await data from an overworked node. In this way system performance may be limited by the slowest executing task within a processor network.

4.3 The Transputer

4.3.1 Transputer Architecture

The hardware realisation of a parallel system, such as described above, requires the use of a number of processor elements with some innate mechanism for passing messages between them. An example of this type of processing element is the transputer processor, developed by Inmos [5].

The transputer is a single chip microcomputer containing a quantity of on-chip memory plus a configurable external memory interface. Transputers have been specifically designed to enable simple interconnection, via communication links, to create parallel processor networks. Each transputer incorporates four standard Inmos links which allow direct connection to any other transputer device.

All transputers used in this project were 32 bit devices, offering up to 4Gbytes directly addressable external memory. The T800 and T805 devices both contain 4 Kbytes of high-speed on-chip RAM and include an internal 64 bit floating point unit. The T414 has 2 Kbytes on-chip RAM and no floating point unit. The basic architecture of these devices is illustrated in Figure 4.1.

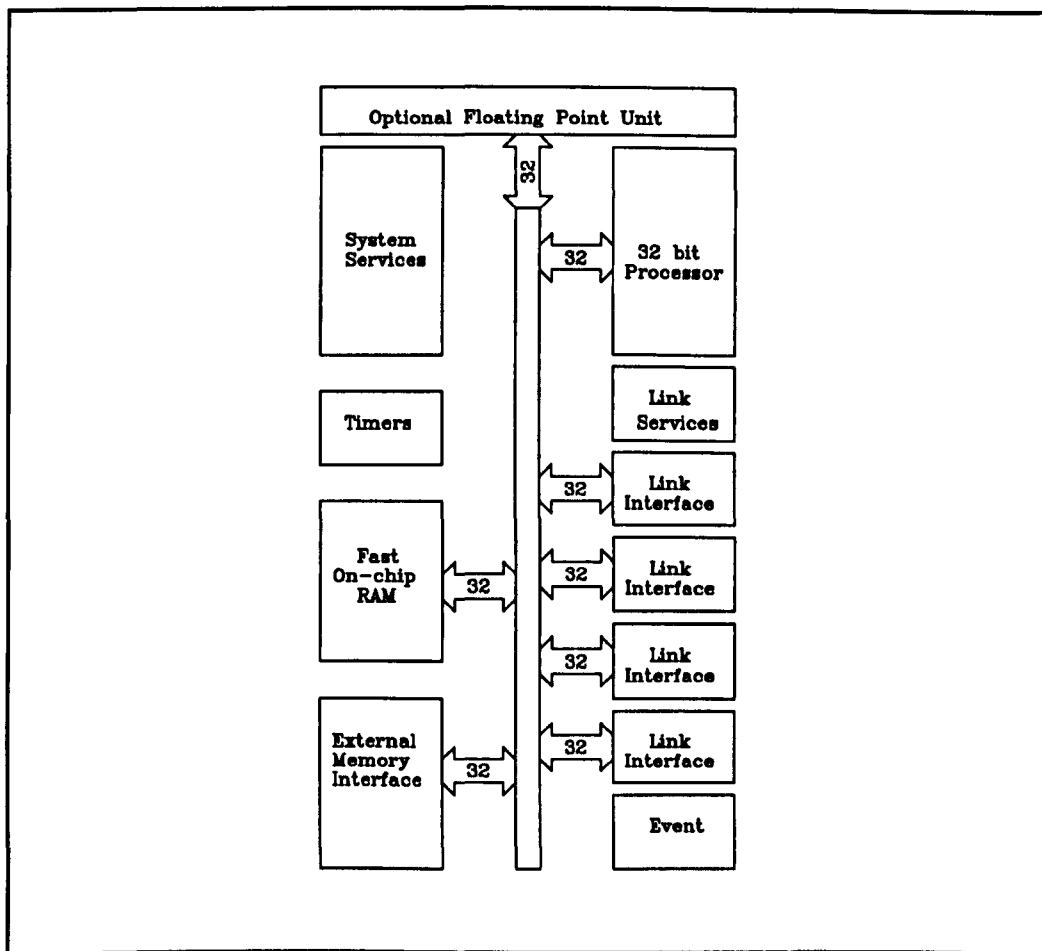


Figure 4.1 Transputer Architecture

At the heart of the transputer is a 32 bit processor with direct access to the fast on-chip RAM. This internal memory can be accessed in a single processor cycle, with a further 4 Gbytes of memory being accessible via the external memory interface.

The processor takes advantage of the fast internal memory by operating a reduced instruction set utilising just six registers. Three of these registers, designated A, B and C, comprise an evaluation stack used for all calculations. The remaining registers are an instruction pointer, a workspace pointer pointing to local variables in memory, and an operand register used in the formation of instruction operands.

The instruction set codes frequently used instructions in a single byte, and provides extended codes for less common instructions. This strategy offers considerable speed increases for commonly used instructions without limiting instruction set size.

The processor allows a degree pseudo parallelism by time-slicing the operation of concurrently executing processes. This sharing of processor time is controlled by a microcoded scheduler. At any given time a process may be executing, waiting in a queue to be executed or inactive waiting for a communication from another process. Inactive processes consume no processor time.

4.3.2 Link Communications

The main feature that distinguishes the transputer from more conventional processors is its ability to communicate with other transputer devices. Communication between transputer devices is achieved by the use of Inmos link interfaces which allow direct connection of devices with no external components. This allows a network of concurrently operating transputers to be constructed simply by connecting links together.

Each transputer provides four Inmos links, consisting of two uni-directional channels, operating at data rates of up to 20 Mbits/s. Communications are unbuffered so that a communication cannot take place until both the sending and receiving processes are ready. This requirement provides the only form of process synchronisation within a network.

Link communications use a simple protocol where each byte is separately acknowledged before the next is sent. Each byte within a message is encrypted as a data packet consisting of two high bits, followed by eight data bits and a low stop bit. The receiving process responds to an incoming data packet by returning an acknowledge packet consisting of a high bit followed by a low bit. The communication protocol increases message throughput by allowing the receiving process to acknowledge receipt of a data packet before the complete message has been received. The standard data and acknowledge packets are illustrated in Figure 4.2.

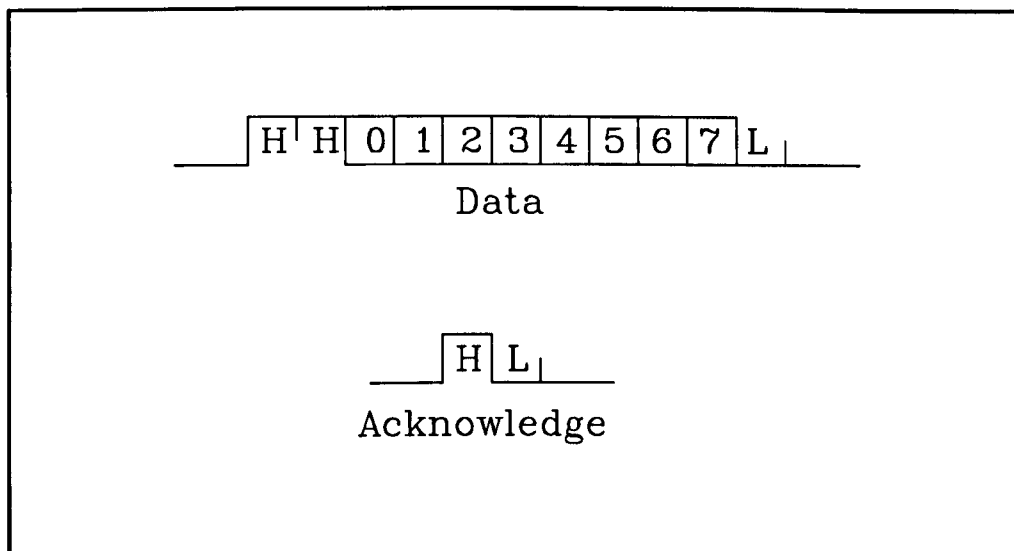


Figure 4.2 Transputer Link Communication Protocol

4.3.3 Link services & interconnection

In order to initialise and operate a network of transputers, three system service lines are provided for system interconnection, these are **reset**, **analyse** and **error**.

The reset line is used to start a transputers initialisation sequence, which consists of internal and external memory configuration routines followed by bootstrapping. Bootstrap data can be set to come either from ROM or from an external link.

The error output line is the logical OR of the internal error flag and the **errorin** input. The transputer can be programmed to halt when the error flag is set. If error and errorin pins of a transputer network are daisy-chained together, any error flagged in the network will be passed up to the master processor.

When the analyse input of a transputer becomes active the device will halt, whilst retaining certain debug information. If the transputer is then reset, the registers will contain information on the processor status immediately before it was halted. In this way a limited debugging function is achieved.

Each transputer has three system service ports, **up**, **down** and **subsystem**. The interconnection of these ports determines how system service signals are propagated through a network. The up port receives reset and analyse signals from its immediate ancestor in the network hierarchy, and propagates them through the down and subsystem ports to its descendant processors. In addition a processor can assert

subsystem reset and analyse signals which are passed to all processors connected to its subsystem port. Error signals received at the down port are propagated up towards the host. Subsystem errors can be read by software but are not automatically passed on up the network hierarchy. This arrangement allows the processor at the top of the system hierarchy to receive all error messages and reset or analyse all processors below it in the network. Similarly, any processor within the network may control all the processors connected to its subsystem port.

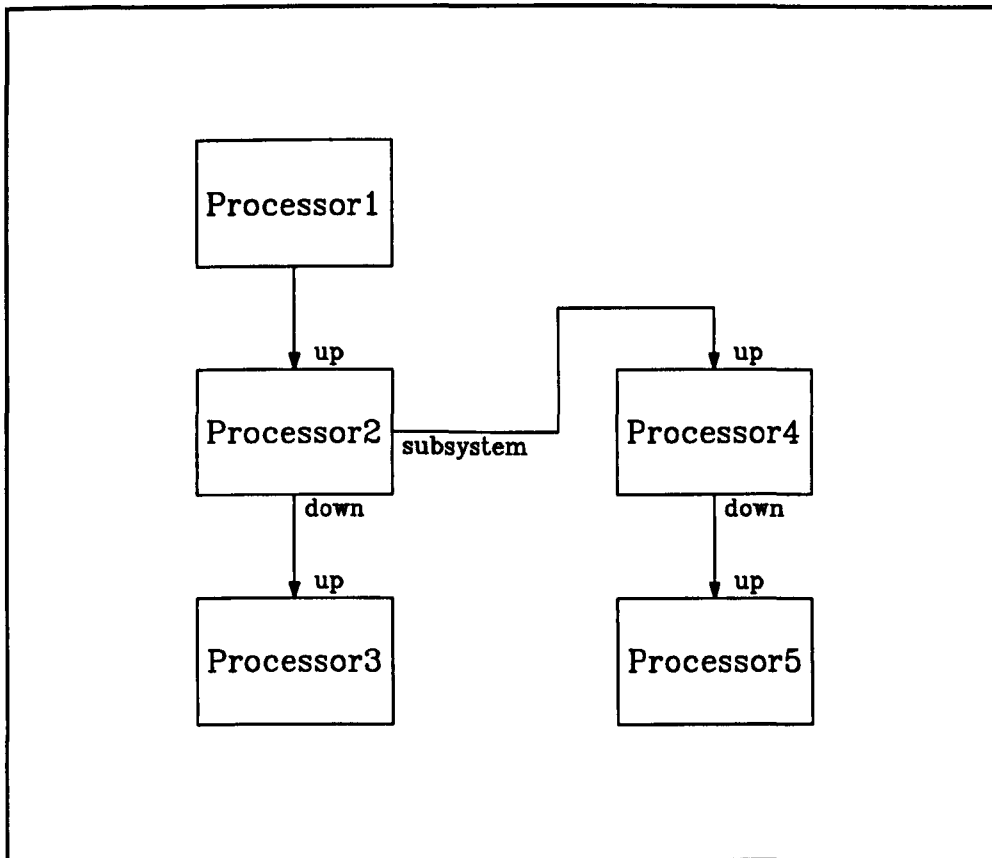


Figure 4.3 Example Showing the Propagation of Transputer Services

This connection scheme is illustrated in Figure 4.3, where Processor 1 is at the top of the network hierarchy and is used to reset and boot all transputers in the system. Processors 4 and 5 comprise a sub-network which may be reset or analysed by Processor 2.

4.4 Interfacing Transputers to External Hardware

During this project it was required to control several pieces of external hardware directly from software. The most simple way of interfacing between a transputer

network and external hardware is to use a standard Inmos link for all communications. In this way there is no conceptual difference in communications between two transputer processors in a network or between a transputer and some external hardware. In terms of system design, the external hardware may be viewed as another node in the transputer network.

4.4.1 The Inmos IMS C011

External hardware may be directly interfaced with any transputer device using the Inmos C011 link adaptor [6]. This device converts between a standard Inmos serial link and either eight parallel input and output lines or a microprocessor bus. The LinkOut and LinkIn lines of the C011 may be connected directly to the input and output ports of an available transputer link. The C011 supports the standard Inmos link communication scheme at data rates of up to 20 Mbits/s

The C011, when operated as an 8 bit parallel interface, uses a simple two wire handshaking protocol to initialise and acknowledge the passing of data packets. During data input to the C011 the IValid line is taken high by external hardware to signify that valid data is present on the eight input data lines. This data is encoded into a standard data packet by the C011 and passed down the LinkOut line. The receiving transputer responds with an acknowledge packet on LinkIn, and once all data has been passed the IAck line is taken high to acknowledge completion of the communication.

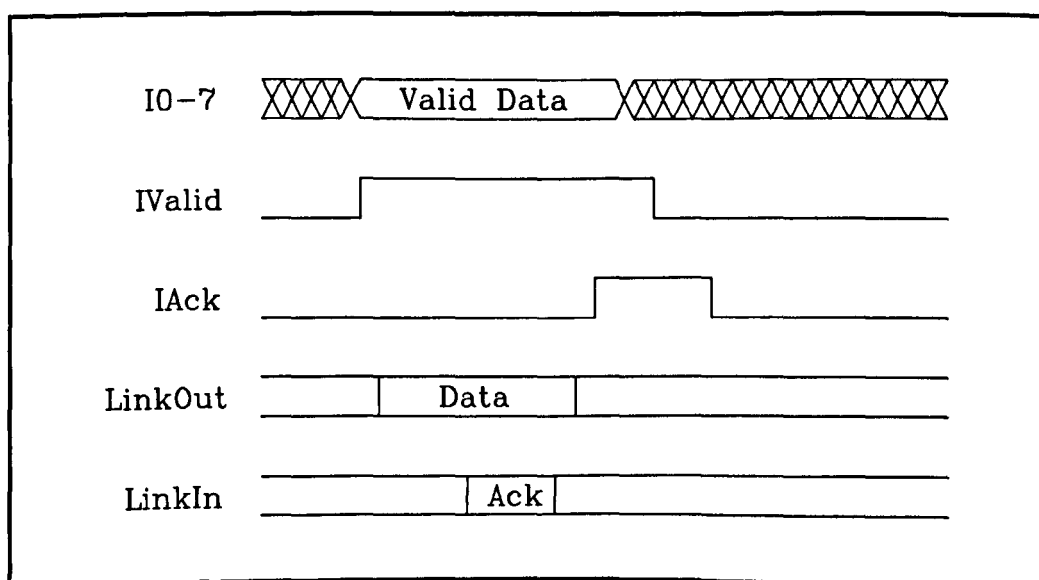


Figure 4.4 Data Input to a CO11 Link Adaptor

A similar mechanism is used by the CO11 for data output from a transputer to external hardware. An incoming data packet on LinkIn is converted to an 8 bit value by the CO11 which then takes QValid high to signify the presence of valid data on the output lines. The external interface hardware may then read data from the Q0-7 lines and respond by taking QAck high to acknowledge receipt of the message. Finally, the communication is completed by sending an acknowledge packet down LinkOut to the transputer.

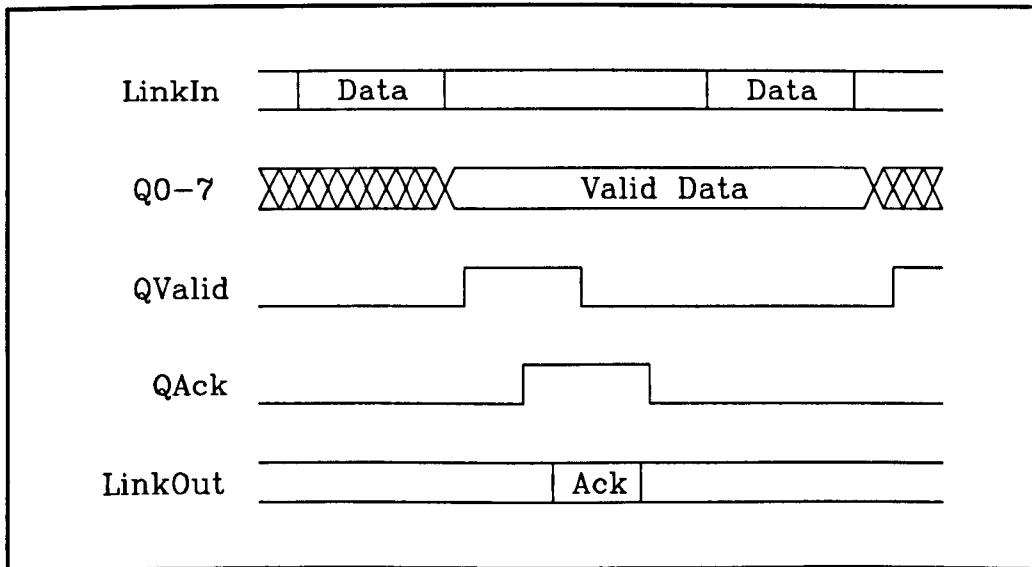


Figure 4.5 Data Output from a CO11 Link Adaptor

4.5 Summary

The basic processing unit used for the control of all hardware and software functions of the film digitisation system was the transputer. The transputer is a single chip microprocessor system designed specifically for the implementation of parallel systems. This chapter contains a brief discussion of the architecture and operation of the transputer processor. Particular attention is paid to the interfacing of transputer devices and external hardware to produce an integrated data acquisition and processing network.

Two reasons were identified for the use of transputers during this research. Firstly, it was required that several separate hardware systems be under software control during image digitisation. To ensure correct operation of the image acquisition system, these different hardware modules were required to operate concurrently. This was most

simply achieved by using a network of parallel processors, such as transputers. Secondly, it was noted that the processing power needed to carry out image processing operations on the high resolution images could most easily be gained by dividing the workload between a number of fast processing elements. The speed and inherent parallelism of the transputer processor made it ideal for use in such a parallel image processing architecture.

References

- 1 A. Carling, "Parallel Processing - The Transputer and Occam", Sigma Press, 1988.
- 2 K. Howson, "A Transputer based Vision System", PhD Thesis, Liverpool University, September 1994.
- 3 Proceedings of the First Conference on Applications of Transputers, Eds. L.Freeman & C. Phillips, IOS Press, 1989.
- 4 A. Rosenfield, "Computer Vision", Advances in Computers, Vol. 27, 1988.
- 5 The Transputer Data Book , Second Edition, Inmos Ltd, 1989.
- 6 IMS CO11 Link Adaptor Engineering Data, Inmos Ltd, 1989.

5. A Radiographic Film Digitisation System

5.1 Introduction

The main element of this project involved the construction of a transputer based image acquisition and processing system. This system was designed to digitise and display high resolution images of industrial radiographic film plates. The film digitisation system was envisaged as the first component in an automatic defect detection system for the interpretation of industrial radiographs.

This chapter contains a detailed discussion of the individual components of the radiographic film digitisation system. Particular emphasis is placed on the operation of the linear CCD¹ array sensors used for image acquisition, since the performance of these sensors and their associated hardware was the key factor in achieving the required image quality.

5.2 System Outline

A block diagram showing the main functional elements of the radiograph digitisation system is given in Figure 5.1. The image acquisition system used a linear CCD array sensor, illuminated by a scanning laser spot mechanism. The radiographic film being imaged was moved across the face of the sensor using a two axis motorised film holder. This direct contact between film and sensor eliminated the need for complex optics.

Two types of commercially available linear CCD array image sensor were evaluated during this project, the Texas Instruments TC104 and the Sony ILX503. Separate drive hardware was developed to run each sensor as the front end of the image acquisition system. A standardised structure was used for the design of each sensor system so that they could be interchanged as plug in modules with the minimum of system modifications.

¹ Charge Coupled Device

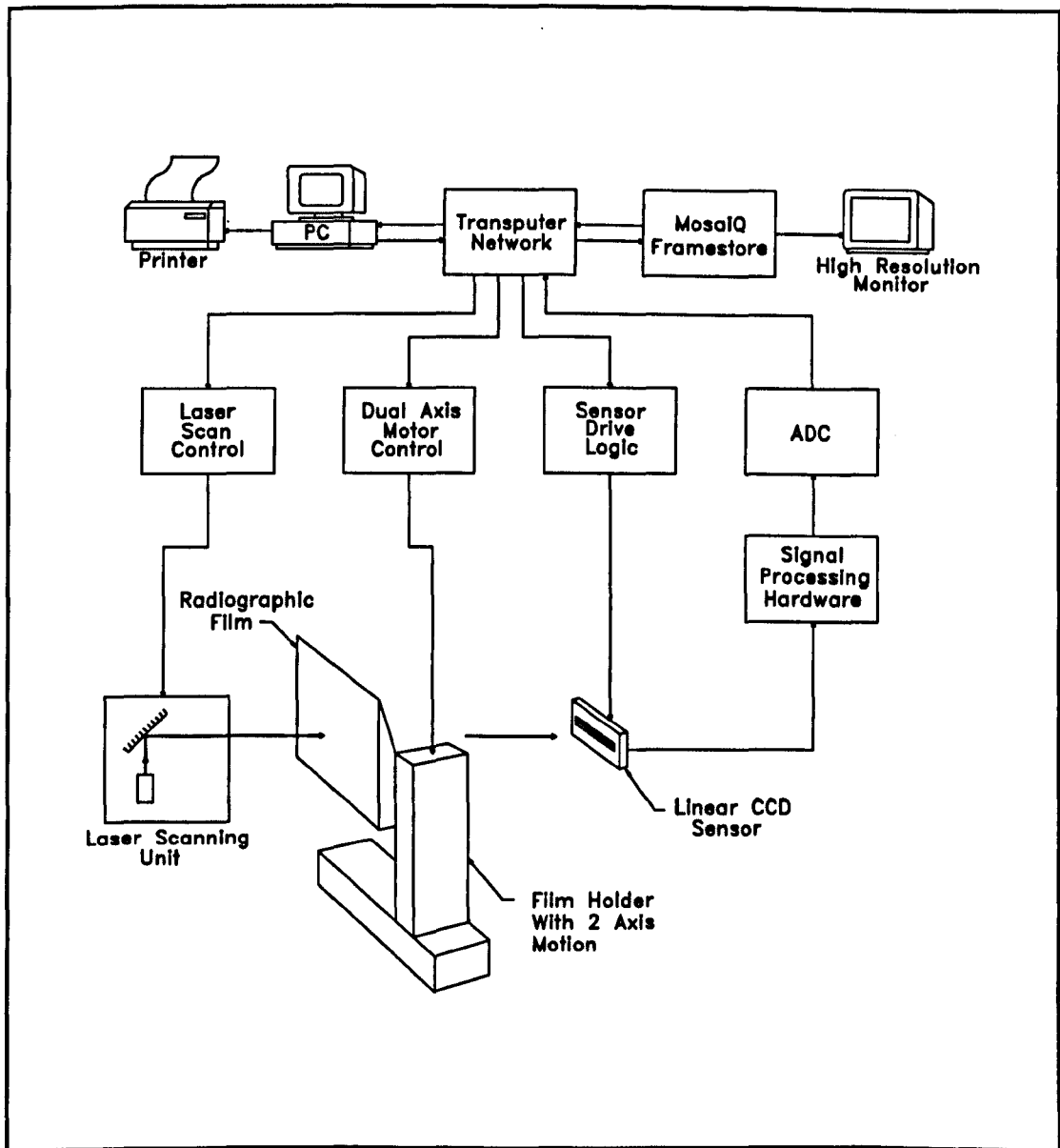


Figure 5.1 Overview of the Radiograph Digitisation System

The raw output from the CCD sensors required a certain amount of signal processing to recover image data from the associated carrier signal. The resulting analogue video signal was then converted to a sequence of discrete digital pixel values using a 12 bit analogue to digital converter.

The image acquisition hardware was interfaced to an array of transputer processors using a standard Inmos transputer communications link operating at 20 Mbits/s. The resulting images were displayed on a high resolution 19 inch monitor via a Quintek MosaIQ transputer frame store.

A Viglen 486DX33 PC¹ was used as a terminal for all operations requiring user interaction with the transputer system. In addition, the file I/O capabilities of the host PC were used for storing and archiving the images produced by the image acquisition system. Hard copy outputs were available using a Hewlett Packard Laser Jet printer, connected to the PC.

5.3 Transputer System

The individual hardware subsystems within the radiograph digitisation system were each controlled by dedicated software processes, executing on transputer processors. This arrangement allowed concurrent operation of each subsystem during image acquisition. Concurrent operation of the laser scanner and image acquisition subsystems was essential, in order to ensure continuous illumination of the sensor during image capture. In addition, concurrency increased system speed and simplified software development.

Each transputer in the network provided four Inmos serial links which allowed direct interconnection with other transputer products. The network was constructed in order to optimise the distribution of hardware and software tasks between the available transputer links. A simplified topology of the transputer network used by the radiograph digitisation system is given in Figure 5.2.

User interaction with the transputer network was carried out via a menu driven server program running on the host PC. The host PC was memory mapped to one of the channel links a T800 transputer. This transputer was defined as the root processor in the network hierarchy. It had executive control over the other processors and controlled message passing around the network.

The laser scanning unit was constructed around a dedicated T414 transputer connected to the root processor. This allowed uninterrupted, independent operation of the laser light source during image acquisition. Similarly, the MosaiQ frame store included a dedicated T805 processor used for image display and manipulation.

¹ Personal Computer

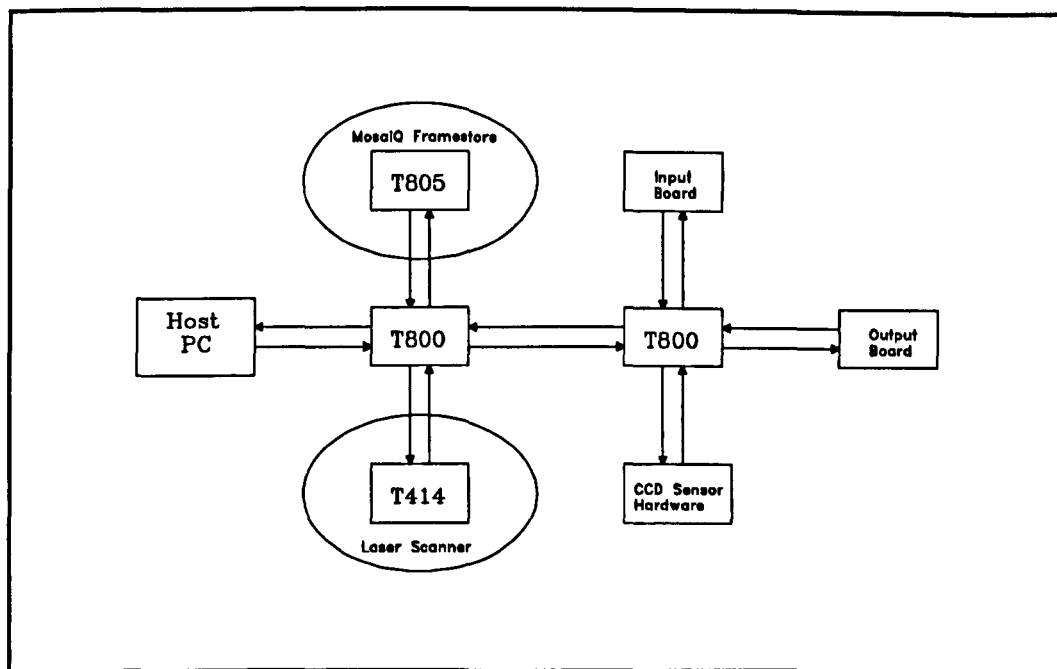


Figure 5.2 Transputer Network Topology

A second T800 transputer was used for low level control of the image acquisition hardware. One link of this transputer was used for the passing of image data from the sensor hardware. The other two links were used for other system I/O via input and output interface boards. These boards passed control signals to and from the sensor drive hardware and the motor control subsystem.

5.4 Texas Instruments TC104 CCD Sensor

The first CCD image sensor under evaluation during this project was the Texas Instruments TC104 [1]. The TC104 uses a virtual phase buried channel structure to produce a sensor with low noise characteristics and simple clocking requirements. The structure of the Texas Instruments TC104 linear array CCD image sensor is shown in Figure 5.3. The layout incorporates two CCD shift registers located either side of an array of 3456 photodiode elements. Photosites are approximately $10.7 \mu\text{m}$ square located with a centre to centre distance of $10.7 \mu\text{m}$.

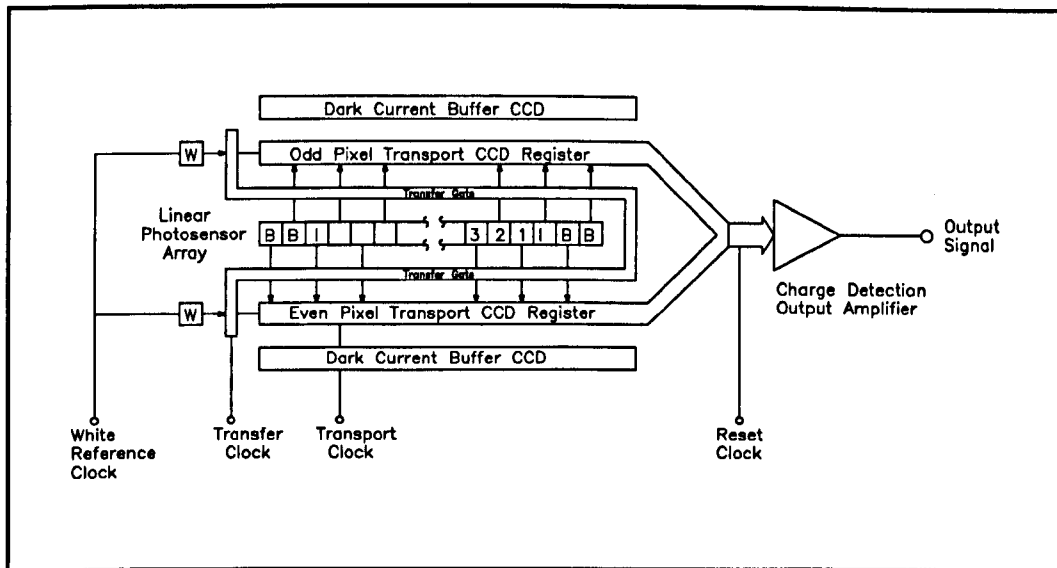


Figure 5.3 Functional Structure of the TC104 Sensor

Typical operating parameters for the Texas Instruments TC104 linear CCD array sensor are given in the following table.

Parameter	Typical Value	Units
Number of Elements	3456	
Element Spacing	10.7	μm
Sensitivity	2	$\text{V}/(\mu\text{J}/\text{cm}^2)$
Gain Nonuniformity	5 %	
Saturation Output	800	mV
Dark Signal Voltage	5	mV
Dark Signal Nonuniformity	10	mV
Total Transfer Efficiency	96.6 %	
Maximum Data Rate	8	MHz

Table 5.1 Typical Performance Characteristics of the TC104 Sensor

5.4.1 Clocking Requirements

The required input waveforms to drive the TC104, together with the resultant output signal are illustrated in Figure 5.4. In order to explain the function of these different clock signals, a brief summary of the sensor operation will be given.

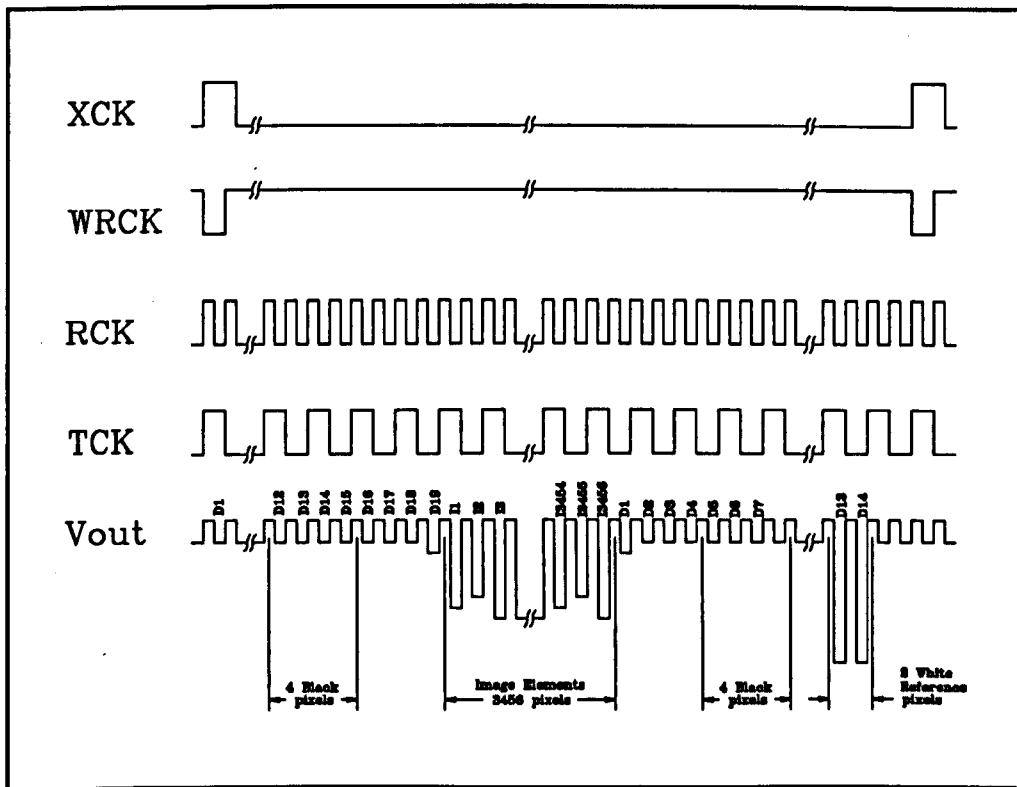


Figure 5.4 Operating Waveforms for the TC104 Sensor

Signal charge accumulates in the sensor elements, at a rate proportional to the incident luminous flux, until the transfer gate is opened, allowing the discrete charge packets to flow into the CCD shift registers. Transfer gate operation is controlled by the Transfer Clock, XCK. The transfer clock period determines the exposure time, during which signal charge is allowed to accumulate.

The transport shift registers move charge packets from the photosites to the output amplifier. In order to achieve the high density of photosites present in the TC104 two transport shift registers are used, one on each side of the photosite array. Alternate charge packets from the array are moved into each shift register, so that odd pixels are transported through one register and even pixels through the other. The transport registers are clocked by the Transport Clock, Tck, which operates at half the pixel data rate. Two outer buffer CCD registers are provided to protect signal charge within the transport registers from contamination by externally generated noise. The virtual phase electrode structure used in the TC104 [2] results in simplified operation since only a single transport clock is required to operate the CCD shift registers.

At each end of the photosite array are a number of elements covered by an opaque layer to provide a black reference level output. A number of isolation elements are situated between the imaging elements and the black pixels, these act as a buffer to prevent signal contamination. At the end of each transport shift register, opposite to the output amplifier, are charge injection diodes which provides a white reference signal. These diodes inject a quantity of charge into each shift register sufficient to cause a saturation output at the output amplifier. These pulses can be used in automatic exposure control circuits. White signal charge injection is controlled by the White Reference Clock, Wrck.

The structure of the shift registers ensures that charge packets are re-combined into their original order before reaching the charge detection amplifier. Charge packets are delivered serially to a pre-charged diode which converts input charge into a voltage signal which is in turn applied to an output amplifier. After the conversion of each charge packet the charge detection diode capacitance is recharged by the Reset Clock, Rck.

The drive circuit shown in Figure A-1 of Appendix A was developed to operate the Texas Instruments TC104 sensor. Three PLD¹ devices were used to derive all the necessary clock signals discussed in the previous section. These TTL² level clock signals were then converted to the high current, high voltage signals required to operate the CCD shift registers, before being applied to the sensor.

5.4.2 Drive Hardware

5.4.2.1 Clock Logic and Exposure Timing

Three ALTERA EP600 PLD [3] devices were used to generate the necessary clock waveforms and provide control outputs for use in the subsequent video signal processing circuits. This clock logic allowed the external variation of sensor exposure

¹ Programmable Logic Device

² Transistor-Transistor Logic

time using a 7 bit control input. An operating pixel data rate of 625 kHz was chosen, which was well within the operating range of the TC104.

The exposure time period was generated by a 7 bit counter, clocked once every 512 ticks of the reset clock, Rck. Hence, for a given reset clock period, T_{RCK} , the maximum exposure time possible, T_{MAX} , is given by:

$$T_{MAX} = 2^7 \times 512 \times T_{RCK} \quad (5.1)$$

The minimum exposure time is determined by the time taken to clock a full data line through the CCD shift register. An entire data line of the TC104 consists of 3489 pixels, and since the minimum exposure time adjustment is 512 pixels, the minimum exposure time is:

$$\text{Minimum Exposure Time} = 7 \times 512 \times T_{RCK} = 3584 \times T_{RCK} \quad (5.2)$$

At the normal operating frequency of 625 kHz, the corresponding pixel clock period is 1.6 μ s. Hence, the drive circuit allowed exposure time adjustment between 5.7 ms and 104 ms.

5.4.2.2 Clock Drivers

The clock signal levels needed to drive the Texas Instruments TC104 sensor are summarised in Table 5.2, below. It can be seen that in addition to requiring high voltage clock signals the CCD transport registers present a high impedance load. The resulting high power clock signals were found to be a major noise source within the drive circuit.

Clock Signal	V _{HI}	V _{LO}	Resistive Load	Capacitive Load
Rck	+5V	-14V	700 k Ω	16 pF
Tck	+5V	-14V	700 k Ω	900 pF
Xck	+5V	-14V	150 k Ω	400 pF
Wrck	+7V	+14V		

Table 5.2 TC104 Input Clock Specifications

National Semiconductor DS0026 MOS¹ clock drivers [4] were used to derive the TC104 clock inputs from TTL voltage levels. The DS0026 is a monolithic high speed MOS interface circuit designed to drive capacitive loads of up to 1000 pF with a maximum switching time of 20 ns. It accepts level shifted TTL inputs and drives the output within 0.5 V of the supply rail voltages.

The DS0026 application notes [5] recommend the use of input capacitive coupling to achieve the required voltage level shift. However, this method was found to degrade clock switching performance to an unacceptable level during high speed operation. Instead, a simple transistor level shifting configuration was used, as shown in Figure 5.5. Rise and fall times of the input clock signal were controlled using a 100 pF capacitor in parallel with the level shifting transistor.

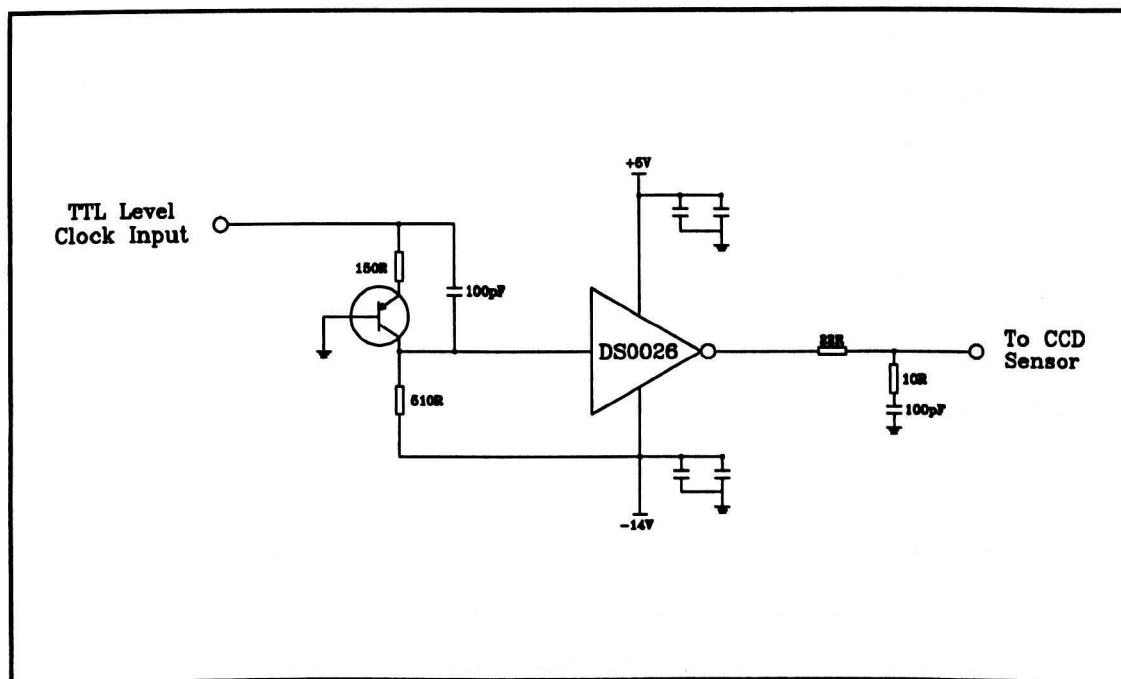


Figure 5.5 Clock Diver Circuit

5.4.2.3 Clock Feedthrough

As noted above, the clock drivers produced a relatively high power output signal. On the rising and falling edges of these clock signals the instantaneous power requirements

¹ Metal Oxide Semiconductor

of the clock drivers were high. This caused significant transient currents to flow in power lines, giving rise to periodic noise pulses. This phenomenon of clock "feed through" was the major contributor to noise pickup in the image capture system.

The clock drivers were not the only components to induce feed through pulses onto the CCD output signal. The clock signals used to control the CCD transfer registers are derived from a series of higher frequency clock signals, these clocks also generated a series of noise pulses on the power supply lines which were visible as lower amplitude pulses on the output signal.

Clock feed through is due to the increased power consumption of components during switching. An effective method of feed through reduction is to increase rise and fall times of clock signals, so that instantaneous power consumption is reduced [6]. This can most simply be achieved by the inclusion of an RC shunt network on high speed clock lines which shunts high frequency current components through a bypass capacitor to ground. This shunt network can be seen on the output line of the DS0026 clock driver illustrated in Figure 5.5.

Experimentation found that inclusion of RC shunt networks on the Rck and Tck lines, close to the clock driver output terminals, was the most effective method of feed through reduction. The lower frequency clock lines did not contribute significantly to system noise.

The digital oscilloscope trace in Figure 5.6 shows the raw output waveform from the TC104 sensor without feedthrough reduction. Although the trace exhibits a certain amount of noise associated with the digital oscilloscope, it is still possible to see the effects of clock feedthrough causing periodic spikes corresponding to the rising and falling edges of the various drive clock signals. These transients cause a "ringing" effect which appears on the output waveform as an overall increase in signal noise.

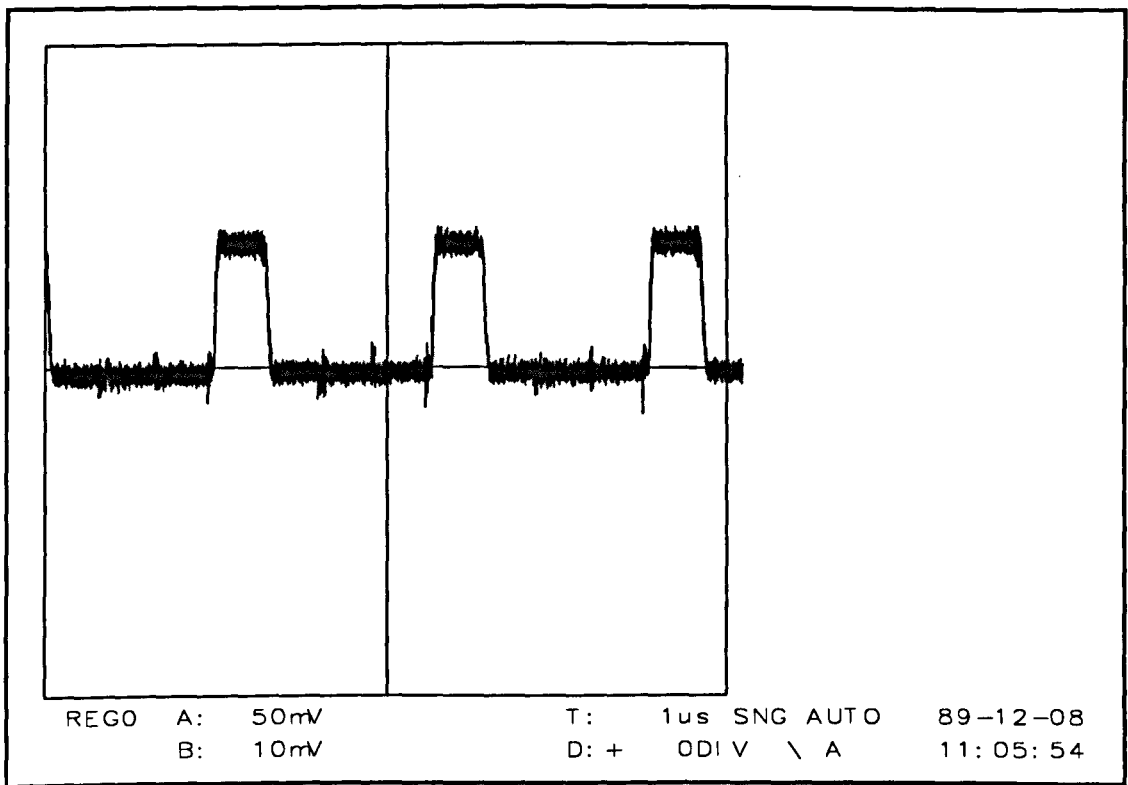


Figure 5.6 TC104 Output Waveform Exhibiting Clock Feedthrough

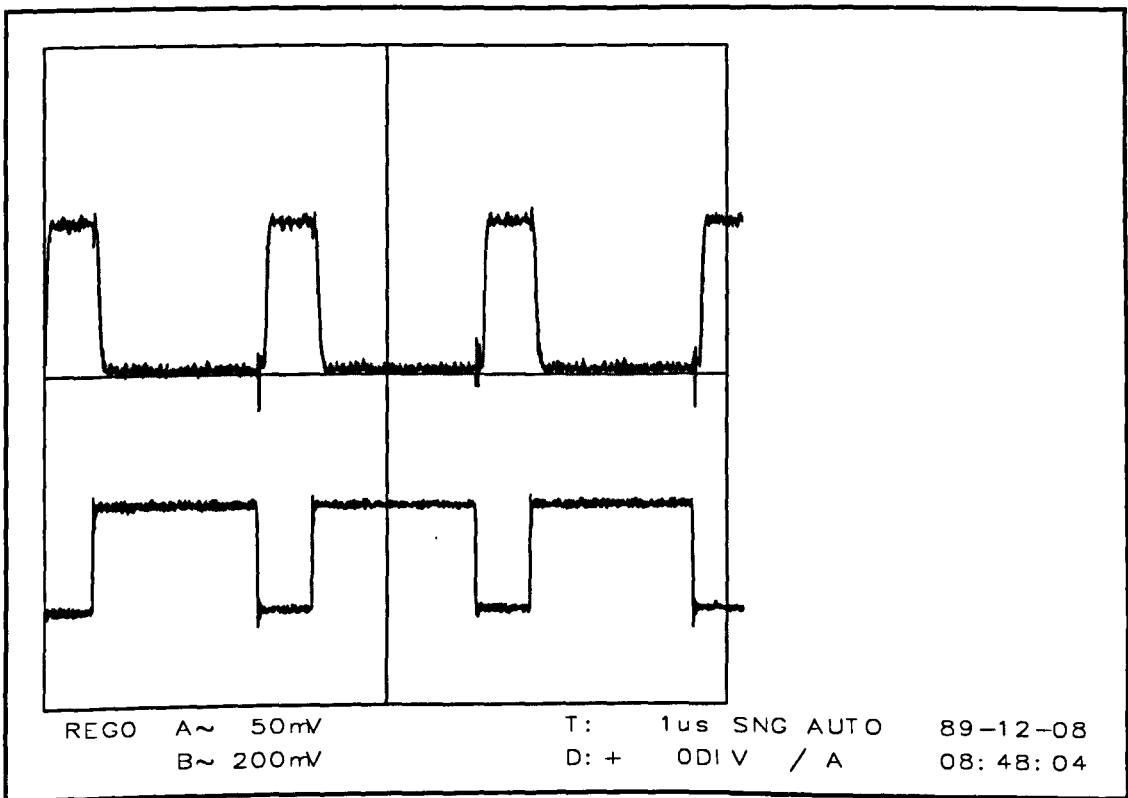


Figure 5.7 TC104 Output Waveform With Feedthrough Reduction

OBJECTIVES

By the end of this course, you should;

Be familiar with using a mouse to manipulate windows and text

Be able to Create and Save Documents

Know how to Edit and Print Documents

Be able to Manipulate Text

Be able to Enhance Text

Know how to Change the Document Format

Be able to use the Spell Checker and Thesaurus

Know how to correct Document Layouts

Be able to use and set-up Tables

INTRODUCTION

MS-Works is a full-featured Integrated computer package including word processing, spreadsheet and database programs designed for creating, revising, editing and printing a wide variety of documents and storing and manipulating data.

Using a word processor is similar to using a typewriter - only much easier. Once a document has been created, it need never be re-typed. Text can be edited and formatted very simply, so that corrections such as fixing a spelling error, to moving an entire section of text can be carried out with minimal effort.

This course is designed to be an introduction to MS-Works. Students need no previous knowledge of the package, but require knowledge of keyboarding skills.

THE KEYBOARD

The keyboard is split into three main sections the Standard QWERTY keypad, the Cursor Control Pad and the Numeric Pad - in addition to these three areas there are usually 10 or 12 Function Keys (F1 - F12)



The term "QWERTY" comes from the first six letters of the keyboard.

The Mouse

A mouse is a piece of hardware that is connected to your computer and can be used as an alternative to the standard QWERTY keyboard.

To use the mouse, roll it on a flat surface, if the mouse reaches the edge of the surface, simply pick it up and position it back in the middle. The mouse has no affect when moved through thin air.

IT CAN BE USED FOR

Selecting text

Moving the insertion point

Selecting and choosing features and options

Selecting, moving and sizing graphics boxes



KEYBOARD AND MOUSE CONT.....

Mouse Terms

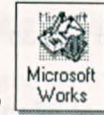
- Point** To move the tip of the mouse pointer on top of an element on the screen
- Click** To quickly press and release the mouse button
- Double Click** To quickly press and release the mouse button twice in succession.
- Drag** Hold down the main mouse button, move the mouse across an area of text, then release the mouse button blocking (or Highlighting) an area of work..
- Release** To Stop holding down a mouse button
- Select** To highlight text
- Choose** To use a mouse to pick an item that begins an action in windows

Mouse Pointer Shapes on Screen

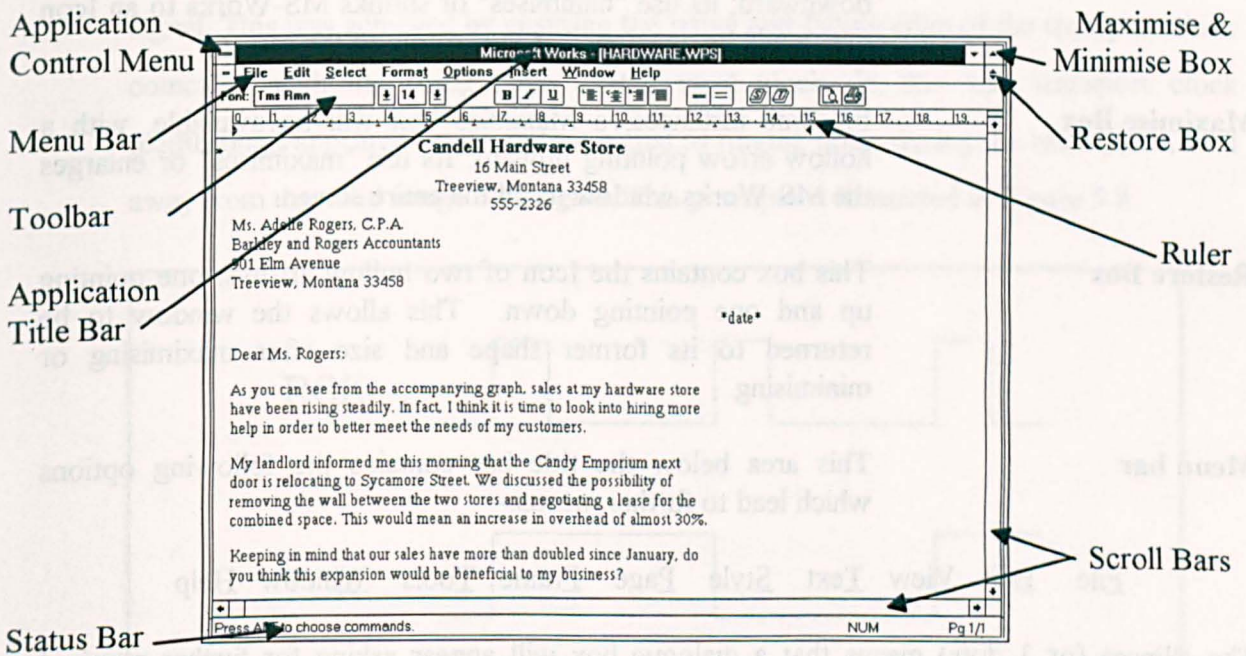
As the mouse is moved, its relative position on screen changes and can change shape depending on it's position. Take a few moments to observe how the shape of the mouse Icon changes as the mouse is moved.

- In the Menu Bar White arrow pointing to the upper left 
- In the Typing Area I-Bar or I-Beam 

STARTING MS-WORKS



Double click the left hand mouse button on the MS-Works icon, this will launch MS-Works, and the screen will be displayed ready for you to create a document.



MS-WORKS WINDOW

Working across the screen from the upper left

Application Control Menu This menu has no text label, it has as Icon symbol of a large bar across it and is sometimes referred to as the space bar menu.

Application Title Bar This bar across the top of the screen identifies the name of the program.

Minimise Box This box contains the Icon of a hollow arrow pointing downward; its use "minimises" or shrinks MS-Works to an Icon which will be positioned at the bottom of the screen.

Maximise Box In some instances, a Maximise box will be available, with a hollow arrow pointing upward. Its use "maximises" or enlarges the MS-Works window to fill the entire screen.

Restore Box This box contains the Icon of two hollow arrows; one pointing up and one pointing down. This allows the window to be returned to its former shape and size after maximising or minimising.

Menu bar This area below the title bar contains the following options which lead to further menus:-

File Edit View Text Style Page Frame Tools Window Help

The ellipses (or 3 dots) means that a dialogue box will appear asking for further required information for MS-Works to complete the command.

Scroll Bars The scroll bars are the grey areas along the right and lower edges of the window. They are used in conjunction with the mouse to move through the document. Each scroll bar has a small scroll box and arrows at each end. At the top edge of the vertical scroll bar is a small, compact black rectangle, which is used for splitting the screen horizontally.

The effects of feedthrough reduction are illustrated by the oscilloscope trace in Figure 5.7. This trace was taken with RC shunt networks included on the Rck and Tck input lines, as described above. The upper waveform in Figure 5.7 shows the output signal from the TC104 sensor and the lower trace shows the Tck signal produced by the DS0026 driver. It is clear from this trace that the periodic noise spikes exhibited in the earlier example are much reduced, as is the overall noise level of the data signal. A certain amount of clock feedthrough is still present during Tck transitions, but this does not obscure valid pixel data.

Since most clock feedthrough was created by the transport clock, the careful timing of Tck transitions helped reduce feed through onto the valid data portion of the output signal. This was achieved by ensuring the rising and falling edge of the transport clock coincided with the rising edge of the reset clock. In this way transport clock feedthrough appeared on the output signal as ringing noise during the reset pulse, well away from the valid image data signal. This principle is illustrated in Figure 5.8.

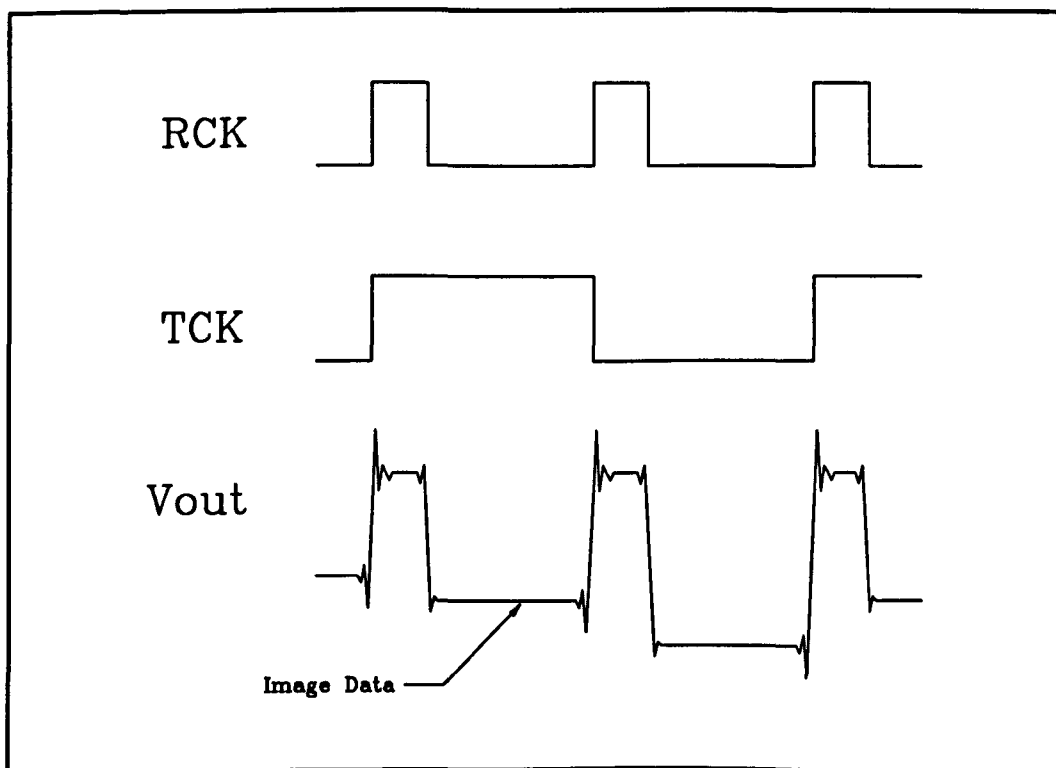


Figure 5.8 TC104 Clock Feedthrough Timing

5.4.2.4 Reduction Of Thermally Generated Noise

CCD image sensors accumulate a quantity of thermally generated charge in sensor elements proportional to the integration time during which charge is allowed to build up. This thermally generated charge appears at the sensor output as a random noise signal [7]. Thermally generated noise may be reduced by cooling of the sensor array.

The total power dissipation of the TC104 is quoted as 112 mW and, hence, cooling the device should be relatively simple. The DS0026 clock drivers used to generate the input signals required by the sensor do, however, consume around 1000 mW with an associated degree of heat dissipation. These drivers must be located close to the sensor array in order to minimise noise generated on the clock lines. Attempts were made to cool the sensor array by the use of heat sinks placed around both the clock driver IC's and the sensor array. Experimentation found little improvement in output noise levels.

The main reason for the lack of success of these heat sinks was that even at room temperature a significant number of thermally generated electrons will be produced in the sensor elements. Ideally, for low noise applications, CCD devices should be cooled to lower temperatures, typically below 0°C [8].

5.4.3 Output Signal Processing

The raw output signal from the TC104 is produced by the action of the charge detection output amplifier. Signal charge packets accumulated within the CCD sensor are delivered serially to the charge detection amplifier which is then recharged, after every pixel, by the reset clock, Rck. The resulting output signal consists of a pulsed voltage waveform, the upper portion of which represents the uniform amplifier reset voltage, onto which is superimposed a series of negative going data pulses.

Under zero illumination conditions, the data pulses on the output waveform represent thermally generated dark signal charge generated in the photosites. Data pulse amplitude increases linearly from this dark level as the incident illumination level is increased. Hence, true image data can only be recovered with reference to some dark level reference signal.

The nominal peak to peak output amplitude, representing saturation luminous input to the TC104, is specified as 0.8 V. The entire output waveform is displaced from zero

volts by the operating voltage of the charge detection amplifier. This DC offset varies greatly due to operating temperature and overall illumination level, but is specified to be greater than 5 V and less than the amplifier supply voltage, V_{DD} (14 V). In practice, during unsaturated operation of the sensor, the DC offset was seen to vary between 7 V and 8.5 V.

Before digitisation of the signal was possible, it was necessary to recover the valid video data from the associated carrier signal. A block diagram showing the signal processing route used to obtain the required analogue positive video signal is given in Figure 5.9. The full circuit used to implement this processing method is shown in Figure A-2 of Appendix A. The individual steps involved during signal processing are discussed in detail below.

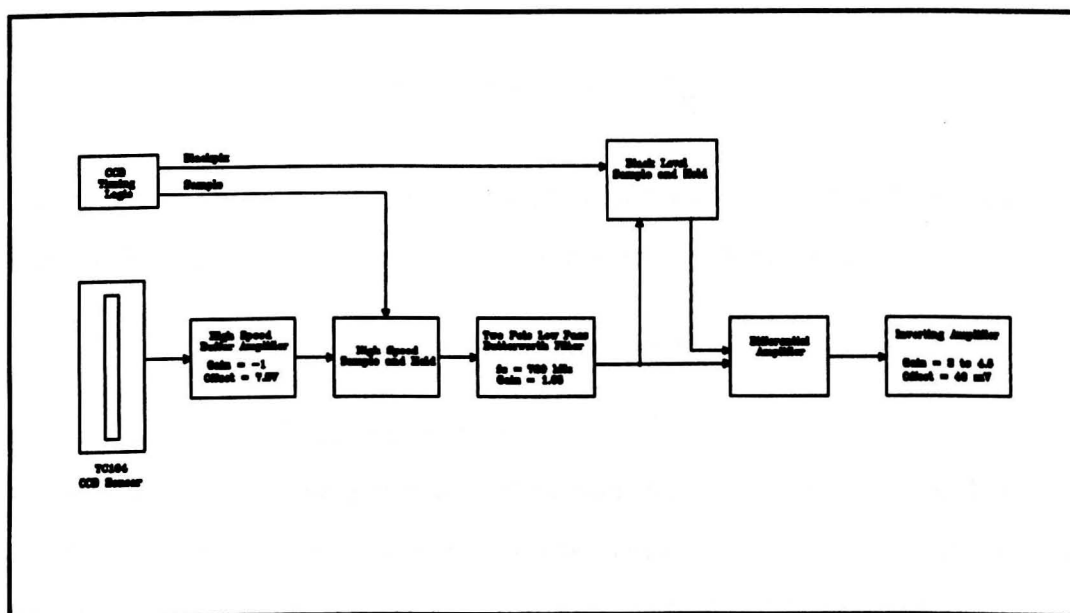


Figure 5.9 TC104 Output Signal Processing Scheme

5.4.3.1 Buffer Amplifier

An Analog Devices AD847 operational amplifier [9], configured as a unity gain inverting amplifier was used as a buffer between the TC104 and subsequent signal processing hardware. The AD847 is a high speed, precision op amp which is internally compensated for unity gain operation. The output slew rate of $300 \text{ V}/\mu\text{s}$ makes it ideal for video applications.

The high speed sample and hold amplifier used as the next stage of the signal processing circuit required an input voltage signal between 0 and 5 V. Texas Instruments recommend that the required level shift to satisfy this condition should be achieved by capacitively coupling the input signal line [10]. However, in view of the high digitisation resolution required by the image acquisition system, this method was found to degrade signal data to a unacceptable level.

Instead, the DC voltage level shift was achieved by applying an offset current to the non-inverting input of the buffer amplifier. The inverting amplifier configuration used allowed large DC offsets, without affecting output voltage gain.

For a given input signal voltage, V_{IN} , with supply voltage, V_S and voltage gain, K , the output voltage of the buffer amplifier is given by the equation:

$$V_{out} = -KV_{IN} \pm V_S(1 + K) \frac{R_3}{R_3 + R_4} \quad (5.3)$$

The circuit design used a unity gain configuration with $V_S = \pm 14V$ and potential divider resistors $R_3 = R_4 = 15 \text{ k}\Omega$. Hence the resultant output voltage was:

$$V_{out} = -V_{IN} \pm 14V$$

5.4.3.2 Data Pulse Sample and Hold

In order to remove amplifier reset pulses from the data signal a Texas Instruments TL1591 sample and hold amplifier [11] was used to sample the valid data portion of each pixel output. The TL1591 is a monolithic sample and hold circuit containing all the necessary components to sample analogue data at rates of up to 15 MHz. The device incorporates a fast buffer amplifier, a digitally controlled diode bridge switch and high impedance output amplifier.

Data sampling is initiated by a TTL level sample pulse to the TL1591. The output voltage then takes 30 ns to settle to within 0.1% of the applied input. The TL1591 then remains in hold mode until another sample pulse is received.

As previously discussed, the phenomenon of clock feedthrough tended to corrupt the TC104 output signal around the rising and falling edges of the reset clock. For this

reason it was desirable to sample the data signal well away from this reset noise. The effect of reset noise was reduced by careful design of the clock signal logic to minimise reset clock pulse width. By reducing the reset pulse duration it was possible increase the data pulse duty cycle and so allow a greater settling time between reset noise pulses. The sample pulse was generated by the clock logic PLDs using a 20 ns delay line. This pulse was timed to occur mid-way between reset clock pulses to avoid the effects of reset feed through. This signal timing is illustrated in Figure 5.10.

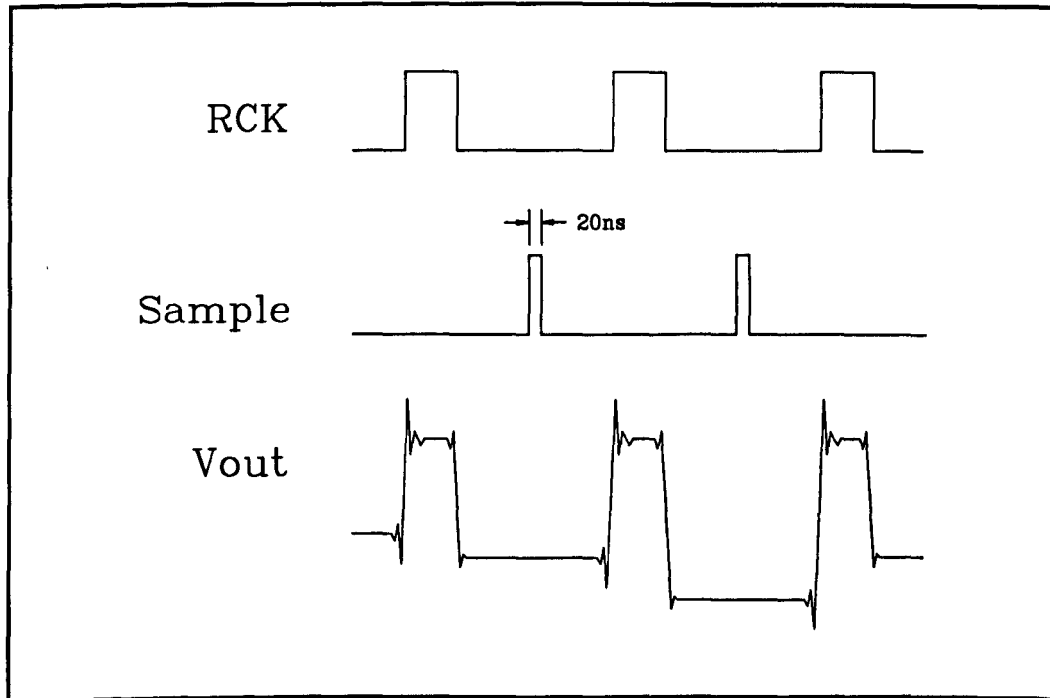


Figure 5.10 Sample and Hold Timing

5.4.3.3 Low Pass Filter

The output signal from the sample and hold amplifier was seen to contain high frequency noise components caused by feedthrough from the CCD drive clocks and the sample pulse. This noise was reduced by the use of a simple two pole Butterworth low pass filter [12] constructed using a NE531 OpAmp[13].

The Butterworth filter configuration was chosen for its sharp frequency cut-off characteristics which allowed the exclusion of high frequency components without attenuating valid pixel data. The 3 dB cut-off frequency, f_c , of the filter was chosen to be double the pixel data rate, f_{DATA} . Experimentation showed that this configuration

reduced feedthrough noise to negligible levels whilst preserving genuine high frequency image data.

$$f_{\text{DATA}} = 625 \text{ kHz}$$

$$f_{\text{C}} = 2 \times f_{\text{DATA}} = 1.25 \text{ MHz}$$

5.4.3.4 Black Level Clamp

The next stage of the signal processing circuit involved black level clamping the signal to a stable reference voltage. This resulted in the output signal from pixels which received no illumination being held at the clamp voltage. In this way the output voltage of each image pixel could be measured relative to its offset from the clamp voltage.

The TC104 provides black level reference pixels before and after each line of image data. The output from these pixels was sampled and used as a reference voltage for the black level clamp circuit. The location of these black level reference pixels was identified using control outputs generated by the clock logic PLDs.

A National Semiconductor LF398 sample and hold amplifier [14] was used to acquire the black level reference voltage. The LF398 is a monolithic sample and hold circuit offering fast, accurate signal acquisition with low hold mode droop rate.

The LF398 requires an external hold capacitor on which to store signal charge during hold mode operation. The choice of hold capacitor is generally a trade off between fast acquisition time and low droop rate. A 2000pF polystyrene hold capacitor was selected which allowed acquisition to 0.01% accuracy within 6 μs , and a hold mode droop rate of 0.02 V/s. At the operating frequency of 625 kHz, this ensured signal droop during a full line output was limited to 0.1 mV.

Black level clamping of the data signal was achieved using a differential amplifier configured to subtract the black level reference voltage from the sensor output voltage. This resulted in pixels under zero illumination being clamped to zero volts output.

A National Semiconductor LM11C precision OpAmp was chosen to construct the differential amplifier because of its excellent stability and CMRR¹ [15]. To ensure accurate operation it was necessary to use precision metal film resistors with 0.1% tolerance, since any resistor mismatch would effect common mode signal rejection. This resistor matching requirement restricted the differential amplifier to a unity gain configuration.

5.4.3.5 Output Amplifier

The resulting output signal from the black level clamp circuit was a negative going video signal with full scale amplitude of approximately 1.5V. This signal required inversion and amplification in order to convert it to the ADC input voltage range of 0 to +5 V. This was done using an NE531 OpAmp configured as an inverting amplifier.

Amplifier gain and offset were separately adjustable using multi-turn potentiometers. The offset range was restricted to just 40 mV to compensate for any errors in the black clamp circuit. A large voltage gain adjustment range of 3 to 4.5 was allowed, to take into account the variation of luminous sensitivity characteristics between individual sensor devices.

5.4.4 Circuit Layout

Despite much effort to reduce noise pick-up from the drive circuitry to the sensor output signal, the original circuit design exhibited an unacceptable signal to noise ratio. In an effort to minimise power supply noise the sensor drive circuit was transferred to a PCB, which offered considerably better grounding than the original prototyping board.

Circuit layout was an important consideration in the reduction of system noise [16]. In particular, provision of ample grounding and power supply de-coupling helped to eliminate much of the noise present on ground and power lines. Other considerations included careful routing of sensitive signal lines to avoid contact with noisy, high level clocks and power supply lines. Efforts were made to achieve a short, straight signal

¹ Common Mode Rejection Ratio

path in order to reduce the opportunities for noise pick-up. The length of all clock lines was also minimised in an effort to reduce their impedance, thereby reducing noise coupling onto adjacent signals.

In order to reduce system noise it was necessary to isolate the high power clock driver stage of the circuit from the sensitive components which process the output signal of the sensor. This was achieved by building the respective circuits on different circuit boards with separate power supplies.

The separation of digital and analogue supplies made a major contribution to noise reduction. The clock driver circuit required a 5 volt supply, by deriving this voltage from the 14 volt analogue supply, rather than the 5 volt digital supply, noise coupling between circuits was avoided.

In addition to generating clock pulses to drive the CCD registers, the sensor logic circuit also produces high frequency control signals used for output signal processing. To avoid these signals coupling noise into the sensitive signal processing stage, the circuit stages were electrically isolated from one another by the use of opto-couplers. High speed HCPL2630 opto-couplers were used, as shown in Figure 5.11.

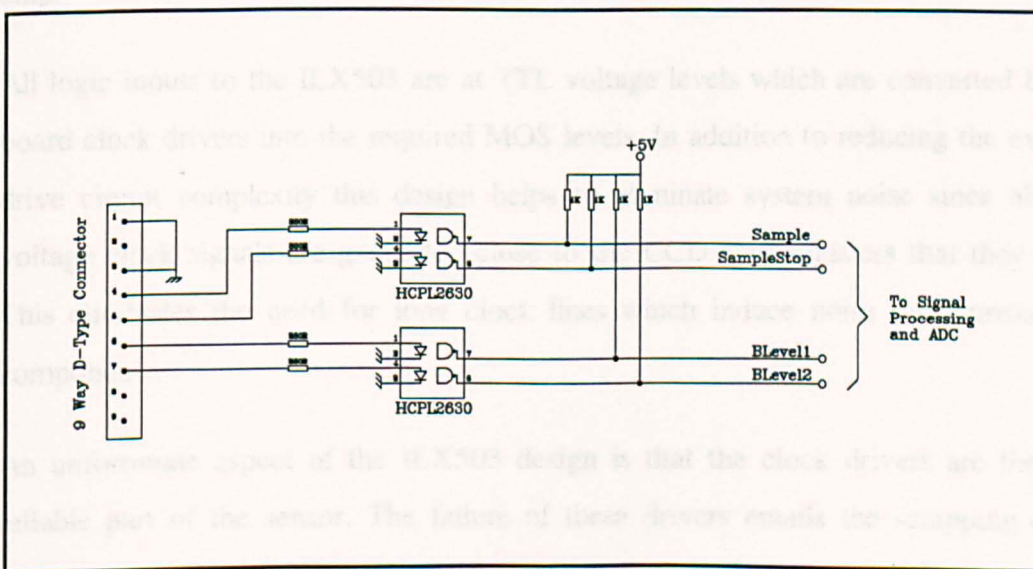


Figure 5.11 Optical Isolation of Sensor Control Signals

Similarly, all exposure timer input lines were optically isolated using ILQ47 optocouplers as shown in Figure 5.12.

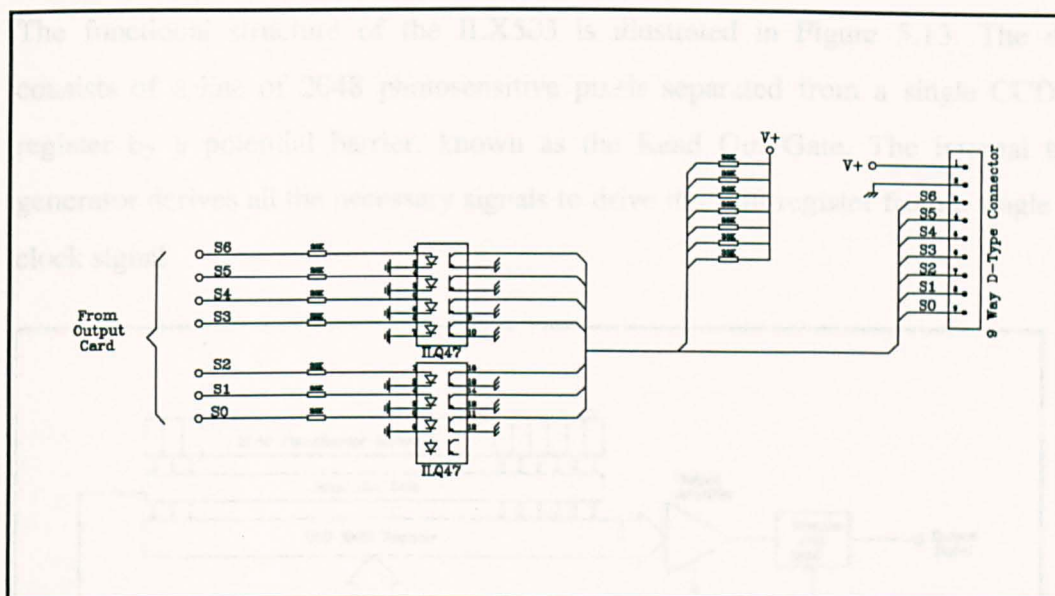


Figure 5.12 Optical Isolation of Exposure Time Inputs

5.5 Sony ILX503 CCD Sensor

The Sony ILX503 is a general purpose CCD linescan sensor containing 2048 photo sensitive elements each measuring $14\mu\text{m}$ square [17]. The ILX503 uses a low noise, buried channel structure and offers simplified operation by incorporating internal logic timing generators, clock drivers and optional sample and hold operation within a single chip.

All logic inputs to the ILX503 are at TTL voltage levels which are converted by on-board clock drivers into the required MOS levels. In addition to reducing the external drive circuit complexity this design helps to eliminate system noise since all high voltage clock signals are generated close to the CCD shift registers that they drive. This eliminates the need for long clock lines which induce noise into surrounding components.

An unfortunate aspect of the ILX503 design is that the clock drivers are the least reliable part of the sensor. The failure of these drivers entails the scrapping of the entire sensor. The ILX503 is particularly sensitive to latch-up, which permanently damages the internal clock drivers and renders the device inoperative. These problems can be reduced by careful handling and operation of the sensor, and are offset by the low cost of device replacement.

The functional structure of the ILX503 is illustrated in Figure 5.13. The device consists of a line of 2048 photosensitive pixels separated from a single CCD shift register by a potential barrier, known as the Read Out Gate. The internal timing generator derives all the necessary signals to drive the shift register from a single input clock signal

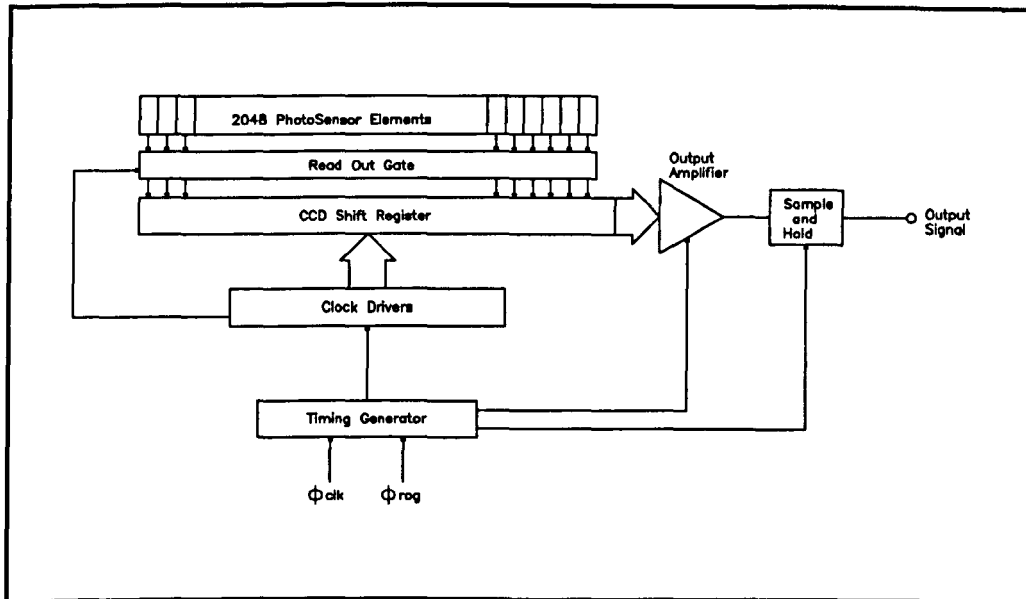


Figure 5.13 Functional Structure of the Sony ILX503 Sensor

Typical performance values for the Sony ILX503 linear CCD array sensor are given in Table 5.1, below.

Parameter	Typical Value	Units
Number of Elements	2048	
Element Spacing	14	μm
Sensitivity	30	V / (lx s)
Gain Non Uniformity	8	%
Saturation Output	2.5	Volts
Dark Signal Voltage	0.5	mV
Dark Signal Non-Uniformity	0.5	mV
Total Transfer Efficiency	97 %	
Maximum Data Rate	5	MHz

Table 5.3 Typical Operating Characteristics of the ILX503 Sensor

5.5.1 Clocking Requirements

The simplest mode of operation for the ILX503 involves using the internally generated reset clock, so that only two logic inputs are required, the read out gate clock, ϕ_{ROG} , and the pixel clock, ϕ_{CLK} . Logic timing waveforms for this mode of operation are illustrated in Figure 5.14.

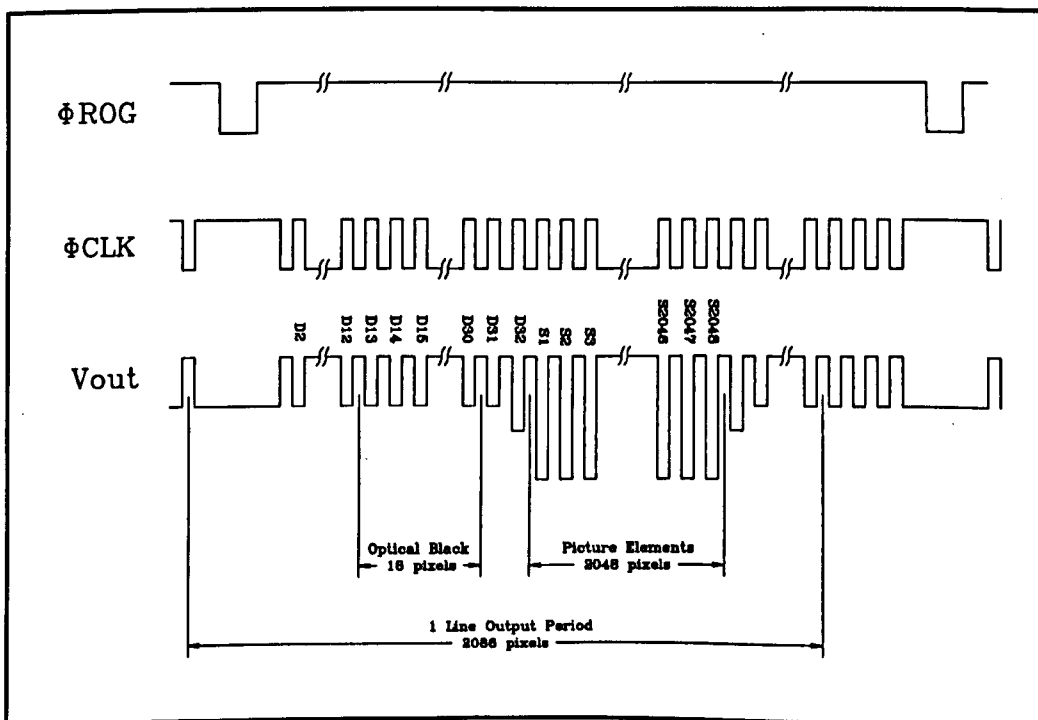


Figure 5.14 Operating Waveforms for the ILX503 Sensor

A single line readout is initiated by a low going pulse on the \emptyset ROG line, which opens the read out gate to allow signal charge to flow from the photosites into the CCD shift register. The design of the ILX503 maintains a high potential gradient across the read out gate in order to induce complete signal charge transfer to the CCD shift register within a finite time. To ensure complete charge transfer across the read out gate the ILX503 data sheet recommends a minimum \emptyset ROG pulse width of 1 μ s.

Signal charge packets are clocked along the shift register to the output amplifier by the pixel clock, \emptyset CLK. The ILX503 requires that \emptyset CLK is held high while the read out gate is open. This suspends the action of the CCD shift register while charge builds up within register elements.

Signal charge accumulates within the photosites at a rate proportional to the incident radiant flux until the read out gate is opened and charge passes to the shift register. Hence, the read out gate period determines the exposure time during which signal charge is allowed to build up. The minimum exposure time is determined by the time taken to clock the previous line of data through the shift register before admitting new signal charge. An entire line readout from the ILX503 consists of 2048 image pixels plus a number of optically black and isolation pixels before and after the image data.

The ILX503 offers the option of either internal or external sample and hold operation. Figure 5.14 illustrates external sample and hold operation, so that the output data signal consists of a series of low going voltage pulses separated by reset clock pulses. During internal sample and hold operation only valid pixel data is sampled, so that the reset pulses are eliminated from the output signal.

5.5.2 Drive Hardware

A drive circuit was developed to allow operation of the ILX503, with the option of external reset mode being selected by a single switch. This circuit is shown in Figure A-3 of Appendix A. Three ALTERA EP600 PLD devices were programmed to produce all the necessary clock signals to drive the ILX503 and provide control inputs to the subsequent signal processing hardware.

The same 7 bit exposure timing architecture was used as described for the TC104 drive circuit. The exposure time input value could be set either internally by on-board

switches or from an external source. The exposure time input source was selected by a single switch which provided the data select input to two 74HCT157 multiplexers.

The minimum exposure time was determined by the time taken to clock a full data line through the CCD shift register. An entire data line of the ILX503 consists of 2086 pixels, and since the minimum exposure time adjustment was 512 pixels, the minimum exposure time was:

$$\text{Minimum Exposure Time} = 5 \times 512 \times T_{\text{CK}} = 2560 \times T_{\text{CK}} \quad (5.4)$$

At the normal operating frequency of 625 kHz, the corresponding pixel clock period is 1.6 μ s. Hence, the drive circuit allowed exposure time adjustment between 4.1 ms and 104 ms.

Internal exposure timing was generally used when running the ILX503 board as a stand alone unit, without software support. External exposure timing enabled remote variation of the sensor system sensitivity from the image acquisition software. Also, when using the scanning laser light source it was necessary to synchronise the scanning period to the exposure timer to ensure equal light exposure during each line of data.

5.5.2.1 External Reset Mode

During normal operation of the ILX503, signal charge present in the output stage is reset before receipt of charge from the subsequent pixel. This operation is controlled by the reset clock signal, $\overline{\text{RS}}$. The reset clock may either be generated by external drive hardware or produced by the internal timing generator. External reset mode is only required when intermittent reset operation, or correlated double sampling (CDS) are used.

Intermittent resetting collects the charge from several adjacent pixels into a single signal charge packet within the output stage. This operation effectively integrates the output signal over a number of pixels, producing increased signal charge under low light conditions, at the expense of reduced spatial resolution.

Correlated double sampling (CDS) is a signal processing technique used to reduce noise levels in CCD sensor output signals. CDS is discussed at length in Section 5.5.3.1.

The ILX503 drive circuit allowed either internal or external reset mode operation to be selected by a single switch. In practice, external reset was only used in conjunction with an implementation of the CDS noise reduction method.

5.5.2.2 Power Supply Considerations

The ILX503 requires two supply voltages, V_{DD1} (+9V) used for the analogue sections of the device and V_{DD2} (+5V) used for the digital timing generator logic. Extensive power supply decoupling was used to minimise system noise. Decoupling was provided by a parallel combination of 10 μ F and 1000 pF capacitors as close to each power supply input terminal as possible. For added noise immunity, good grounding was ensured by constructing the ILX503 drive circuit on a Eurocard prototyping board with three separate supply and ground planes.

The ILX503 is particularly susceptible to SCR¹ latch-up, a potentially destructive state causing high power conduction between power supply and ground [18]. Latch up occurred when gate inputs were driven beyond supply rail voltage during power-up and power-down. In order to prevent latch up of the sensor, it was necessary to ensure that V_{DD2} never exceeded V_{DD1} . This condition was satisfied by deriving V_{DD2} from V_{DD1} using a 7805 voltage regulator.

5.5.3 Output Signal Processing

An output signal processing circuit was developed in order to condition the video output from the ILX503 sensor prior to digitisation. A block diagram of the signal processing route used is given in Figure 5.15, and the full circuit implementation is shown in Figure A-4 of Appendix A.

¹ Silicon Controlled Rectifier

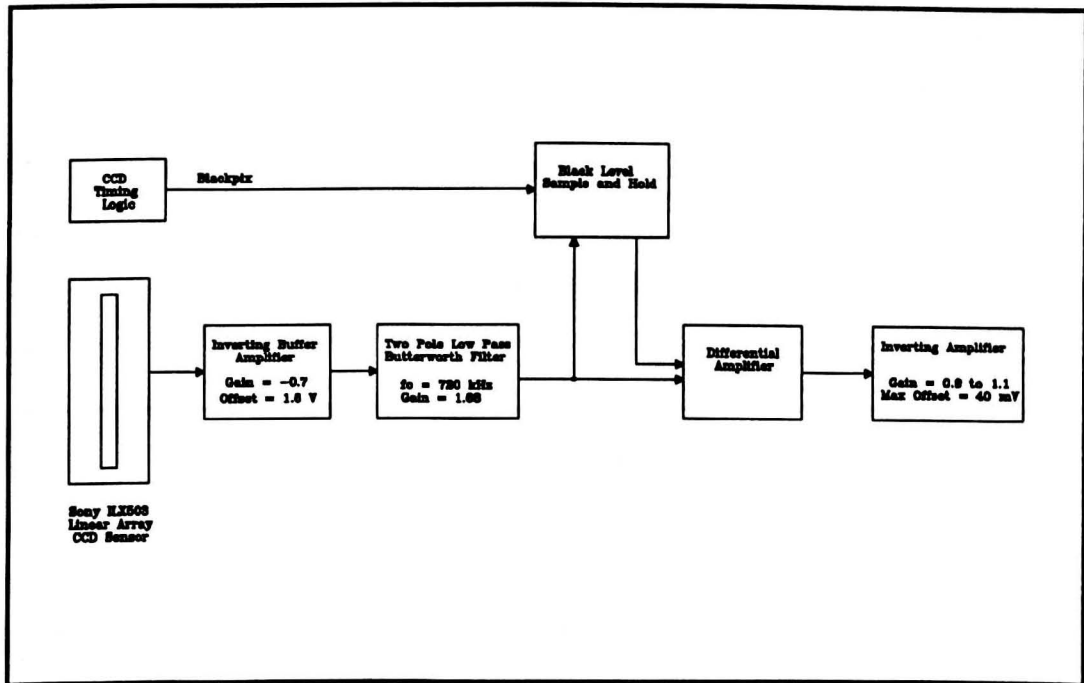


Figure 5.15 ILX503 Output Signal Processing Scheme

It can be seen that the signal processing method used for the ILX503 sensor is basically a simplified version of that used for the TC104. The main difference between the circuits involves the use of internal sample and hold mode for the ILX503, thereby removing one of the signal processing stages.

The raw output from the ILX503 was a negative going video signal, with a full scale range between approximately +0.2V and +5V. Hence, the overall system gain of the output signal processing circuit was much reduced from that used by the TC104.

5.5.3.1 Correlated Double Sampling (CDS)

As mentioned previously, the ILX503 offered the option of external reset mode for use with the noise reduction method of correlated double sampling. CDS consists of sampling both amplifier precharge and valid video output levels and differencing the two signals [19] [20]. The technique relies on the fact that the noise voltages on both the sampled signals have a strong correlation in phase and will cancel out at the output of the differential amplifier.

Figure 5.16 shows reset clock, $\overline{\text{ORST}}$, and output signal, V_{out} , waveforms sampled on a digital storage oscilloscope during external reset mode operation of the ILX503. The

output signal exhibits three distinct portions for each pixel, corresponding to reset, precharge and data signals.

The reset pulse is the result of clock feedthrough from the reset clock signal, which recharges the charge detection amplifier after each pixel output. The precharge signal represents the amplifier reset level. Finally, signal charge is shifted into the output stage, causing the corresponding low-going data pulse.

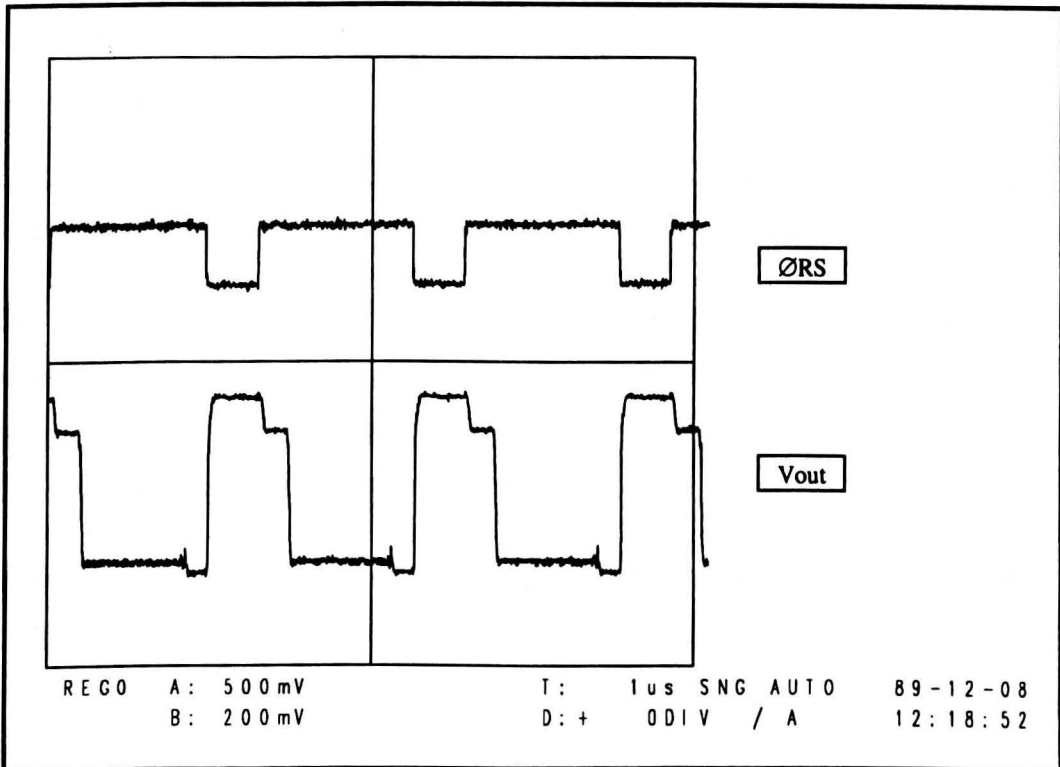


Figure 5.16 External Reset Mode Operation of the ILX503

Several variations of CDS were attempted before settling on the method shown in Figure 5.17. This method helps to reduce sample pulse feed through by using a single sample pulse to sample both the amplifier reset voltage and the valid output data. The same sample pulse feed through appears on both signals and is hence cancelled out at the differential amplifier output.

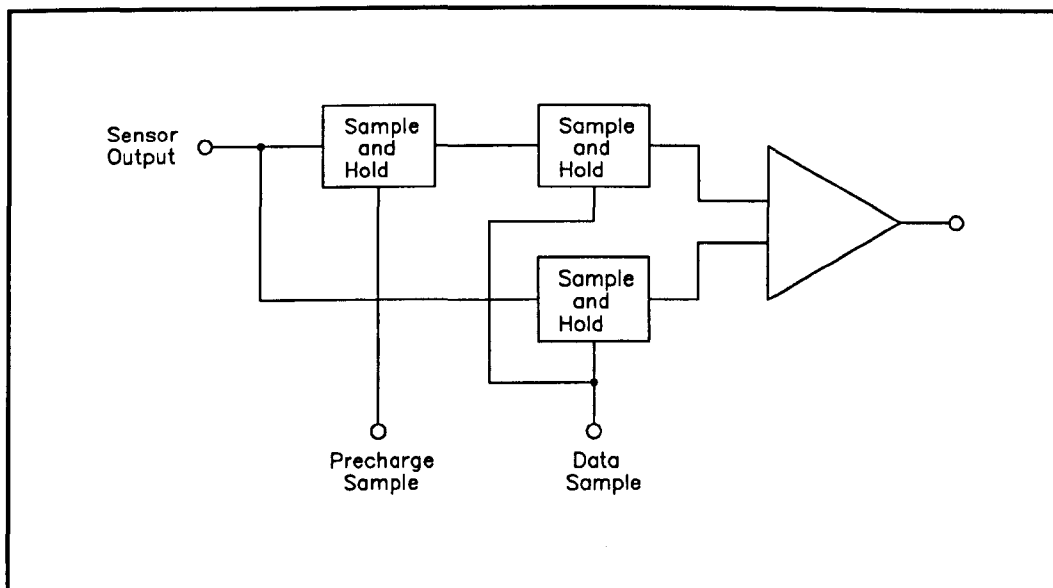


Figure 5.17 Correlated Double Sampling

The CDS circuit was implemented using a Sony CXA1439M IC [21]. This is a surface mount chip developed specifically for CDS noise reduction of CCD arrays. The sampling of precharge and data voltages were triggered respectively by the logic input SHP and SHD, generated by the CCD timing logic.

It was found that reset clock feedthrough onto the narrow precharge section of the CCD output signal tended to increase the noise voltage of the precharge sample. This effect overshadowed any benefits gained by the use of CDS.

The CDS technique was eventually discarded, since no appreciable improvements to output signal to noise ratio were achieved. However careful circuit design ensured that the output signal noise level, without the application of CDS, could be reduced to negligible levels.

5.6 Data Acquisition System

5.6.1 Analogue to Digital Conversion

The output from the analogue signal processing portion of each sensor circuit was an optically black clamped voltage waveform with 0 to 5 V full range voltage swing and a pixel rate of 625 kHz. This analogue signal was then sampled and digitised to produce a series of discrete digital pixel values prior to interfacing with the digital processing hardware.

Although the frame store used in this project was only capable of displaying images at a resolution of 8 bits per pixel, a greater digitisation resolution was chosen to allow digital contrast enhancement. A 12 bit ADC¹ was used in order to maximise the contrast information in the captured signal, without sampling spurious noise variations

Since the output frequency and voltage range of each sensor was identical, the same ADC was used for digitisation of pixel data from either of the two CCD sensors used in this project. The analogue input signal to the ADC was selected using a single jumper switch. Careful design of the control logic timing ensured compatibility between the two separate sensor systems without the need for any additional modifications.

Analogue to digital conversion was achieved using an Analog Devices AD1671 [22]. The AD1671 is a monolithic 12 bit ADC incorporating a high speed sample and hold amplifier and a precision voltage reference on a single chip. The design of the AD1671 uses a hybrid architecture consisting of subranging flash converters and digital error correction to achieve a combination of speed and accuracy. A maximum conversion time of 800 ns ensures a throughput rate of up to 1.25 MHz. The input voltage range can be configured for either unipolar (0 to +5 V or 0 to +2.5V) or bipolar (± 5 V or ± 2.5 V) operation.

The operation of the AD1671 is controlled by two TTL logic signals as illustrated in Figure 5.18. A conversion cycle is initiated by the rising edge of the ENCODE pulse, which should remain high for between 20 and 50 ns. When conversion is complete the ADC responds by taking the DAV (Data Available) line low. The DAV pulse duration is specified to be between 150 and 300 ns. Data from the previous conversion is guaranteed valid on the falling edge of DAV and is replaced by data from the new conversion on the rising edge of DAV.

¹ Analogue to Digital Converter

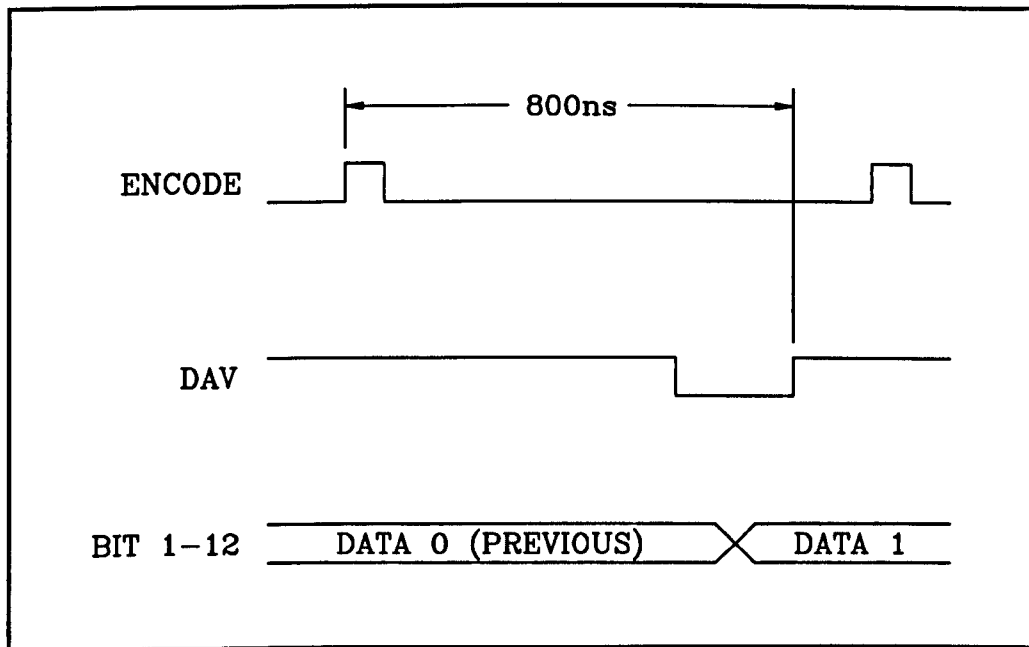


Figure 5.18 Operation of the AD1671 Analogue to Digital Converter

The AD1671 provides an out of range (OTR) output to signify when the analogue input signal exceeds the selected voltage range. This signal, along with the most significant bit (MSB) output, may be used to indicate low or high overflow. Under certain circumstances, either type of overflow may occur during luminous saturation of the CCD array sensor. The interface hardware used the OTR and MSB signals to flag a saturation error to the system software.

For an ADC sampling pixel data at a digitisation resolution of N bits per sample, the number of possible grey scale values taken by the digital output is given by :

$$\text{Number of Grey Scales} = 2^N$$

Hence, a 12 bit ADC quantises an analogue input into one of 4096 possible grey scale values. The threshold voltage, V_{TH} , between each digital output code of an N bit ADC with a full scale voltage range between V_{MAX} and V_{MIN} is ideally half the input voltage level required to trigger the least significant bit output.

$$V_{TH} = \frac{1}{2} \left[\frac{V_{MAX} - V_{MIN}}{2^N} \right] \quad (5.5)$$

For a 12 bit ADC with input range 0 to +5V, this equation gives a threshold voltage of 0.6 mV. Signal voltage levels below V_{TH} will cause no change in digital output code. This digital quantisation threshold level places a practical limit on the signal noise level that will be detectable at the ADC output.

5.6.2 Interface Circuit

Digital pixel data was passed directly to the input channel of a T800 transputer via an Inmos serial link operating at 20 Mbits/s [23]. Parallel data from the ADC was converted to the required serial link data format using an Inmos CO11 link adaptor [24] controlled by an Altera EP600 PLD. The PLD device was programmed using Algorithmic State Machine logic representation to implement the simple CO11 handshaking protocol. The interface circuit used is shown in Figure A-5 of Appendix A.

Each ADC conversion was initiated by an ENCODE pulse generated by the PLD. This pulse was timed to ensure the analogue pixel data was sampled midway between each pixel transition. This timing avoided sampling clock feedthrough noise that occurred between valid pixel data, as illustrated in Figure 5.19. The ENCODE pulse width was adjusted to approximately 30 ns using a simple RC delay network.

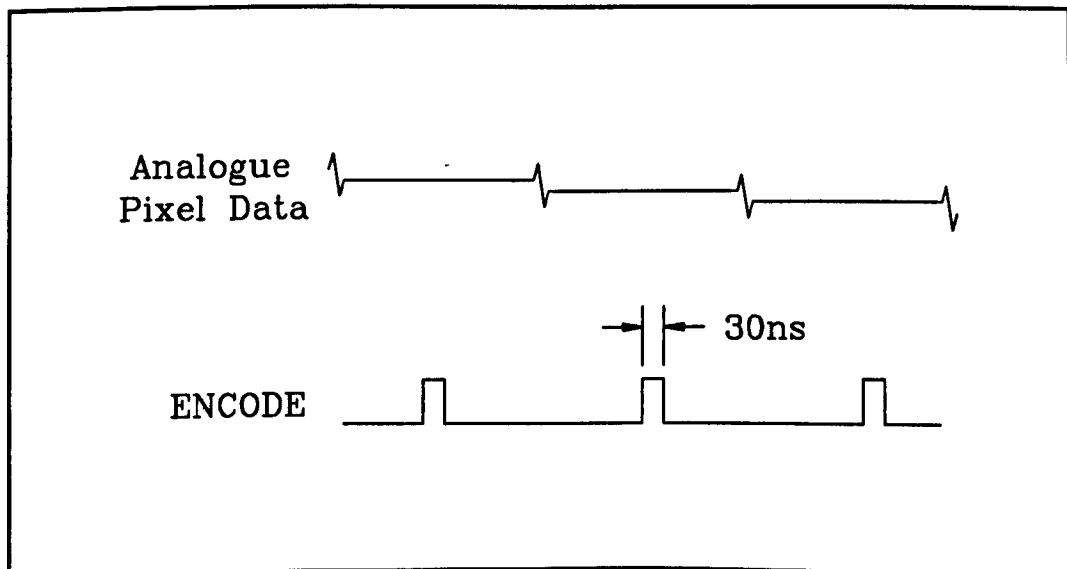


Figure 5.19 Encode Pulse Timing

As specified in the previous section, after a conversion new data from the ADC is guaranteed valid on the rising edge of the DAV signal. Pixel data and the OTR flag

were latched using 74HCT574 octal D-type latches, triggered on the rising edge of the DAV pulse.

The standard Inmos link protocol passes data as a stream of 8 bit data packets. This required that the 12 bit parallel data produced by the ADC needed to be divided into two bytes for separate transmission down the serial link. The first byte consisted of the eight least significant data bits from the ADC, the second byte was the four most significant data bits, followed by the out-of-range (OTR) flag and three zero bits. This data format is illustrated in Figure 5.20.

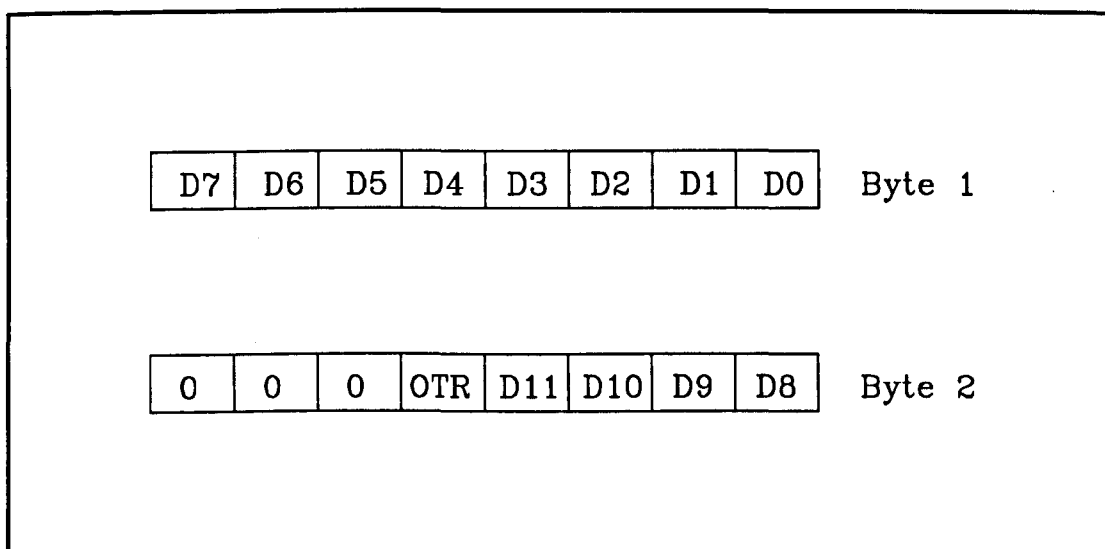


Figure 5.20 Pixel Data Format

Acquisition of a line of pixel data was initiated from software by sending a single byte down the Inmos link to the interface hardware, which responded by acknowledging receipt of the message. The logic then waited for the beginning of the next full data line, signified by the rising edge of the BLACKPIX signal.

Pixel data was then sent down the Inmos link, low byte first, by holding IVALID high until the receiving process responded with an IACK pulse. This process was repeated until the end of the data line was signified by the ENDSCAN pulse. A simplified waveform representation of the interface logic timing is given in Figure 5.21.

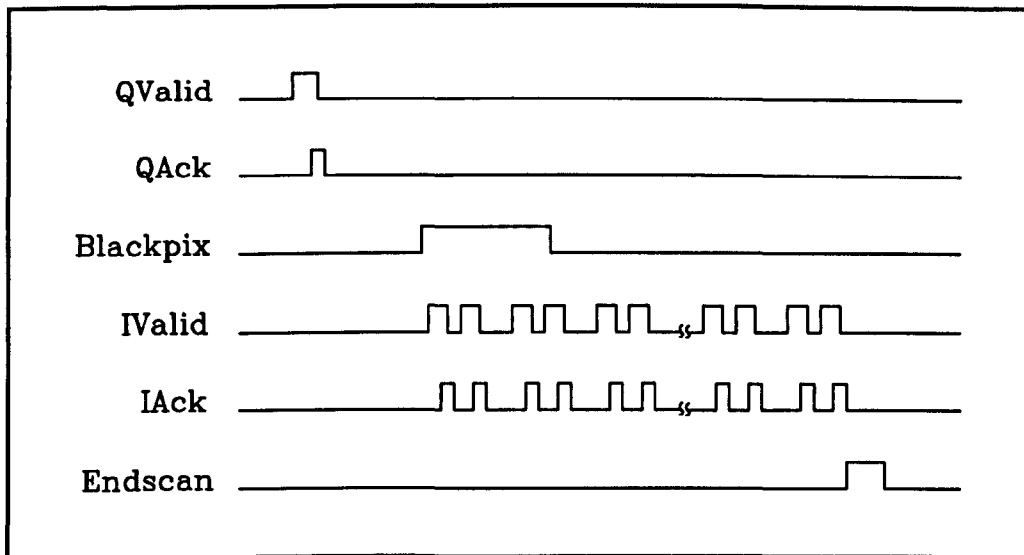


Figure 5.21 Interface Circuit Timing Waveforms

The data rate of the image acquisition system was limited by the 20 Mbits/s transmission speed of the Inmos serial link. The link protocol codes each output data byte as an 11 bit data packet on LinkOut. The receiving process responds with a two byte acknowledge packet which occurs on LinkIn during the transmission of the data packet. This communication protocol sets a theoretical minimum transmission rate of 550 ns per byte.

The maximum achievable pixel rate, using the interface circuit described above, was found to be around 700 kHz, or 1.4 μ s for the transmission of each two byte pixel. In order to achieve these high data rates it was necessary to clock the state machine logic at 10 MHz, twice the input clock speed of the CO11.

5.7 Image Display

5.7.1 MosaiQ Frame Store

Image display was controlled by a Quintek MosaiQ monochrome frame store [25]. The MosaiQ is an image capture and display board constructed around a T805 transputer. It provides 1 Mbyte of main program memory with an additional 1 Mbyte of video memory, allowing storage and display of multiple high resolution images. The MosaiQ provides an 8 bit monochrome image plane and a separate overlay plane for graphics, text and cursor display.

The MosaiQ engine software provides a comprehensive library of parallel C functions for capture and display of video images, overlay text and graphics and interactive image manipulation via a mouse controlled cursor.

In addition to supporting several standard video display formats, the MosaiQ engine software allows explicit specification of the video timing parameters to provide customised display formats. A display resolution of 1024 by 512 pixels was chosen to provide the high quality images required for this project. To achieve this high resolution format it was necessary to increase the pixel update rate above that provided by the standard MosaiQ hardware.

The existing pixel clock rate of 25 MHz used by the MosaiQ board was increased by substituting a 33 MHz crystal to provide the master pixel clock signal. The resulting image display format is summarised in Table 5.4.

Lines	512
Columns	1024
Pixel Rate	33 MHz
Line Rate	28 kHz
Frame Rate	50 Hz
Interlace	No

Table 5.4 Image Display Format Used by the MosaiQ Framestore

5.7.2 Colour Monitor

Images produced by the image acquisition system were displayed on a Mitsubishi C6922 19 inch colour monitor [26], which provides high resolution image display with excellent geometric stability to ensure no image distortion. The monitor accepts a standard RGB video input plus separate vertical and horizontal sync signals. The use of a colour monitor allowed pseudo colour palettes to be displayed, which can be used to enhance the contrast definition of monochrome images.

5.8 I/O Control

During operation of the image acquisition system it was necessary to provide software control of the image sensor control logic and the dual axis motor system. The necessary control signals were interfaced to the transputer processor network using the two I/O boards described below.

5.8.1 Input Card

Input signals to the image acquisition software were provided via a 32 channel transputer input card, developed at Liverpool University by Michael Houghton. The input card used a CO11 link adaptor to interface between a standard Inmos transputer serial link and 32 parallel data input lines. All input lines were optically isolated to prevent noise coupling from external hardware.

Input signals were read by addressing one of four 8 bit ports. The protocol used by the input card involved first writing a single byte port address to the CO11 and then reading the relevant 8 bit input data from the link.

5.8.2 Output Card

Output signals from software to external hardware were produced by a 32 channel transputer output card, developed at Liverpool University by Michael Houghton. The output card used an Inmos CO11 link adaptor to write data to 32 optically isolated parallel data lines.

The protocol used by the output card involved sending a two byte message down the transputer link. This message consisted of a single byte addressing one of the four output channels, followed by an 8 bit output value.

5.9 Motor Subsystem

In order to build up an image using a linear array sensor it is necessary to provide relative movement between the sensor and the object being imaged. The optical arrangement used in this project required that the sensor be stationary with respect to the scanning laser light source whilst moving the radiographic film.

Two dimensional motion of the radiographic film was achieved using two stepper-motor-controlled linear translation stages, mounted perpendicular to one another. Each translation stage allowed precise, repeatable motion over an extended travel distance.

The motor subsystem was under the software control of a dedicated transputer process with all communications being carried out via the I/O cards described in the previous section. A block diagram of the motor subsystem is shown in Figure 5.22.

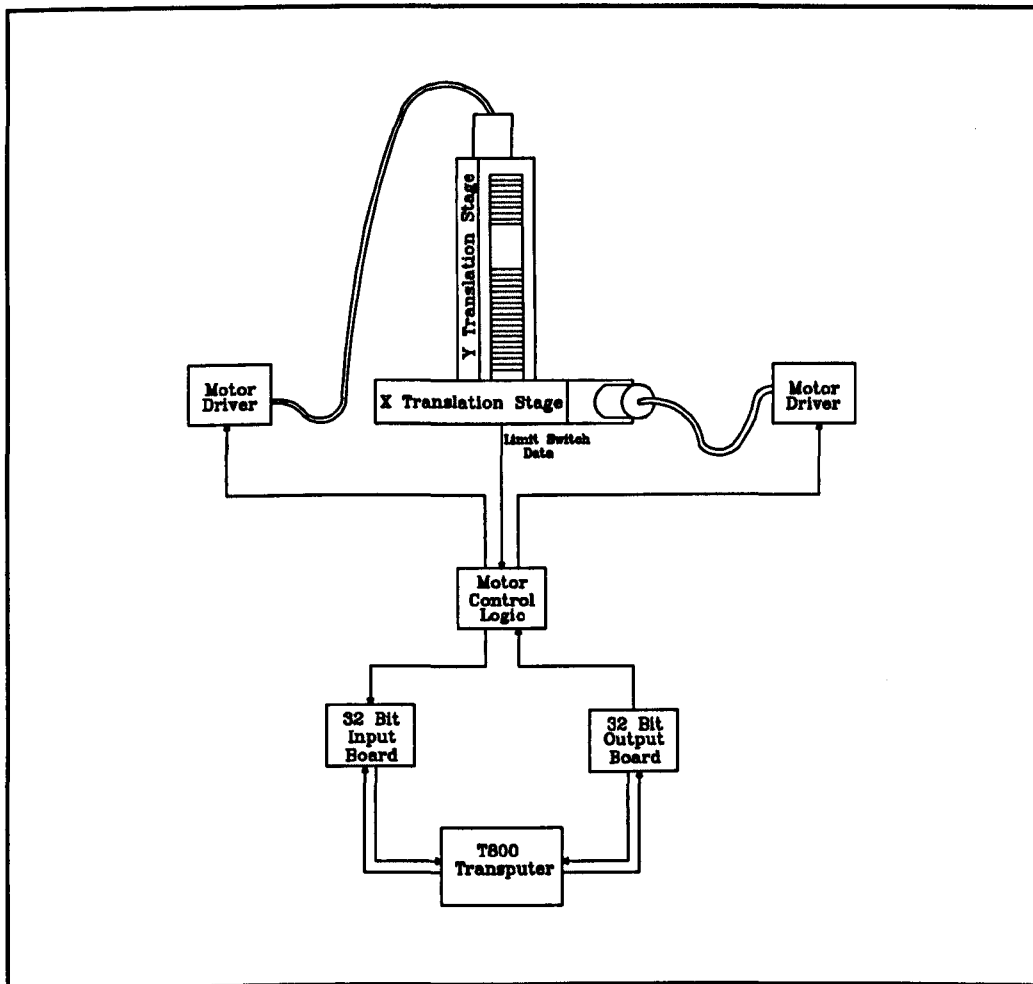


Figure 5.22 Overview of the Motor Subsystem

5.9.1 Vertical Motion

5.9.1.1 Stepper Motor and Controller

The CCD sensors used in this project were mounted so that the array of sensor elements ran horizontally. Hence, in order to build up a two dimensional image it was necessary to scan the radiographic film vertically across the sensor array. In order to produce a uniform sampling grid the film had to be moved in steps equal to the sensor

array spacing. The pixel array spacing of the two sensors used in this project is shown in Table 5.5.

Sensor	Pixel Spacing
TC104	10.7 μm
ILX503	14 μm

Table 5.5 Sensor Pixel Spacings

The extremely close spacing of sensor elements within the CCD sensors required precise control of the vertical scanning mechanism to give a linear step resolution of less than a micron.

Vertical motion of the film holder was produced by a Photon Control NanoStepper motor [27] driven by a Compumotor A Series Microstep Controller [28]. Microstepping drives vary the input current sinusoidally between motor phases to provide intermediate positions between those of a conventional stepper system. This drive method offers smoother motion in addition to a dramatic increase in step resolution.

The Compumotor A-Driver provides a variety of resolutions up to 25,000 steps per revolution, which when used with a 1mm pitch leadscrew equates to 0.04 microns per step.

5.9.1.2 PTS 1000 Linear Translation Stage

Rotary motion from the stepper motor was converted into linear motion using a Photon Control PTS 1000 linear translation stage. The PTS 1000 uses a precision ground leadscrew arrangement to provide smooth, repeatable linear travel of the sliding carriage. The use of preloaded bearings essentially eliminates backlash even under high loading conditions.

The PTS1000 is constructed as a sealed unit to prevent encroachment of dirt into the leadscrew. A series of mounting holes are provided to allow flexible stacking with other translation stages from the Photon Control range to achieve multi-axis motion.

Limit switch sensors are provided to indicate the full extent of travel in each direction. The operating characteristics of the PTS 1000 are summarised in Table 5.6 below.

Maximum Travel	150 mm
Tracking Accuracy	1 micron
Leadscrew Pitch	1 mm

Table 5.6 PTS1000 Translation Stage Specifications

5.9.2 Horizontal Motion

5.9.2.1 Stepper Motor & Controller

Horizontal motion of the film holder was produced using a Compumotor A57-83 stepper motor [29]. This motor was driven by a Digiplan 1054 half stepping controller capable of an angular resolution of 400 steps/revolution [30]. When used in conjunction with a 1 mm pitch leadscrew this arrangement allowed a minimum step distance of 2.5 μm .

The horizontal motor system was only used for relatively large movements, when positioning the film holder prior to capturing a vertical scan line. Hence, the minimum step resolution was not such an important factor as with the vertical stage.

5.9.2.2 PTS 1600 Linear Translation Stage

A Photon Control PTS 1600 linear translation stage was used to provide accurate linear motion in the horizontal axis. The PTS1600 uses precision leadscrew arrangement to produce precise, repeatable motion over a 300 mm travel distance. It is basically a larger, more robust version of the PTS1000, described in the previous section.

The operating characteristics of the PTS 1600 are summarised in Table 5.7, below.

Maximum Travel	300 mm
Tracking Accuracy	1.5 microns
Leadscrew Pitch	1 mm

Table 5.7 PTS1600 Translation Stage Specifications

The entire vertical translation stage was mounted on the sliding carriage of the PTS1600 using a specially designed mounting bracket. This arrangement produced two axis motion of the film holder in a plane parallel to the surface of the linear CCD array sensor.

5.9.3 Motor Control Logic

Simple motor control logic was implemented to drive both motor stages and ensure neither linear stage was driven past its full travel. This logic, constructed around an Altera EP310 PLD [31], is shown in Figure A-6 of Appendix A. The circuit also included warning LEDs¹ to indicate when either positioning stage had been driven to its limit.

Control of the motors was achieved by supplying direction and clock signals to the encoder inputs of the relevant motor driver. Each step of the motors required a single input clock pulse. This clock and direction data was produced from software using the transputer output card described previously.

The limit switch logic operated by inhibiting clock pulses to the motor indexer if the resultant motion would drive the linear translation stage beyond its full travel distance. In addition, limit switch data could be read by software via the transputer input card.

The Compumotor stepper driver unit accepts standard TTL logic level inputs, enabling direct connection to the motor control logic. However, the Digiplan 1054 driver required a simple interface circuit to convert inputs to CMOS logic levels. This was provided by using HCT7407 buffers with open collector outputs on the relevant output lines.

5.10 Laser Scanner

The choice of light source for the image capture system was governed by the need for uniform, collimated, high intensity illumination across the face of the linear CCD sensor. A conventional, incoherent arc light source of the required power, with

¹ Light Emitting Diode

associated collimating optics, would cost several thousand pounds [32]. Instead, a scanning beam laser source was used, which provided excellent imaging qualities at low cost.

As was noted in Chapter 2, it is a general characteristic of industrial radiographs that areas corresponding to solid, parent metal are characterised by very dark images, typically with optical densities of 4 and above. Consequently, a high radiant power light source is necessary to ensure adequate imaging of these areas. It should be noted, however, that this light intensity is only required over the small area of the sensor elements.

The use of a laser light source offered several advantages over a conventional incandescent beam. Laser diodes emit highly ordered, coherent light from a small radiant source area [33]. This enabled simple collimating optics to produce a high intensity beam with small diameter and low divergence. The monochromatic nature of laser light eliminated any defocusing caused by chromatic aberration.

The method of illumination consisted of scanning a powerful, collimated laser spot across the photosensitive elements of the CCD sensor, using two rotating mirrors to deflect the beam. A block diagram of the laser scanning subsystem is shown in Figure 5.23.

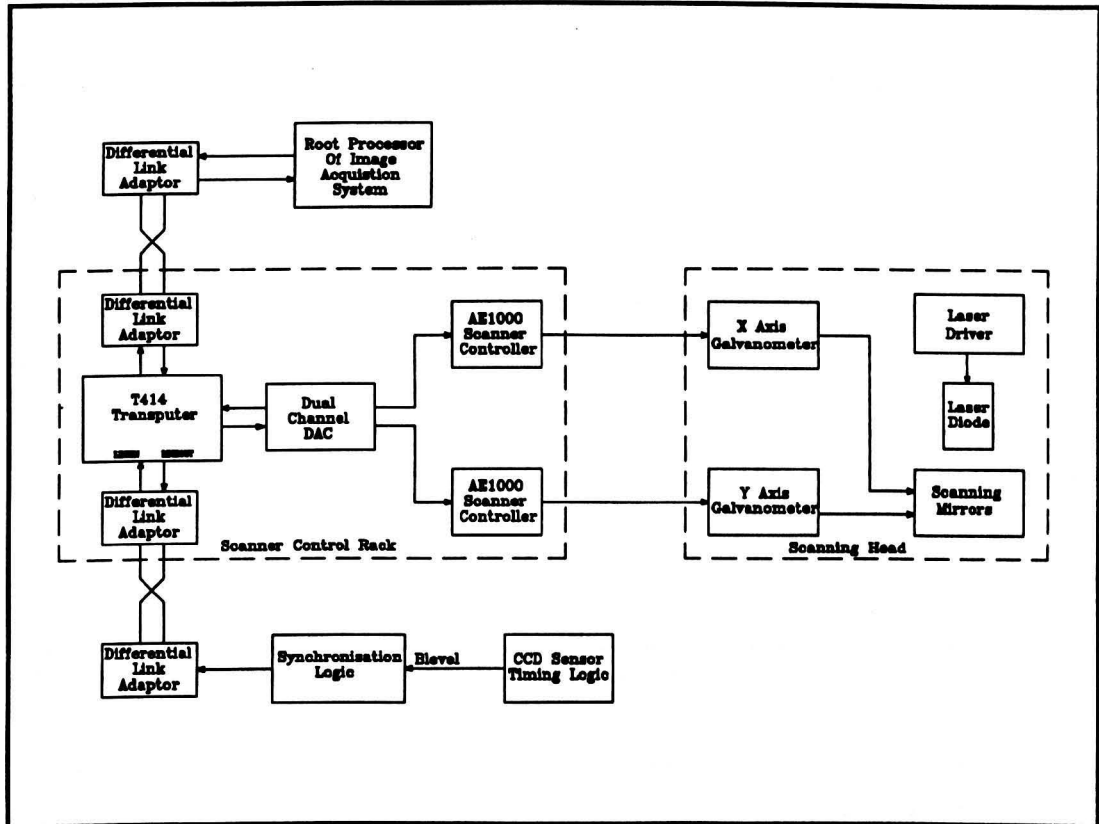


Figure 5.23 Overview of the Laser Scanning Subsystem

The scanning mechanism was under direct software control from a T414 transputer. The scanning motion could be synchronised to the exposure period of the CCD sensor using an input from the sensor timing logic.

Positional control of the scanning mirrors was achieved by the output of two 14 bit values from software to a dual channel digital to analogue converter (DAC). These values, converted to analogue voltage signals, controlled the angular position of the two mirrors mounted on precision galvanometer scanners. The geometry of the scanner was such that vertical and horizontal deflection of the beam were controlled by separate galvanometers.

5.10.1 Laser Source

The illumination source used by the image capture system was a Phillips CQL801 laser diode [34]. This device is a low cost laser source emitting visible red light at a wavelength of 675 nm. The maximum beam power of the laser diode was 15 mW, which proved to be ample for imaging even the most dense areas within a radiograph. Light from the laser source was focused into a beam using simple collimating optics.

A primary criteria during laser diode selection was the spectral sensitivity characteristics of the CCD sensors used for image acquisition. Spectral sensitivity curves, plotting photosensitivity against wavelength of incident light, for the two sensors used in this research are given in [1] and [17]. The wavelength corresponding to peak spectral response of each sensor is given in Table 5.8. By reference to this data it can be seen that the source wavelength of 675 nm is well within the operating range of both sensors.

Sensor	Peak Spectral Response Wavelength
TC104	700 nm
ILX503	500 nm

Table 5.8 Sensor Spectral Sensitivity Characteristics

The structure of the CQL801, illustrated in Figure 5.24, consists of two separate diodes, one a laser source and the other a photosensitive monitor diode. The monitor diode is used by the power supply to regulate the luminous output of the laser diode.

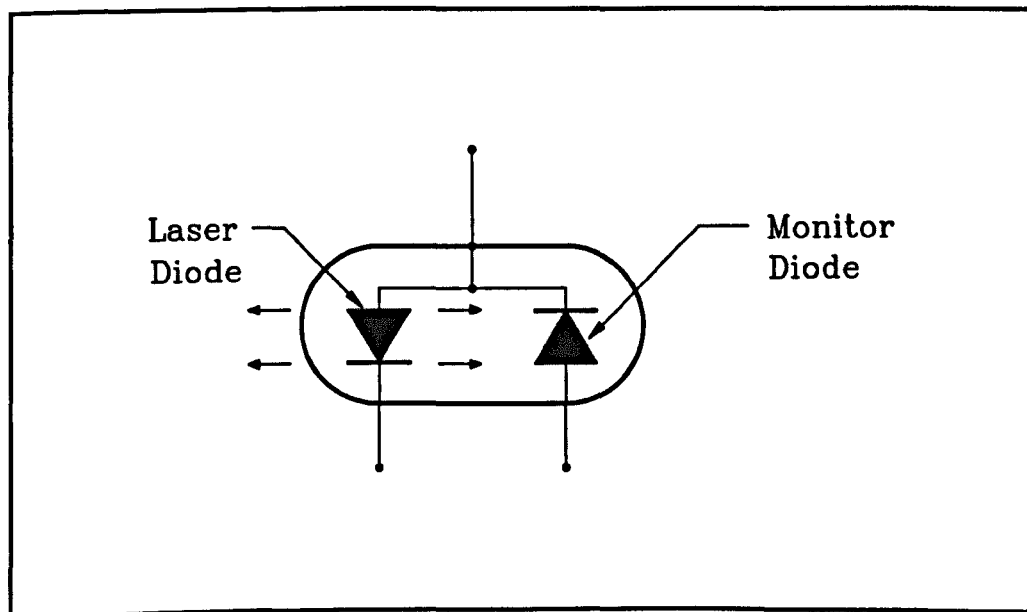


Figure 5.24 CQL801 Laser Diode Structure

The laser source was driven by a Hero laser power board, which allowed adjustment of output beam power via a potentiometer. A certain amount of intensity adjustment was necessary, depending on the density of the radiograph being imaged. The output beam power was measured by connecting an ammeter in line with the monitor diode.

5.10.2 Galvanometer Scanners

Positional control of the laser spot was by means of a dual axis galvanometer scanning system. This arrangement used two rotating mirrors to control the deflection angle of the light beam. A simplified geometry of the laser scanning head is shown in Figure 5.25. The rotation axes of the two mirrors were positioned so that the lower galvanometer controlled horizontal motion of the laser spot, and the upper galvanometer controlled vertical motion.

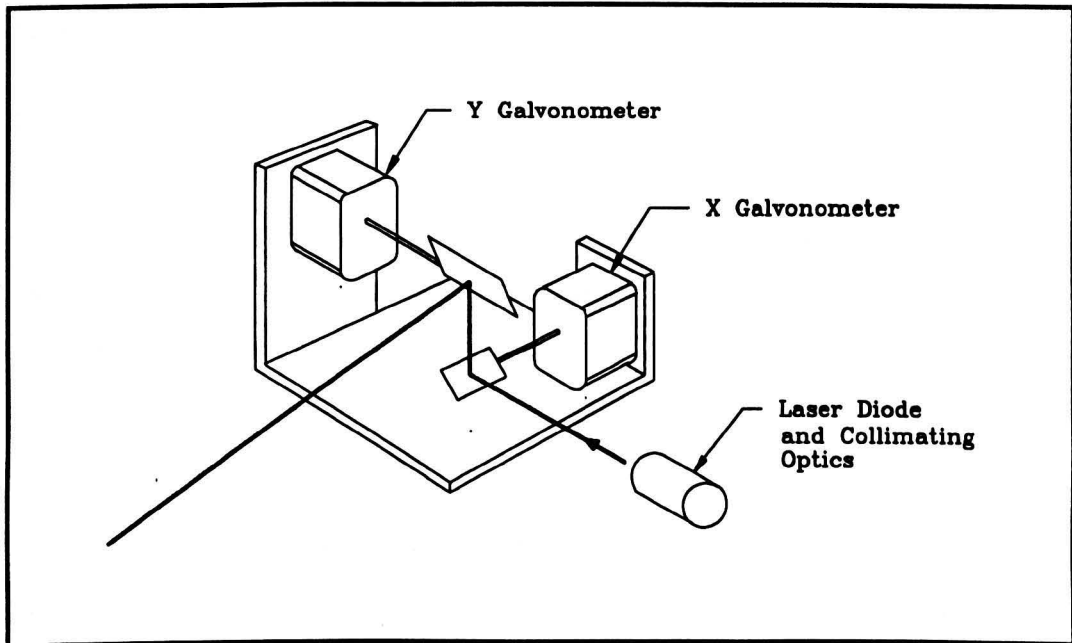


Figure 5.25 Laser Scanning Head Geometry

The scanner used two General Scanning G138DT galvanometers [35]. These galvanometers provide precise, high speed servo control of rotation over a maximum excursion angle of 76° . The low inertia and high torque of the system gave excellent dynamic response characteristics, allowing scan speeds of up to 200 scans per second. This corresponds to a full angle scan in around 5 ms, which is compatible with the short exposure times used by the CCD sensors.

5.10.3 AE1000 Scan Controllers

Control of the two galvanometer scanners was achieved using General Scanning AE1000 analogue controller boards [36]. These provided precise closed loop PID control of the galvanometers, with thermal regulation to maintain an operating temperature of 40°C .

The controllers respond to an analogue input signal proportional to a requested angular position. Full range excursion of the galvanometers corresponds to an input voltage swing of ± 5 volts. The controllers produce an error feedback signal by comparing this input request to a position output from the galvanometers.

In order to ensure uniform illumination over the full length of the CCD sensor, it was necessary to produce a highly uniform scan speed. This condition was achieved by careful calibration of the feedback control parameters as detailed in the AE1000 operating manual.

5.10.4 Dual Channel DAC

The scanning system used two 14 bit AD7538 digital to analogue converters [37] to generate the ± 5 volt input signals to the controller modules. The AD7538 is a monolithic CMOS¹ DAC converter which accepts TTL or CMOS logic level inputs and provides a corresponding current output. The maximum output current settling time of the AD7538 is 1.5 μ s, which was judged to be comparable to the response time of the galvanometer control system.

A dual channel, transputer controlled DAC module, designed at Liverpool University by Rui Yu [38], was constructed. This module used an Inmos CO11 link adaptor to interface between the DACs and a standard transputer serial link operating at 10 Mbits/s. The required bipolar analogue voltage was derived from the current output of the DACs using an application circuit given in the AD7538 data sheet.

The angular rotation of each galvanometer scanner was directly proportional to the digital input to the DAC. Hence, the resulting angular position, θ , of the galvanometer can be derived from the digital resolution of the system, N , the maximum rotation angle of the galvanometer, L , and the digital input value, k , as follows.

$$\theta = \frac{L}{2^N} k \quad (5.6)$$

¹ Complementary Metal Oxide Semiconductor

Hence, using a 14 bit DAC the minimum angular resolution of the system, θ_{\min} , was given by the equation.

$$\theta_{\min} = \frac{L}{16384}$$

Since only a restricted area was to be scanned by the laser spot, the full 76° rotation of the galvanometers was not required. As previously noted, the full scale input to the galvanometer control module was rated at ± 5 volts. In order to restrict the travel of the galvanometer and gain greater resolution over the operating range of the scanner, the DAC circuit was calibrated to give a full scale output of ± 3.75 V. This equated to an angular travel distance of approximately 57° . A calibration procedure for the DAC circuit is given in [36].

The DAC module implemented a simple communication protocol consisting of a three byte message sent by the controlling transputer process. The first byte of this message specified the DAC channel to be addressed, corresponding to either the vertical or horizontal galvanometer scanner. The next two bytes consisted of the 14 bit position data, least significant byte first. The DAC channels were designated as follows:

DAC Channel	Scan Axis
0	Horizontal
1	Vertical

Table 5.9 DAC Channel Protocol

5.10.5 Exposure Time Synchronisation Logic

As previously discussed, the output from a CCD photosensor is proportional to the integral of the incident light input during the exposure period. To ensure uniform image illumination, it was necessary to provide equal incident light power onto the film image during each exposure period. This condition was achieved by synchronising the laser scanner apparatus with the CCD sensor timing.

The start of each exposure period was indicated using the black reference pixel signal from the CCD sensor drive circuit. This signal was provided as an input to the

controlling transputer process via an Inmos serial link. The transputer used this input to trigger the laser scanning software at the start of each sensor integration period.

The synchronisation logic employed an Inmos CO11 link adaptor to allow direct communication with the transputer hardware over a standard Inmos serial link operating at 10 MBits/s. A single Altera EP600 PLD was used to implement the link protocol. The exposure time synchronisation circuit is shown in Figure A-7 of Appendix A.

The luminous input to the sensor system was controlled by varying the number of sweeps made by the laser spot during each exposure period. The maximum luminous input was determined by the number of sweeps possible during a single exposure period.

5.10.6 Differential Link Adapters

Standard Inmos links are designed for direct connection of transputer devices over transmission distances of less than 300 mm [39]. However, the transputer controlled scanning laser system was housed in a remote rack system, away from the main transputer network, so that direct connection was not possible.

In order to ensure data integrity over this increased transmission distance a differential link adaptor was used to buffer communications at the input and output of Inmos links connected to the laser scanning unit. Each differential link adaptor converted from the TTL level Inmos serial link to a standard RS422 differential link, capable of transmission of data over distances of up to 30 m [40].

Two differential links were connected to the laser scanning unit, one from the main transputer network and one from the exposure time synchronisation logic. The differential link arrangement is illustrated in Figure 5.26.

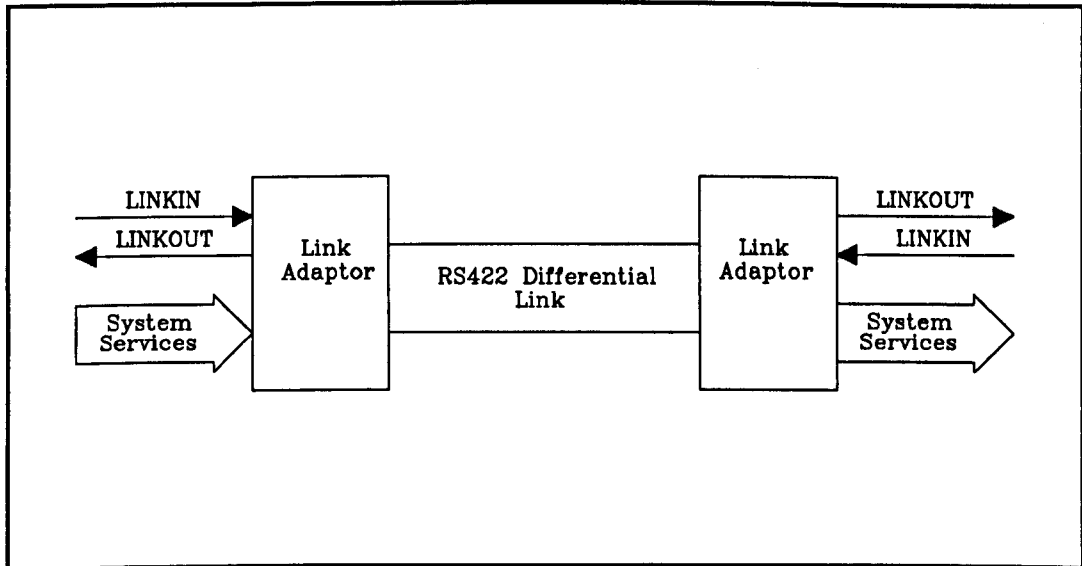


Figure 5.26 Extending Transmission Distance using Differential Link Adapters

5.11 Mechanical Apparatus

The experimental set up used for image acquisition involved moving the radiographic film across the face of the CCD sensor in discrete repeatable steps, whilst holding the sensor stationary with respect to the scanning laser light source. Building up images of an extended area required both horizontal and vertical motion of the film. To preserve the high spatial sampling resolution attained by the sensor array, it was necessary to ensure this scanning mechanism was exceedingly stable, so that errors were not introduced by vibration or misalignment of the apparatus. The apparatus also required some simple method of firmly mounting radiographic films in the scanning apparatus so that films could be changed quickly and easily.

This scanning mechanism required the design and construction of several pieces of mechanical apparatus. The main design criterion for this apparatus involved the reduction of any unwanted relative motion between components which might result in image blurring. The scanning apparatus was built in the mechanical workshop at the Electrical Engineering Department of Liverpool University. The assembled apparatus is shown schematically in Figure 5.27.

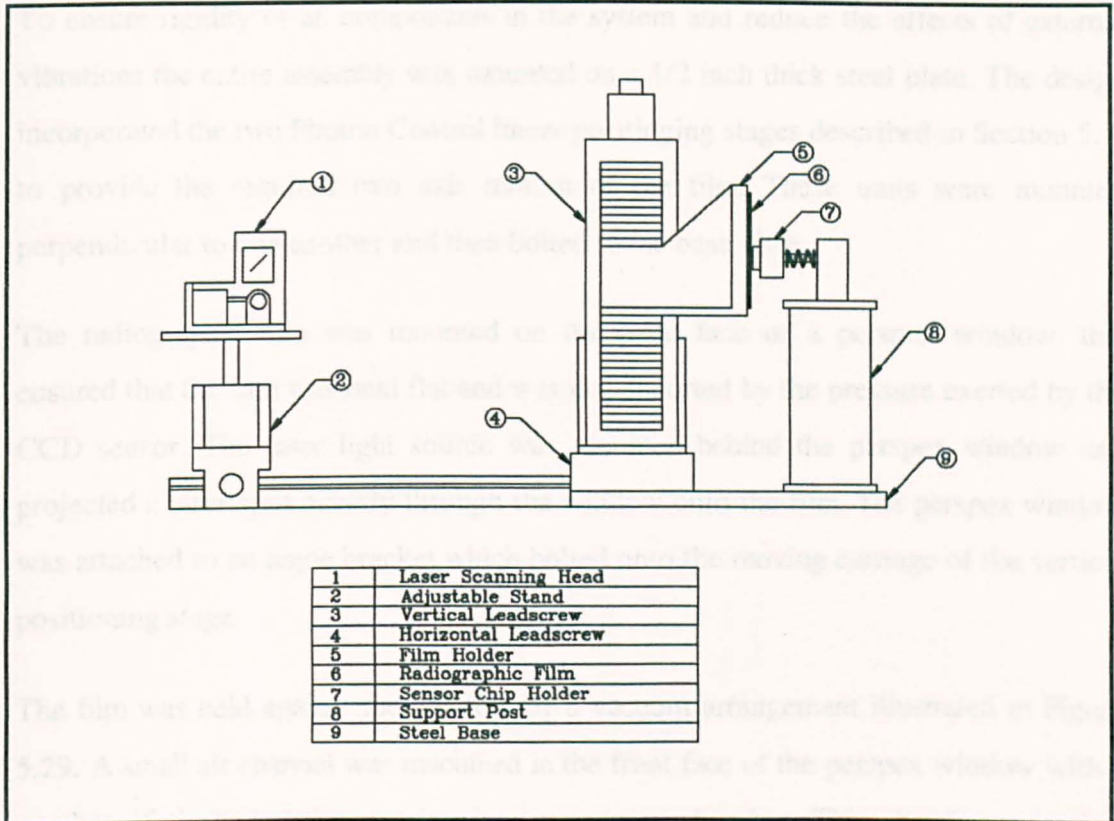


Figure 5.27 Schematic View of Experimental Apparatus

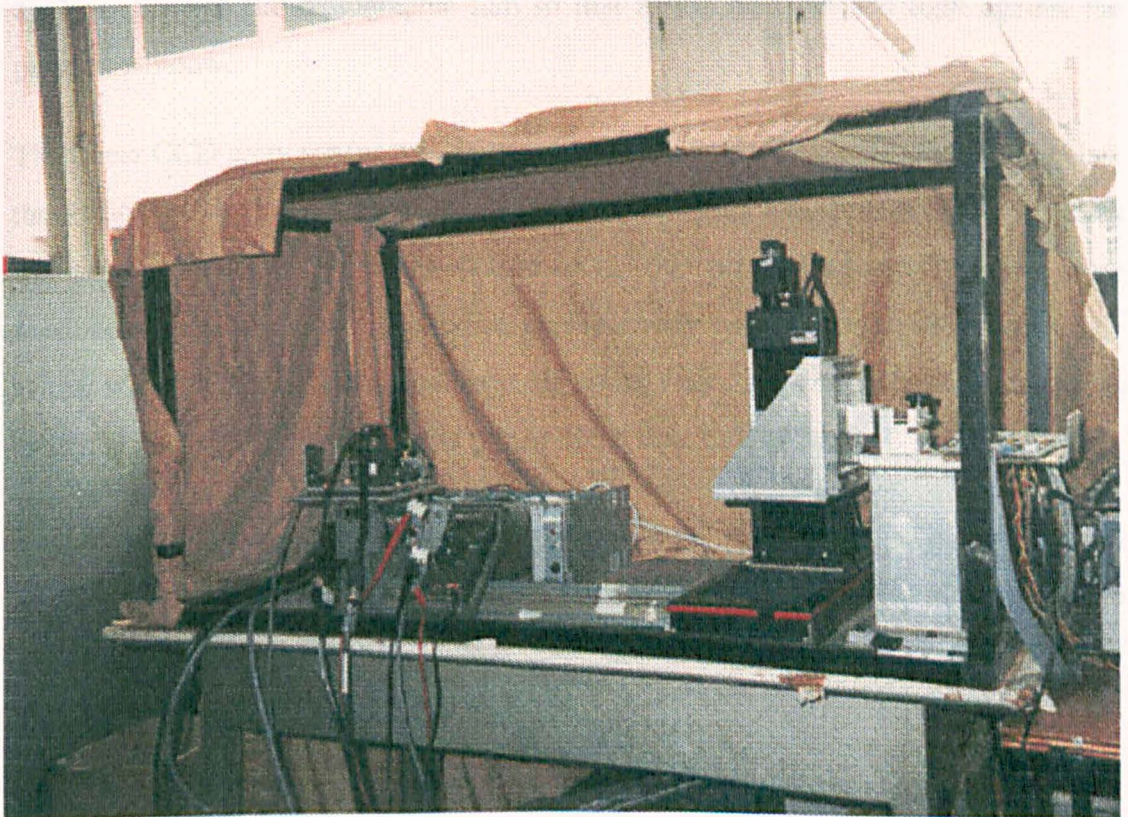


Figure 5.28 The Assembled Radiograph Digitisation Apparatus

To ensure rigidity of all components in the system and reduce the effects of external vibrations the entire assembly was mounted on a 1/2 inch thick steel plate. The design incorporated the two Photon Control linear positioning stages described in Section 5.9, to provide the required two axis motion of the film. These units were mounted perpendicular to one another and then bolted to the base plate

The radiographic film was mounted on the front face of a perspex window, this ensured that the film was held flat and was not distorted by the pressure exerted by the CCD sensor. The laser light source was mounted behind the perspex window and projected a laser spot directly through the window onto the film. The perspex window was attached to an angle bracket which bolted onto the moving carriage of the vertical positioning stage.

The film was held against the window by a vacuum arrangement illustrated in Figure 5.29. A small air channel was machined in the front face of the perspex window with a number of through holes passing into a vacuum chamber. This chamber was then connected to the inlet of a vacuum pump. This arrangement effectively evacuated the air from beneath the radiographic film so that it was held flat and tight against the perspex window.

The linear CCD array sensor was mounted in a chip holder assembly which sat against the radiographic film. A single chip holder mechanism was constructed with slots machined in to accept either the TC104 or ILX503 sensor package. The slots were machined so that the sensor protruding approximately 1 mm from the chip holder. The sensor was glued into the chip holder using epoxy resin. The chip holder assembly was mounted on a rigid post which was bolted to the steel base plate.

The sensor was held against the radiographic film by two compression springs. This arrangement ensured that film and sensor always remained in intimate contact despite any distortion of the window face. A locking mechanism allowed the sensor to be retracted during film replacement.

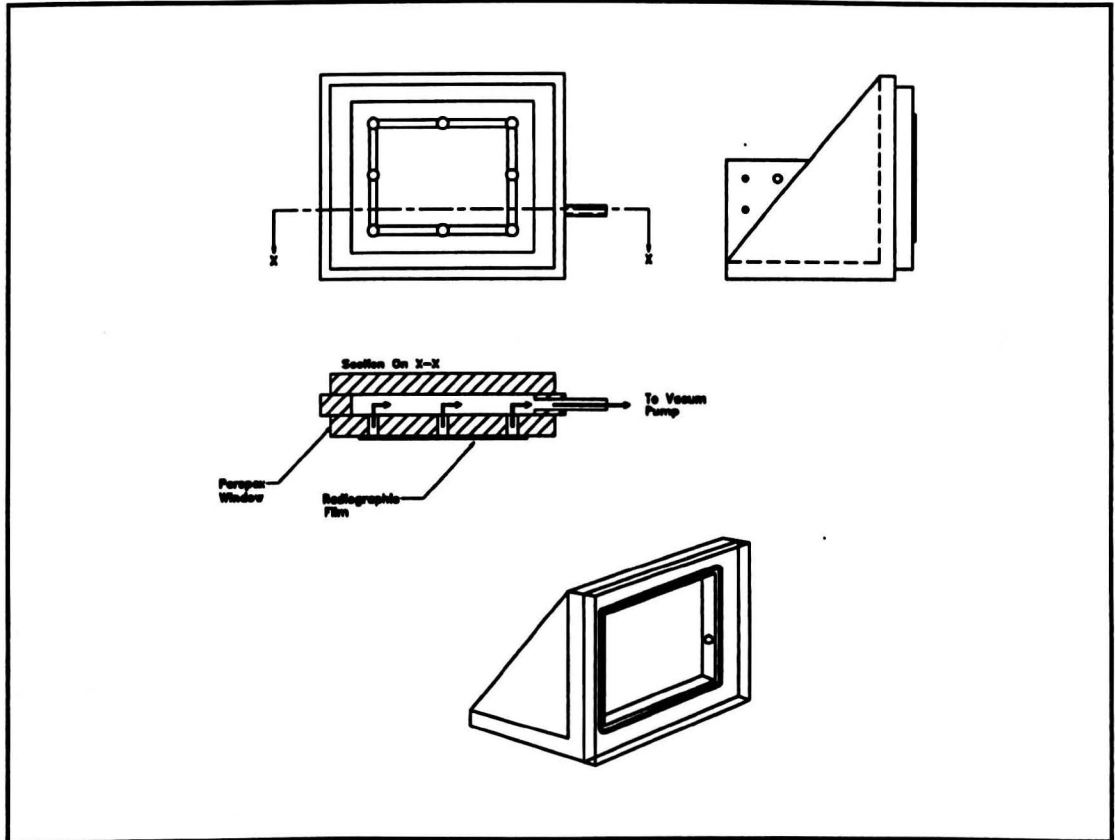


Figure 5.29 Perspex Window Assembly

The laser scanning head was mounted on an adjustable stand which allowed three dimensional motion to ensure uniform illumination of the sensor. The laser scanner was adjusted so that the undeflected laser beam intersected the plane of the CCD sensor elements.

5.12 System Construction and Layout

5.12.1 Main System Rack

It will be apparent from the foregoing chapter that the radiograph digitisation system consisted of many, diverse hardware components. The need for communication between different system components necessitated the passing of a large number of control signals.

In an effort to simplify component interconnection most of the electronic hardware constructed during this project was housed in a 19 inch rack mounting system. The rack accepted standard 3U Eurocard circuit boards, which plugged into a series of 96-way DIN41612 sockets. The reverse side of these sockets provided wire wrap pins which were used for connections between circuit boards in the rack. This arrangement

allowed easy retraction of individual circuit boards for modification without having to dismantle the system.

A large number of external signals required interfacing with the rack system, these included sensor output and control signals, exposure timing data lines, motor control signals, limit switch data from the linear positioning stages and transputer link and system service lines. Most of these connections were made via D-type sockets in the back of the rack cabinet.

The components mounted in the rack cabinet were :

- The TC104 Signal Processing Board
- The ADC and Interface Circuit Board
- The I/O Control and Motor Logic Board
- A T800 Transputer Module (TRAM)
- A 32 Channel Transputer Input Board
- A 32 Channel Transputer Output Board
- Associated Power Supplies

5.12.2 Sensor Hardware

The raw CCD sensor output video signal was extremely susceptible to noise pick-up due to cross talk with the high power drive clock lines. This phenomenon limited the practical connection distance between the CCD sensor and its associated drive hardware, typically to less than 20 cm. For this reason the CCD drive circuit hardware was placed as close to the sensor as possible and not mounted inside the rack. Connections between each sensor and its drive circuit were made using a short ribbon cable with standard DIL header sockets at either end.

As discussed in Section 5.4.4, the noise problems encountered with the Texas Instruments TC104 sensor necessitated the separation of sensor drive hardware and output signal processing circuits. These circuits were constructed on separate circuit boards and powered from separate supplies. The TC104 drive circuit was constructed on a PCB board, designed using Cadstar Schematic Capture and PCB Design software supplied by Redac [41].

The raw TC104 output signal was buffered using a transistor configured as an emitter follower before being connected to the signal processing circuit via a co-axial cable. To avoid unwanted transmission losses and noise pick-up this cable was kept as short as possible. All sensor control signals passed between the drive circuit and the rack were connected via opto-isolators to eliminate noise coupling between the circuits.

The TC104 signal processing circuit was constructed on a 19 inch Eurocard prototyping board, mounted in the rack cabinet. This board provided four separate power supply and ground planes in order to reduce noise effects. All sensitive analogue stages of the circuit were powered by a dedicated ± 14 V supply.

The basic design of the Sony ILX503 sensor eliminated much of the noise pick-up associated with high power clock lines by the use of on-chip clock drivers. This design reduced the problems associated with noise signal coupling between the sensor drive circuit and the sensitive signal processing circuit. Consequently it was decided to build both the sensor drive and signal processing circuits on a single Eurocard prototyping board.

5.12.3 Laser Scanning Rack

The scanning laser system was designed as a self contained unit controlled by an on-board transputer. A separate rack was constructed to contain all the laser scanning sub-system components. The control inputs into the laser scanning rack were provided via two Inmos transputer links, one from the main transputer system and one from the exposure timing synchronisation circuit. All the components in the laser scanning unit were of standard 3U Eurocard format with DIN41612 96 way edge connectors. The laser scanning sub-system contained the following components :

- A Differential Link Adaptor
- A T414 Transputer Module
- A Dual Channel DAC
- Two AE1000 Scan Control Modules
- Power Supplies

5.12.4 Host PC

User interaction with the transputer system was achieved via a host PC. Two of the available expansion slots in the PC motherboard were occupied by transputer boards, these were :

- A T800 Transputer Module
- A T805 Transputer-Based MosaiQ Framestore

5.13 Summary

This chapter gives a detailed account of the design and construction of the radiographic film digitisation system developed during the course of this research. The system used a linear array CCD image sensor with very high pixel density, to ensure the required image resolution was attained. A direct contact scanning method was used in order to eliminate the need for complex optics and thereby minimise the associated image distortion. The image was built up by moving a film image across the face of the sensor in discrete steps using a two axis positioning system. Image illumination was provided by a scanning laser spot mechanism projected onto the face of the sensor, through the radiographic film. The analogue output signal from the CCD array was digitised with a resolution of 12 bits/pixel. The resulting image data was displayed on a high resolution 19 inch monitor using a Qintek MosaiQ transputer framestore. A host PC was used for all user interaction with the image acquisition hardware and the hard disk of the PC was used to store all digitised image data.

References

- 1 "Texas Instruments Optoelectronics and Image Sensors Data Book", Texas Instruments Corporation, 1992, pp 65-76.
- 2 Jaroslav Hynecsek, "Virtual Phase Technology: A New Approach to Fabrication of Large Area CCDs, IEEE Trans. Electron Devices **ED-28**, June 1981, pp 483-489.
- 3 Altera EP600 Series EPLDs, Altera Data Book, Altera Ltd 1990, pp 49-58.
- 4 "DS0026 5 MHz Two Phase MOS Clock Drivers Data Sheet", National Semiconductor, 1989.
- 5 B Siegel, "Applying Modern Clock Drivers to MOS Memories", National Semiconductor Application Note 76, October 1975.

- 6 "Linear CCD Operation at 10 MHz", Texas Instruments Technical Note, 1989
- 7 D. F. Barbe & S. B. Campana, "Advances in Image Pick Up and Display Vol. 3", Academic Press, 1977, pp 185-194.
- 8 D. F. Barbe, "Charge Coupled Devices", Proc IEEE **63**, 1975, pp38.
- 9 AD847 High Speed, Monolithic OpAmp Data, Analog Devices Linear Products Data Book, p 2-319, 1991.
- 10 "CCD Output Signal Processing", Texas Instruments Technical Note, 1990.
- 11 TL1591 Sample and Hold Circuit for CCD Imagers Data Sheet, Texas Instruments Ltd, 1989.
- 12 P. Bowron & F Stephenson, "Active Filters for Communications and Instrumentation", McGraw Hill Ltd, 1979, pp 163-179.
- 13 NE531 High Slew Rate Operational Amplifier, Phillips Semiconductor Linear Products Data Book, 1990.
- 14 LF398 Monolithic Sample and Hold Circuit Data Sheet, National Semiconductor Ltd.
- 15 LM11C Operational Amplifier Data Sheet, National Semiconductor Ltd.
- 16 Brian Jones, "Electronics for Experimentation and Research", Prentice Hall International, 1986, pp 126-145.
- 17 "Sony CCD Cameras and Peripherals Data Book", Sony Semiconductor, 1992, pp 913-922.
- 18 P Horowitz and W Hill, "The Art of Electronics", Cambridge University Press, 1989, pp150-151.
- 19 T. Murphy, "Correlated Double Sampling", Fairchild Weston Application Note, 1987.
- 20 "Sony Linear Sensor Application Notes", Sony Semiconductor, November 1992.
- 21 CXA1439M Monolithic CDS IC for CCD Applications, Sony CCD Cameras and Peripherals Data Book, 1992, pp 713-720.
- 22 AD1671 12 bit Monolithic A/D Converter, Analog Devices Data Converter Reference Manual Vol. 2, 1992.
- 23 The Transputer Data Book, Inmos Ltd, 1989 , pp 39-40.
- 24 IMS C011 Link Adaptor-Engineering Data, Inmos Ltd 1989.
- 25 MosaiQ Frame Grabber Board User Manual, Quintek Ltd, 1990.

- 26 C-6922 16" Colour Monitor User Manual, Mitsubishi Ltd.
- 27 Photon Control Linear Positioning Product Manual, Photon Control Ltd, 1994.
- 28 A Series Microstep Motor Driver Manual, Compumotor Ltd.
- 29 A57-83 Stepper Motor Manual, Compumotor Ltd.
- 30 1054 Stepper Motor Drive Instruction Manual, Digiplan Ltd.
- 31 Altera EP310 Series EPLDs, Altera Data Book, Altera Ltd, 1990, pp 33-40.
- 32 Light Sources and Monochromators Catalogue, Oriel Corporation, 1992.
- 33 J. Bellis, "Lasers: Operation, Equipment, Applications and Design, McGraw Hill Books, 1980, pp 25-28.
- 34 CQL801 15 mW Visible Laser Diode Data Sheet, Phillips Ltd, 1994.
- 35 G138DT Galvanometer Scanner User Manual, General Scanning Ltd, 1991.
- 36 AE1000 Scanner Control User Manual, General Scanning Ltd, 1991.
- 37 AD7538 14 bit DAC, Analog Devices Data Converter Reference Manual Vol. 1, 1992.
- 38 R. Yu, "A 3 Dimensional Vision System for Remotely Operated Vehicles", PhD Thesis, University of Liverpool, August 1994.
- 39 "Transputer Technical Notes", Inmos Ltd, 1989, pp26-41.
- 40 EIA Standard EIA-STD-RS422.
- 41 Cadstar PCB Design Software Version 7 User Manual, Racal-Redac Ltd, 1992.

6. Sensor System Performance

6.1 Introduction

This chapter deals with performance tests carried out on the two CCD sensors evaluated during this project, the Texas Instruments TC104 and the Sony ILX503. These tests included assessment of thermally generated dark signal and the overall noise signal levels. This is followed by a short discussion of some relevant theory concerning the measurement of resolving power in optical systems. These ideas are then used to assess certain system transfer characteristics of the image acquisition system, including line spread function and modulation transfer function. Finally the relative merits of the two sensor systems are discussed, leading to the selection of the Sony ILX503 as the sensor used in the final prototype of the image acquisition system.

6.2 Dark Signal and DSNU

As discussed in Chapter 3 a proportion of the output signal from a CCD image sensor will always be the result of thermally generated dark signal charge accumulating within the photosites. To take into account the effects of this dark signal charge, both of the sensors used during this project provided a number of black reference pixels which were covered by an opaque metal layer. The output signal from these pixels was therefore the result of thermally generated charge alone. As previously discussed, these black reference pixels were used by the black clamp circuit to calibrate out the effects of dark signal charge and ensure zero output under zero illumination conditions.

In order to evaluate the effects of dark signal charge, tests were carried out on the sensor system under zero illumination conditions. These tests involved integrating the dark signal charge over a long exposure period and measuring dark signal output as a function of exposure time. The exposure time was varied between the limits set by the drive circuit logic, these limits may be determined using equations (5.1) and (5.2). For an operating output pixel rate of 625 kHz, the maximum exposure time, T_{MAX} , and minimum exposure time, T_{MIN} , are given below.

$$T_{MAX} = \frac{2^7 \times 512}{625 \times 10^3} = 104 \text{ms}$$

$$T_{\text{MIN}} = \frac{7 \times 512}{625 \times 10^3} = 5.7 \text{ms}$$

The grey scale level (gsl) output of each pixel within the sensor array was recorded under dark conditions for exposure times of T_{MAX} and T_{MIN} . Dark signal output was then expressed as the average digital output taken over the entire sensor array. Consider an N element array with individual dark current output values, I_0, I_1, \dots, I_{N-1} . The average dark current, I_{avg} is then given by :

$$I_{\text{avg}} = \frac{\sum_{i=1}^N I_i}{N} \quad (6.1)$$

Integration Time	Average Output Grey Scale Level	
	TC104	ILX503
104 ms	0.1742 gsl	0.061 gsl
5.7 ms	0.1431 gsl	0.084 gsl

Table 6.1 Average Dark Signal Characteristics

It can be seen from these results that the effects of thermally generated dark current are minimal and may be neglected, since on average they amount to much less than a single grey scale level in the output signal. The results gained using the TC104 sensor show that increasing the exposure time by a factor of eighteen produced only a 0.0311 gsl difference at the output, which represents around 0.00075 % of full scale output. The average dark signal output from the Sony ILX503 actually decreased for the longer exposure time, which indicates that some other noise source obscured the effects of thermally generated dark current.

The dark signal voltage produced by each pixel within a sensor array varies slightly due to individual photodiode characteristics. This effect is characterised by the dark signal nonuniformity (DSNU). The effect of DSNU is to produce random fluctuations of individual black level pixel outputs, around a mean black level output. As discussed above, the black clamp circuit effectively set this average dark signal output to zero.

During digitisation, any pixel with output voltage below this average dark signal will be quantised to zero, so that some of the effect of DSNU is eradicated.

In order to measure DSNU a small offset voltage was applied to the black clamped output signal before digitisation. This had the effect of increasing the average black level output so that both positive and negative dark signal fluctuations were recorded. The dark signal nonuniformity was then calculated as the maximum difference in grey level output measured across all the sensor elements. This principle is illustrated in Figure 6.1.

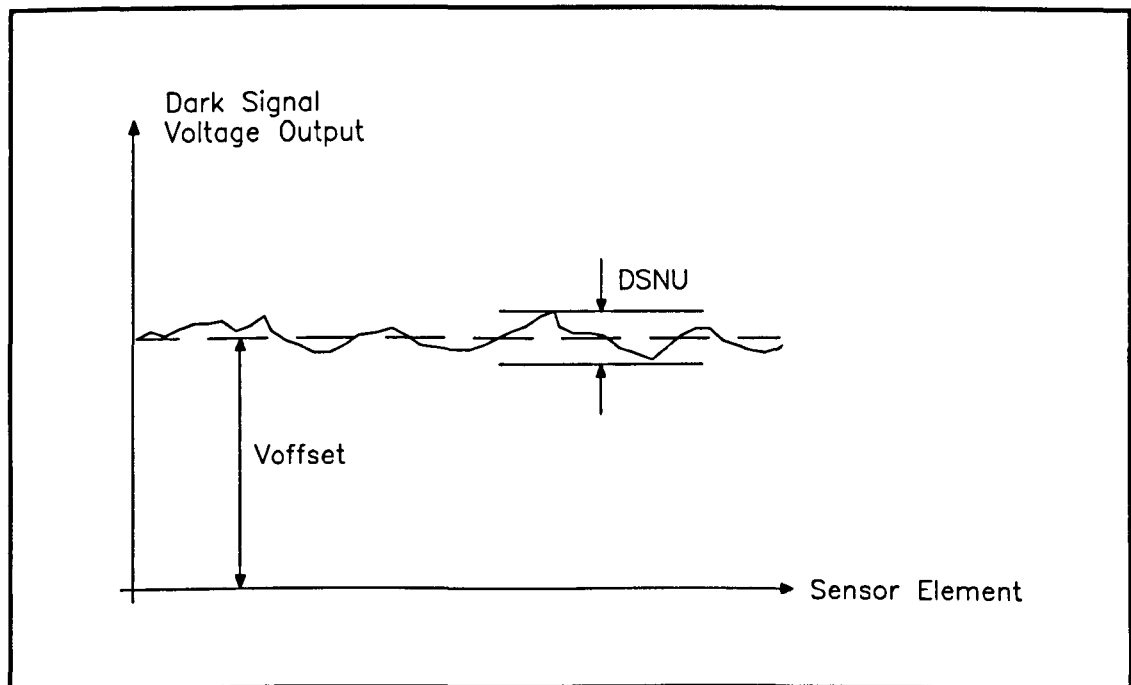


Figure 6.1 Measuring Dark Signal Nonuniformity

The results of these tests on dark signal nonuniformity are given in Table 6.2, which records DSNU in terms of the grey scale level (gsl) outputs produced by the image acquisition system. It can be seen that DSNU causes very little effect on the output signal level, typically less than 1 % of the full scale intensity output. Hence the effects of DSNU on output image quality are negligible.

Sensor	Dark Signal Nonuniformity	Error Due to DSNU
TC104	32 gsl	0.8 %
ILX503	15 gsl	0.4 %

Table 6.2 Dark Signal Nonuniformity Characteristics

6.3 Sensor System Noise

A major factor affecting system performance was signal noise produced by the CCD sensors and their associated hardware. This noise tended to obscure legitimate pixel data and produce a random variation of image contrast data.

The potential sources of noise within a circuit are extremely varied and in general many factors each contribute some component to the overall noise signal. Noise sources may be divided into two groups, those generated internally within the electrical circuit, and those due to interference from external sources. Internal noise sources may be due to the intrinsic properties of circuit elements or may be picked up by some coupling mechanism from signals generated in other parts of the circuit. External noise sources may include such things as equipment generating electromagnetic radiation, mains power supply fluctuations, changes in ambient light and temperature conditions and mechanical vibrations. Truly random or "white" noise exhibits a uniform spectral distribution, so that the noise power density is the same at all frequencies. However, many noise sources within a circuit generate periodic signals at well defined frequencies.

Several sources of reducible noise were identified within the image acquisition system, and much effort during this project was concentrated on the reduction of signal noise levels. A major contributor to signal noise was feedthrough produced by the high power drive clocks used to operate the CCD sensor shift registers, steps taken to minimise this problem are discussed at length in section 5.4.2. Circuit construction was also a key factor in the reduction of noise, details of this are discussed at length in Chapter 5. It was found that a considerable amount of noise was generated by electromagnetic interference from the energised coils of the two stepper motors. This problem was reduced to some degree by the provision of common ground connections

between the casings of both motors, the steel base of the scanning mechanism assembly and the main power supply ground.

The pixel output from each element of the sensor array consisted of some signal voltage, with an associated noise component overlaying it. During digitisation this analogue signal was quantised to a discrete grey level value by the ADC. Digitisation was achieved using a 12 bit ADC, accepting analogue inputs within the range 0 to 5 V. Equation 5.4 gives the threshold voltage, V_{TH} , necessary to cause an output change of one least significant bit from an N bit ADC with input voltage range between V_{MAX} and V_{MIN} .

$$V_{TH} = \frac{1}{2} \left[\frac{V_{MAX} - V_{MIN}}{2^N} \right] = \frac{1}{2} \left[\frac{5 - 0}{2^{12}} \right] = 0.6\text{mV}$$

Any noise signal in excess of this threshold voltage would cause spurious codes to be generated by the ADC causing a consequent loss of digitisation fidelity. In practice reducing noise levels below V_{TH} proved to be difficult since inherent noise sources such as power supply fluctuations and component characteristics set a limit on attainable signal to noise ratio.

The effects of noise on the output video signal from the image capture system were assessed in order to estimate the attainable digitisation fidelity of the resultant images. This simple analysis considered the imaging system as a “black box” rather than concentrating on individual noise sources.

Noise sources caused the grey level output from each sensor element to fluctuate randomly with time around some mean value. The effect of signal noise on both the sensors evaluated during this project are illustrated in Figure 6.2 and Figure 6.3. These examples show graphics screen outputs produced by the system software plotting two data lines scanned consecutively, under constant lighting conditions. In the absence of signal noise these two data lines would be identical. It can be seen from these plots that in general the Texas Instruments TC104 sensor produced a higher level of noise than the Sony ILX503. Reasons for the poor performance of the TC104 will be discussed later in this chapter.

Although the intensity plots shown in Figure 6.2 and Figure 6.3 illustrate the effect of signal noise they do not give a meaningful measure of noise intensity. Since signal noise is a time variant property, it was necessary to measure the variation of signal output over an extended period, in order to fully assess the effects of noise. This was done by monitoring the output signal under uniform illumination conditions over a total of 500 scans of the sensor array. In this way any variation in signal output between individual data lines was directly attributable to noise.

To give some idea of signal noise levels, the peak to peak noise variation was assessed for each element within the sensor array. Consider a single element, i , in an N element sensor array. If the maximum intensity recorded during the test period for that element was \max_i , and the minimum intensity was \min_i , then the peak to peak noise level, P_{toP}_i , of the given pixel element is given by :

$$P_{toP}_i = \max_i - \min_i$$

Table 6.3 summarises the peak to peak noise characteristics of the TC104 and ILX503 image sensors, measured over 500 successive scans. The peak noise variation refers to the maximum signal noise fluctuation, measured across the entire length of each sensor array. The mean noise level was calculated as the average value of peak to peak noise taken across the full sensor array. For an N element sensor array the mean noise level is then given by the following equation.

$$\text{Mean Noise Level} = \frac{\sum_{i=1}^N P_{toP}_i}{N}$$

Sensor	Peak Noise Variation	Mean Noise
TC104	110 gsl	85.6 gsl
ILX503	32 gsl	13.0 gsl

Table 6.3 Peak to Peak Sensor Noise Measured over 500 Scans

Obviously, the peak to peak noise measures the maximum fluctuation in grey level output, which may not be representative of the overall noise level. In practice the

average noise level will be significantly lower than the peak to peak level, and is often represented as the root mean square (rms) output. Assuming that noise is truly random, the rms noise level may be estimated as approximately one eighth of the peak to peak noise [1].

Figure 6.4 and Figure 6.5 show the variation of peak to peak noise level across the length of the TC104 and ILX503 sensors respectively. It can be seen that peak to peak noise in the ILX503 appears to vary randomly about a mean of approximately 13 grey scale levels. From this it can be observed that the noise produced by this device appears to be invariant with pixel position. In contrast, the TC104 noise characteristics exhibit a definite variation with sensor element position. Further investigation of this phenomenon revealed that noise output from the TC104 was a function of incident light intensity and the measured trend was due to slight variations in overall illumination level.

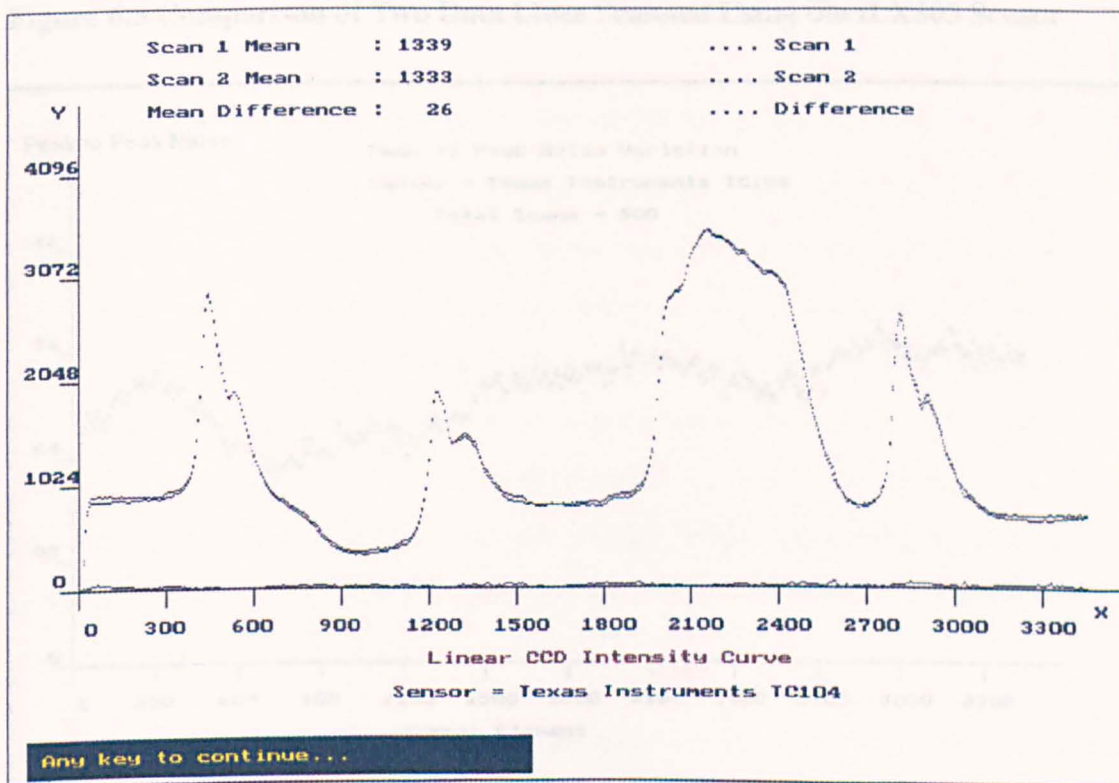


Figure 6.2 Comparison of Two Data Lines Scanned Using the TC104 Sensor

Figure 6.4 Peak to Peak Noise Characteristics of the TC104 Sensor

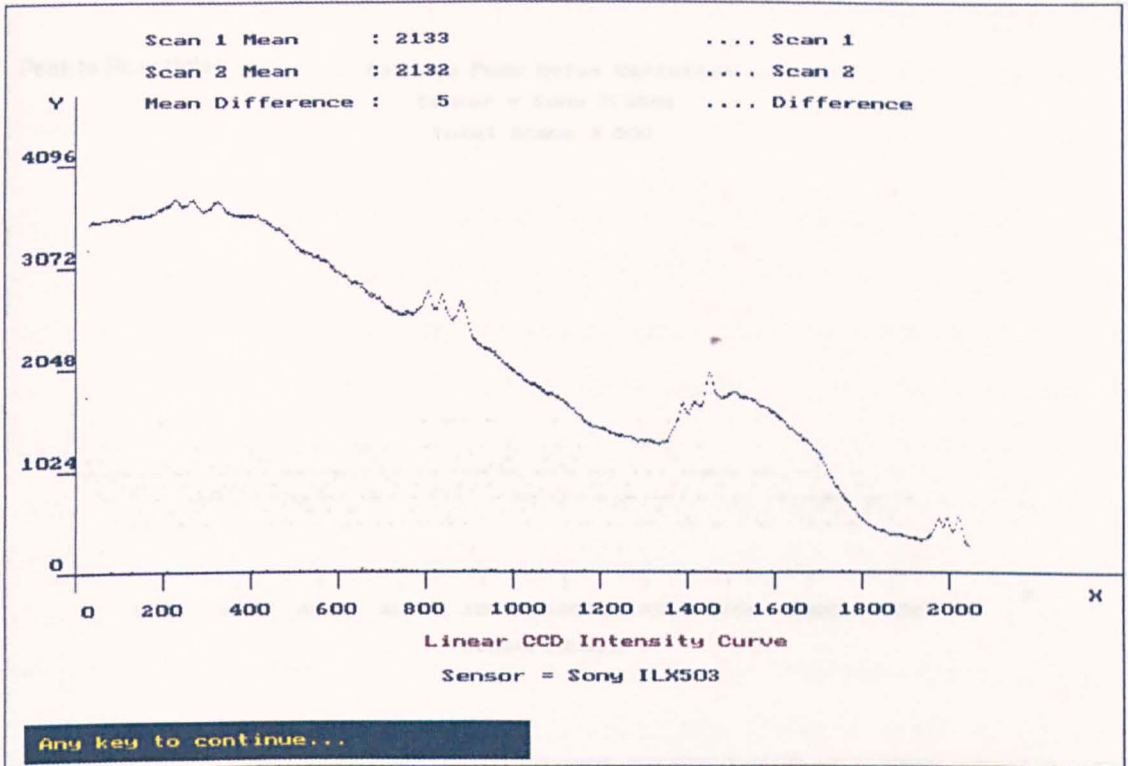


Figure 6.3 Comparison of Two Data Lines Scanned Using the ILX503 Sensor

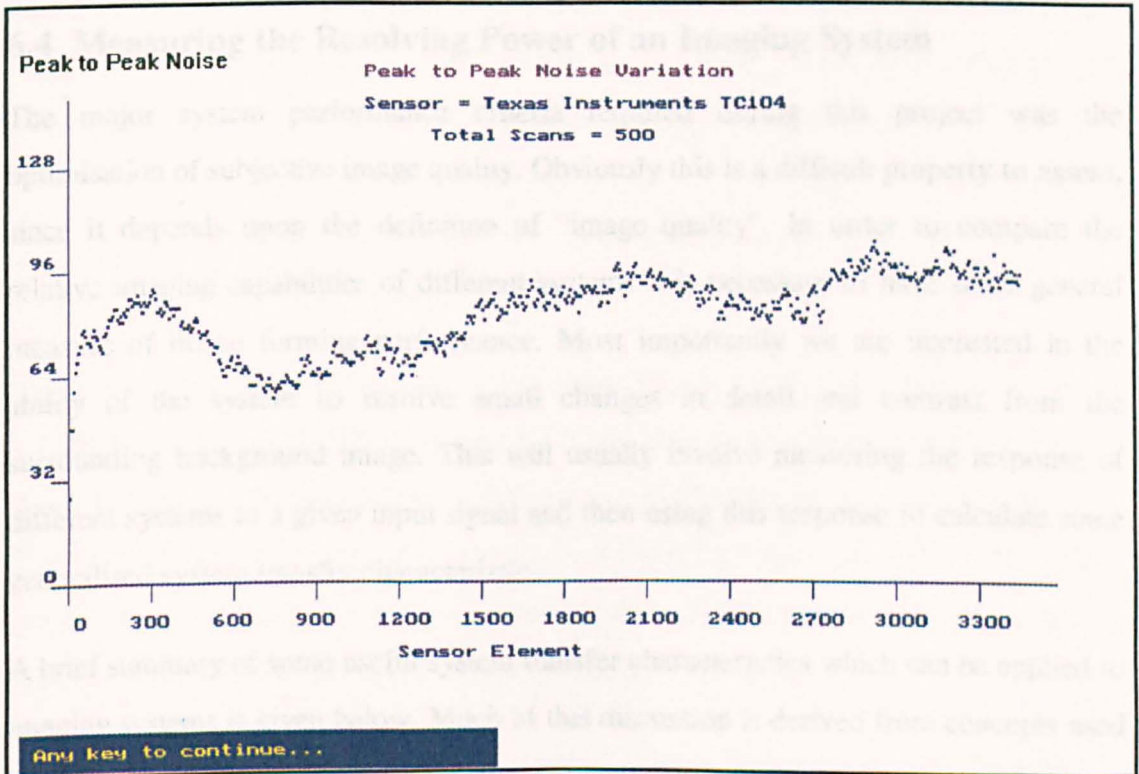


Figure 6.4 Peak to Peak Noise Characteristics of the TC104 Sensor

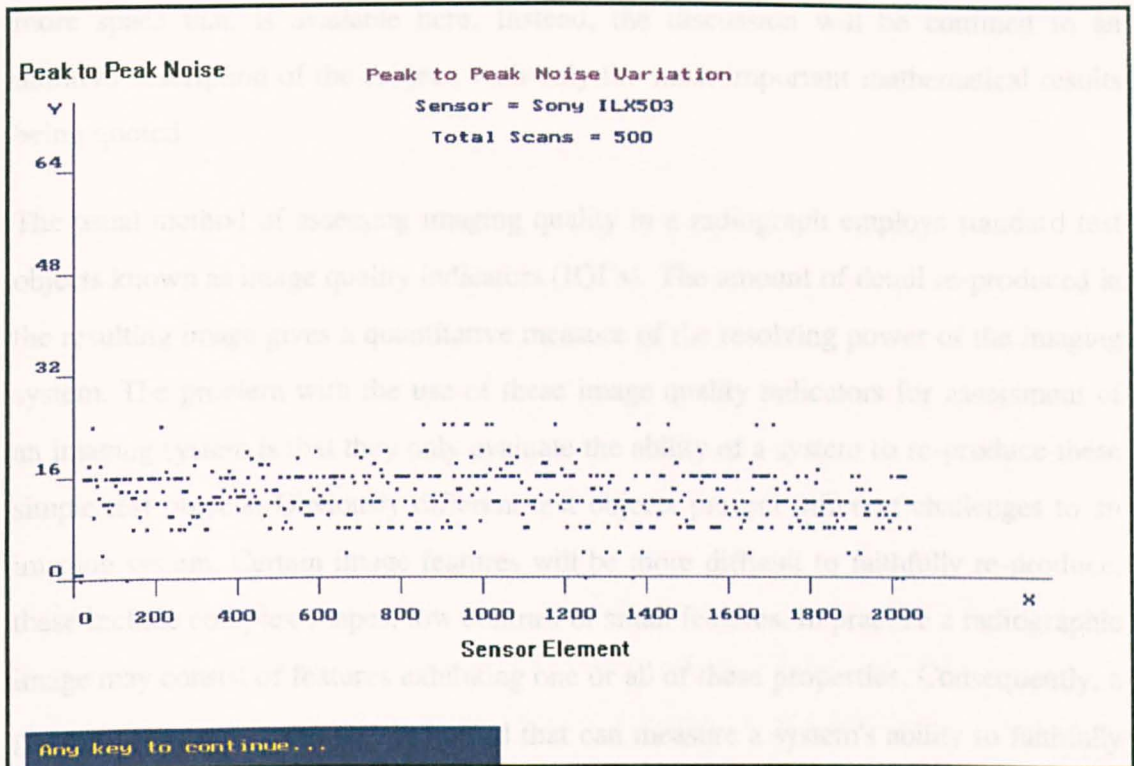


Figure 6.5 Peak to Peak Noise Characteristics of the ILX503 Sensor

6.4 Measuring the Resolving Power of an Imaging System

The major system performance criteria required during this project was the optimisation of subjective image quality. Obviously this is a difficult property to assess, since it depends upon the definition of “image quality”. In order to compare the relative imaging capabilities of different systems it is necessary to have some general measure of image forming performance. Most importantly we are interested in the ability of the system to resolve small changes in detail and contrast from the surrounding background image. This will usually involve measuring the response of different systems to a given input signal and then using this response to calculate some generalised system transfer characteristic.

A brief summary of some useful system transfer characteristics which can be applied to imaging systems is given below. Much of this discussion is derived from concepts used in general communication theory, which are well documented in the relevant literature [2] [3]. These ideas are then applied to the particular case of an imaging system, using theory developed for the assessment of optical systems and radiographic imaging apparatus [4] [5]. A rigorous mathematical proof of these concepts would require

more space than is available here. Instead, the discussion will be confined to an intuitive description of the subject, with only the most important mathematical results being quoted.

The usual method of assessing imaging quality in a radiograph employs standard test objects known as image quality indicators (IQI's). The amount of detail re-produced in the resulting image gives a quantitative measure of the resolving power of the imaging system. The problem with the use of these image quality indicators for assessment of an imaging system is that they only evaluate the ability of a system to re-produce these simple test objects. Obviously different test objects present different challenges to an imaging system. Certain image features will be more difficult to faithfully re-produce, these include complex shapes, low contrast or small features. In practice a radiographic image may consist of features exhibiting one or all of these properties. Consequently, a general system characteristic is needed that can measure a system's ability to faithfully image an arbitrarily shaped object.

The first step in producing this generalised system characteristic is to reduce the complex internal behaviour of the system to a simple "black box". Rather than considering the mechanics of the signal passing through the system we are only interested in the ability of the system to convert an input signal to some resultant output. The system transfer characteristic must be able to define the resultant output signal produced by any given input.

In this analysis it should be noted that the imaging system under consideration converts a radiant energy distribution in the object plane into a two dimensional pixel array in the image plane. The object plane is defined as a flat surface co-incident with the front face of the linear CCD array as it scans across the radiographic film. The image plane, rather than being a physical surface, is defined by a digital array of pixel values held in system memory and displayed on a video monitor.

In any practical imaging system there will be a certain degree of blurring between the object and image. This effect tends to degrade the reproduction of small low contrast features. In the radiograph digitisation system developed during this project, factors such as optical imperfections, external noise pick up, photosite blooming, charge

packet transfer loss and finite spatial sampling limitations all lead to this spreading of intensity data across the image plane.

6.4.1 The Point Spread and Line Spread Functions

The simplest theoretical test object which can be applied to an imaging system is a point source of radiant energy. In an ideal system the image formed of a point source object will itself be a point in the image plane. Practical systems will exhibit a degree of image degradation such that the sharp input energy point in the object plane appears as an energy distribution. This resultant intensity distribution is referred to as the point spread function.

In order to simplify the following analysis certain assumptions will be made about the system under consideration. Firstly, it was noted that, within the operating range of a CCD sensor, the voltage output varies linearly with input light intensity, as discussed in chapter 3. This property may be defined as constituting a linear system in which the effects of separate intensity components in the object are additive in the resulting image. It was also assumed that if an object were moved across the object plane the shape of the resultant image would be unchanged. This property is known as isoplanism. The sensor arrays used in this project may reasonably be considered to be continuous with no significant gaps between sensor elements, and any spatial quantization error will typically be less than half the sensor element width, so that the assumption of isoplanism was considered to be reasonable.

In principal any two dimensional image may be represented as the result of an infinite number of point sources radiating at different intensities. Each of these point sources in object space will produce a corresponding point spread intensity distribution in the image. The output image may then be considered as the aggregate of all these point spread intensity distributions. If the point spread function is position invariant and the response of the system is linear then the resultant image is then given by summing these individual point spread functions taken over the entire image plane. This process of multiplying each point in object space by the system point spread function and then summing them together to produce the output intensity image is known as convolution. This convolution property illustrates that the point spread function is

general system transfer function, since it can be used to fully describe the conversion of an arbitrary input intensity distribution into a resultant intensity distribution in object space.

Consider the imaging system shown in Figure 6.6, which converts some intensity distribution in object space, $i(\xi, \eta)$, into a corresponding intensity distribution, $I(x, y)$, in image space, where (ξ, η) and (x, y) are co-ordinates in object and image space respectively. If the imaging system has a characteristic point spread function, $A(x, y)$, then the resultant output image intensity distribution may be calculated using the convolution integral:

$$I(x, y) = \int \int_{-\infty}^{\infty} i(\xi, \eta) A(x - \xi, y - \eta) d\xi d\eta \quad (6.2)$$

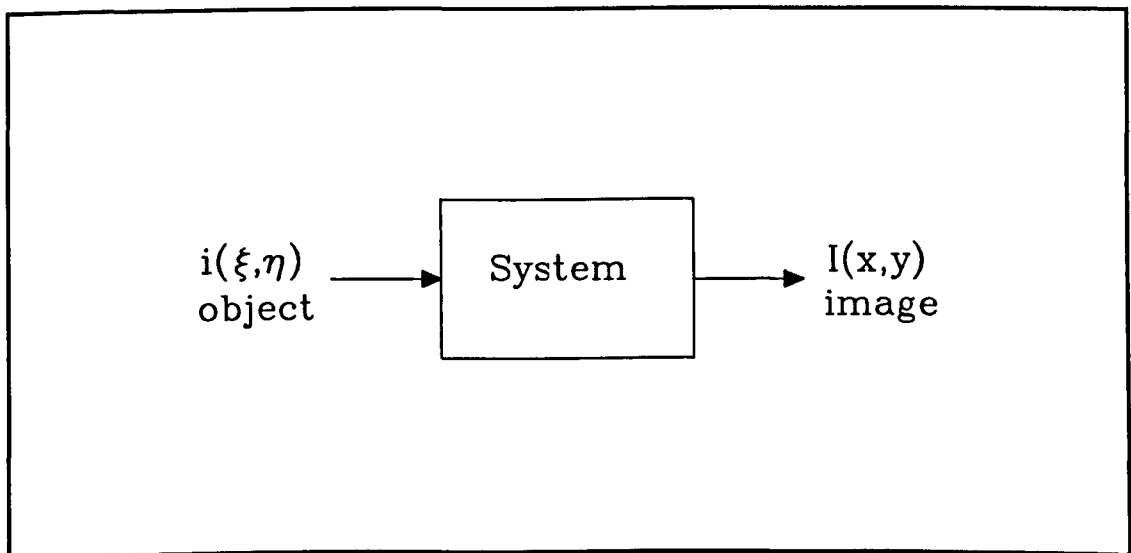


Figure 6.6 System Transfer Representation of an Imaging System

In general the point spread function will be a two dimensional intensity distribution which is not be symmetrical about its central point. Where a system does exhibit a symmetrical point spread function it is known as an isotropic system. For the special case of an isotropic system the point spread function is identical at all orientations around its central point and may be fully defined by a simple one dimensional distribution, known as the line spread function. Since the image scanning system developed in this project operates by taking a series of one dimensional samples over a uniform, two dimensional sampling grid, the assumption of isotropism may be

reasonably made. Hence the system transfer characteristics may be completely defined by the line spread function.

6.4.2 Measurement of Point and Line Spread Functions

To directly measure the point spread function of a system would require the imaging of an infinitely small point source. In a system with rotational symmetry, this problem may be simplified to that of imaging an infinitely narrow slit, to produce the line spread function. However, such a test object can obviously not be realised in practice.

In practice the size of this slit should be very much smaller than the minimum feature size to be imaged, so that slit dimensions do not affect the resultant intensity distribution. This requirement raises a number of practical obstacles to direct measurement of the line spread characteristics of an imaging system. The production of a sufficiently thin slit is difficult, and if such a slit were produced the amount of light radiating through it would be exceedingly small, so that a bright image would not be formed. In addition, diffraction effects at the slit would tend to cause distortion of the image.

Fortunately, an indirect method of measuring line spread function, which avoids these practical problems, has been proposed [6]. This method uses a simple intensity step object to produce a unique system transfer characteristic. This intensity distribution may be represented mathematically by the function

$$i(\xi) = \begin{cases} i_1 & -\infty < \xi \leq 0 \\ i_2 & 0 \leq \xi < \infty \end{cases}$$

Differentiation of this function yields a point impulse function of intensity $(i_2 - i_1)$. Similarly, differentiation of the resultant step function produced in the image plane will give the characteristic system line spread function. In mathematical notation the normalised line spread function, $A(x)$, is given by the equation.

$$A(x) = \frac{dI/dx}{(dI/dx)_{\max}} \quad (6.3)$$

Where $(dI/dx)_{\max}$ refers to the maximum gradient of the intensity step image.

Simple tests were carried out on the two sensors under evaluation during this project, to assess their characteristic line spread functions, and hence give some measure of their ability to resolve detail. These tests involved imaging a discrete intensity step object and then differentiating the resultant intensity distribution in the image to arrive at the line spread function.

Since we required measurement of the response of the entire assembled imaging system, tests were carried out using standard optical arrangement used during image acquisition, as described in Chapter 5 of this thesis. This ensured that any sources of image degradation encountered during image acquisition were also included in the measurement of the line spread function. The intensity step object was simply produced by placing an adjustable slit mechanism in front of the CCD sensor. The edges of this slit were machined flat to ensure a sharp change of incident light intensity. The slit was opened wide enough to produce two separate intensity step images separated by a distance of several millimetres, either of which could be used to calculate the line spread function.

The response of the TC104 and ILX503 sensors to an intensity step input are plotted in Figure 6.7 and Figure 6.8 respectively. These graphs plot the output digitised grey level value produced by each individual pixel element within the region of the step discontinuity.

It can be seen from these plots that the step response of the TC104 sensor produced an image exhibiting smearing of intensity data across approximately 10 sensor elements. The step response of the ILX503 sensor was seen to produce a significantly sharper image of the discontinuity. The resultant intensity distribution of the image extended over approximately 5 sensor elements, with the maximum intensity gradient being considerably higher than that measured for the TC104. From these results it can be seen that the ability of the ILX503 to resolve a sharp intensity step is considerably better than that of the TC104.

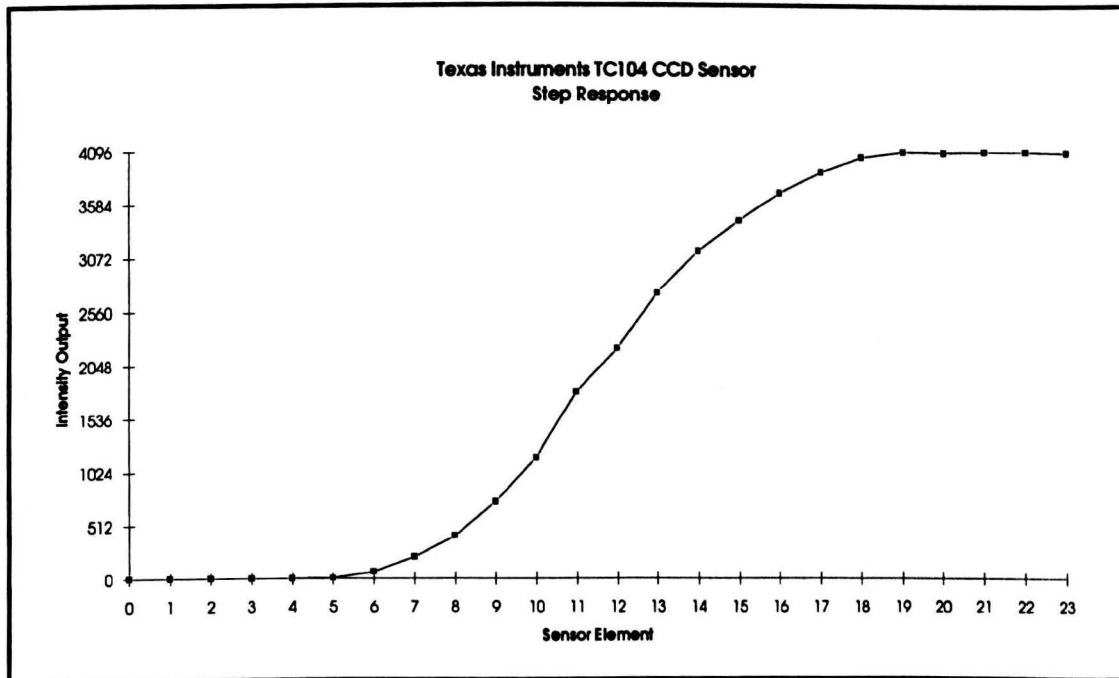


Figure 6.7 Response of the TC104 Sensor to an Intensity Step Input

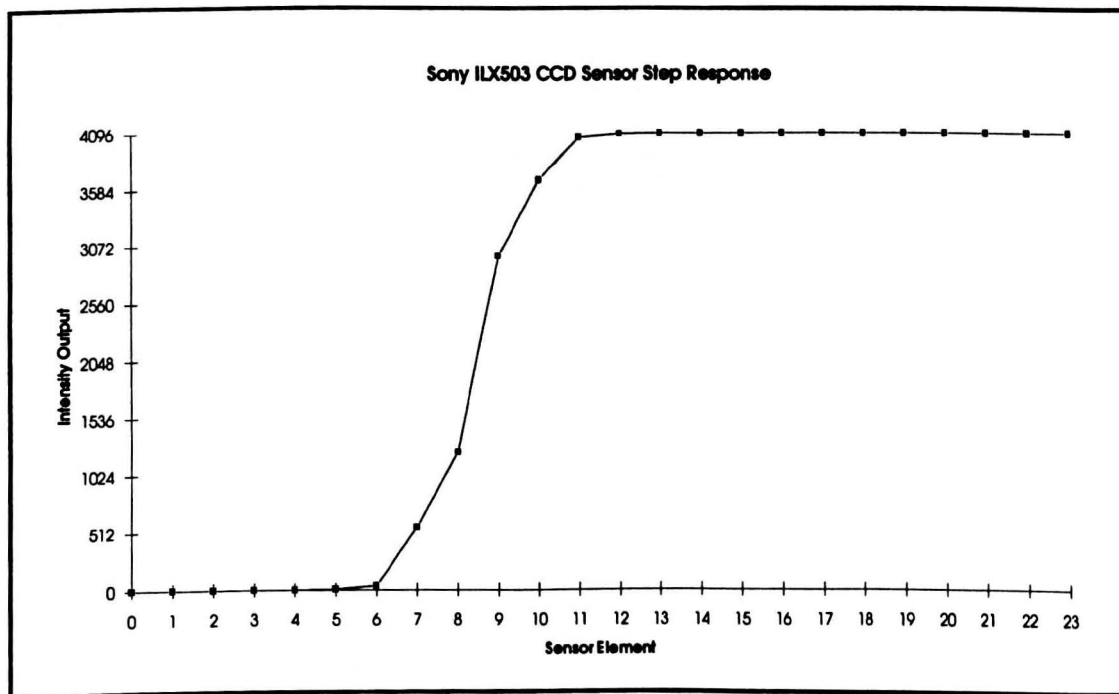


Figure 6.8 Response of the ILX503 Sensor to an Intensity Step Function

The step responses of each sensor were used to calculate their characteristic line spread functions, using Equation (6.3). This data is plotted for the ILX503 and TC104 sensors in Figure 6.9 and Figure 6.10 respectively. These graphs plot normalised intensity output against distance measured along the sensor element array. The use of a

distance scale eliminated any discrepancy between results caused by a difference in sensor element spacing. It should be noted, however, that this distance scale does not represent the actual size of the displayed image, since that is dependent on the pixel size used by the monitor display.

Again it was noted that the resolving power of the TC104 was poor compared to that of the ILX503. The line spread function may be considered as the response of the imaging system to a one dimensional point source radiating at unit intensity. From this standpoint it is obvious that the ILX503 would produce a significantly sharper image of the point than the TC104.

Several sources of image unsharpness were identified which would lead to a degradation of line spread function, these included optical effects, spatial quantisation errors, charge transfer inefficiency, high frequency signal attenuation in electronic hardware and data signal noise. A brief discussion of these different phenomena will be given to assess their effect on overall image sharpness.

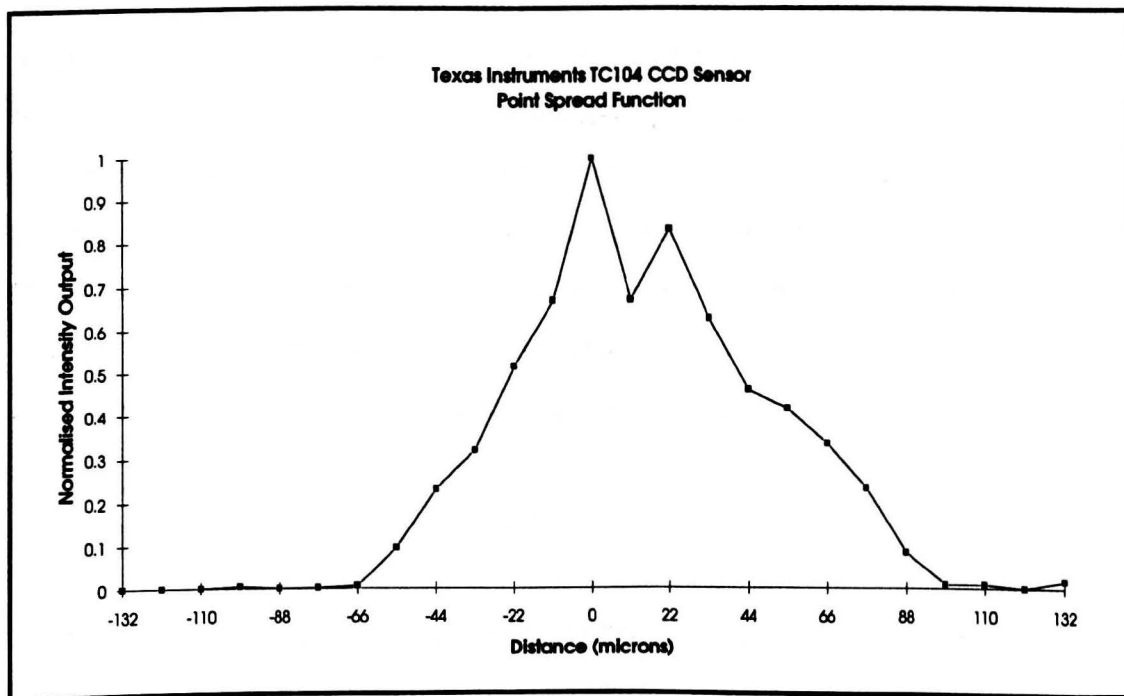


Figure 6.9 Calculated Line Spread Function for the TC104 Sensor

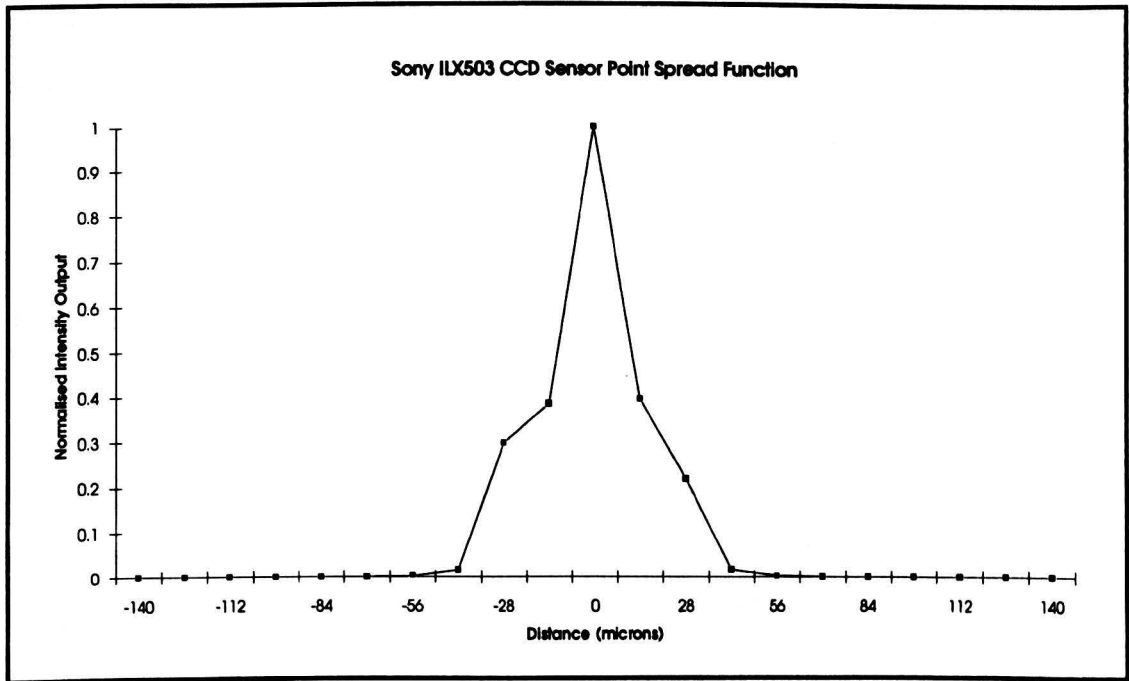


Figure 6.10 Calculated Line Spread Function for the Sony ILX503 Sensor

Optical unsharpness due to geometrical effects in the experimental set up lead to a defocusing of the imaged discontinuity. This situation is illustrated in Figure 6.11, where a collimated laser beam of diameter, d , is scanned across a sharp discontinuity formed by placing an opaque block of thickness, t , directly below the deflecting mirror in contact with a row of light sensitive elements. As the deflecting mirror scans the beam over the test object, an area of unsharpness of width, δ , is formed due to a shadow of its upper edge being cast onto the sensor elements. If the stand off distance from the laser source to the sensor element plane is D , then the deflection angle of the mirror, ϕ , producing maximum unsharpness is given by equation (6.4).

$$\tan^2 \phi = 1 - \frac{2d}{D} \quad (6.4)$$

Simple geometry then gives the area of unsharpness, δ , using equation (6.5).

$$\tan(90 - 2\phi) = \frac{\delta}{t} \quad (6.5)$$

The experimental apparatus used for measuring line spread involved a beam approximately 1 mm wide, mounted 1 metre from the sensor array. The distance from

the upper edge of the slit to the top of the sensor was approximately 4 mm with an additional distance of 0.8 mm from the top of sensor window to the surface of the photosensitive elements. Using these values and solving for ϕ gives :

$$\phi = 44.97^\circ$$

The maximum geometrical unsharpness is then :

$$\delta = 4.8 \mu\text{m}$$

The unsharpness component due to geometrical effects is thus equivalent to less than half the sampled pixel size of either sensor. Similar arguments hold true for the case of imaging a radiographic film image, although the value of object thickness, t , then refers to the thickness of the film emulsion which will typically be less than $200 \mu\text{m}$ [1]. This equates to a total geometric unsharpness during the radiograph scanning process of approximately $0.2 \mu\text{m}$.

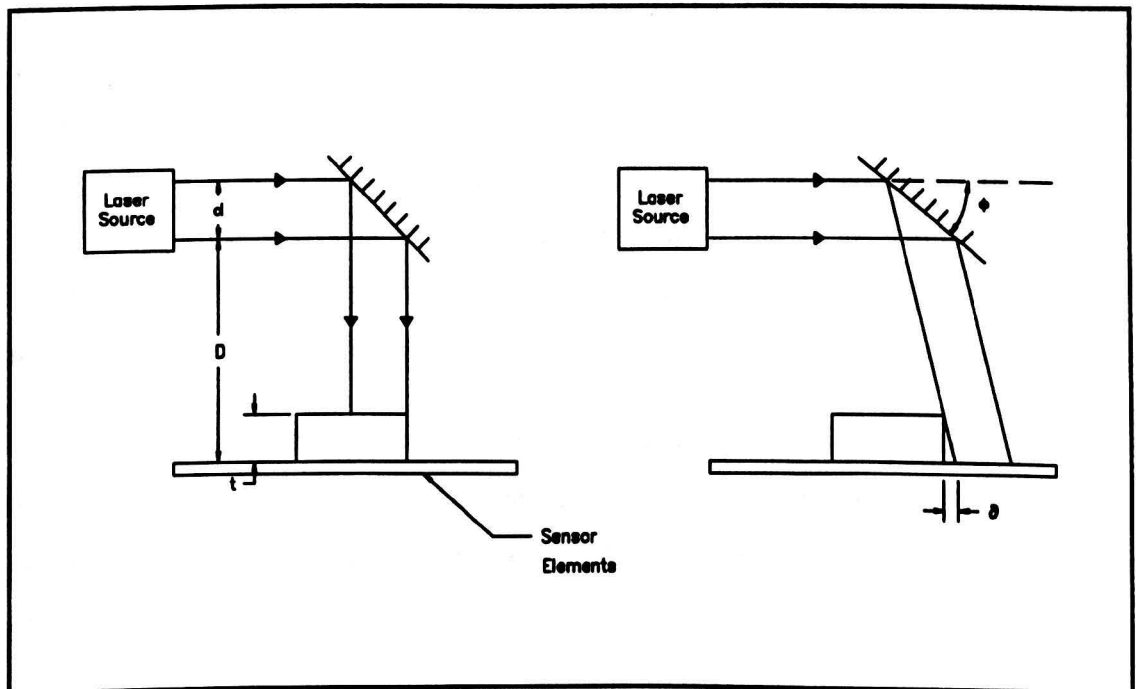


Figure 6.11 Geometric Unsharpness Effects During Line Spread Measurement

A further loss of detail resolution was caused by spatial quantisation of the continuously varying intensity pattern of the test object into discrete pixels. This was

due to the fact that luminous energy falling anywhere within a sensor element was accumulated together into a single pixel data packet.

Charge transfer inefficiency within the CCD shift registers of the sensor also lead to a defocusing of sharp image data. This problem was caused by incomplete transfer of signal charge packets between adjacent potential wells in the CCD, leading to a “smearing” of image data between pixels. Charge transfer inefficiency was limited by the design of the sensor and was unavoidable.

Another source of image defocusing was that finite slewing rates of op amps used in the signal conditioning circuit and efforts to filter out noise spikes led to an associated attenuation of high frequency pixel data. Although this phenomenon was minimised by careful circuit design and choice of components, examination of voltage waveforms produced by the signal conditioning circuits indicated that it was still significant.

Finally, there was a loss of detail resolution in the image due to the presence of spurious noise signals. In the case of the TC104 sensor it was seen that random noise signals overlying the legitimate pixel data tended to compromise its ability to resolve small low contrast features. This phenomenon was illustrated by the fact that the measured line spread characteristics for the TC104 were not particularly repeatable, since noise signals interfered with measured intensity levels. In general the calculated shape of the TC104 line spread function varied widely between measurements, whilst that of the ILX503 was reasonably constant.

The line spread function measured the combined action of all the effects discussed above, so that a true system characteristic was found. Unfortunately this method provided no data concerning the relative contribution of different sources to the overall image unsharpness.

6.4.3 The Modulation Transfer Function

Both the point spread function and line spread function are unique system transfer characteristics which can be used to calculate the response to any conceivable object intensity distribution. However, they give little direct information about how a system responds to changes in object size and contrast. Although they are powerful

conceptual tools the point and line spread functions prove exceedingly cumbersome when used to predict the output produced by any but the most simple input signals.

A more flexible approach analyses the way sinusoidal signals are reproduced by a system. Probably the most useful measure of an imaging system's ability to reproduce both contrast and detail resolution is given by the Modulation Transfer Function (MTF). The MTF uses frequency domain techniques to describe the ability of an imaging system to transmit signals of varying spatial frequency.

General communications theory has accumulated a large body of information concerning the transmission through a linear system of signals which vary periodically with time. This knowledge can be brought to bear on the problem of image data transmission through an optical system by considering an image as the sum of a series of sinusoidal spatial frequency components of varying amplitude.

The concept of spatial frequency comes as a direct analogy to time base frequency. Consider an intensity pattern that varies periodically with distance, the spatial frequency of the object may be described in terms of the number of intensity cycles per unit distance, in just the same way as a time varying signal may be described by the number of cycles performed per unit time.

Classical Fourier theory describes how any periodic signal may be represented by its characteristic spectral components [7]. This argument holds true whether it is applied to signals varying with respect to time or distance. Hence, any image may be represented as the sum of an infinite series of sinusoidal components. This concept leads to two separate but equivalent representations of an image either as an intensity distribution in the spatial domain or as an amplitude spectrum in the spatial frequency domain.

A convenient method of representing the transfer characteristics of a linear system involves the concept of modulation transfer, as illustrated in Figure 6.12. The modulation of a sinusoidally varying signal is defined as the ratio of its amplitude to its mean value. In general any linear system will convert a sinusoidal input of a given frequency into a sinusoidal output signal of the same frequency but exhibiting some

change in modulation and shift in phase. In the special case where the system under consideration is isotropic then no phase shift will occur, and the system response may be fully described by the change in modulation of the sine wave.

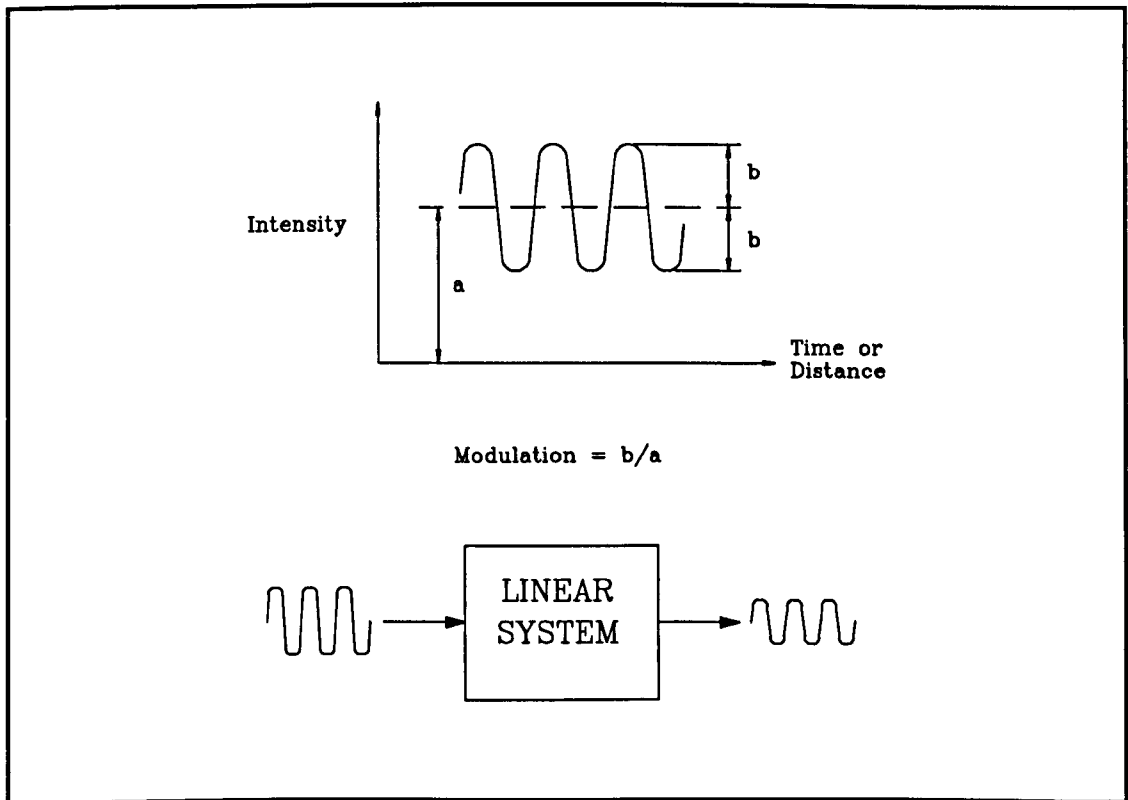


Figure 6.12 Modulation Transfer in a Linear System

This behaviour gives rise to the concept of the modulation transfer factor. Consider the case of a sinusoidally varying intensity pattern, with modulation, M_{object} , being reproduced by a linear imaging system. If the resultant image of the test pattern is a sinusoidally varying intensity distribution with modulation, M_{image} , then the modulation transfer factor, T , of the system at that frequency is defined as follows:

$$T = \frac{M_{\text{image}}}{M_{\text{object}}}$$

The modulation transfer factor will generally vary with frequency so that the amount of amplitude data reproduced by the system will tend to decrease as spatial frequency increases. This phenomenon is characterised by a loss of contrast definition when imaging objects of small dimensions. From this standpoint an imaging system may be considered as a filter of spatial frequency signals, tending to suppress high frequency

data. This concept leads to the definition of the modulation transfer function, MTF, which simply expresses the modulation transfer factor as a function of spatial frequency, R .

$$\text{MTF} = f(R)$$

The great utility of the MTF lies in the fact that if the modulation and frequency characteristics of the input image are known, the resultant image may be predicted by simple multiplication by the MTF rather than the complicated convolution procedure required by the spatial domain methods discussed earlier.

It can be shown mathematically that the MTF of a system can be calculated by taking the two dimensional Fourier transform of the point spread function [8]. Hence the point spread function and MTF are equivalent system transfer characteristics expressed in the spatial and frequency domains respectively. When considering the special case of an isotropic system, the point spread function will be rotationally symmetrical so that the modulation transfer function may be calculated by one dimensional Fourier transformation of the line spread function. The result of this Fourier transformation is a periodic function, symmetrical about its central point. The physical significance of this function requires certain discussion in order to understand the results of this MTF analysis.

Any solid state photosensor array may be considered as a carrying out uniform spatial sampling of a continuously varying light intensity function. Nyquist sampling theory states that, in order to avoid aliasing, the sampling rate must be at least twice the frequency of the highest frequency component in the waveform being sampled. This condition gives rise to the Nyquist frequency which defines the highest frequency component which may be faithfully sampled by a data acquisition system. For an optical sensor the spatial sampling frequency, f_{sensor} , is given by the reciprocal of the sensor element spacing, Δx . The Nyquist frequency, f_n , is then given by re-stating equation (3.8), as follows.

$$f_n = \frac{f_{\text{sensor}}}{2} = \frac{1}{2\Delta x}$$

The TC104 sensor has a pixel element spacing of 10.7 μm , which corresponds to a Nyquist frequency of 46.7 cycles/mm. Similarly the ILX503 sensor element spacing of 14 μm gives a Nyquist frequency of 35.8 cycles/mm.

Given an input line spread function, $f(x)$, consisting of a one dimensional array of N discrete pixels, the discrete Fourier transform, $F(u)$, is given by.

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) \exp[-j2\pi ux / N] \quad (6.6)$$

$$x = 0, 1, 2, \dots, N-1;$$

$$u = 0, 1, 2, \dots, N-1;$$

Where x and u are indices relating to discrete data samples taken in the spatial and frequency domain respectively. The resolution of the spatial frequency axis, Δu , for a spatial sampling resolution, Δx , is then given by the equation :

$$\Delta u = \frac{1}{N\Delta x}$$

$F(u)$ will generally be a complex function consisting of a real component, $R(u)$, and an imaginary component, $I(u)$. For convenience this complex function is often expressed as the Fourier spectrum, $|F(u)|$, which is the absolute magnitude of $F(u)$.

$$|F(u)| = \sqrt{R^2(u) + I^2(u)}$$

The modulation transfer characteristics of the two sensors used in this project were calculated, as described above, by the discrete Fourier transformation of their line spread functions. The resulting modulation transfer functions for the TC104 and ILX503 sensors are plotted in Figure 6.13 and Figure 6.14 respectively.

The MTF curves derived for each sensor show that the modulation transfer factor reaches unity as it intercepts the spatial frequency axis. This means that uniform intensity areas are reproduced with no loss of modulation between the object and the image. This condition will always be fulfilled since an object with zero spatial frequency exhibits no contrast modulation. As spatial frequency is increased towards

the Nyquist limit, there is a consequent degradation of image modulation. The value of MTF at the Nyquist frequency represents the amount of contrast modulation retained in the image under the most demanding imaging conditions encountered by the sensor system.

Examination of these MTF curves gives quantitative information regarding the amount of contrast data retained in the image as spatial frequency of the object is varied. The modulation transfer function for the TC104 sensor reveals that contrast modulation in the image falls quickly as a function of spatial frequency. It can be seen that 50 % of contrast modulation is lost when imaging objects with spatial frequency of around 5.8 cycles/mm, this corresponds to imaging a sinusoidal intensity pattern with a 170 μm repetition period. At the Nyquist frequency, contrast modulation in the image is only around 5 % of that exhibited by the object.

The MTF characteristics for the ILX503 sensor show a much improved ability to resolve image detail. At least 50% of contrast modulation data is retained when imaging spatial frequencies at or below 11.2 cycles/mm which corresponds to imaging a sinusoidal intensity pattern with an 89 μm repetition rate. At the Nyquist limit approximately 30 % of contrast modulation data is retained. From the above discussion it is clear that the ability of the ILX503 sensor to resolve small, low contrast features is significantly improved from that of the TC104.

The ASME Code relating to radiograph digitisation includes a stipulation that all such equipment must be tested to evaluate its modulation transfer characteristics [9]. However, this document gives no guidelines concerning the required MTF which should be attained. Studies commissioned by the U.S. Nuclear Regulatory Commission concerning the digitisation of industrial radiographs have lead to the recommendation that any radiograph digitisation equipment should retain at least 33 % of contrast modulation at spatial frequencies at or below 5 cycles/mm [10]. Examination of the MTF curves plotted for both the ILX503 and TC104 sensors reveal that this condition is satisfied for both sensor systems, with the ILX503 retaining approximately 85 % of contrast modulation at 5 cycles/mm.

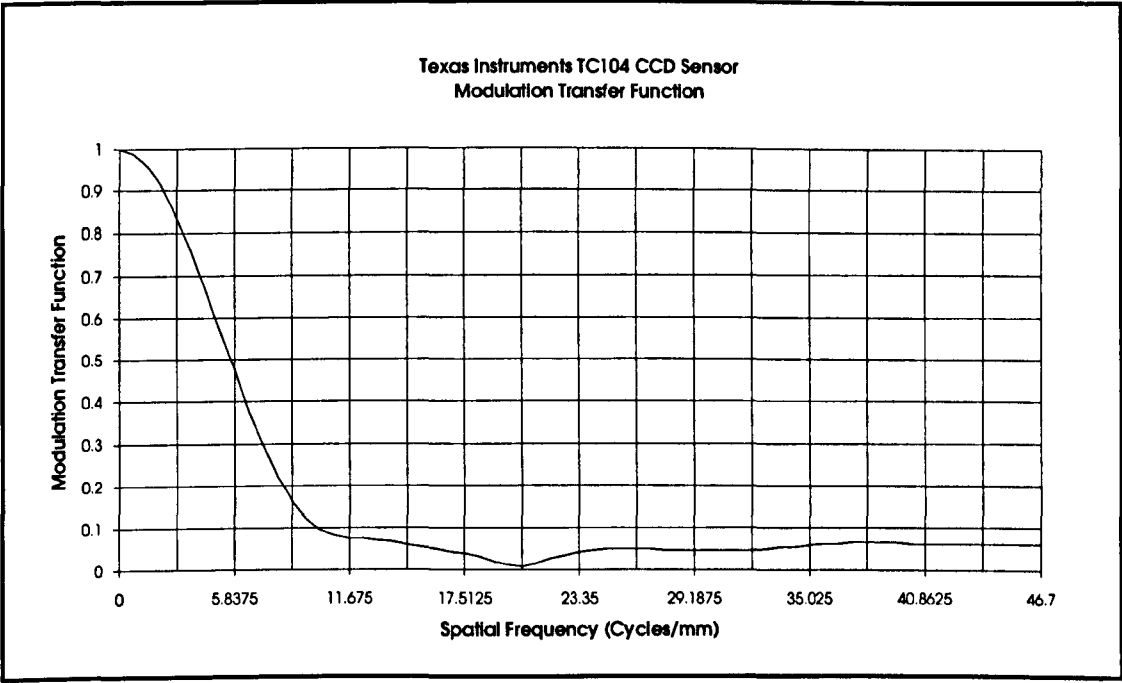


Figure 6.13 Calculated Modulation Transfer Function for the TC104 Sensor

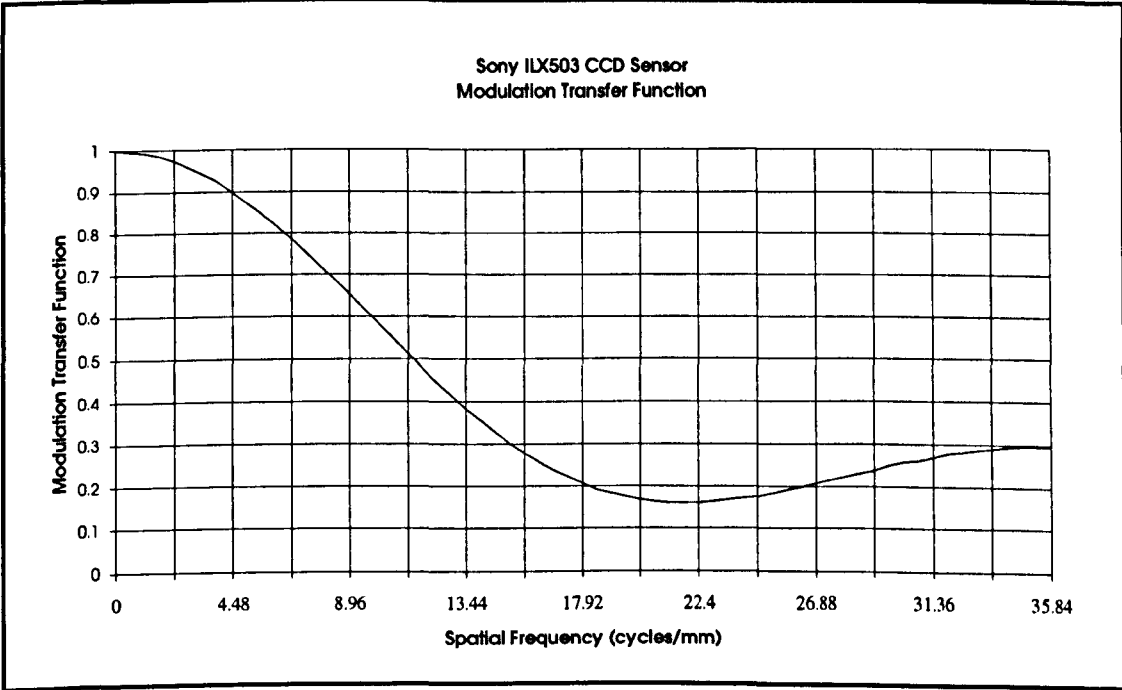


Figure 6.14 Calculated Modulation Transfer Function for the ILX503 Sensor

6.4.4 Resolving Power of Radiographic Films

Many researchers have used the methods described in the preceding sections to analyse the resolving power of various radiographic imaging techniques [6][11]. It will be

instructive to compare the findings of this work with the performance of the CCD sensors assessed above.

Figure 6.15 and Figure 6.16 show line spread and modulation transfer functions calculated for a medium speed radiographic film under ideal conditions. The major factors causing loss of image resolution in these systems are inherent and geometric unsharpness. Hence, this data gives a reasonable indication of the optimum imaging ability of modern radiographic film. Obviously, the actual resolving power of any radiographic technique will depend heavily on geometrical arrangement, beam power and source size.

The data given for the frequency response of radiographic films shows the point spread function extending over a much wider area than that measured for either sensor system. Similarly MTF characteristics for radiographic films indicate a much reduced ability to resolve small, low contrast features than that exhibited by the CCD sensors.

Although ideally we would wish for no loss of image contrast at high spatial frequency, all practical imaging systems will exhibit some degree of defocusing. However, it can be stated that significantly more contrast information is lost during the radiographic imaging process than during subsequent digitisation of the image.

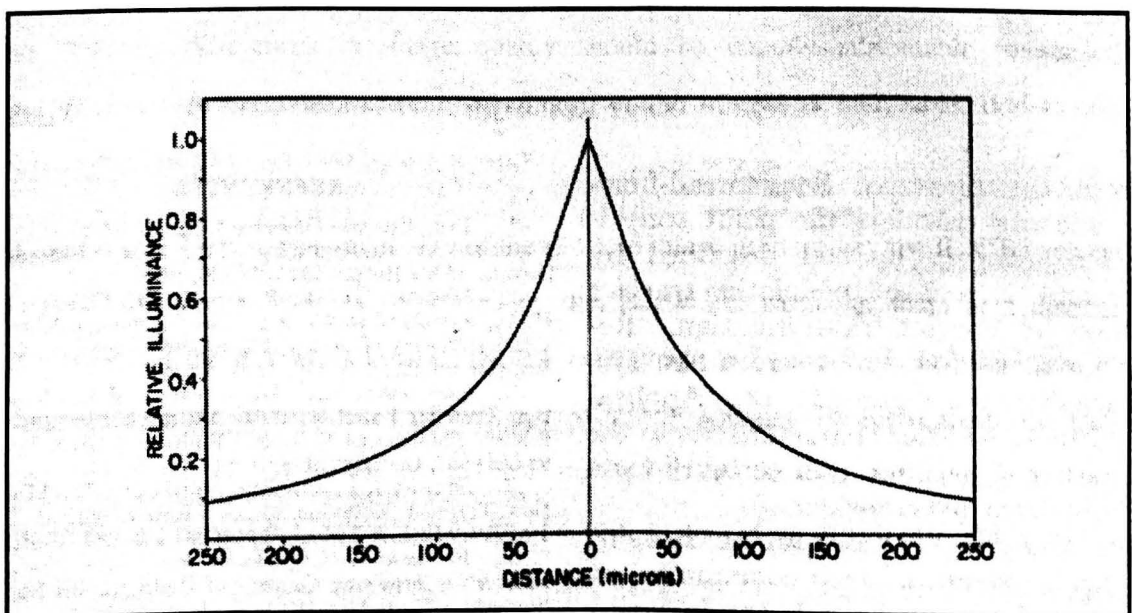


Figure 6.15 Normalised Line Spread Function of a Medium Speed Radiographic Film

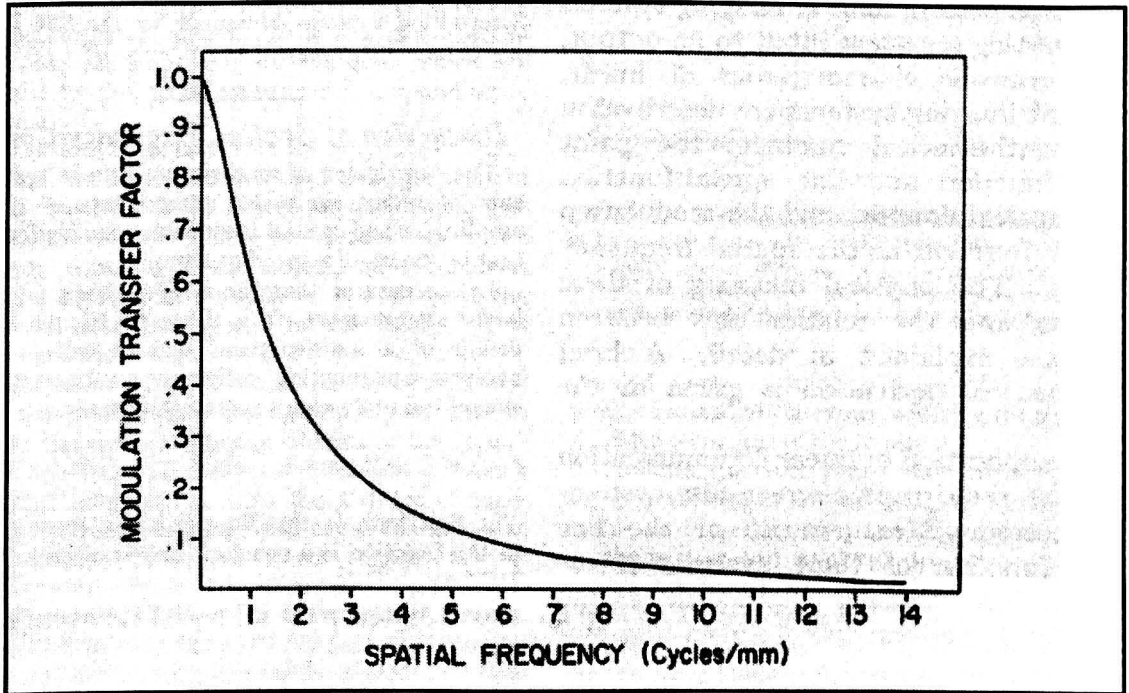


Figure 6.16 Modulation Transfer Function Calculated from Figure 6.15

6.5 Test Pattern Images

The final performance characteristic of any imaging system will necessarily involve the assessment of the images it produces. Figure 6.17 and Figure 6.18 show example test pattern images produced by the image acquisition system using the TC104 and ILX503 sensors. The optical arrangement used during the digitisation of each image was identical, so that any differences in image quality should be directly attributable to sensor performance. A slight difference in resolution of the images is noticeable, due to the different pixel spacings used by the sensors.

It can be seen that the general quality of the test card image produced by the ILX503 sensor is noticeably better than that produced by the TC104. In particular there is a distinct defocusing of the thin white lines on the test card image in Figure 6.17. The thickness of these lines on the original test card was approximately 300 μm . By reference to the step response of the TC104 sensor, plotted in Figure 6.7, it can be seen that imaging a sharp intensity change will lead to the spreading of across an area of up to 100 μm . This effect is clearly visible on the resulting image. In contrast, the image produced by the TC104 sensor does not exhibit a noticeable defocusing of sharp intensity changes.

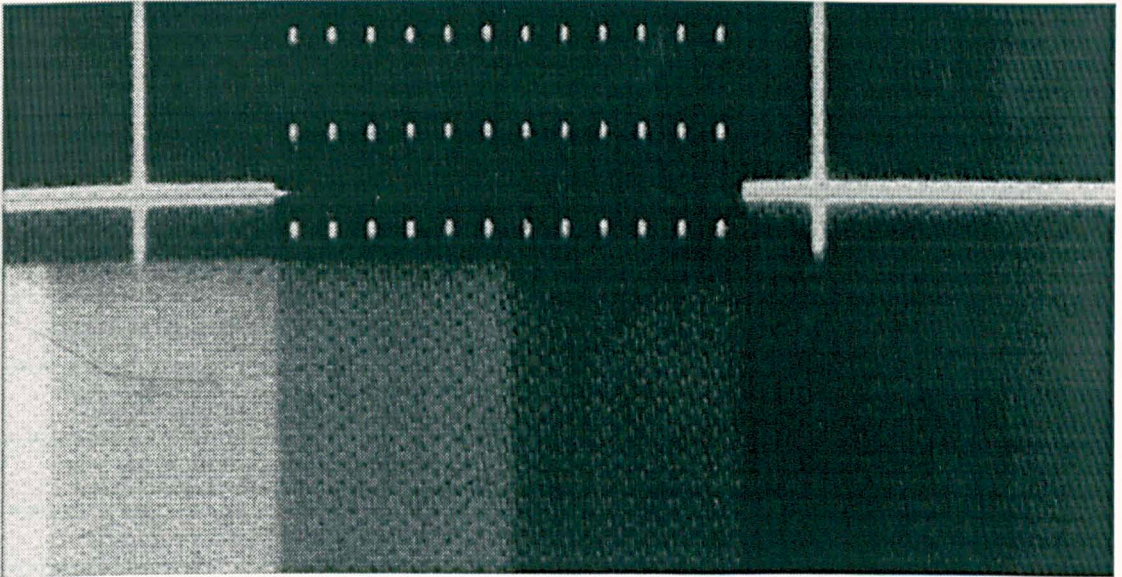


Figure 6.17 Test Pattern Image Produced by the TC104 Sensor

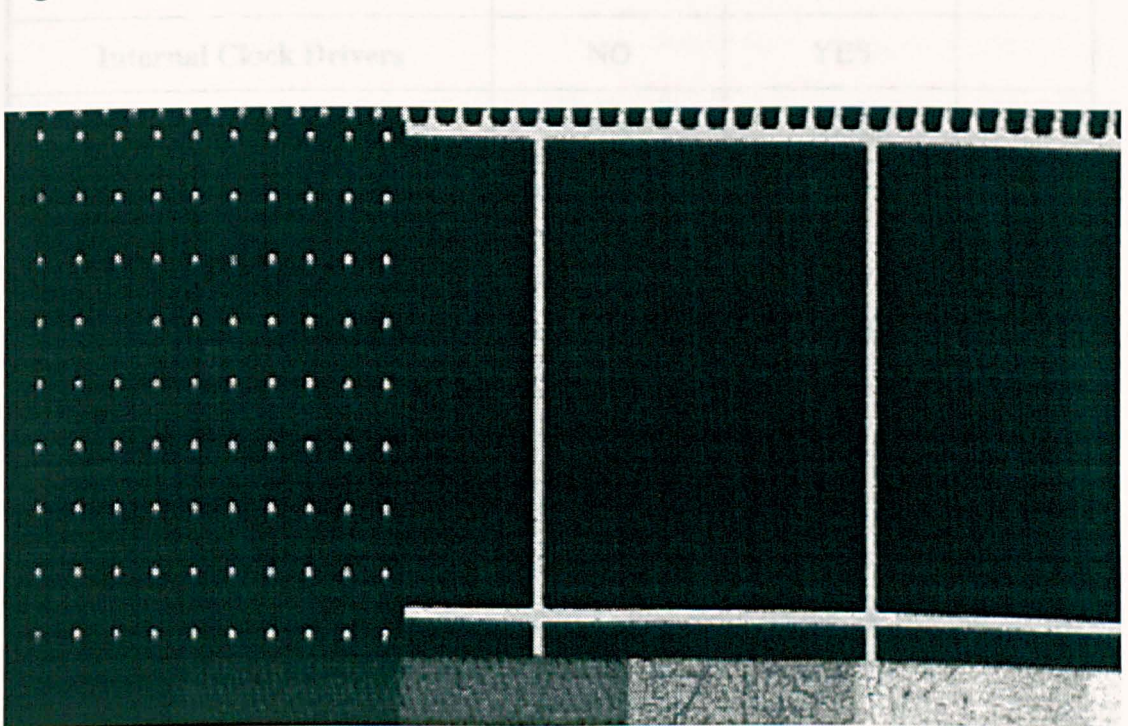


Figure 6.18 Test Pattern Image Produced by the ILX503 Sensor

6.6 Comparison of the TC104 and ILX503 Sensors

The two linear CCD array sensors evaluated during this project were the Texas Instruments TC104 and the Sony ILX503. Both are general purpose linear CCD array sensors designed for applications including facsimile equipment, bar code reading and document scanning. A comparison of some important operating parameters of the two sensor systems is given in Table 6.4.

Parameter	TC104	ILX503	Units
Pixels	3456	2048	
Pixel Spacing	10.7	14	μm
Effective Length	37	28.5	mm
Saturation Output Voltage	0.8	4.8	V
Sensitivity @ $\lambda = 675 \text{ nm}$	13.8	30	V/(lx.s)
Peak to Peak Noise	86	13	gsl
Signal to Noise Ratio ¹	48:1	315:1	
Price	£112-50	£27.50	
Internal Clock Drivers	NO	YES	
Internal Sample and Hold	NO	YES	

Table 6.4 Comparison of Sensor Parameters

The Texas Instruments TC104 sensor had the advantage of having a closer pixel spacing over a longer scan length than the Sony ILX503. The smaller pixel size of the TC104 sensor gives it a 31 % higher spatial sampling resolution than that of the ILX503. In theory this should result in an improved ability to resolve small objects. However, assessment of the modulation transfer functions of the two sensor systems indicated that, at high spatial frequencies the ILX503 sensor performed better than the TC104. The main reason for this appeared to be that noise on the TC104 output signal obscured small changes in contrast data.

Although the TC104 sensor offered a greater number of sensor elements per scan line it should be noted that working with images 3456 pixels wide is not particularly convenient for image processing and display applications. The display format used by the MosaiQ frame store consisted of 1024 columns by 512 lines. When scanning large areas at high resolution, the size of the resultant image data files soon exceeded the

¹ relative to peak to peak noise

display resolution of the monitor. In order to display all or part of a scanned image it was necessary to zoom in and out, so that only a proportion of the total image pixels were displayed at any one time. When working with the TC104 sensor this zooming process required either the use of complicated software algorithms or the need for cropping the edges of an image, so that a portion of the image was not displayed. The ILX503 pixel format, consisting of 2048 elements, considerably simplified the software required for image display and manipulation.

In the TC104 drive circuit all high power clock signals used to drive the CCD shift registers had to be derived from TTL logic levels using MOS clock drivers. These devices caused problems with drive clock feedthrough and were found to be the most inherently noisy component of the TC104 drive circuit. Another problem encountered was that the DS0026 clock drivers used in this circuit were prone to failure during prolonged use. Failure of these devices sometimes resulted in the destruction of the TC104 sensor.

A major advantage of the ILX503 sensor was the provision of internal clock drivers, so that all logic inputs to the sensor were at standard TTL levels. This design eliminated the need for any external clock driver hardware, considerably simplifying circuit design. Deriving high power clock signals close to the CCD shift registers that they drove ensured that there was no need for long clock lines, which would tend to induce noise signals into surrounding components. Hence, the use of internal clock drivers in the ILX503 sensor simplified the external drive circuit and improved output signal noise immunity.

The ILX503 sensor also offered the option of internal sample and hold operation. This feature effectively eliminated one of the output signal conditioning operations required to produce the required video input signal to the ADC.

Another important factor in the assessment of these sensors was their cost and availability. Both sensors are readily available from stockist within the UK, but it can be seen from Table 6.4 that the Sony ILX503 is considerably cheaper.

The ILX503 sensor exhibited higher sensitivity and saturation output voltage characteristics than the TC104. The signal to noise ratio (SNR) of the raw output voltage signal from either sensor may be defined as the ratio of full scale voltage output to the peak to peak noise voltage. Consequently, for a given noise level the output SNR is proportional saturation output voltage. This meant that the ILX503 output signal was less sensitive to noise contamination than that of the TC104.

From the discussion of sensor noise characteristics in Section 6.3 it is apparent that the amount of noise associated with the TC104 output signal was considerably higher than that from the ILX503. Any noise signal in excess of the ADC threshold voltage lead to a loss of digitisation fidelity and a consequent decrease in the ability to re-produce small changes in image contrast. When considering the relatively demanding imaging problem of radiograph digitisation and interpretation, any loss of image fidelity was obviously unwelcome. From this reasoning it is clear that the imaging performance of the TC104 sensor system was unacceptable for acquiring image data at 12 bits per pixel.

The preceding discussion indicates that the ILX503 sensor offered both improved performance and simplified drive circuit requirements than the TC104. It should be evident from these points that the ILX503 was much better suited to the requirements of the radiograph digitisation system. For these reasons the remaining work carried out during this project concentrated on using the Sony ILX503 sensor to acquire high quality images from radiographic film.

6.7 Summary

This chapter describes the measurement of performance characteristics of the two types of CCD image sensor evaluated during this research. These sensors were the Sony ILX503 and the Texas Instruments TC104. Tests involved measurement of noise levels, thermally induced dark signal and modulation transfer functions.

Both systems were seen to exhibit insignificant levels of thermally generated dark signal noise. This was due to black level clamping hardware used during image acquisition. However, the signal to noise ratio, relative to peak to peak noise, of the

ILX503 sensor was found to be 315:1, as compared to a figure of 48:1 for the TC104 sensor. Modulation transfer analysis revealed that the Sony ILX503 sensor preserved 85 % of contrast modulation data at a spatial frequency of 5cycles/mm. This compared favourably with a modulation transfer factor of just 50 % exhibited by the TC104 sensor at the same spatial frequency. As a consequence of these tests the Sony ILX503 sensor was selected for use in the final prototype imaging system due to its superior imaging properties.

References

- 1 H. W. Ott, "Noise Reduction Techniques in Electronic Systems", Second Edition, Wiley, 1988, p. 242.
- 2 A. E. Karbowiak, "Theory of Communication", Oliver and Boyd Ltd, 1969.
- 3 A. B. Carlson, "Communications Systems : An Introduction to Signals and Noise in Electrical Communication",
- 4 E. W. H. Selwyn, "Fourier Theory of Detail Reproduction", in Applied Optics and Optical Engineering, Vol. 2, Ed. R. Kingslake, Academic Press, 1965, pp 175-191.
- 5 A. Notea, "Evaluating Radiographic Systems using the Resolving Power Function", NDT International 16 5. October 1983, pp 263-270.
- 6 K. Rossman, "Image Forming Quality of Radiographic Screen Film Systems : The Line Spread Function", American Journal of Roentgenology 90, July 1963, pp 178-183.
- 7 D. C. Champney, "Fourier Transforms and their Physical Applications", Academic Press, 1973.
- 8 K. Rossman, "Point Spread Function, Line Spread Function and Modulation Transfer Function-Tools for the Study of Imaging Systems", Radiology 93, August 1969, pp 257-272.
- 9 ASME Boiler and Pressure Vessel Code, Section 5, Subsection A, Article 2, Appendix III - Digital Image Acquisition, Display and Storage for Radiography and Radioscopy, American Society of Mechanical Engineers, July 1992, pp 35-36.

- 10 "Review and Evaluation of Technology, Equipment, Codes and Standards for Digitisation of Industrial Radiographic Film", U. S. Nuclear Regulatory Commission, Washington D. C. May 1992.
- 11 R. H. Morgan, L. M. Bates, U. V. Gopalarao and A. Marinaro. "The Frequency Response Characteristics of Xray Films and Screens", American Journal of Roentgenology **88**, 1962, pp 426-440.

7. Software Development

7.1 Introduction

As described in Chapter 5, the radiograph digitisation system relied on a network of transputer processors to provide software control over the different components used for image capture, display and processing. All transputers within the network were programmed using the 3L Parallel C language. The transputer network was interfaced to a host PC which provided an interactive software environment for user control of the system. The host PC also carried out image archiving and system I/O functions which were programmed using the Borland Turbo C++ language.

This chapter describes the software methods used to program and configure the transputer network and its associated hardware components. The chapter is divided into three sections. Firstly, a brief discussion of the concepts underlying the programming of parallel systems is given since this represents a considerable conceptual jump from the methods used for programming sequential systems. An overview of the 3L Parallel C language and its configuration language utility is then given. Finally, details of the programming and configuration of the radiograph digitisation system are discussed.

7.2 Programming Parallel Systems

A parallel program running on a network of transputers may be viewed as a series of concurrently executing sequential processes which communicate data over channels [1]. Each process is a self-contained unit, whose only interaction with other processes is via channel communications. A channel is a uni-directional communication line linking an input port of one process to an output port of another process. A communication can only take place when both the sending and receiving processes are ready. This requirement provides a simple method of synchronisation between processes.

More than one software process can operate on a single processor, with each process being assigned a portion of private, local memory. The number of processes assigned to a processor is only limited by the available memory of the device, and the execution

speed required by the system. There is no limit to the number of channels linking processes on the same processor. A process may be either a software module or a piece of hardware, since its internal operation is hidden from other processes in the network. Conceptually, internal software channels and external hardware channels are identical.

Processes may be distributed between any number of separate processors, with remote processes communicating via hardware channels. The number of channel links between processors is only limited by the number of external links provided by the hardware architecture.

This model is illustrated in Figure 7.1. In this example two independent software processes execute concurrently on one processor, with their only means of communication being via a software channel. Both these processes are connected via hardware channel links to a single process operating on a separate processor.

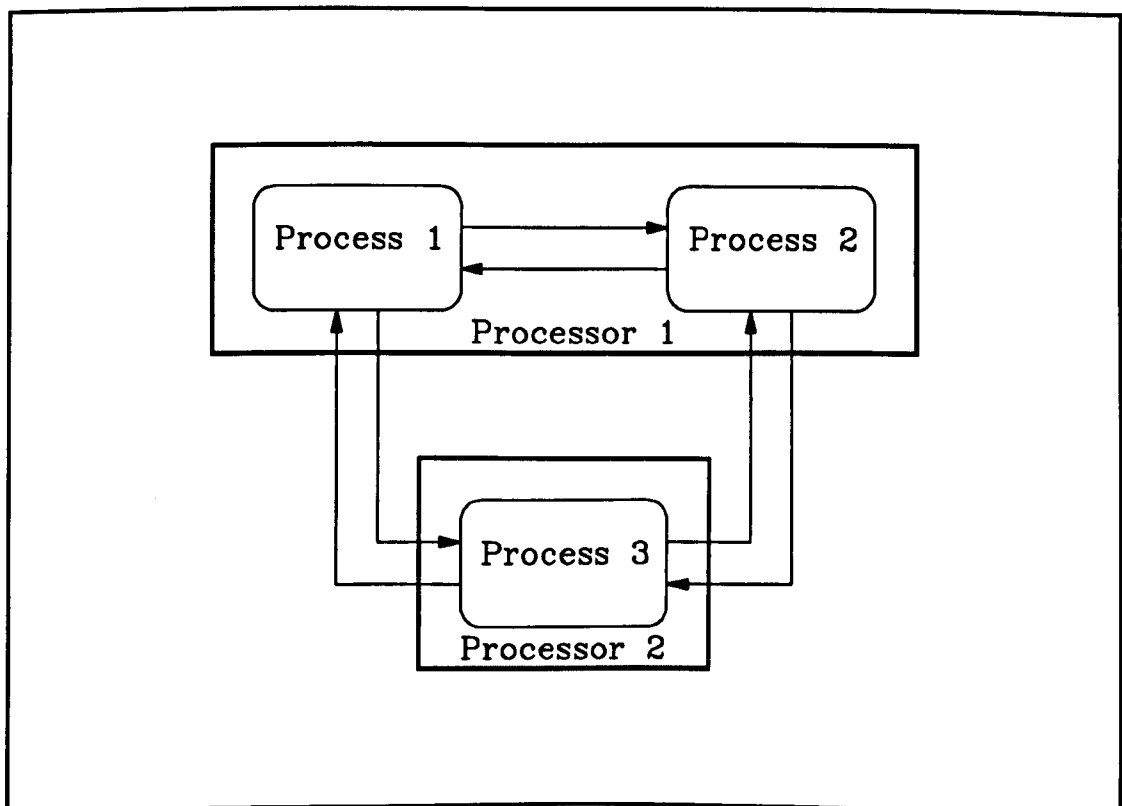


Figure 7.1 Software Model of Intercommunicating Processes

7.3 Overview of the 3L C Language and Associated Tools

All transputers in the network were programmed using the 3L Parallel C Language [2]. Parallel C was designed specifically for use with the Inmos transputer processor and supports all the features needed to build a concurrently operating transputer network. It is syntactically similar to conventional sequential C, with the addition of message passing functions to control link communications between processes.

The 3L C compiler implements a parallel version of the C language as defined by Kernighan and Ritchie [3]. C is a relatively low level, general purpose programming language which has been applied to a wide range of applications. Since the basic C data types correspond closely to those supported by the processor hardware on which it executes, the language is ideal for low level operations such as pointer and address arithmetic.

C provides a very limited set of operators and program flow control constructs, so that a basic implementation of a C compiler is both simple and compact. This basic language building block is supported by a rich set of function libraries which may be linked with program code. This arrangement produces a flexible program structure in which the programmer includes only those function libraries required by the application. The compact size and optimal coding of C produces a very efficient language which produces code with very little difference in execution speed than its machine code equivalent.

The 3L C language provides most of the function libraries common to standard sequential C with the addition of a number of functions giving access to the parallel processing capabilities of the transputer. The 3L C software package consists of a compiler and associated tools used to build and debug concurrently operating transputer networks. The following section contains a brief discussion of some of the elementary programming concepts which distinguish Parallel C from its sequential counterpart. It is not intended as a language reference, but merely gives a flavour of the scope of the language.

7.3.1 The 3L C Language

The 3L C compiler implements the parallel programming model of communicating sequential processes outlined in Section 7.2. A parallel application may be built from a series of software tasks which communicate via channels. A task is essentially a software “black box”, which may in turn be constructed from a number of sub-tasks. Each task has a number of input and output ports through which messages are received and sent. These ports are connected together by communication channels to form an intercommunicating network.

Internal communications channels are implemented simply by copying data between memory areas used by the communicating tasks. External channels are implemented by bi-directional Inmos serial links, which may be used to connect between any devices in the transputer family. Each processor may use a maximum of four external links to other devices.

Both hardware and software channels are pointed to by a single word identifier in memory known as the channel word. The address of the channel word determines whether it points to an internal or external channel. Internal channel words may be stored anywhere within memory, whereas external channel words are held in a reserved area at the bottom of the transputer address space.

3L Parallel C defines a new data type, **CHAN**, to store channel word identifiers. This data type, along with functions designed to access the transputers communications facilities, are defined in the header file **<chan.h>**. Each task definition file must include a declaration of these channel words if it is to communicate with any other task. The code fragment below illustrates the syntax used by parallel C to declare channel word identifiers.

```
#include <chan.h>
```

```
main(in_ports, ins, out_ports, outs)  
int ins, outs;  
CHAN *in_ports[], out_ports[];  
{  
    /* program code */  
}
```


The arrays `in_ports[]` and `out_ports[]` are a list of pointers to channel word identifiers. The integers `ins` and `outs` define the number of input and output ports used by the task. These variables are passed to the task as arguments to its main function by the configuration utility software. They may either point to internal or external link addresses depending on the information held in the configuration file. This arrangement allows a network configuration to be altered without having to re-compile the code for individual tasks.

In order to implement a concurrently operating network of communicating tasks some mechanism for message passing must be developed. Software functions designed to allow message passing between tasks are defined in the header file `<chan.h>`. For instance a single byte message may be passed between tasks using the following pair of functions.

```
chan_out_byte(char ch, CHAN *chan);  
chan_in_byte(char *ch, CHAN *chan);
```

The `chan` parameter in these functions is a pointer to a channel word identifier, and must point to one of the channels declared in the tasks main function argument list.

To avoid situations where a process may “hang” while it waits to communicate, it is sometimes convenient to time out the communication if it is not completed within a specified period. This may be achieved the following pair of functions.

```
chan_out_byte_t(char b, CHAN *chan, int timeout);  
chan_in_byte_t(char *b, CHAN *chan, int timeout);
```

In the above examples communications will be aborted if they have not been completed within `timeout` ticks of the timer associated with the calling task. The tick rate of the timer depends on the priority level assigned to each task. The timer associated with urgent tasks operates with a period of 1 μ s, while the timer period for non-urgent tasks is 64 μ s. Similar functions are also provided for the passing of words and variable length messages, either with or without a timeout period.

In some situations an input message to a task may emanate from one of a number of channels, depending on the relative execution speeds of other tasks in the network. In

this case some method is needed to detect which channel becomes ready to communicate first. Functions designed to handle this sort of contention are defined in the header file `<alt.h>`. For instance, the following function suspends program execution until a communication is attempted on one of the channels specified in its argument list.

```
int alt_wait( int n, CHAN *chan0, CHAN *chan1, ... , CHAN *chann);
```

chan0, chan1,...chann are pointers to the **n** channels on which a communication may be expected. The **alt_wait** function monitors these channels until a communication is attempted on one of them. It then returns a value, in the range 0 to n-1, indicating which channel is trying to communicate.

The operation of the concepts discussed in the previous sections may be illustrated by the simple program code listing below. This task accepts a single byte message from either one of two input channels, depending on which channel attempts to communicate first. The input number is then doubled and returned to the task from which it was sent. If this communication is not completed within 1000 ticks of the associated timer it will be aborted. This operation is repeated until an output communication fails, at which point the task terminates.

```
#include <chan.h>
```

```
#include <alt.h>
```

```
main(CHAN *in_ports[], int ins, CHAN *out_ports[], int outs)
```

```
{
    int in_channel;
    int ok;
    char number, twice;

    do{
        in_channel = alt_wait(2, in_ports[0], in_ports[1]);
        chan_in_byte(&number, in_ports[in_channel]);
        twice = number * 2;
        ok = chan_out_byte_t(twice, out_ports[in_channel], 1000);
    } while(ok);
}
```

7.3.2 Configuration Utility

In addition to the parallel C compiler described above, the 3L C software package contains a configuration language utility which is used to map individual software tasks onto the available processors within a physical network [4]. Each process is separately compiled and linked to produce code for individual executable tasks, contained in a task image file with the extension “.b4”. A number of these tasks are then configured into an intercommunicating network. This allows parallel networks to be expanded or re-configured without altering the code within individual processes.

The output from the configuration utility is a “bootable” file consisting of a short bootstrapping routine followed by executable code. Bootable code is downloaded from the host PC, via a server task, to the transputer network. The configuration utility routes executable code to a destination processor and distributes available memory between processes.

Each process in a network has a designated priority, either “urgent” or “not urgent”. Urgent processes execute uninterrupted until they require communication with an adjacent process, whereas non-urgent processes execute either until a communication is required or they are descheduled. Processes waiting for a communication require no processor time. The priority of each process within a network is defined in the configuration file, and consequently it may be changed without needing to re-compile the associated program code.

The user defined configuration file specifies the following information.

- A list of all the processors making up the physical network.
- The physical connection of hardware links between processors within the network.
- A list of all the software tasks contained in the application, including the location of the task image (.b4) file, the memory requirements of the task, the number of input and output ports used by it and its priority level.
- The placement of particular tasks onto processors within the physical network

- The logical connection of communication channels between software tasks within the network.

7.4 Configuring the Processor Network

The distribution of software tasks within a physical processor network is an important factor in ensuring the efficient implementation of a parallel application. Since transputer link communications are unbuffered, a communication cannot take place until both the sending and receiving processes are ready. Consequently, poor software design, and network configuration can result in the operation of several processes being suspended while they await a response from some overworked piece of software.

7.4.1 Hardware Configuration

The hardware configuration of the system refers to the way in which individual hardware components were connected together to produce a communicating network of processors and peripheral hardware. The radiograph digitisation system constructed during this project operated a network of transputer processors which performed several distinct functions including low level control of image acquisition hardware, storage and display of images produced by the sensor hardware and image processing and analysis functions.

Software control of peripheral hardware was accomplished using Inmos communication link adaptors, which allow direct connection between transputer processors and external hardware. The use of standard Inmos link protocol for communications between all components within the network ensured that the internal operation of each task was hidden from all other components. Hence, the operation of each component, whether hardware or software, was completely defined by the interaction of input and output signals at its communication ports.

The different hardware components comprising the radiograph digitisation system are shown in Figure 7.2. The diagram indicates the link channel identifiers, numbered 0 to 3, used to connect between hardware components.

Peripheral hardware modules were declared as processor nodes in the user defined configuration file, in the same way as a transputer processor would be. The only

difference between a transputer processor and a piece of peripheral hardware was that the hardware had no program code associated with it. These hardware modules were referred to as “dummy” processors, since the network software did not distinguish them from real transputer processors.

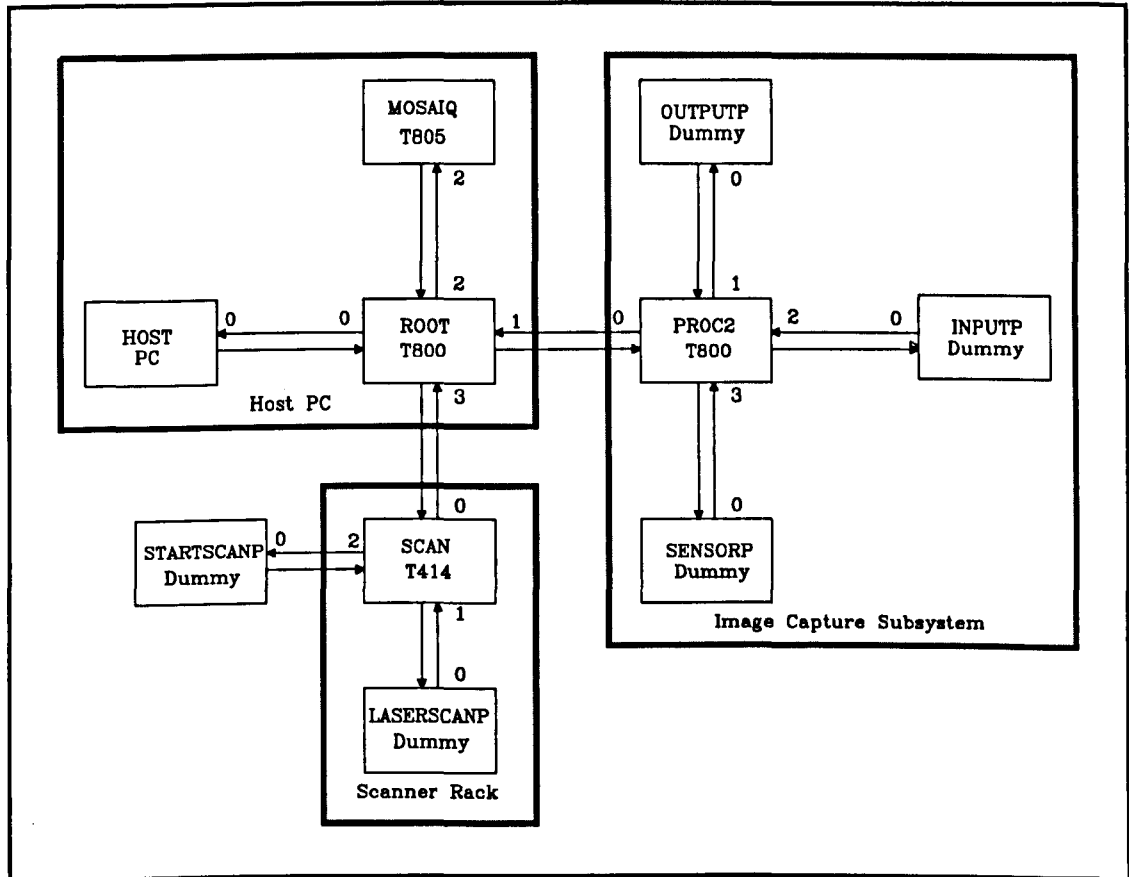


Figure 7.2 Hardware Configuration of the Radiograph Digitisation System

User access to the network was via a host PC, which was used to download compiled code and handle all user I/O functions. The host PC was interfaced to the processor network by memory mapping between the PC address space and the link interface of a T800 transputer, designated the “root” transputer. This transputer was mounted in one of the available expansion slots on the PC motherboard. All communications between the host PC and the processor network were handled by the root processor software which acted as a multiplexer, distributing messages around the network.

All images captured by the system were displayed using a MosaiQ frame store board, controlled by a dedicated T805 transputer. The MosaiQ was also mounted in an

expansion slot on the PC motherboard. However, it was not directly interfaced with the PC, but was connected to one of the available links of the root processor.

Software control of the image capture subsystem was provided using a second T800 transputer, designated “proc2”, mounted within the main system rack cabinet. This transputer was directly connected to the root processor. One channel link from proc2 was used to implement a dedicated interface with the image sensor system. This interface read lines of digitised data from the sensor hardware using an Inmos link interface operating at 20 Mbits/s. The remaining two links from this transputer were connected to 32 bit transputer input and output cards. Signals from these cards were used to control the two axis motor system and to provide control inputs to the image sensor hardware.

A separate rack system was constructed to hold all the components of the laser scanning subsystem. Software control of this module was provided by a T414 transputer, referred to in the configuration file as “scan”, connected to the root processor. The laser scanning software provided control inputs to the dual axis scanning hardware, via an Inmos link adaptor. A synchronisation input to the laser scanner unit was also provided from the sensor drive logic using an Inmos CO11 link adaptor connected to an available link of the T414 transputer.

7.4.2 Software Configuration

The software configuration of the network refers to how individual software tasks were mapped onto the available physical processors. Communications between tasks on different processors could only take place using one of the hardware links provided by the transputer processors. However, no such restriction was placed on the number of software channels used to connect tasks resident on the same processor. The software configuration of the radiograph digitisation system is illustrated in Figure 7.3.

All hardware modules were defined as dummy tasks, whose operation was identical to software tasks, but which required no program code to be downloaded during bootstrapping.

Software processes controlling hardware which requiring uninterrupted execution during image acquisition were distributed around the network in such a way as to ensure no contention for processor time. This ensured that the laser scanning software carried on operating during image acquisition so that uniform illumination was maintained.

A specially written server task was implemented on the host PC which allowed communication between the processor network and software running on the PC. During bootstrapping this task was responsible for downloading bootable code to the processor network. Subsequently, the host server task handled all communications between the PC and the transputer network.

A single software task, defined in the task image file "rootask.b4", executed on the root processor. This task performed two main functions depending on whether the system was being used for image acquisition or image processing and analysis. It acted as a multiplexer task during image acquisition, controlling the passing of data and control signals around the network. During image processing operations the root processor acted as the master processor in a distributed image processing network.

Three software tasks, used to control the image capture subsystem, were resident on the transputer designated "proc2". The software task defined in the file "capture.b4" was used to control the acquisition and processing of raw pixel data from the image sensor hardware. This task was assigned high priority so that it executed uninterrupted during capture of a line of pixel data. The task "IO.b4" handled all system I/O to the 32 bit input and output cards. This provided low level software control over the motor subsystem and the CCD sensor drive hardware. Finally, the software task "subtask.b4" acted as a multiplexer task during image acquisition, passing data and control signals between the root processor and the external hardware modules. It also acted as part of the distributed processing system during image processing operations.

The T805 transputer on the MosaiQ frame store board was used to run a single software task, controlling the display of all images produced by the system. The MosaiQ engine software distributed with the frame store provided a set of image display and manipulation functions programmed in Parallel C [5]. This software was

provided with its original source code to allow user adaptation of the system software. A certain amount of modification of the MosaiQ engine software was required in order to fix bugs contained in the original code and to allow a higher display resolution than was offered by the standard MosaiQ set up.

The laser scanning software also consisted of a single task running on the T414 transputer board. This allowed the scanning mechanism to operate independently of all other software during image acquisition. Each scan of the laser spot was initiated by a single input signal from the sensor synchronisation hardware.

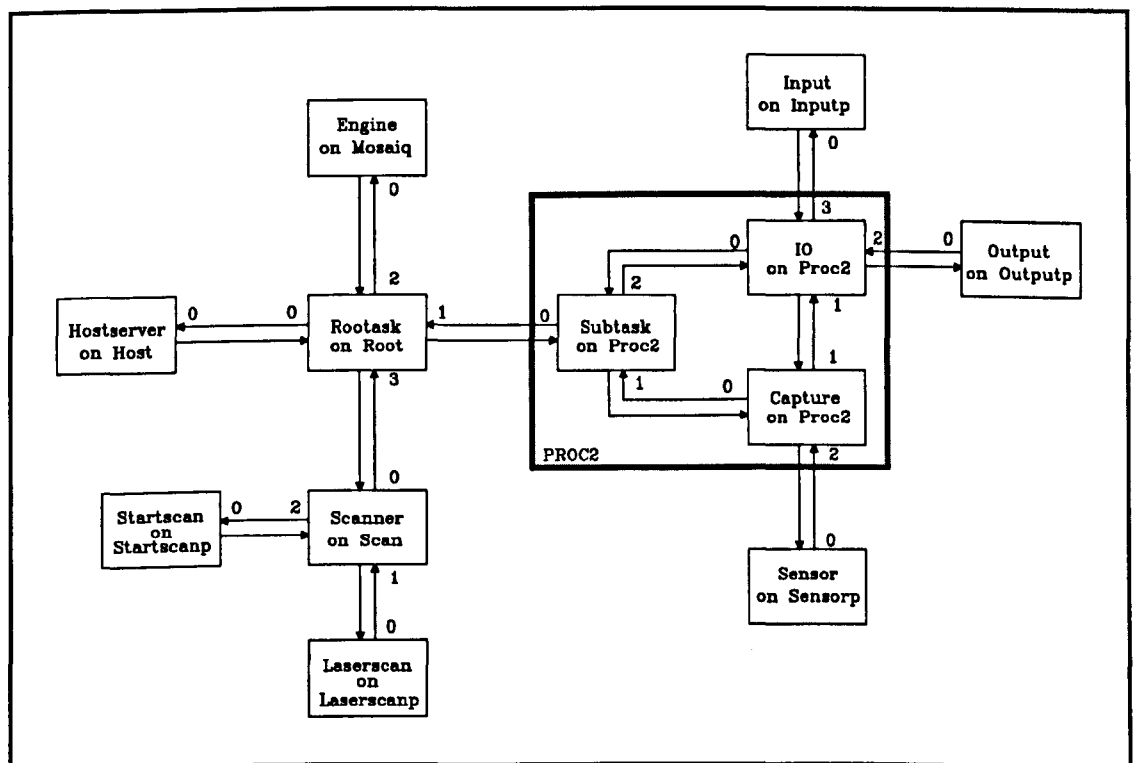


Figure 7.3 Distribution of Software Tasks onto Physical Network

7.4.3 System Services

The connection of system services signals in the processor network defined the network hierarchy. This determined how error messages were propagated up the network and reset and analyse signals were distributed downwards. A simple connection scheme was used in which all processor were connected “down” from the root transputer, with peripheral hardware modules being connected as subsystems. All peripheral hardware modules only required a reset input line and ignored subsystem analyse and error signals.

This connection method ensured that reset signals generated by the root processor were propagated down the network to all hardware components, whilst analyse signals were only passed to its descendent transputers. Error signals from any of the processors were propagated up the network hierarchy, but errors generated by peripheral hardware were ignored.

Any individual processor in the network could generate a subsystem reset signals in software which was used to reset hardware modules connected to its subsystem port. Hence, peripheral hardware could be reset after an error without having to reboot the entire processor system.

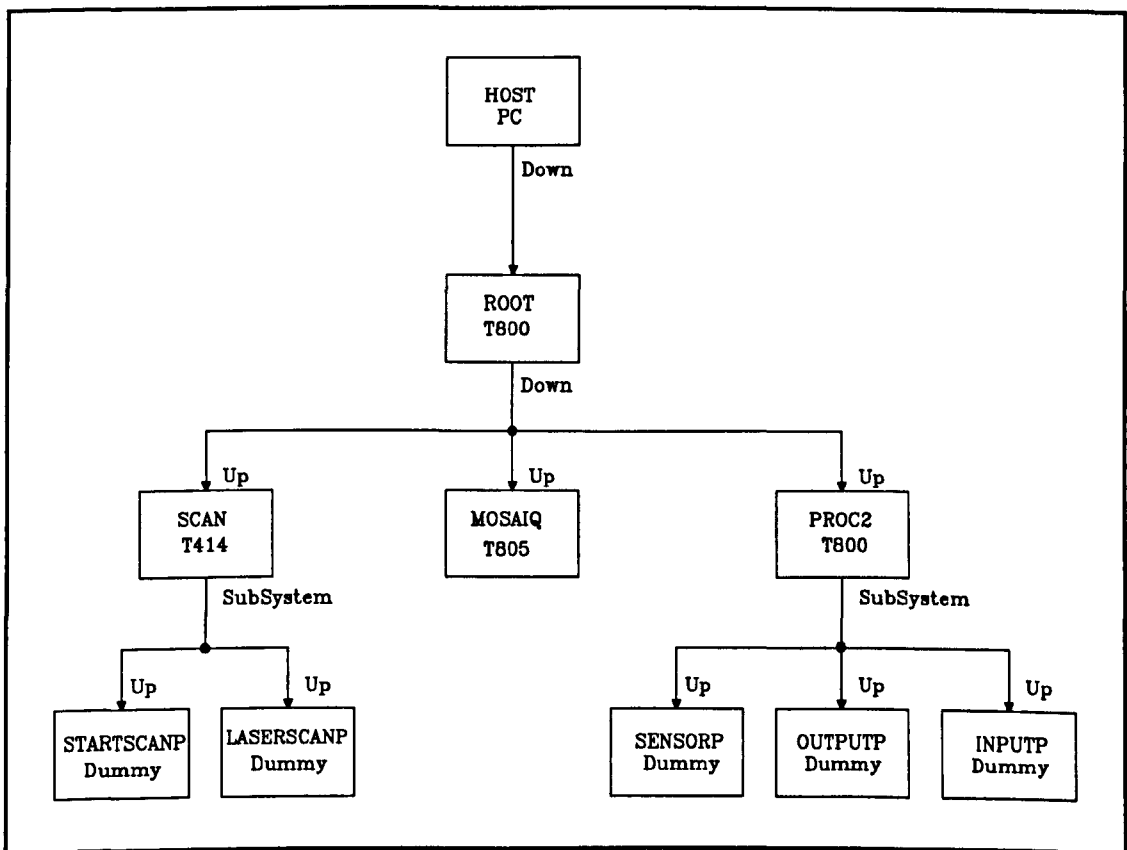


Figure 7.4 System Services Connection Scheme

One problem with this system was that when more than one piece of external hardware was connected to a transputers subsystem port they could not be reset independently. This resulted in situations where resetting one component affected the operation of other components in the system. For instance, if the sensor subsystem was reset at the start of a data acquisition cycle, both the input and output board were also reset.

Resetting the output board cleared all the output data signals to the sensor drive hardware, effectively setting the exposure time count back to zero. To avoid this type of problem it was necessary to closely monitor the effects of any subsystem reset operations.

7.5 Host Link Server

Standard 3L Parallel C provides a limited library of I/O functions to enable user interaction with the transputer network via the host PC. This library is not extensive enough to provide the kind of flexible user interface required for this project. In particular, no graphics functions are provided, so that calls to MS-DOS graphics functions are necessary to produce all but the most simple screen output.

It was decided to write a separate user interface module to run concurrently on the host PC. This user interface software was written in Borland Turbo C++, which provides a rich library of file and graphics I/O functions [6]. This language is an object oriented extension of the basic C language and includes all of the features present in standard implementations of C.

Normally all I/O between the network and the host is handled by a dedicated server task, known as the “afserver”, running on the host. To implement the required user interface it was necessary to write a new server routine, known as “hostserver”. This server task was responsible for booting the transputer network, downloading executable code and then handling any data passed between the host and the root processor.

As discussed in Section 7.4.1 the root processor of the network was a T800 transputer, which was resident in one of the available bus slots of the host PC motherboard. The bottom eight words of this transputer’s RAM¹ space were reserved for the channel control words to its four external links [7]. The host interface to the root transputer was implemented by memory mapping from the PC address space to

¹ Random Access Memory

the channel control words for link0 of the root processor. This memory mapping scheme is shown in Table 7.1.

The sequence of operations necessary to operate a transputer network involves first resetting the network, then waiting for the memory configuration sequence to execute, before downloading bootable code to the root processor. Bootable code consists of a short bootstrapping routine followed by executable program code. The bootstrap data defines the required network configuration and routes executable code to a destination processor.

PC Address	Read/Write	Description
\$150	Read	Host Link Input Data
\$151	Write	Host Link Output Data
\$152	Read	Host Link Input Status
\$153	Read	Host Link Output Status
\$160	Write	Host system Reset
\$161	Write	Host System Analyse
\$162	R/W	Host DMA Enable
\$163	Read	Host System Error

Table 7.1 Memory Mapping Scheme Used for Host Link Interface

During normal execution of the system, the server task simply acted as a message passing interface between the transputer and PC environments. A simple protocol was implemented where each message was preceded by a single byte codeword followed by a variable length set of arguments.

7.6 Menu Driven Software Front End

The standard “afserver” task supplied with the 3L Parallel C software package uses the host PC simply as a terminal for user interaction with the transputer network. This involves taking user input via the keyboard and relaying output messages either to the PC monitor or to disk. Such a system makes little use of the functionality offered by a modern PC.

The development of the new server task allowed additional software to run on the host PC in parallel with the software running on the transputer network. This software created a flexible, menu driven environment for user interaction with the radiograph digitisation system. This environment offered mouse control over a series of pull-down menus and dialogue boxes, giving easy access to the extensive range of operations carried out by the system. Sample screen outputs from the system software front end can be seen in Figure 7.5 and Figure 7.6.

One limitation of the Turbo C++ compiler used for programming this PC environment was the absence of any routines giving direct control over mouse functions. Instead, mouse operations were performed by calling 8086 software interrupt codes. These interrupt codes were used to access the internal registers of the 80486 processor used by the host PC. The 80486 processor uses interrupt code 51 for access to all mouse hardware [8].

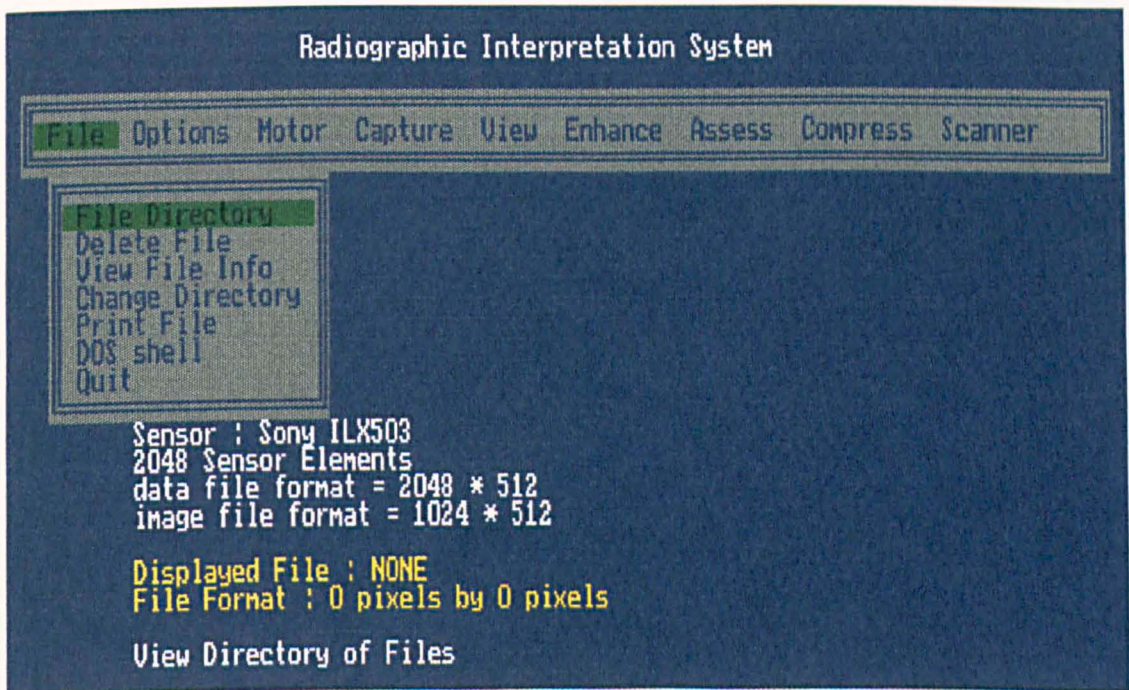


Figure 7.5 The Main Menu Screen of the System Software Environment

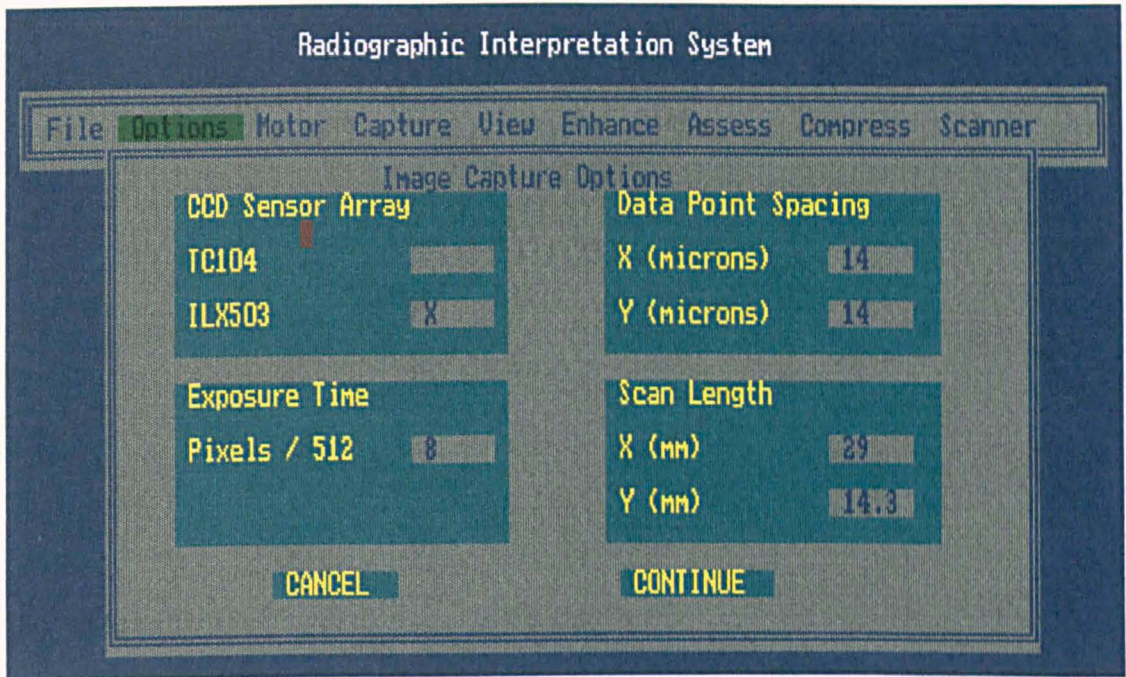


Figure 7.6 Sample Sub-Menu Screen of the System Software

The C data type **REGS** defines a structure for representing register values from processors in the 8086 processor family. 8086 interrupt codes can be called from within a C program using the **int86** function defined in the header file **<dos.h>**.

The mouse pointer was also used, in conjunction with the MosaiQ engine software to allow interactive analysis and manipulation of images displayed by the system. This function was achieved by monitoring mouse movements and translating them into pixel co-ordinates within the displayed image. These co-ordinates were then passed to the MosaiQ software which used the overlay image plane to display a moving cursor on the display monitor. This function allowed the user to measure features within an image or pin point areas of interest requiring further enhancement.

The software operating on the host PC also carried out several other functions, mainly concerned with low level file and printer operations. These functions are discussed in more detail in the relevant sections of this thesis, and are summarised in the following list.

- File Handling Utilities
- Image Archiving
- Data Compression
- Image Printing

7.7 Summary

In Chapter 7 details are given of the programming and configuration of the transputer network at the heart of the film digitisation system. This chapter includes a discussion of the principles of programming parallel systems and describes the development of an integrated software environment to provide control of all system functions from a PC based front end.

References

- 1 C. A. R. Hoare, "Communicating Sequential Processes", Prentice Hall Ltd, 1988.
- 2 3L Parallel C Version 2.1 User Guide, 3L Ltd. 1989.
- 3 Brian W. Kernighan & Dennis M. Ritchie, "The C Programming Language", Prentice Hall Ltd, 1978.
- 4 3L C Configuration Language Reference Manual, 3L Ltd. 1989.
- 5 MosaiQ Engine Software User Manual, Quintek Ltd, 1991.
- 6 Borland Turbo C++ Version 1.01 Programmers Reference Manual, Borland International, 1990.
- 7 The Transputer Data Book, Second Edition, Inmos Ltd. 1989, pp 217-218.
- 8 J. Reid, "Newnes PC User's Pocket Book", Butterworth-Heinemann Ltd, 1991, pp 169-181.

8. Image Capture Performance and Software

8.1 Introduction

This chapter discusses the methods and software algorithms used by the system during image capture and display. In designing this software emphasis was placed on producing a flexible system capable of digitising a wide variety of different film images with the minimum of adjustment by the user. In particular software algorithms were devised to extend the dynamic range of the system and to compensate for nonuniformity effects produced by the sensor and the optical system. Several sample images are used throughout this chapter to illustrate the operation and performance of the image acquisition system.

8.2 Image Format

When building up a two dimensional image it was important to choose appropriate scanning dimensions and spatial resolution to reproduce the required image detail without having to process excessive amounts of data. The selection of data file format was a trade off between several factors including sampling area, spatial resolution, data file size, image acquisition time and the monitor display resolution. In addition, image format was highly dependent on sensor dimensions and pixel size.

The approach used for image acquisition allowed the sampled spatial resolution to be varied to allow either quick, preliminary scans over a large area, or slower, high resolution scans over a previously identified area of interest. The image was built up from a series of vertical scans by the linear sensor array across the radiographic film. Horizontal spatial resolution was adjusted by varying the interval between sampled pixel outputs from the sensor element array. The vertical scan step size was always equal to the sampled pixel resolution, so that a symmetrical sampling grid was built up. This sampling method, allied to the direct contact image scanning method, allowed direct mapping between dimensions in object and image space.

This method of image scanning is demonstrated in Figure 8.1, where the shaded pixels represent data sampled at half the spatial resolution of the CCD sensor elements. It can be seen that this produces a four fold decrease in the number of acquired data points.

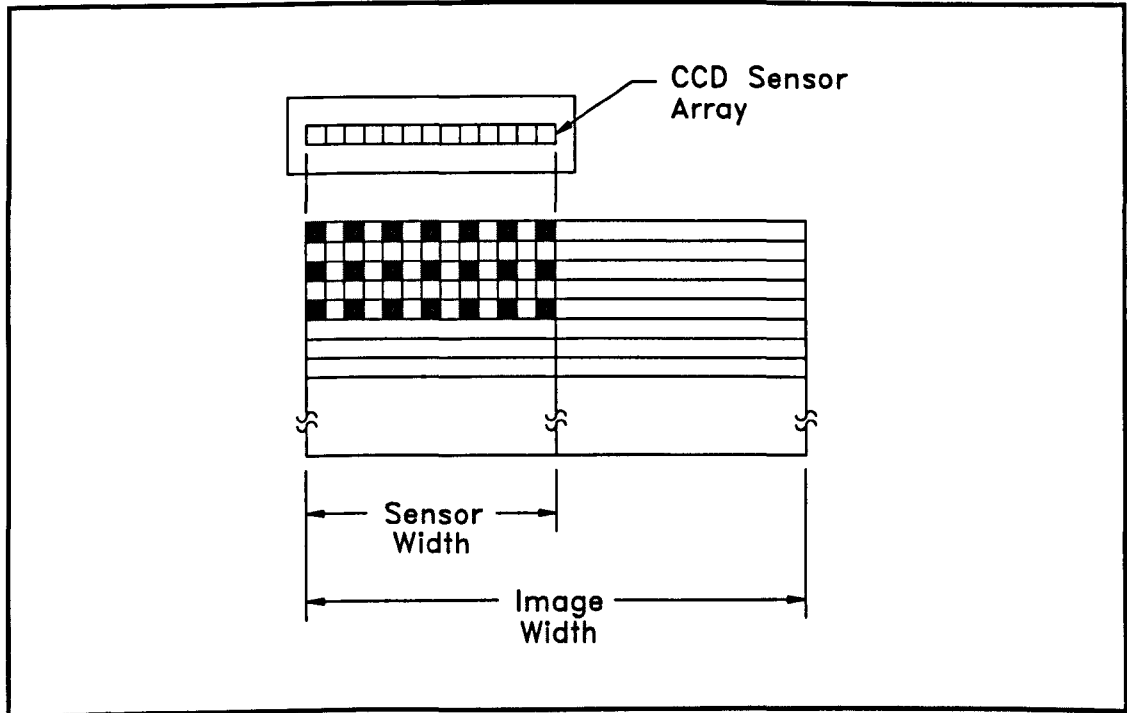


Figure 8.1 Method used for Scanning an Image

The maximum size of images displayed using the MosaiQ frame store was limited to 1024 pixels wide by 512 pixels high. Hence the scanned area was usually limited to some multiple of this display format, having an aspect ratio of 2:1. The image capture software allowed the option of either explicitly defining the dimensions of the scanned area and the sampled spatial resolution or choosing one of a number of pre-defined image formats. The pre-defined scan formats used for the ILX503 sensor are shown in Table 8.1.

Scanned Area		Pixel Format			
Width	Height	Columns	Lines	Spatial Resolution	File Size
28.7 mm	14.3 mm	2048	1024	14 μm	4 Mbyte
57.3 mm	28.7 mm	2048	1024	28 μm	4 Mbyte
28.7 mm	14.3 mm	1024	512	28 μm	1 Mbyte

Table 8.1 Scan Formats used for the Sony ILX503 Sensor

Examples of these different image formats are illustrated in Figure 8.2 and Figure 8.3, which show test card images captured by the film digitisation system. The first of

these images shows a section of the test card sampled at $14\ \mu\text{m}$ intervals, over an area of $28.7\ \text{mm}$ by $14.3\ \text{mm}$. This image was sampled using a single vertical scan of the sensor array across the film image. The captured image was displayed by the system on a 19 inch monitor, so that it was magnified and appeared approximately 16 times larger than the original image.

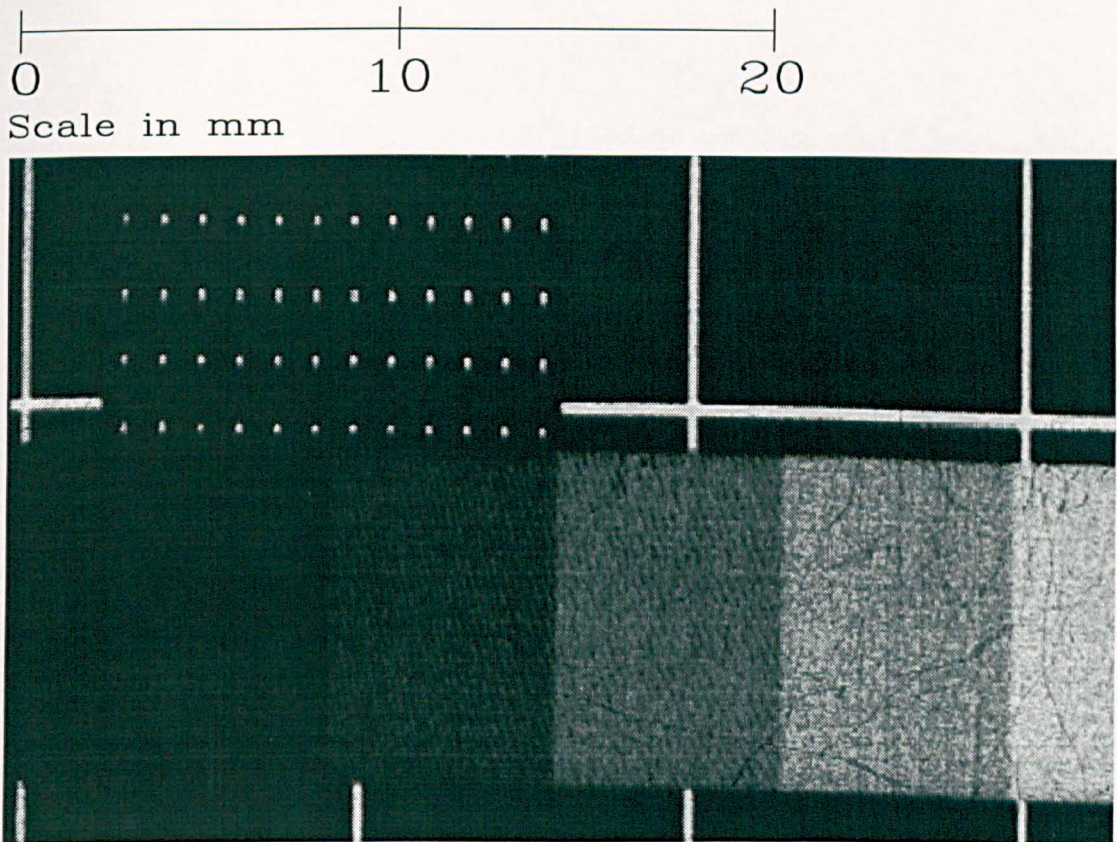


Figure 8.2 Test Card Image Captured at 2048 pixels by 1024 pixels by $14\ \mu\text{m}$ Format

Figure 8.3 shows a test card image sampled at $28\ \mu\text{m}$ intervals, over an area of $57.3\ \text{mm}$ by $28.7\ \text{mm}$. Since the ILX503 sensor array is only $28.7\ \text{mm}$ wide, it was necessary to undertake two separate vertical scans of the film in order to build up the full image. A slight misalignment of the two scans can be seen running down the centre of the resulting image. The reason for this misalignment arose from the difficulty in defining the exact extents of the sensor array, since the Sony ILX503 data sheet simply quotes a nominal sensor spacing of $14\ \mu\text{m}$. It was expected that this effect could be eliminated by experimentation to determine the exact travel distance

required by stepper motors between scans. The misalignment was measured on the digitised image and found to corresponded to less than $300\ \mu\text{m}$ on the original image.

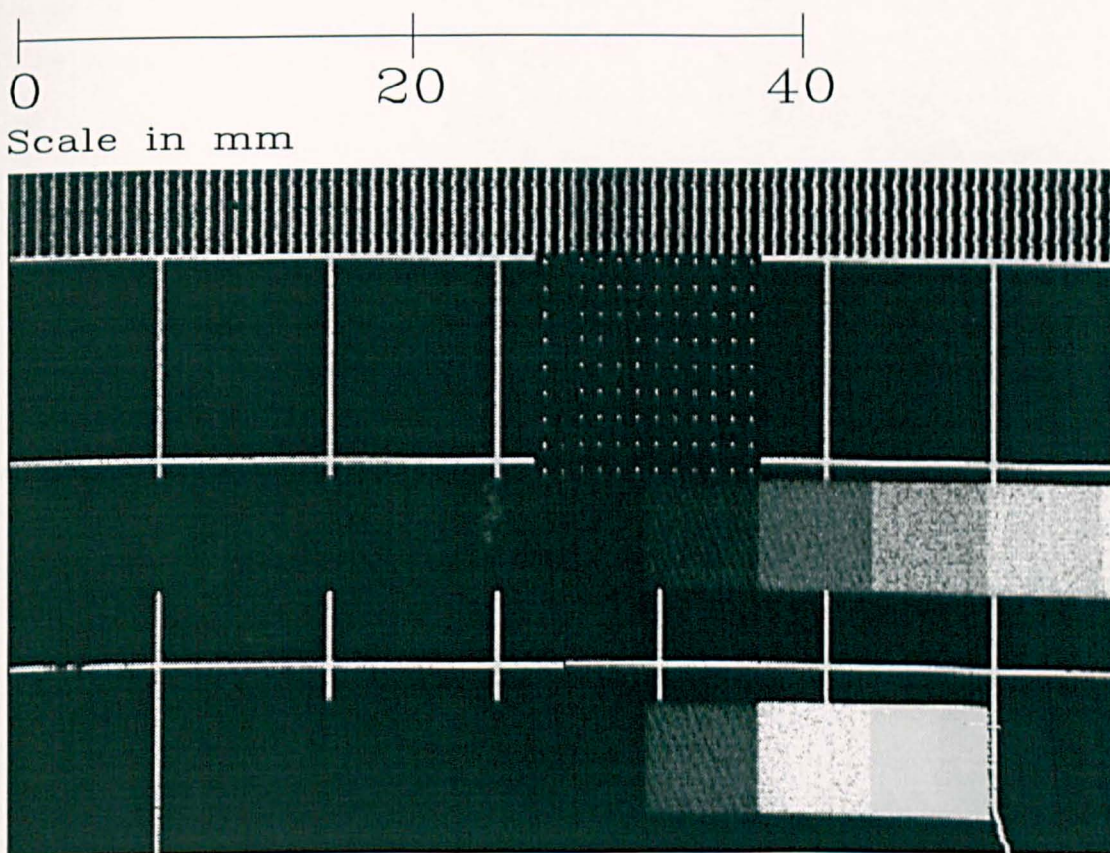


Figure 8.3 Test Card Image Captured at 2048 pixels by 1024 pixels by $28\ \mu\text{m}$ Format

8.3 Image Capture

Acquisition of a line of pixel data was initiated by sending a single byte message from software to the sensor hardware. The interface circuit then responded with a message consisting of data pixel values plus a number of leading and following dummy pixels. As discussed in Section 5.6, each pixel of this input data message was coded as two bytes consisting of a 12 bit intensity value and a leading Out Of Range (OTR) bit.

The image capture software located the image pixels within the data message by counting and discarding the leading dummy pixels. The number of dummy pixels depended on the type of sensor used. Data integrity was checked by examining the output from the black reference pixels contained in this dummy signal. During correct operation the output value from these pixels should be zero.

If required, spatial sampling resolution was reduced by discarding a proportion of the pixels in each line of data. The standard formats used during this project required sampling at pixel intervals of either one, two or four. The resulting pixel array was then passed to the host PC for file storage.

A hard copy of the captured image could then be obtained using a laser printer. The printer software allowed the image to be displayed either in landscape or portrait format and be printed either directly to the printer or to an EPS¹ file. All printed images included a brief summary of archive information including the associated data file name, the resolution of the image and the name of the operator.

When reading a line of pixel data the transputer software first had to be synchronised with the first pixel of a new line. On average, the time delay between receiving a request for data and reading the first pixel of a new line was one half of the exposure period. This meant that doubling the sensor exposure time increased the overall data acquisition time by a factor of three. The full data acquisition cycle performed by the image capture software consisted of identifying the first pixel of a new data line then reading pixel data from the hardware system, checking data integrity, compensating for calibration data and passing the data back to host PC for storage, before finally moving the stepper motors to next scan position.

The time taken to acquire a complete image depended on the sampling resolution, the image dimensions and the exposure time used by the sensor. As a rough guide to system speed, a 2048 by 1024 pixel image captured using an exposure time of 20 ms took approximately 70 seconds to digitise. This corresponds to around 70 ms per scanned line. The discrepancy between the exposure time and the time taken to acquire a line of pixel data was due to various processing overheads, and was independent of the exposure time.

¹ Encapsulated PostScript Format

8.3.1 Sensor Saturation

Light saturation of the CCD sensor was a major source of image distortion. During unsaturated operation of the CCD sensors the output signal voltage rose linearly as a function of incident light exposure. At higher exposures sensor saturation caused loss of image contrast in bright areas, and in extreme cases, signal charge “blooming” from adjacent pixels caused contamination of darker areas.

The designs of both CCD sensors used during this project attempted to eliminate blooming by ensuring that the output charge detection amplifier saturated before the signal charge held within potential wells of the CCD transfer registers overflowed to adjacent pixels. Just prior to saturation the characteristic gain of the amplifier began to decrease, causing a loss of contrast resolution. Increasing the incident light exposure above the saturation level had no effect on the output voltage produced by the sensor. Increasing light exposure still further eventually resulted in blooming, but this only occurred under very bright conditions.

Although this design helped to eliminate blooming it did produce a loss of image contrast at high exposure levels. This effect was not always immediately apparent since it often caused the “clipping” of high intensity peaks, producing areas of uniform brightness which were difficult to distinguish from legitimate data. To eliminate the undesirable effects of sensor saturation it was necessary to devise some method of extending the dynamic range of the image capture system

In order to avoid sensor saturation an image capture algorithm was developed that reduced incident light exposure when imaging bright areas. The ADC used for pixel digitisation was calibrated so that a full scale output signal corresponded to the exposure level just before sensor saturation, when linearity was lost. Hence, a saturation input to the sensor triggered the out of range (OTR) error flag on the ADC.

The image acquisition software responded to this OTR flag by reducing the incident light exposure and re-scanning the image line. Incident light exposure was controlled by a combination of laser diode output power, sensor exposure time and the action of the scanning laser spot. For a given laser power, incident exposure was altered in discrete steps by adjusting the number of scans across the sensor elements by the laser

spot during each exposure period. The corresponding exposure time was adjusted to match the time taken by the scanning action of the laser spot.

This algorithm resulted in each scan line having an associated brightness factor, corresponding to the number of scans completed by the scanning laser during the exposure period. This brightness factor was stored as the first element in each line of pixel data. When displaying this data the pixel intensity values were normalised to take account of changes in exposure between individual scanned lines.

Unfortunately, the out of range flag on the ADC developed a fault so that this method of eliminating sensor saturation could not be tested. However, the algorithm is reasonably simple and, once a new ADC is available, the system should require very little modification to get it working

8.3.2 Laser Scanning Algorithm

The laser scanning software provided a ramped voltage input to the horizontal galvanometer in order to project a laser line across the sensor elements. The PID controllers used to drive the scanning galvanometers were calibrated to produce a uniform scan speed. Acceleration and deceleration of the scanning spot at each extremity of the projected line resulting in bright spots being formed. Obviously, it was essential that these bright spots did not occur in the vicinity of the CCD sensor array, since this would result in nonuniform illumination of the image.

The laser power supply provided a relay switch to allow remote operation from software. However the relay response was too slow to allow the laser diode to be switched on and off between scan movements. Instead, to avoid unwanted light exposure, the laser line was extended well past each end of the sensor element array.

The laser scanning algorithm produced an 18° full angle deflection of the laser spot, corresponding to a projected line length of approximately 85 mm. The time taken to complete a single scan of the laser spot was approximately 5.8 ms.

8.3.3 Sensor Calibration

As discussed in Section 3.4 of this thesis, one the major sources of imaging anomalies produced by CCD sensors, under unsaturated illumination, is gain nonuniformity.

Gain nonuniformity is a consequence of each individual photosensitive element within an imaging array having a slightly different characteristic signal gain. Another source of error in a practical image scanning system is illumination nonuniformity. This may be caused by light source characteristics or shading in the optical path. Both these nonuniformity effects, to a first order approximation, increase linearly with overall illumination level.

Tanaka [1] proposed a hardware method of compensating for CCD sensor gain nonuniformity. This method consists of a two step approach, first calibrating the imaging system under uniform illumination, then applying this calibration data to the CCD output signal during image capture. The calibration procedure used a proportional gain control system to force the output signal to a constant value under uniform illumination conditions. The resulting series of proportional gain constants derived for each separate pixel during calibration were stored as digital values in memory. These constants were then applied to signal data during image capture operation, in order to compensate for gain nonuniformity. Since this calibration method uses a simple proportional gain compensation factor, it can also be used to compensate for light source nonuniformity and shading effects caused by attenuating elements within the optical path.

A software implementation of this calibration procedure was developed and tested in an effort to improve output image quality. The software algorithm applied a weighting factor to each pixel of raw image data to compensate for nonuniformity and shading effects. These weighting factors served to equalise the signal gain of each imaging element within the CCD array.

Consider an array of N photo sensitive elements receiving a uniform incident light exposure, E . The resulting signal voltage from an individual pixel element, n , with characteristic signal gain, $k(n)$, is digitised to an intensity value, $I(n)$. Neglecting dark signal noise effects, equation (3.3) becomes :

$$I(n) = k(n).E \quad (8.1)$$

If a weighting factor, $G(n)$, is applied to each pixel intensity value to produce a transformed intensity, $T(n)$, then :

$$T(n) = I(n).G(n) \quad (8.2)$$

Under uniform lighting conditions this array of weighting factor should act to produce a uniform transformed intensity distribution for all values of n . Hence, during calibration all values of $G(n)$ should be chosen to produce a constant reference output. A convenient value for this reference output is the average intensity output of the untransformed data set, I_{MEAN} .

$$I_{MEAN} = \frac{\sum_{n=0}^N I(n)}{N} \quad (8.3)$$

The required weighting factor then becomes :

$$G(n) = \frac{I_{MEAN}}{I(n)} \quad (8.4)$$

The calibration procedure involved applying a uniform luminous input to the system, by placing a neutral density filter between the light source and the CCD sensor. Equation (8.4) was then applied to the resulting intensity distribution to produce an array of weighting factors, $G(n)$. This calibration data was saved in the form of a transformation look up table. During normal operation of the image capture system Equation (8.2) was used in conjunction with the calibration data to produce an array of compensated pixel values, $T(n)$.

The effects of this calibration procedure are illustrated in Figure 8.4, which compares two lines of digitised data produced by the sensor system, one calibrated and one uncalibrated. In this example light source nonuniformity was exaggerated in order to accentuate the effects of system calibration. This was done by slightly tilting the scanning laser line, so that one side of the sensor array received brighter illumination than the other. Both data lines were imaged under the same lighting conditions, with no object present between the light source and the sensor.

The uncalibrated scan line in Figure 8.4 clearly illustrates the effects of both sensor gain nonuniformity and light source nonuniformity. Gain nonuniformity tends to produce small variations in sensor output between each individual sensor element,

whereas light source nonuniformity produces a gradual variation of the overall illumination level. This had the effect of producing a rather noisy output signal. In comparison, the calibrated scan line was seen to produce an almost uniform intensity output across the entire length of the sensor. Variation of the overall illumination level, by changing the output power of the laser source, produced similar results.

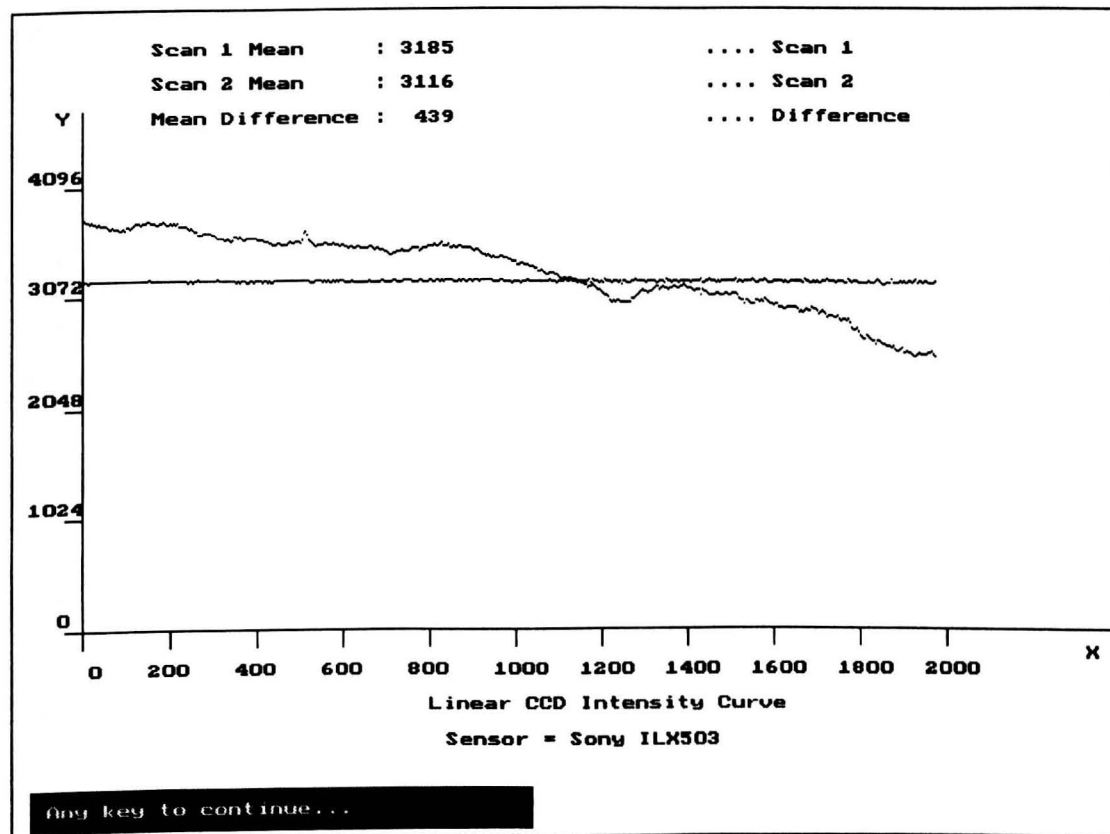


Figure 8.4 Compensation for Light Source Nonuniformity

The benefits of system calibration are particularly noticeable when attempting to image low contrast features. In such cases nonuniformity effects from both the sensor and the light source may obscure legitimate image data. This effect was simulated by placing a thin piece of wire in front of the sensor, using the same lighting conditions as in the previous example. Figure 8.5 illustrates the results of this test procedure.

It can be seen that the image of the wire produced by the uncalibrated system is distorted by nonuniformity effects, so that it is difficult to distinguish from the underlying noise signal. Although it is clear that some feature is present in the image, the actual shape of the wire cannot be accurately assessed. However, the calibrated

scan clearly resolves the wire from the background image and shows that the image has a regular intensity profile.

In practice, when the scanning laser system is correctly adjusted, the effects of light source nonuniformity are not so pronounced as in the previous examples, so that calibration is not always necessary to produce acceptable images. However, a slight misalignment of the scanning beam can significantly degrade light source uniformity and reduce image quality. The simple calibration procedure, which only requires a few seconds to complete, ensures that this effect is always eliminated.

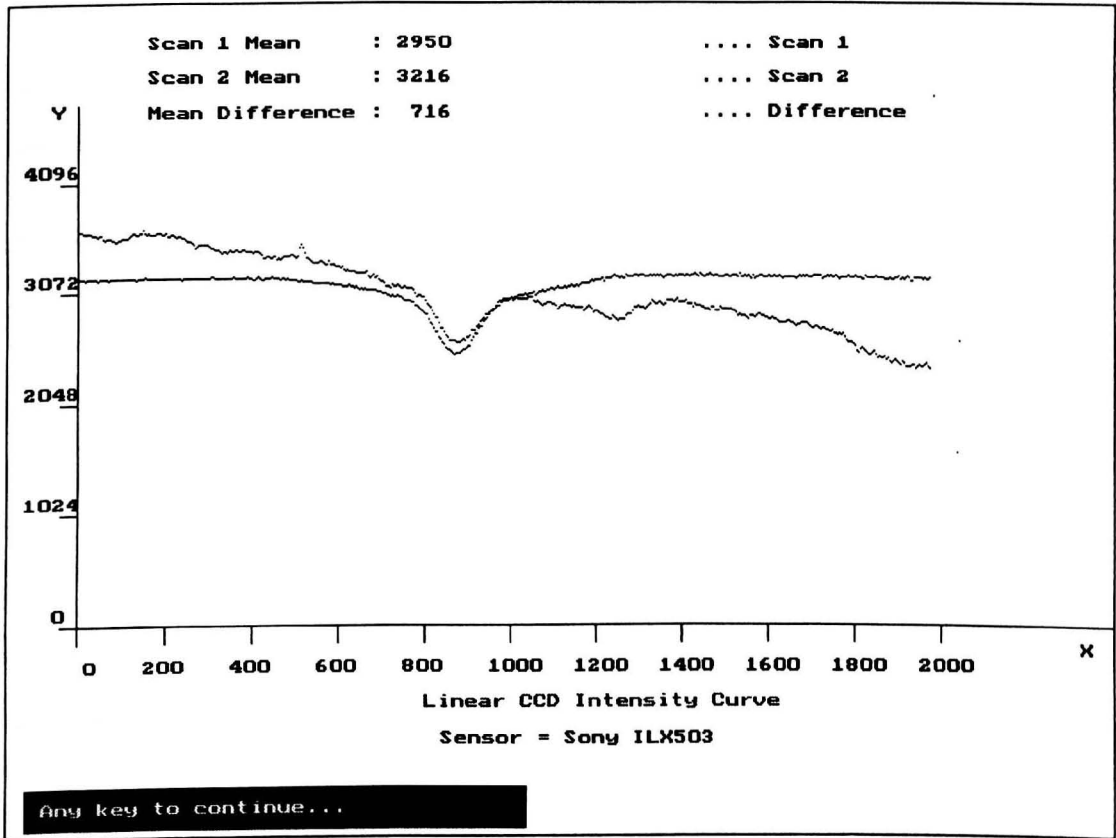


Figure 8.5 Feature Enhancement due to System Calibration

The effects of sensor gain nonuniformity are reasonably minor, typically less than 5 % of the output intensity level. This effect is not usually visible on a raw image produced by the system. However, this noise becomes more pronounced when certain image enhancement algorithms, such as edge enhancement, are applied to the image. These algorithms rely on differential operators which tend to accentuate abrupt changes in contrast, so that the effects of noise are amplified. Under these conditions system calibration may produce a significant increase in the quality of the processed image.

8.4 Image Display

Digital image data captured by the radiograph digitisation system at a contrast resolution of 12 bits/pixel, was displayed using a MosaiQ frame store. The MosaiQ was only capable of displaying images at a resolution of 8 bits/pixel. This necessitated a reduction in contrast data, and an associated loss of image fidelity. However, it should be noted that the human visual system is only able to distinguish somewhere between 100 and 200 separate grey level graduations at any one time, with any smaller variations being perceived as a uniform intensity. Since an 8 bit image quantises contrast data into 256 different grey scale levels, the images produced by the system were judged to be acceptable. The additional contrast data produced by the system was used by the image processing software to enhance low contrast areas within an image.

The MosaiQ frame store was driven by a software engine operating on a dedicated, T805 transputer. The image scanning software passed lines of raw image data to the MosaiQ processor as the image was built up. This data was transformed by the MosaiQ software to an array of 8 bit pixel values which were displayed on the monitor.

The amount of detail displayed on the monitor was also limited by the screen resolution of 1024 pixels by 512 pixels. When scanning large areas, the dimensions of the captured pixel array often exceeded the display format of the monitor. In this case the image display software scaled the image in order to display the entire scanned area. Once the capture operation was complete, scaled images could be viewed at full resolution by zooming in on a selected area. Zooming was controlled interactively using a mouse controlled cursor to window an area of interest on the monitor screen.

An example of this zoom feature is shown in Figure 8.6, which shows an expanded view of the test card pattern previously shown in Figure 8.2. The dotted grid pattern shown in this image has a spacing of 1 mm between each dot.

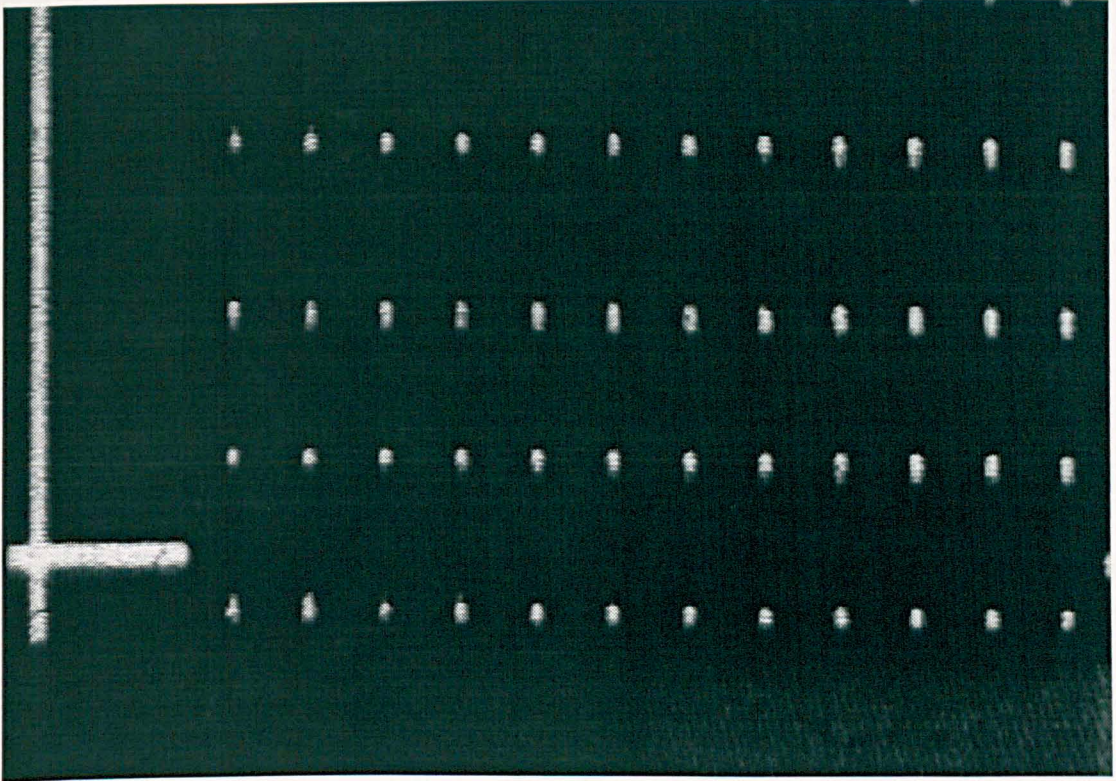


Figure 8.6 Zoomed Image Showing the Top Left Corner of Figure 8.2

8.5 Image Archiving

Any industrial radiographic image is of no practical use unless some information is known about the object being viewed. It is standard practice to place lead identification markers, containing a weld identification number, on an object being radiographed. These markers are effectively opaque to ionising radiation so that they appear in the processed radiograph as bright symbols which are clearly discernible from the surrounding image. However, the amount of data that can be displayed by this method is severely limited by the size of the image.

In practice a great deal more information is required to accurately interpret a radiographic image. When considering radiographic images of welded joints, relevant documentation usually includes details of the components being joined, the welding technique used and the method of radiography used. All this information must be supplied with the image in order to give an a comprehensive knowledge of what is being viewed. In conventional radiography this information is entered on a standard documentation form which must be filed along with each set of images. Obviously

this may lead to problems if a radiograph becomes separated from its information sheet.

Detailed specifications of the minimum documentation which should accompany any radiographic image are given in many industrial standards, such as BS2600, AESS6001 and ASME V [2][3][4]. The documentation system used in this project was based on a standard information sheet format used by BNFL¹. This documentation sheet comprises three separate sections giving information on the components being imaged, details of the welding procedure and the radiographic technique used.

Details relating to the two components being joined included the component material, its dimensions and location. Weld data included the weld identification number, welder identification and the welding technique used. Relevant information concerning the radiographing process included the identity of the radiographer, the radiographic technique used, the type and strength of radiation source used, details of any screen or filter, beam geometry, film type, film density limits, measured IQI sensitivity and the type of processing used. A separate section of data was included, concerning the film digitisation process. This data included the image file name, the operators name, the time and date of digitisation, the type of sensor used, details of any post processing applied to the image and the dimensions and sampling resolution of the image.

Whenever a new image was digitised, the system prompted the user for all the relevant documentation data. This data was then stored by the system as a series of data fields at the start of the image file. The archival system used a variable size file header so that extra documentation data fields could be added to the system without affecting the compatibility of older files. The data input routine allowed the user to scroll through the data fields and edit the selected information.

¹ British Nuclear Fuels plc.

Whenever an image was retrieved from a file the associated archive information was read by the system. A summary of the characteristics of the currently displayed image were then displayed on the main menu screen. The full documentation for any image file produced by the film digitisation system could be viewed or printed by selecting the appropriate option from the main system menu and then selecting the required file name from a list of image files contained in the current directory.

Documentation data was automatically appended to any new image file produced as the result of some image processing operation on an existing image. This ensured that all archive information was preserved in the new image file.

The system software also allowed annotations to be displayed with the image. These comments were displayed in the overlay plane of the MosaiQ frame store, so that they were superimposed over the digitised image. The position of these annotations was selected interactively using the mouse to position a cursor within the displayed image. These comments were also stored within each image file.

8.6 Digitisation of Industrial Radiographs

Since the aim of this project was to produce high quality digital images from industrial radiographic films, the most significant measure of performance involved the assessment of these digitised images. This section presents a selection of digitised images reproduced from radiographic films. All the radiographs considered in this section relate to welded joints in steel, either between pipes or plates.

The very high spatial sampling resolutions used by the system caused data files to be extremely large when scanning extended areas. Processing and displaying such large amounts of data is complicated and time consuming, so that initial tests were carried out on relatively small areas. The small sampling area of these images leads to a degree of difficulty in assessing them, since they cannot be viewed in the context of surrounding image features. For this reason a certain amount of explanation of the images presented in this section will be required.

Industrial radiographic film images tend to contain a large range of contrast data. In particular, joint geometry often causes the edges of a radiograph to receive very little exposure to ionising radiation, so that film density in this area approaches the base fog

level of less than 0.5. Densities measured around the centre of the radiograph may often be as high as 4. This means that certain areas in a radiograph may be around 3000 times brighter than other areas.

The limited dynamic range of the sensor frequently caused dense areas to appear completely black, whilst the edges of the radiograph appeared white. In such cases very little contrast data could be discerned from the image. Since an entire line of pixels were scanned together, it was not always possible to vary the exposure time of the sensor between these different areas. The method used to eliminate these problems involved masking off particularly bright areas of a radiograph using masking tape, so that just the dense area of interest around the centre of the frame was visible. Although this procedure eliminated a certain amount of data from each image, it should be noted that standard radiographic procedure requires several images to be taken of a welded joint at regular intervals along its length, with only the central area of each frame being used for weld evaluation. In practice, the procedure of masking off bright areas is similar to that used by inspectors when viewing particularly dense radiographs.

The problem imposed by the limited dynamic range of the CCD sensor is illustrated by the image in Figure 8.7, which shows a section of a welded pipe joint. This image was acquired using a single vertical scan of the CCD sensor array across a radiographic film. It can be seen that the bright area at the right hand side of the image caused saturation of the sensor, so that some contrast information was lost. Reduction of the exposure time to avoid this saturation would result in a loss contrast in the dark area on the left hand side of the image.

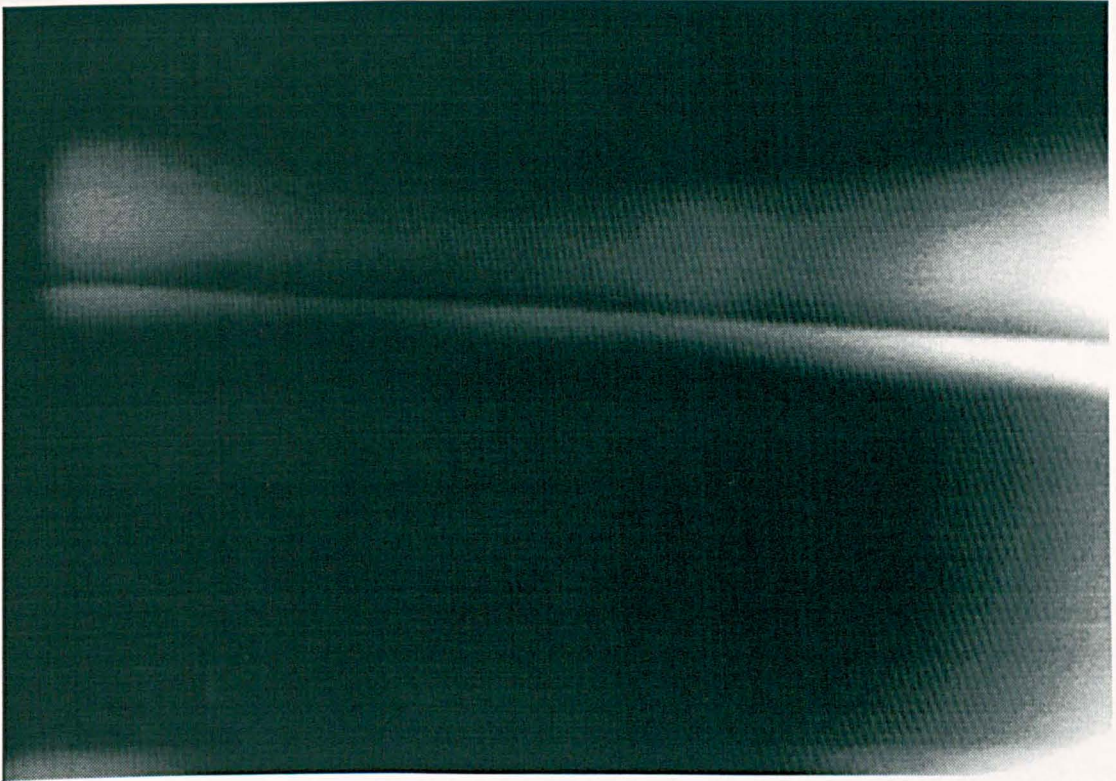


Figure 8.7 Welded Seam Showing Sensor Saturation at Right Hand Edge

During image digitisation, the radiographic film was mounted on a perspex window and held in place by a vacuum seal. Although raw perspex is initially transparent, it is a relatively soft material and is particularly susceptible to scratching which gives it a dull surface finish. For this reason it is not a particularly appropriate material for optical applications. This became apparent when attempting to image low density radiographs, when it was found that scratches on the perspex screen were reproduced in the digitised image. The effect could not be eliminated by calibration of the system, since the perspex screen moved as the radiograph was scanned across the sensor, so that the image distortion varied across the scanned image.

The effect of these scratches on the perspex screen can be seen in Figure 8.8, which shows a digitised image of a T joint between two pipes. The scratches and blemishes seen on the image are not present on the original radiograph, but were picked up during the digitisation process. Note the bright letter “A” at the centre of the image, which is an identification marker, produced on the image using a lead numeral in contact with the radiographic film. Since lead is virtually impervious to ionising radiation, the film directly beneath the lead numeral exhibits hardly any film blackening. It can be seen that the image of the pipe is of a similar density to that of

the lead numeral. This indicates that the density of the radiograph in this area is approaching the base fog level, so that very little contrast information can be discerned.

To eliminate the problem caused by scratches on the perspex screen it was proposed to construct a new screen out of glass. Glass is many times harder than perspex and consequently does not scratch so easily. Unfortunately, since glass is hard and brittle it is very difficult to machine into complex shapes, such as that used in the design of the perspex screen. In the light of this discussion it can be seen that some new design for a glass screen, or other film support apparatus, is required.

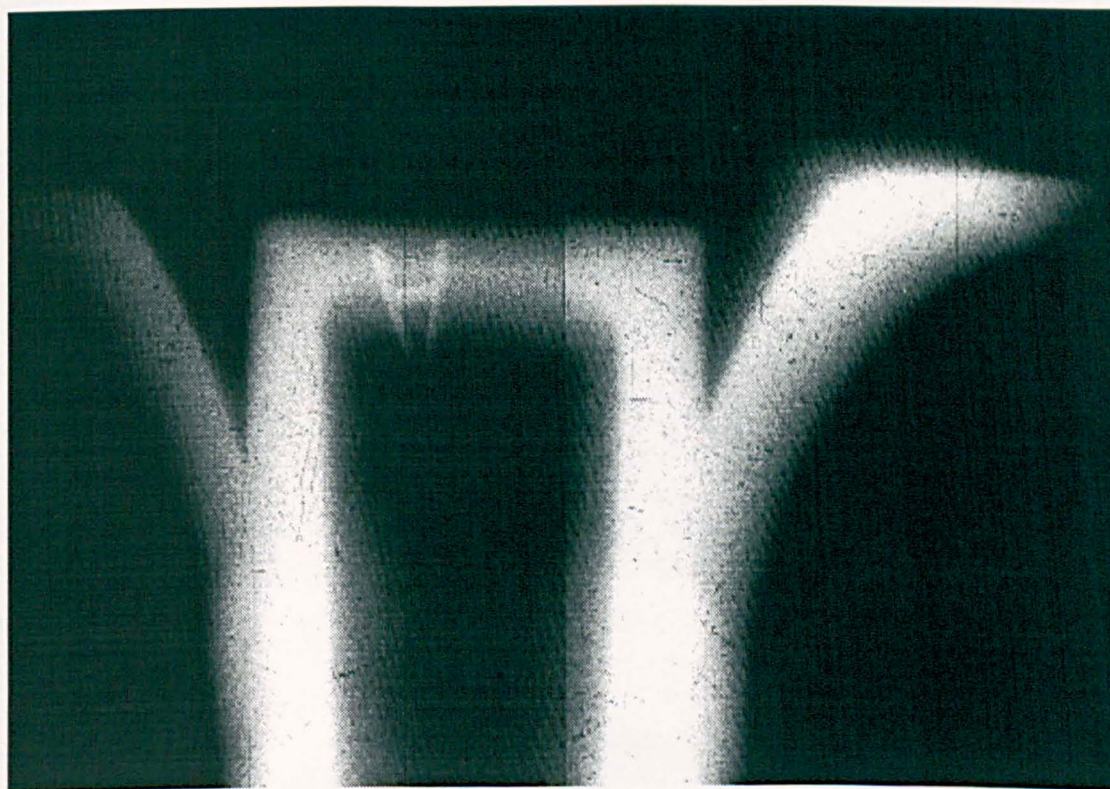


Figure 8.8 Image of a Pipe Joint Showing Scratches on Perspex Screen

Figure 8.9 shows part of a vertical welded seam between two plates. The welded seam is thicker than the surrounding metal and consequently absorbs more of the incident radiation, so that it appears as a bright area running down the centre of the image. Two small dark areas, of roughly circular dimensions, can be made out within the seam. The larger of these two features appears on the right hand edge of the seam and the other can be seen just above it in the centre of the seam. These dark areas represent defects within the weld material, probably caused by some slag inclusion or cavity.

The image in Figure 8.9, especially when printed, appears rather dark so that defects are difficult to resolve from the background image. However, the large amount of contrast data acquired by the 12 bit digitisation system gave plenty of scope for brightness correction, using image processing techniques. It was found that simple contrast enhancement techniques made the two defects clearly visible. Details of this work will be discussed at length in the next chapter.



Figure 8.9 Vertical Welded Seam With Inclusion

8.7 Summary

This chapter has discussed the software algorithms used during image capture and display. It was shown that image quality could be improved by a simple calibration routine which compensated for nonuniformity effects exhibited by both the sensor and the light source. A method of preventing sensor saturation and thereby extending the dynamic range of the sensor system was also proposed. Example images digitised by the system were also included in order to illustrate the quality of images produced during digitisation.

References

- 1 S. Tanaka, "A Need for Nonuniformity Correction in Solid State Sensors", SPIE Proceedings, Vol. 350, 1982, pp 275-282.
- 2 BS2600, "Radiographic Examination of Fusion Welded Butt Joints in Steel Plates", British Standards Institution, 1983.
- 3 Atomic Energy Standard Specification AESS6001, "Radiography of Welds in Plate, Sheet, Bar or Tube", U. K. Atomic Energy Authority, 1983.
- 4 ASME Boiler and Pressure Vessel Code, Section V, "Nondestructive Testing", American Society of Mechanical Engineers, 1992.

9. Image Enhancement

9.1 Introduction

The aim of this project was to produce digitised images from radiographic films in order to assist the operator in the task of image interpretation. The radiograph digitisation system produced an output image digitised at 12 bits per pixel. Consequently, the amount of intensity data contained within each image was vast. However, much of this data was superfluous, or served to conceal relevant image features. This chapter describes the methods of digital image enhancement which were used to recover visual information from the raw images.

The field of image enhancement is large, and many good introductions to the subject can be found in literature [1] [2]. These techniques have been applied to numerous imaging problems, including industrial and medical inspection systems [3] [4]. The techniques investigated in this research concentrated on interactive image enhancement to assist the operator in image interpretation. These techniques sought to recover as much useful information from an image as possible by removing spurious noise, increasing image contrast and accentuating relevant features.

This chapter outlines the basic concepts and techniques used in digital image enhancement and demonstrates how they were applied to the digitised radiographs produced by the image acquisition system. In addition details are given of how a network of transputers was used to implement these enhancement techniques in parallel, thereby significantly reducing processing time. The image enhancement techniques are illustrated by a number of example images which were created and processed during the course of this research.

9.2 Image Processing Fundamentals

9.2.1 A Digital Image Model

A digital image consists of a two dimensional pixel array, each element of which refers to a sampled intensity value at some point in object space. Individual pixel elements within an image may be addressed relative to their horizontal and vertical offsets from some arbitrary origin point. Stated formally, a digital image, I , is a

discrete, two dimensional light intensity function whose value at some spatial coordinates (x,y) is given by:

$$I = f(x, y)$$

The domain of I is dependent on the spatial resolution of the image under consideration. The spatial resolution of the radiograph digitisation system was limited by the display format of the MosaiQ frame store, whose maximum resolution was 1024 columns by 512 rows. Similarly, the range of I is determined by the digitisation resolution used to capture and display images. The radiograph digitisation system developed during this project used a 12 bit ADC, which quantised intensity values to one of 4096 possible grey levels. Hence, the function $f(x,y)$ is defined within the following limits.

$$0 \leq f(x,y) < 4096$$

$$0 \leq x < 1024$$

$$0 \leq y < 512$$

9.2.2 Image Enhancement Overview

The aim of any image enhancement algorithm is to transform some input image in such a way as to accentuate some desirable feature or property of the image whilst preserving the general appearance of relevant image features. In terms of the feature enhancement techniques used in this research, this process involves manipulation of image intensity values in order to maximise the amount of the useful information visible to the viewer.

In general these techniques were concerned with adjusting the distribution of pixel intensity values within an image rather than changing the underlying image structure. This necessitated the use of low level, pixel orientated algorithms. The large amount of pixel data contained within each image presented a high computational workload. For instance, a single 1024 by 512 pixel frame acquired by the radiograph digitisation system at a digital resolution of 12 bits per pixel occupied 1 Mbyte of processor memory. Hence, in order to limit processing time, the image processing techniques were confined to relatively simple algorithms. In addition, the application of parallel

algorithms using an array of transputer processors served to reduce to overall processing time needed to enhance an image.

9.2.3 Pixel Mask Operators

Many of the feature enhancement techniques used during this project involved modifying each individual pixel value in the image, based on some interaction with its neighbours. In mathematical terms, this involves transforming the input image, $f(x,y)$, into some enhanced image, $g(x,y)$, using the neighbourhood operator, T .

$$g(x,y) = T[f(x,y)]$$

The neighbourhood operator, T , is defined over a restricted sub-set of the input image, centred on the pixel being modified, with its value at all points outside this range being zero. The centre of this pixel mask is then moved across the input image, applying the operator at each pixel location. The most common type of neighbourhood oriented pixel operation involves using a pixel mask (also known as a filter or template) which effectively isolates a small area of the input image and applies some algorithm to the pixels contained within it. Although pixel mask operators may be any shape, by far the most common configuration is a square.

Consider the generalised 3 x 3 pixel mask shown below. The pixel being modified lies at the centre of the mask, and has co-ordinates (x,y) . A series of coefficients w_1, w_2, \dots, w_9 are applied to the input image pixels which lie within the mask boundary. The choice of coefficients determines the function of the mask operator.

w_1 $(x-1,y-1)$	w_2 $(x,y-1)$	w_3 $(x+1,y-1)$
w_4 $(x-1,y)$	w_5 (x,y)	w_6 $(x+1,y)$
w_7 $(x-1,y+1)$	w_8 $(x,y+1)$	w_9 $(x+1,y+1)$

The neighbourhood pixel operator, T , may then be expressed by the generalised operation:

$$\begin{aligned}
T[f(x,y)] = & w_1f(x-1,y-1) + w_2f(x-1,y) + w_3f(x-1,y+1) \\
& + w_4f(x,y-1) + w_5f(x,y) + w_6f(x,y+1) \\
& + w_7f(x+1,y-1) + w_8f(x+1,y) + w_9f(x+1,y+1)
\end{aligned}$$

If a square pixel mask operator with sides N pixels long, is applied to an image it is necessary to evaluate N^2 separate multiplications and additions at each location within the image. Obviously as mask size is increased the processing overheads soon become impractical. For example, the simple 3×3 convolution mask operator defined above requires 9 separate multiplications and additions at each pixel location within the image. When applied to a 1024 by 512 pixel image, this algorithm results in the evaluation of nearly 5 million mathematical operations. For this reason most commonly used mask operators are restricted to 3×3 , 5×5 or 7×7 dimensions.

9.3 Image Enhancement Techniques

9.3.1 Noise Reduction

Any digital image will contain a certain amount of spurious noise data, picked up during image acquisition. This noise is characterised by single "rogue" data points whose value varies widely from that of the surrounding data. Hence noise pixels appear as points within an image which are clearly defined from their surroundings.

The most simple method of noise reduction in digital images involves local pixel averaging using the pixel masks operator shown below.

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

This operator calculates the average value of all the pixels contained within the mask area, and applies this value to the central pixel. An associated effect of local pixel averaging is that legitimate high frequency data is also attenuated. Hence the smoothing of image noise also results in the blurring of object edges, leading to a loss of image definition. For this reason, any noise reduction technique should be used with great care.

This blurring effect may be reduced by using the technique of median filtering as suggested by Tukey [5]. This algorithm replaces each pixel with the median value of all pixels contained within a predefined mask area. The median of a data set is calculated by sorting all the values in ascending order and selecting the middle value. In this way any pixel value which is significantly larger or smaller than its neighbours will be discarded. Hence, median filtering preserves edge sharpness whilst removing isolated noise spikes.

9.3.2 Image Sharpening

Image sharpening is a term used to describe operations which enhance contours within an image so that the edges of features appear more pronounced than in the original image.

Any practical imaging system will be subject to degradation of resolving power which tends to produce some degree of image blurring. This effect may be regarded as an attenuation of high spatial frequency data. Image blurring is an averaging, or integrating, process in which sharp image points are diffused over a localised area. In contrast, image sharpening may be considered as the inverse of blurring, tending to concentrate diffuse features into sharp points by some differential process. In other terms, this process may be regarded as high frequency emphasis. However it should be noted that image noise is characterised by sharp, high frequency components. Hence, any sharpening operation will also tend to accentuate these noise pixels. For this reason any image sharpening algorithm must be applied with care so that it does not generate an excessive noise component.

The most simple differential operator that can be applied to an image involves calculating the difference between adjacent pixels in both the horizontal and vertical direction. These first derivative operators, Δ_x and Δ_y , may be expressed as :

$$\Delta_x[f(x, y)] = f(x, y) - f(x - 1, y)$$

$$\Delta_y[f(x, y)] = f(x, y) - f(x, y - 1)$$

Applying this pair of equations to an image is analogous to using the pixel mask operator shown below.

0	-1
-1	2

$$\Delta x + \Delta y$$

A useful image property used in many image processing techniques involves calculating the second order derivative, often known as the Laplacian operator, $\nabla^2 I$. An effective method of reinforcing image edges involves subtracting the Laplacian of an image from the original image data [6]. It has been shown that this process can be achieved using the 3 x 3 convolution mask shown below. [7].

0	-1	0
-1	5	-1
0	-1	0

$$I - \nabla^2 I$$

9.3.3 Contrast Enhancement

A common obstacle to accurate image interpretation involves images which, for some reason, exhibit low contrast. In these images most of the pixel data is concentrated within a small range of intensity values so that image features are indistinguishable from the surrounding background. Low intensity images are usually caused by some deficiency in the image acquisition process. Poor image illumination may result in dark images where most image data consists of low intensity pixels. Similarly, excessive light exposure during image acquisition may result in bright images containing very few dark pixels.

Contrast enhancement techniques seek to adjust the distribution of intensity levels within an image in such a way as to more evenly spread intensity data across the full range of pixel values. This operation involves a pixel mapping function in which all pixel values in the input image are transformed to produce a new image with increased contrast. This mapping function, T , transforms an input pixel value, r , into some new pixel value, s , using the relationship :

$$s = T(r)$$

If the original image is quantised into L separate grey levels, with values ranging from zero to $L-1$, then the transformation function, $T(r)$, must map each pixel in the input image to some new value within the same range.

$$0 \leq r \leq L-1$$

$$0 \leq s \leq L-1$$

9.3.3.1 Linear Contrast Enhancement

The most simple method of increasing image contrast involves a linear contrast "stretch". This algorithm works by identifying a subset of grey levels within the untransformed image which contains most of the contrast information. This restricted set of pixel values is then stretched using a linear mapping function so that it extends over the full grey level range. All grey levels not within the subset are transformed to either black (zero intensity), or white (maximum intensity).

This simple linear contrast enhancement algorithm is only effective when the majority of grey scale values in the input image lie within a relatively small range. This condition is not always satisfied in low contrast images. Sometimes grey scale values within an image are concentrated into two or more separate ranges, representing different low contrast objects. Images in this class may be enhanced using more complex, composite transformation functions consisting of a series of linear stretches. However, this composite linear stretch method is very cumbersome to implement and is rarely used.

9.3.3.2 Histogram Equalisation

A useful method of analysing the grey level distribution within an image involves plotting an image histogram. Image histograms evaluate the number of image pixels occupying each grey scale level within the digitisation range. Low contrast images are typified by image histograms containing sharp peaks over a restricted range of grey scale values, whilst other grey levels outside this range occur very infrequently.

Consider the case of an unprocessed image in the form of a pixel array of dimensions $M \times N$, quantised into L discrete grey levels. Histogram equalisation aims to

redistribute the grey level values within an image so that an equal number of pixels occur at each grey level. If an image exhibits this ideal "equalisation" then the number of pixels, n_r , having any grey level value, r , will be :

$$n_r = \frac{MN}{L} \quad (9.1)$$

In the case of the 4096 grey level, 1024 x 512 pixel images displayed by the film digitisation system, the number of pixels at each grey level will ideally be 128.

The actual distribution of grey levels within an image may be represented using the cumulative frequency distribution, where the number of pixels having a grey level less than or equal to some value, r , is given by the cumulative frequency, t_r .

$$t_r = \sum_{j=0}^r n_j$$

It can be shown that the ideal pixel distribution given in (9.1) can be approximated by using the following transformation function [8].

$$T(r) = \frac{L t_r}{MN} \quad (9.2)$$

In practice, each transformed grey level must be an integer value in the range from zero to L , so that equation (9.2) becomes.

$$s = \max \left[0, \text{round} \left(\frac{L t_r}{MN} \right) - 1 \right] \quad (9.3)$$

where $\text{round}(x)$ represents the closest integer to x and $\max[]$ selects the largest value from the list enclosed by the brackets.

9.3.4 Edge Detection

A vital stage in the process of viewing and understanding an image involves the extraction of primitive shapes from the continuously varying image intensity distribution. Visual perception tends to respond most strongly to feature boundaries, or edges. These edges may be defined as localised changes or discontinuities in the intensity of an image.

The detection of edges may be considered as a basic method of feature extraction. As such it is often the first stage of a high level vision system. Although the stated aim of this research was to enhance images for inspection by a human observer, some basic methods of edge detection were investigated in order to assess the suitability of such images for input into a machine recognition and assessment system.

The most simple class of edge detection techniques involve the use of first order differential operators to create an intensity gradient map of the image, from which edges can be located. A second class of edge detection operators attempt to locate intensity gradient maxima by the zero crossing property of the second order derivative. Examples of such methods include the Laplacian Operator described earlier, and the Canny Operator [9]. Work carried out during this project only considered the first class of edge gradient detection operators.

A common approach to edge detection is to use two separate pixel mask operators sensitive to edges running in the vertical and horizontal directions. These operators yield a horizontal gradient component, G_x , and a vertical gradient component, G_y . The simplest such gradient operator consists of the 2 x 2 convolution templates proposed by Roberts [10].

0	-1
1	0

 G_x

-1	0
0	1

 G_y

A more sophisticated set of operators are the 3 x 3 Sobel edge detection filters illustrated below.

1	2	1
0	0	0
-1	-2	-1

Horizontal Sobel Operator, G_x

-1	0	1
-2	0	2
-1	0	1

Vertical Sobel Operator, G_y

The resultant intensity gradient of the image is then calculated from the horizontal and vertical gradient components as follows.

$$G(x, y) = \sqrt{G_x^2 + G_y^2}$$

Once this gradient data has been evaluated, some threshold operation must be applied, in order to distinguish between legitimate edges and background intensity fluctuations. The setting of this threshold value is crucial to the efficient detection of image edges and will generally vary depending on the characteristics of the image.

The Sobel operator also allows some estimation of the edge direction. The direction of the detected edge will tend to run normal to the direction of maximum gradient, so that it will be given by :

$$\theta(x, y) = \tan^{-1} \left[\frac{G_y}{G_x} \right]$$

Edge detection has been investigated by a great many researchers and many more sophisticated detectors have been proposed in recent years with comprehensive treatments of the subject existing in literature [11].

9.4 Software Implementation

9.4.1 Parallel Image Processing

Due to the vast amount of intensity data held in a single image frame captured by the film digitisation system, any algorithm applied to these images represented a significant processing workload. When carried out sequentially, even simple image processing algorithms required a large amount of processor time. Clearly significant advantages could be gained by the use parallel processing techniques to divide the computational workload between a number of processors [12] [13]. The approach taken in this research used a network of transputers to construct a parallel image processing system.

The image enhancement algorithms used in this project were essentially low level operations, concerned with the manipulation of individual pixel values rather than some global representation of the image structure. Hence, these operations could conveniently be divided into independent processing tasks, requiring little or no shared data. In practice, this division of labour was organised by splitting each image into a series of sub-images and distributing the pixel data between processors, so that each processor operated on a restricted section of the image.

The architecture of the transputer network allowed for no shared memory areas between software tasks executing on different processors. Hence, all pixel data had to be passed around the network using relatively slow Inmos link communications [14]. To minimise processing speed, algorithms attempted to minimise the need for intercommunication of pixel data between processes.

The structure of the distributed image processing network used during this project is illustrated in Figure 9.1. The master processor in the network, a T800 transputer with 2 Mbyte of RAM, was connected directly to the host PC. All image data files were stored on the hard disc of the PC and downloaded to the transputer network for processing. The master processor was responsible for the equal division of the pixel data around the network so that the processing workload was evenly distributed and bottlenecks were avoided.

A MosaiQ frame store board, based around a T805 transputer, was used to display all the images produced by the system. The MosaiQ board was not used for image processing functions, but was dedicated to controlling image display. The master processor was the only transputer which communicated directly with the MosaiQ frame store, so that it was responsible for the passing of all processed pixel data to the display software.

A second T800 transputer acted as a worker task, taking data and instructions from the master processor, operating on the data, and passing the results back for display and storage. This hardware structure allowed the processing workload to be equally shared between the two T800 transputers. However, all software was designed in such a way that additional software processes, identical to that of the worker task, could be added to the network configuration in order to increase processing speed, with very little need for code modification.

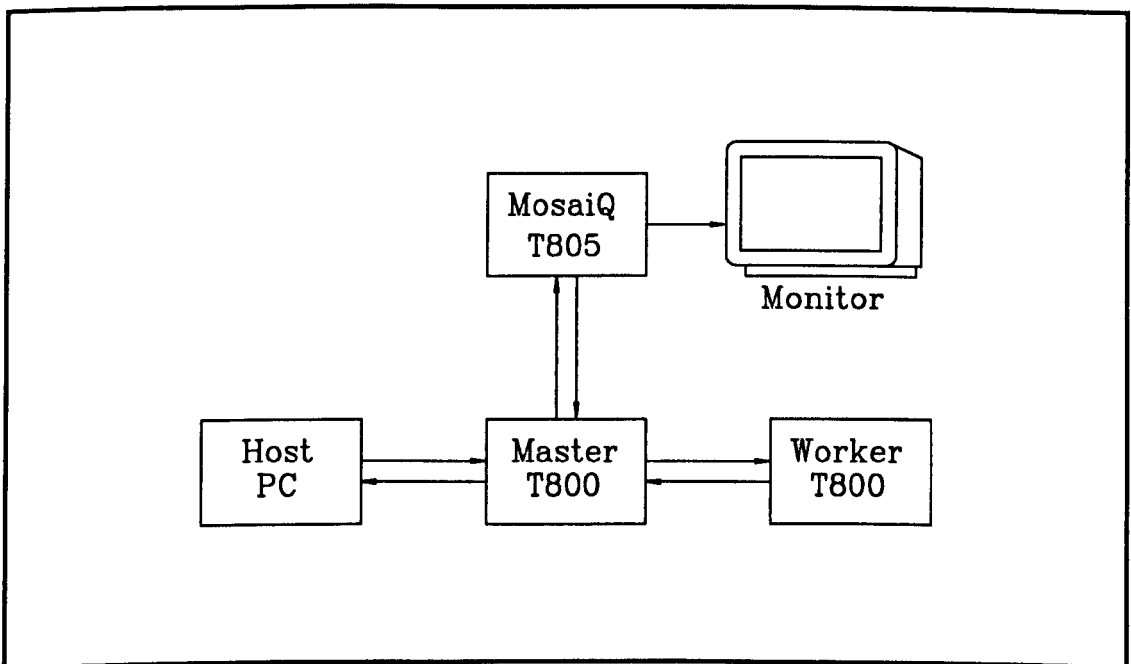


Figure 9.1 Distributed Image Processing Hardware Configuration

To avoid the processing of redundant pixel data the image format used by the system was restricted to 1024 pixel by 512 pixels, since this format represented the maximum display resolution of the MosaiQ frame store. Larger images were zoomed or cropped to this format using the image display software discussed in the previous chapter.

When applying a pixel mask operator to a single pixel within an image it is necessary to know the values of several neighbouring pixels. At the edges of an image certain of these neighbouring pixels are not defined, so that the pixel mask operator cannot be applied. This normally results in a small, unprocessed border region being left around the edge of the image. When considering a distributed image processing network, such as used in this project, it should be noted that each sub-image will produce an associated, unprocessed edge. Hence, in order to ensure complete processing of each image it was necessary to duplicate certain edge pixels held on separate processors so that the sub-images overlapped and no areas were left unprocessed.

Ideally the division of a computational task between n processing elements should reduce the overall processing time by a factor of $1/n$. In practice overheads associated with the communication of pixel data between processors will limit the attainable increase in speed. For instance, it was found that a parallel version of the histogram equalisation routine, using the two transputer system described above, produced a 42% reduction in the total processing time over the same algorithm executing on a single transputer.

9.4.2 Pixel Buffer

Any image processing operation requires access to pixel data held in memory. In the case of the image processing system developed during this project, all pixel data was accessed via the relatively slow external memory interface of the transputer. The large number of such memory accesses required to carry out simple enhancement algorithms represents a major processing overhead. As discussed in the previous section, many low level image enhancement algorithms rely on the application of pixel mask operators which adjust the value of each pixel in the image based on some interaction with a small number of neighbouring pixels. The application of a simple 3×3 pixel mask operator requires that each pixel value should be accessed nine times. In view of this situation, it is clear that the implementation of any such algorithm should attempt to minimise access to external memory.

A method was developed to ensure that, during the application of these pixel mask operators, each image pixel was only read from external memory once. This method used a series of buffer arrays to hold pixel data in fast internal memory during

processing. The transputer architecture ensured that each internal memory access was carried out in a single processor cycle, so that the overall processing time was drastically reduced.

This technique may be applied to any size of pixel mask operator, with a buffer line of pixel data being required for each horizontal row of the pixel mask. In the case of a 3 x 3 pixel mask operation, access to three lines of pixel data was required in order to process each horizontal line of the image. After processing each line of the image the bottom buffer array was updated with a new line of pixel data.

9.5 Results

9.5.1 Image Smoothing and Sharpening

Software code was written to achieve image smoothing using both 3 x 3 pixel neighbourhood averaging and median filtering. The benefits of these image smoothing operations were not immediately evident, since the original images did not appear particularly noisy and the results of smoothing were barely visible.

The advantages of smoothing image noise became apparent when applying differential operators to achieve effects such as sharpening or edge detection. These operators are sensitive to the sharp pixel gradients produced by isolated noise pixels, so that the effects of noise are accentuated. The application of some form of image smoothing before these differential operators was found to drastically decrease the amplification of noise pixels.

It was noted that the small area of the 3 x 3 pixel averaging mask, compared to the size of image features, ensured very little blurring of feature edges whilst effectively reducing image noise. The median filtering technique performed slightly better at eliminating noise than did neighbourhood averaging, however this increase in performance was gained at the expense of added code complexity and processing time.

Software code was also written to implement the Laplacian image sharpening technique using a 3 x 3 pixel mask operator. As discussed earlier, this image sharpening algorithm uses a second order differential operator to detect and enhance sharp intensity gradients within an image. Due to the very high spatial sampling

resolution of the image capture system developed during this project, even small features extended over a large number of pixels. Hence, feature edges were characterised by only moderate intensity gradients which were difficult to detect using such a small pixel operator. In contrast, the 3 x 3 Laplacian operator was very sensitive to the sharp, localised intensity gradients produced by isolated noise pixels.

It was clear that such a small area filter was incapable of sharpening image edges whilst rejecting spurious noise. Many formulations for the construction of larger edge enhancement filter masks have been proposed [15]. Obviously the size of such filters is a compromise between noise immunity and computational expense. As an alternative to the use of large convolution masks, edge sharpening may often be achieved more efficiently using frequency domain filtering methods [16] [17].

9.5.2 Histogram Equalisation

When producing industrial radiographs, much of the test specimen often consists of areas of uniform thickness and density which result in an even film blackening. These areas will appear on the processed radiograph as large, dark regions containing very little useful contrast information. The absorption characteristics of many relevant image features will tend to vary only slightly from these dark areas, so that they are difficult to distinguish from the background image. Hence, a typical radiographic image may often appear rather dark and exhibit very little contrast information within the area of interest.

A typical image produced by the radiograph digitisation system is shown in Figure 9.2. Although the general shape of the object being radiographed is visible, it appears particularly dark, so that it is difficult to distinguish individual features from the background image. This situation is emphasised by reference to the intensity histogram plotted in Figure 9.3, which plots the frequency with which each grey scale level occurs within the image shown in Figure 9.2. It can be seen that the vast majority of pixel data is concentrated in the low range of intensity values. Clearly, this image would benefit from some form of contrast enhancement to emphasise image features. All the contrast enhancement methods illustrated in this section have been applied to the image shown in Figure 9.2.

The results of applying histogram equalisation to this image can be seen in Figure 9.4. It can be seen that the image appears significantly brighter and the general shape of the welded seam has been enhanced, so that more detail is visible to the eye. The operation of this contrast equalisation technique is illustrated by the modified histogram shown in Figure 9.5. It can be seen that intensity data is now much more evenly distributed across the entire grey scale range, especially in the high intensity region. It should be noted that low intensity data in the transformed image is concentrated into a small number of grey scale values which do not show up on the plotted histogram because they exceed the plotted range of frequency values. The application of this algorithm to a 1024 by 512 pixel frame required 4.9 seconds to complete.



Figure 9.2 Low Contrast Image

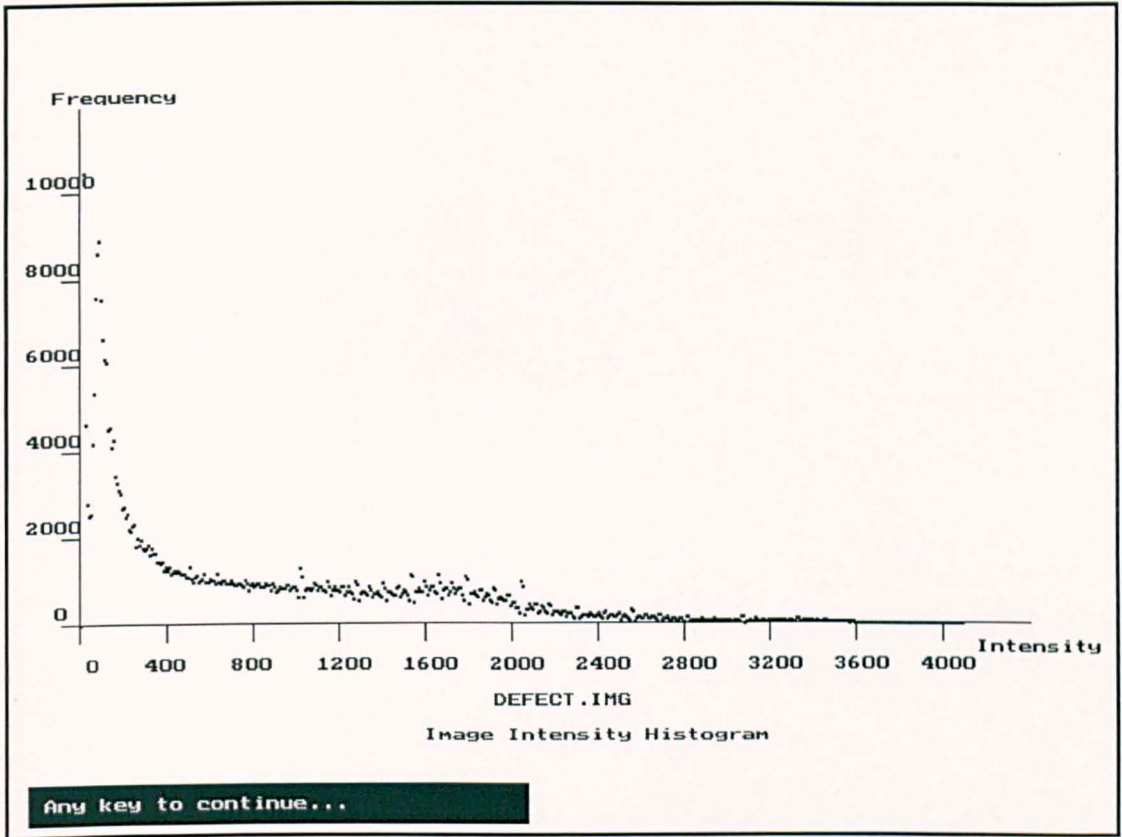


Figure 9.3 Intensity Histogram for the Image Shown in Figure 9.2

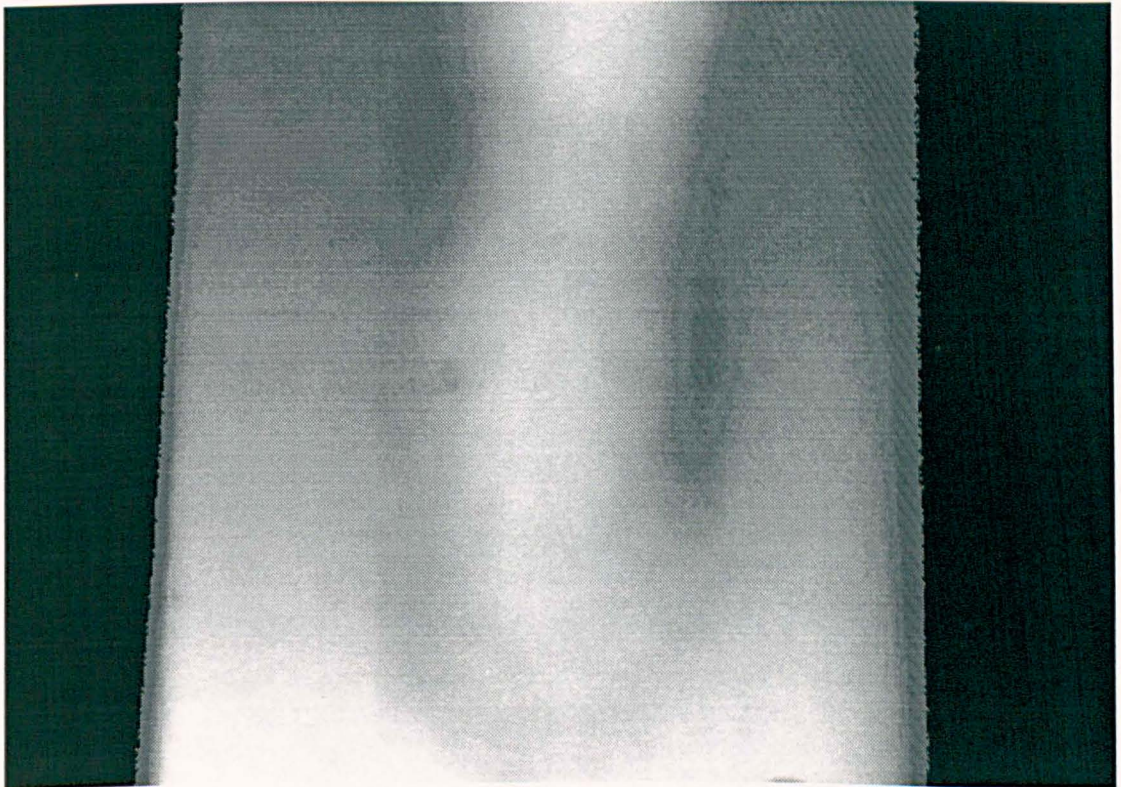


Figure 9.4 Contrast Enhancement by Histogram Equalisation of Figure 9.2

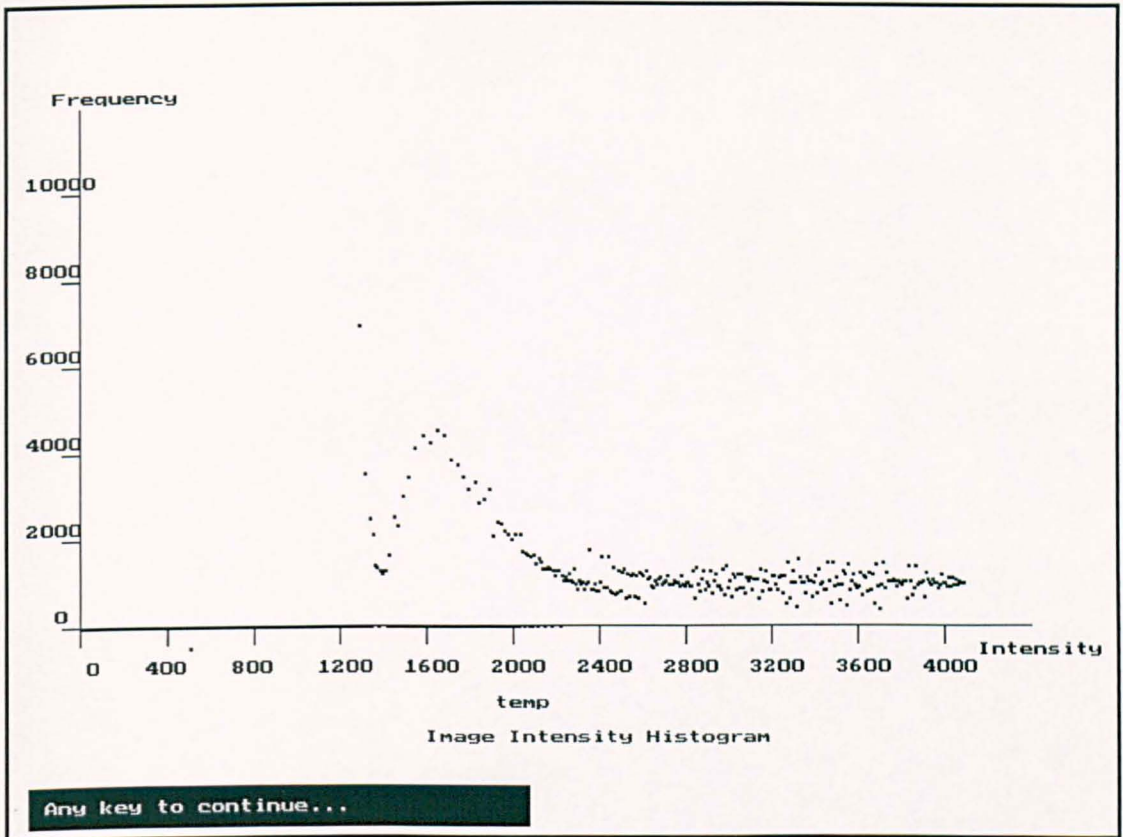


Figure 9.5 Intensity Histogram for the Image Shown in Figure 9.4

The previous example attempted to optimise the distribution of intensity information across the entire image. However, we are often interested in maximising contrast within a small region of the total image. For instance, we may wish to resolve a low contrast feature from an otherwise well balanced image. This point is illustrated by considering the dark, circular area at the centre of the image in Figure 9.2. It can be seen that histogram equalisation of this image did not significantly increase the contrast between this feature and its background. Instead, some method of local contrast equalisation was required.

Software code was written allowing the user to specify a small region of the image and apply the histogram equalisation algorithm to the pixels contained within this area. This sub-image was selected by using a mouse controlled cursor to define the extents of a window around the low contrast image features. The result of applying this processing method are shown in Figure 9.6. The specified window was centred around the dark areas at the centre of the image. It can be seen from the processed

image that the contrast of the specified features was improved so that they were clearly visible from the background image.

Although the preceding example did produce considerable enhancement of contrast within the specified window, it produced an undesirable discontinuity around the edges of this area. To avoid this effect, the pixel transformation function derived from histogram analysis of the specified sub-image was applied to the entire frame, as shown in Figure 9.7. It can be seen that this algorithm greatly enhanced the designated image features, whilst maintaining image continuity. However, since the contrast equalisation routine only considered a small subset of the total image, this technique resulted in light and dark intensity saturation outside the area of interest.

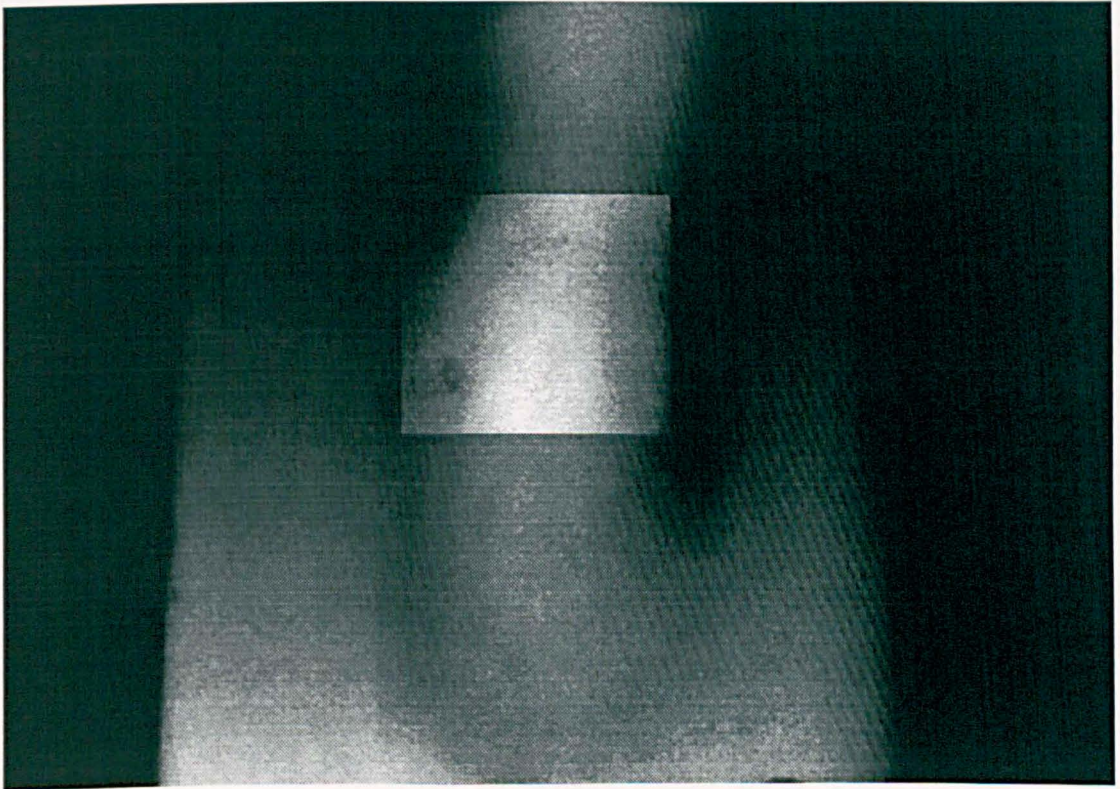


Figure 9.6 Histogram Equalisation Within a Specified Window

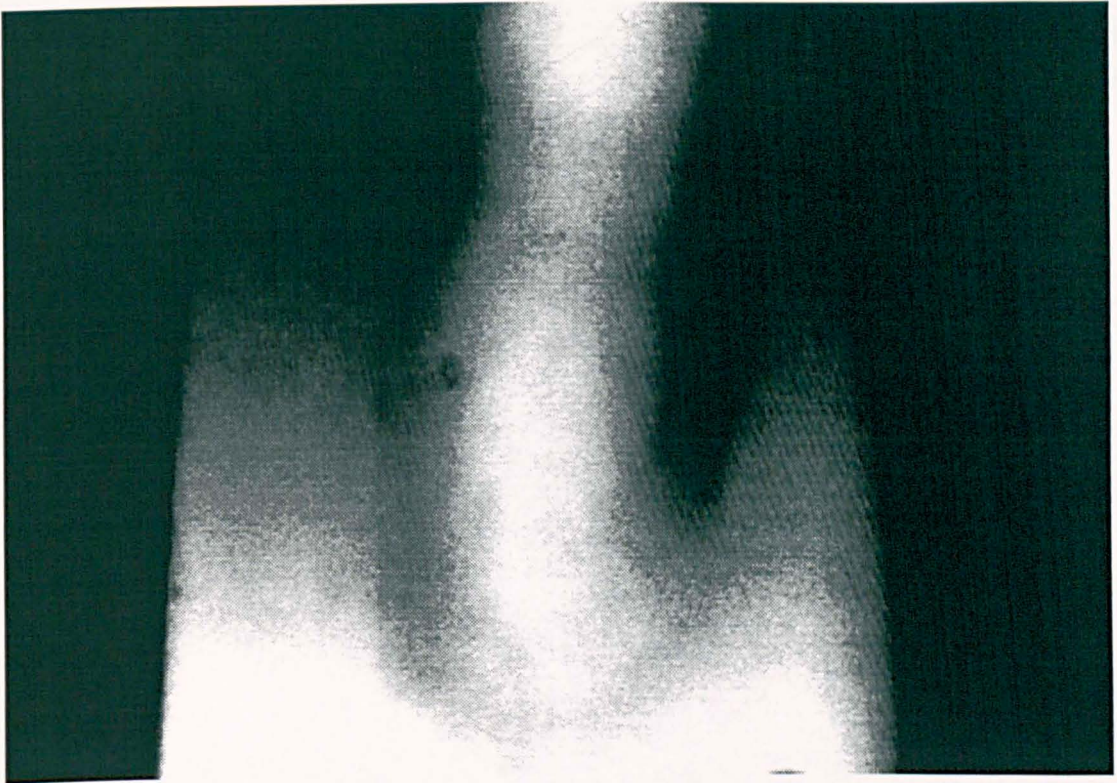


Figure 9.7 Window Equalisation Applied Across Entire Image

9.5.3 Linear Contrast Enhancement

Software code was also written to implement the technique of linear contrast enhancement. Initially this software automatically analysed the image histogram in order to identify the subset of grey levels in which most contrast information was concentrated, and then applied the linear stretch algorithm to this subset of pixels.

However, the benefits of applying this algorithm were found to vary significantly depending on the characteristics of the original image. In many images there were large areas of redundant dark data so that the vast majority of pixels did not significantly assist in image recognition. These dark pixels tended to predominate, so that more significant, high intensity pixels were ignored by the histogram analysis software. This resulted in the loss of contrast in bright image features when the contrast stretching algorithm was applied.

A more flexible technique involved direct specification of the linear pixel mapping function by the operator. This technique allowed interactive enhancement to maximise the contrast within a specified area of the image. Appropriate grey scale limits for the

mapping function were selected using histogram plotting software to assess the intensity distribution within a user defined image window.

The following examples of linear contrast enhancement all apply to the image shown in Figure 9.2. By examination of the intensity histogram for this image, shown in Figure 9.3, it can be seen that very few pixels have an intensity greater than 2400, so that these grey scale levels were effectively redundant. Figure 9.8 shows the results of applying a linear contrast stretch using the pixel mapping function illustrated in Figure 9.9. This function maps all pixels with intensities less than 20 to black, and all pixels with intensities above 2400 to white. The remaining data was then stretched to extend across the entire intensity scale. The resulting image shows a pronounced increase in overall image contrast so that the features in the centre of the frame are much more apparent.

Figure 9.10 illustrates how local contrast can be improved at the expense of overall image quality. In this example the image was enhanced by linear contrast stretching using the transformation function shown in Figure 9.11. This transformation function was chosen by examining the range of intensities around the central area of the image and stretching the grey scale values between these limits. It can be seen that this example produced increased contrast around the centre of the image but resulted in an attenuation of the high intensity data at the bottom of the frame.



Figure 9.8 Result of Linear Contrast Stretch of the image in Figure 9.2

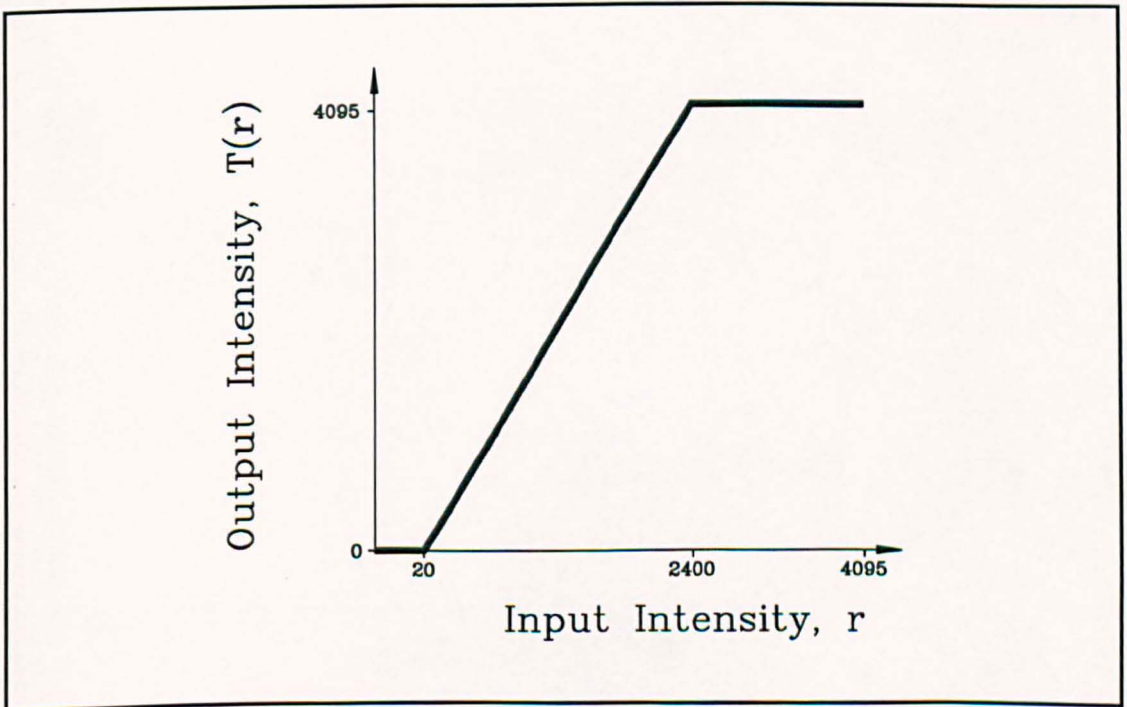


Figure 9.9 Transformation Function used in Figure 9.8



Figure 9.10 Linear Contrast Stretch

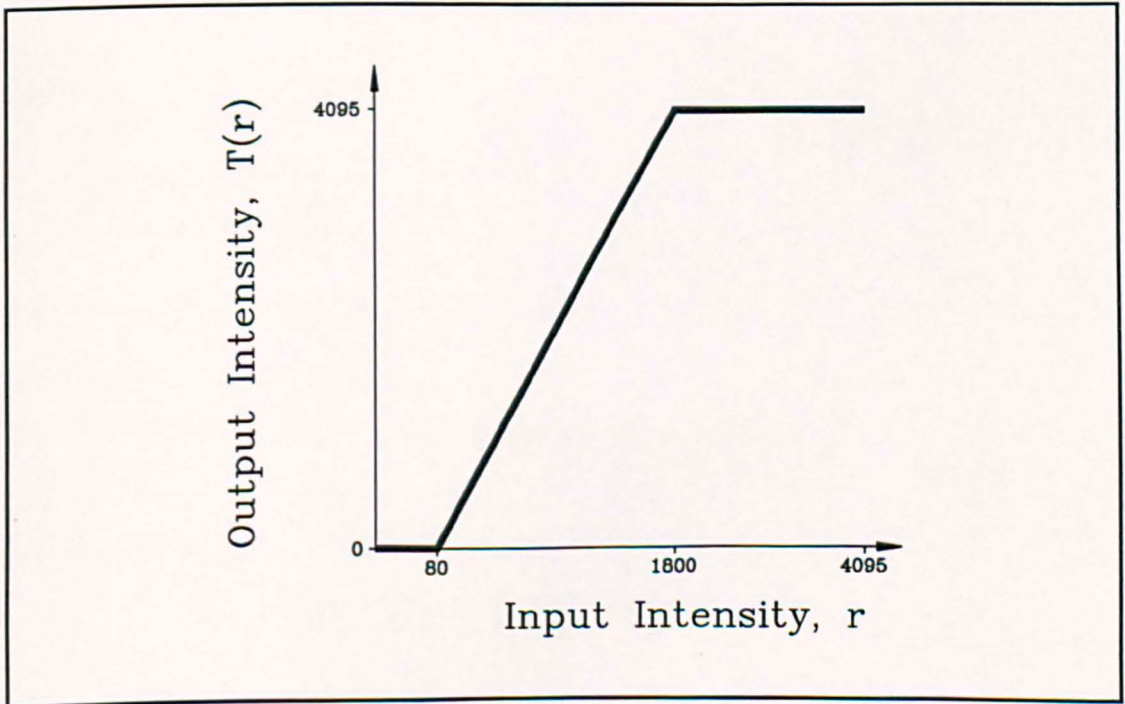


Figure 9.11 Transformation Function used in Figure 9.10

9.5.4 Edge Enhancement

Figure 9.12 shows a radiographic image which has undergone contrast enhancement by histogram equalisation. A dark circular feature is clearly visible in the centre of this

image. Tests were carried out in an attempt to locate and enhance the edges of this feature using a 3 x 3 pixel Sobel operator.

As discussed previously, differential operators such as edge detection templates tend to respond to sharp changes in image intensity, thereby accentuating the effects of spurious noise fluctuations. This effect was compounded by the fact that contrast enhancement techniques tended to stretch the intensity variations within an image, so that noise fluctuations were amplified. In an effort to reduce the effects of noise some form of image smoothing was applied to the equalised image before edge detection. A simple 3 x 3 pixel averaging operator was found to be adequate and had the advantage of executing faster than the median filter algorithm.

An edge enhancement method was adopted which used the Sobel operators to detect the edges of features and then emphasised these areas by subtracting edge data from the original image. The effect of this operation was to darken feature edges, thereby making them more visible. The results of this edge enhancement are shown in Figure 9.13. Although the edges of the central feature were successfully detected, a certain amount of false contouring can be seen in other parts of the image, due to noise effects. Clearly these false edges would be eliminated if the pixel template being used was large enough to reject the effects of isolated noise pixels.

Several approaches have been proposed to eliminate the detection of false edges due to the effects of noise. Larger versions of the Sobel templates has been used successfully to detect edges in low contrast images [18]. Other researchers such as Marr and Hildreth have proposed the use of large window Gaussian filters for image preparation prior to edge detection [19] [20]. However, it should be noted that the computational overheads of applying larger template operators increases as the square of the template size, so that these methods represent a significant processing workload.

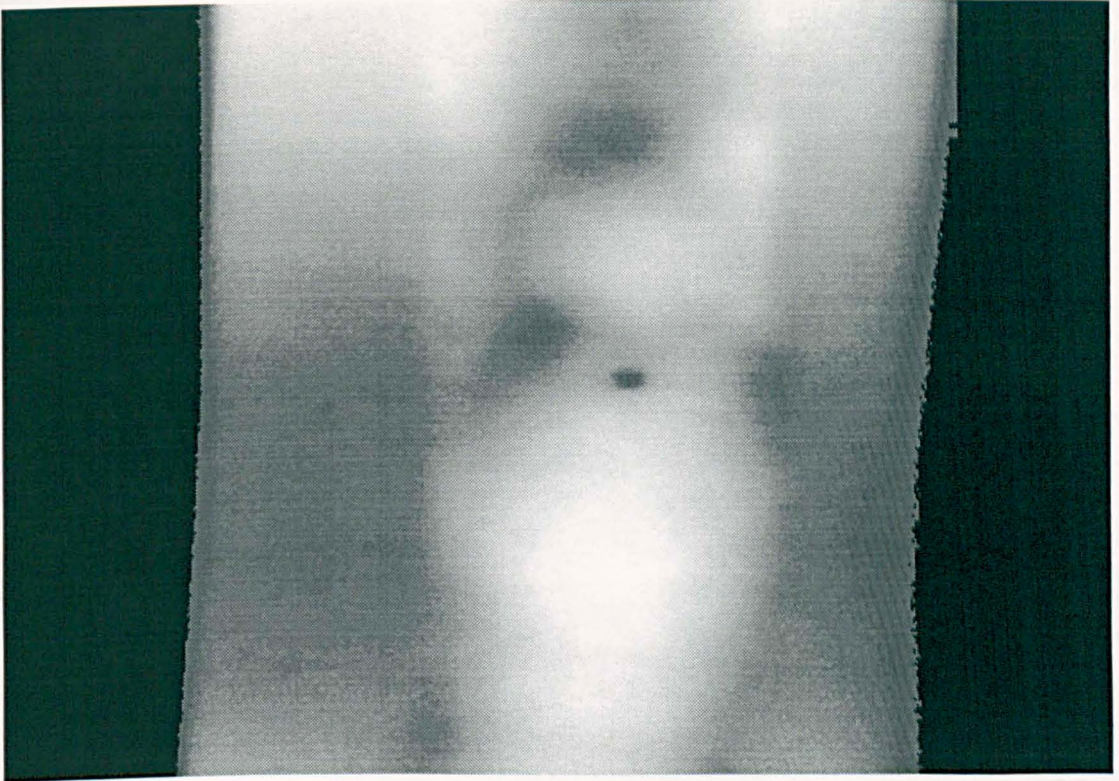


Figure 9.12 Image with Circular Feature at Centre

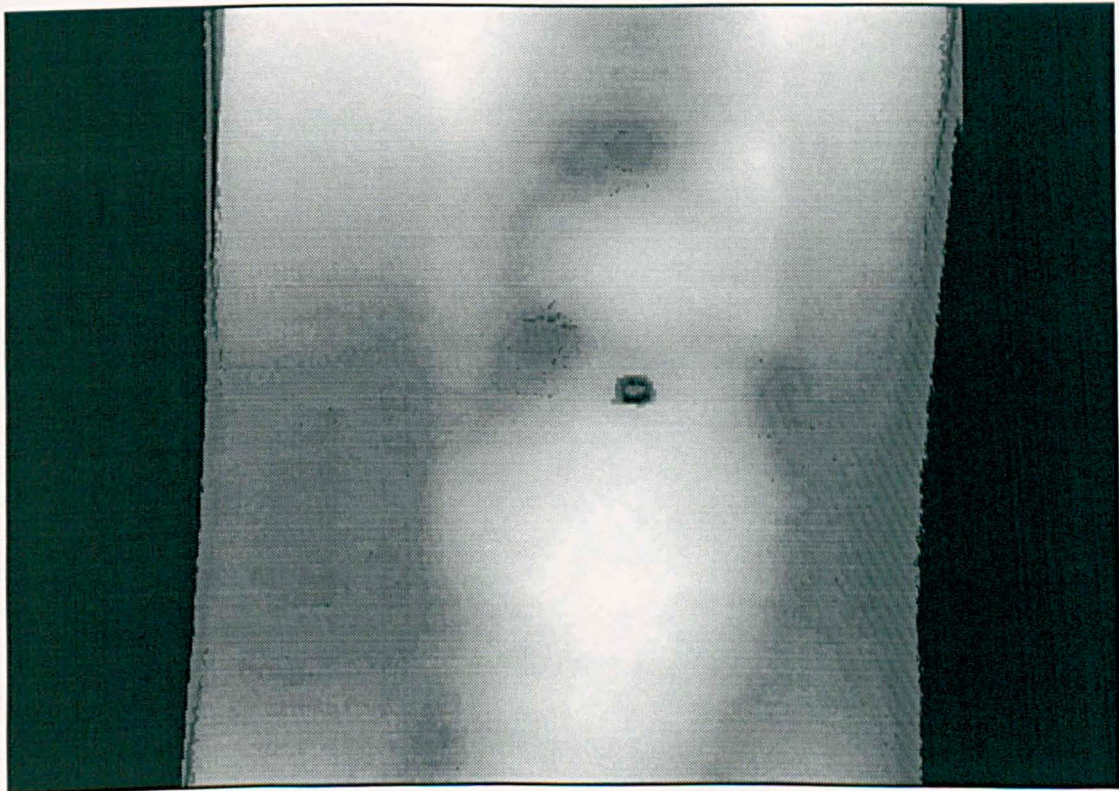


Figure 9.13 Result of Edge Enhancement on Image shown in Figure 9.12

9.6 Summary

A number of image enhancement algorithms were applied to the digitised radiographic images produced during this project. The aim of these algorithms was to accentuate image features, thereby assisting the operator to correctly interpret images. These methods were implemented in software using a distributed transputer processing system.

In general it was found that the high quality of the unprocessed images prevented significant benefits from being gained by the application of filtering operations for image smoothing or sharpening. Both these operations led to a degradation of legitimate pixel data without any visible improvements in image quality. However, methods of contrast enhancement, either by simple linear intensity stretching or by histogram equalisation methods were demonstrated to be particularly effective in accentuating relevant features within images.

A simple edge detection algorithm was implemented in order to illustrate its potential for feature extraction as the first stage of a high level image processing system. It was noted that, due to the very high pixel sampling density used during image acquisition, the application of large area edge detection masks would be required in order to resolve feature edges from background intensity changes.

References

- 1 H. C. Andrews. "Monochrome Digital Image Enhancement", *Applied Optics* **15**, 1976, pp 495-503.
- 2 E. Hall et al. "A Survey of Pre-processing and Feature Extraction Techniques for Radiographic Images", *IEEE Trans. Comput.* **20** 9, 1971, pp 1032-1044.
- 3 W. Lin, C. Dinghua & X. Qinghong, "Study on Automatic Detection and Identification of Defects in Welds by Microcomputer", *Proceedings of the International Conference on Quality and Reliability in Welding*, Hangshou, China, September 1984.
- 4 H. J. Trussell, "Processing of X-Ray Images", *Proceedings of the IEEE*, **69** 5, 1981, pp 615-625.

- 5 J. W. Tukey, "Exploratory Data Analysis", Addison-Wesley 1976, pp 205-236.
- 6 L. Kovaszny & H. Joseph, "Processing of Two Dimensional Patterns by Scanning Techniques", Science **118**, 1953, pp 475-477.
- 7 A. Rosenfeld & A. Kak, "Digital Picture Processing", Vol. 1, Academic Press, 1982, pp 237-250.
- 8 E. L. Hall, "Almost Uniform Distributions for Computer Image Enhancement", IEEE Trans. Comput., vol. C23, 1974, pp 207-208.
- 9 J. F. Canny, "A Computational Approach to Edge Detection", IEEE Trans. Pattern Anal. Machine Intell. PAMI-8 6, 1986, pp 679-698.
- 10 L. G. Roberts, "Machine Perception in 3-D Solids", in Optical and Electro-Optical Information Processing, Ed. J. Tipper et al. MIT Press, 1965.
- 11 V. Torre & T. Poggio, "On Edge Detection", IEEE Trans. Pattern Anal. Machine Intell. PAMI-8, 1986, pp 147-163.
- 12 V. Chaudhary & J. Aggarwal, "Parallelism in Computer Vision: A Review" in Parallel Algorithms for Machine Intelligence and Vision", Springer Verlag, 1990.
- 13 "Applications of Transputers 3", Eds. T. Durrani, W. Sandham, J. Soroghan & S. Forbes, IOS Press, 1991, pp 519-570.
- 14 "The Transputer Data Book", Second Edition, Inmos Ltd. 1989, pp 36-39.
- 15 J. M. S. Prewitt, "Object Enhancement and Extraction" in Picture Processing and Psychopictorics", Eds. B. Lipkin & A. Rosenfeld, Academic Press, 1970, pp 75-149.
- 16 R. C. Gonzalez and P. Wintz, "Digital Image Processing", Addison-Wesley, 1987, pp 146-152.
- 17 R. J. Arguello, H. Sellner & J. Stuller, "Transfer Function Compensation of Sampled Imagery", IEEE Trans. Comput. **C-21**, 1972, 812-818.
- 18 R. Yu, "A Three Dimensional Vision System for Remotely Operated Vehicles", PhD Thesis, University of Liverpool, August 1994.
- 19 D. Marr & E. Hildreth, "Theory of Edge Detection", Proceedings of the Royal Society of London **207**, 1980, pp 187-217.
- 20 A. Huertas & G. Medioni, "Detection of Intensity Changes with Subpixel Accuracy using Laplacian-Gaussian Masks", IEEE Trans. Pattern Anal. Machine Intell. PAMI-8, 1986, pp 651-664.

10. Image Data Compression

10.1 Introduction

The very high resolution of the images produced during this project necessitated the storage of a huge amount of pixel data. The size of these image files dictated that any image archival system would be liable to encounter problems with data storage and transmission.

For instance, a digitised image of a 28 mm by 14 mm section of radiographic film, captured at a spatial sampling resolution of 14 μm and a digital resolution of 12 bits/pixel, will occupy 4 Mbytes. Hence, a single, uncompressed image file may easily exceed the storage capacity of a standard floppy disk. Similarly, the typical hard disc space of a modern PC is in the order of a few hundred Mbytes, so that the archival of a large number of image files will quickly exceed the available storage space.

It is clear that the storage and transmission of such large image files would benefit from some form of data compression. Many such compression methods have been proposed in recent years [1] [2]. The successful implementation of these techniques will help to reduce the cost of high integrity digital storage media and decrease the time and expense associated with transmission of large amounts of data.

Some forms of compression are specific to image data whilst others may be applied to any type of coded information. Image data is particularly suited to compression, since it tends to consist of large areas containing pixels of the same, or similar, intensity. The information content in these areas is low, and can often be coded in a more efficient manner than the standard pixel map representation.

Compression methods may be classed as either "lossless" or "lossy". Lossless compression algorithms ensure that all the data contained in the original image is retained in the compressed form, so that an exact reconstruction of the image can be made with no loss of data. Lossy compression methods seek to selectively discard data which is superfluous to the perceived quality of an image. The performance of lossy algorithms is highly dependent on the amount of image degradation that can be tolerated by the viewer.

The relevant industrial standards concerning radiographic film digitisation [3] and real time radiography [4] provide for the use of appropriate image compression methods, but make no mention of which types of compression are acceptable. However, a report into radiographic film digitisation commissioned by the U. S. Nuclear Regulatory Commission recommends that all compression methods applied to such images should be lossless [5].

In view of the very high fidelity of images produced during this project, it was considered that no loss of image quality was acceptable. For this reason all the methods of data compression investigated in this research were lossless.

This chapter discusses the work carried out on image data compression during this project. A number of relevant compression methods are discussed in some detail. Details are then given of comparative tests carried out to assess the effectiveness of various methods of lossless data compression. The most successful of these methods was found to involve a combination of difference coding and adaptive Huffman coding. Software was then developed to apply this compression algorithm to the images produced by the radiographic film digitisation system. Typical compression ratios of around 4:1 were achieved.

10.2 Lossless Compression

10.2.1 Entropy Coding

Entropy coding methods of data compression rely on the fact that certain symbols in a data set occur more frequently than others. In an uncompressed image each pixel value will be represented by a fixed length bit code, the length of which will depend on the digitisation fidelity of the imaging system. Consider an image consisting of a sequence of pixels quantised to one of N discrete intensity levels. The number of data bits, n , used to represent each pixel in the image is given by :

$$n = \text{Log}_2(N) \quad (10.1)$$

If an equal number of pixels in the uncoded image occupy each grey scale level from zero to $N-1$, then this method of fixed length coding will be optimal. However, if certain grey scale values are more likely to occur than others, then these values could be encoded more efficiently, using short length code words.

Entropy coding refers to a process in which variable length code words are assigned to source symbols depending on how often they occur in the original data set. By choosing appropriately short code words to represent the most frequently occurring objects within a data set, the average code word length may be reduced. The resulting data set will then consist of a series of different length code words, with short codes representing much-used data and long code words representing rarely occurring data.

The degree of data compression attainable by entropy coding is dependent on the frequency distribution of grey levels within the original image. For instance, if a small number of grey level values occur very often they can be assigned short code words, resulting in a high degree of compression. However, if pixel values are evenly distributed across digitisation range, then entropy encoding will not yield significant benefits.

The degree of order within a data set can be measured statistically by the entropy function. Shannon [6] showed that, a data set consisting of N different symbols, having individual probabilities of occurrence, p_0, p_1, \dots, p_{N-1} , has an entropy, H , given by :

$$H = \sum_{i=0}^{N-1} -p_i [\text{Log}_2(p_i)] \quad (10.2)$$

In practical terms the entropy of a sample can be regarded as the average number of bits needed to uniquely code each source symbol. Hence, the maximum attainable compression is limited by the entropy of the original data set. In general the entropy of an image captured at n bits per pixel will vary within the range from zero to n . An entropy of zero corresponds to an image in which all pixels have the same grey scale value, whilst an entropy of n corresponds to an image in which all grey scale levels have equal probability of occurrence.

The amount of data compression achieved is measured by the compression ratio, which is simply the ratio of the original size of the image to the size of the compressed image. When considering entropy coding methods, the compression ratio may be more conveniently expressed as the ratio of the number of bits per pixel in the original

image to the average number of bits per pixel in the coded image. Hence the maximum possible compression ratio, CR_{\max} , is given by :

$$CR_{\max} = \frac{n}{H} \quad (10.3)$$

10.2.1.1 Huffman Coding

Clearly entropy coding can produce significant compression of an image file, provided that an appropriate set of codes can be developed. When assigning such code words to a data set it should be remembered that, since code words are of variable length, they should be uniquely decodeable. This condition requires that no short code group should be duplicated at the start of a longer code group.

Several entropy coding schemes have been proposed, including work by Shannon and Fano [6] and Rice [7]. However, an optimal method of code construction was found by Huffman [8]. Huffman coding, when properly implemented, gives average code word lengths approaching the entropy limit for the data set.

The method of assigning Huffman codes to a data set may be illustrated by the example shown in Figure 10.1, where the data set to be coded consists of a four symbol alphabet, designated X_1 , X_2 , X_3 and X_4 .

Huffman coding requires that the frequency distribution of the data set is known before code words are assigned, so that the first step in the coding process involves calculating probabilities for all the source symbols present in the data set. The data symbols are then sorted in order of decreasing probability of occurrence. For the example data set, these probabilities are shown in the second column of Figure 10.1.

The two least frequently occurring symbols are then assigned single digit codes and their combined probability is calculated. In this case the least probable symbol, X_4 , is assigned a one and X_3 a zero. The combined probability of X_3 and X_4 is shown to be 0.25. The code book is then re-sequenced to include this combined probability and the two least occurring symbols are again given single digit codes. In this way a tree structure is built up with each fork of a branch representing a binary digit in the eventual Huffman code.

The process continues until a combined probability of one is reached. The Huffman code of each symbol is then found by tracing the branches of the tree back to the symbol and reading off a binary digit at each fork. These code words are shown in the right hand column of Figure 10.1. Once Huffman codes have been assigned to each possible source symbol the compression process is simply a matter of substituting the code words into the source data.

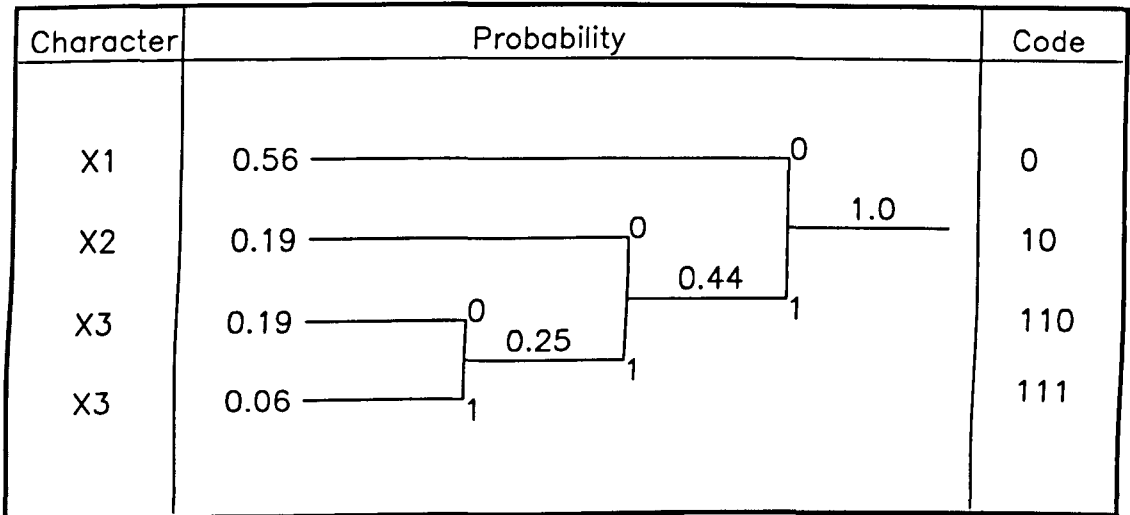


Figure 10.1 Huffman Coding Tree Structure

If each symbol in the original data set is now coded using the Huffman code book developed in the preceding example, the entropy of the coded data may be calculated as the average number of bits per symbol needed to code the entire data set. For a code book consisting of N variable length code words with lengths, n_0, n_1, \dots, n_{N-1} and probabilities of occurrence, P_0, P_1, \dots, P_{N-1} , the entropy will be :

$$H = \sum_{i=0}^{N-1} [n_i \cdot P_i] \quad (10.4)$$

The effectiveness of the Huffman coding process can be assessed by comparing the results of equation (10.4) with the entropy limit calculated using equation (10.2), and the fixed length code word size calculated using equation (10.1).

Coding Method	Bits Per Symbol
Fixed Length Code Words	2
Huffman Coding	1.69
Entropy Limit	1.62

Table 10.1 Efficiency Comparison of Huffman Coding

10.2.1.2 Implementing a Practical Huffman Coding Scheme

In practice, if there are many possible source symbols to be encoded, the code book size will quickly become very large, with exceptionally large codes representing seldom occurring source data. When the data set consists of a large alphabet of rarely occurring symbols these code word lengths can quickly exceed the word length of the computer on which the algorithm is implemented. These long codes are of little practical use but occupy a large amount of processing time and memory.

These problems may be avoided by only assigning codes to the most frequently occurring source symbols. Less common symbols are then represented by a special identifier code, plus the full, uncoded symbol. Although this results in a slight reduction in the achievable compression ratio the effect will be minimal if an appropriate cut-off point is selected for the code book size.

When implementing Huffman coding on a computer it is necessary to consider the method of data representation. A microprocessor operates on data objects that are of a fixed length, for instance the eight bit byte. Huffman codes, rather than consisting of these simple fixed length objects are a continuous stream of variable length codes. This situation requires that, once individual codes have been assigned to each source data symbol, it is necessary to fit these variable length codes into the fixed length data objects used by the processor. The coded data can then be regarded as a continuous stream of code words rather than a series of discrete data objects. This process is known as bit packing.

Finally it should be remembered that encoded data is meaningless unless it can be decoded back to its original form. This decoding process requires that the code book

used during compression is known. Hence, the first part of any compressed data file should contain the code book which was used to compress the data.

10.2.1.3 Adaptive Huffman Compression

In order to produce an optimum Huffman code table it is necessary to know the probability of occurrence of each source symbol within a data set. This condition necessitates analysis of the raw data and construction of a new set of Huffman codes for each set of data to be compressed. This process can prove time consuming and impractical, given the huge amounts of data to be analysed within an image. Another problem associated with this approach is that data cannot be coded sequentially as a continuous stream. Instead it must be completely scanned once, to build up the code table, before being coded.

A more acceptable solution would be to use the same code table for several sets of data. However, using a code table of fixed composition results in the compression ratio being very sensitive to changes in input data composition. In certain cases this may mean that frequently occurring symbols are assigned long code words, leading to an overall expansion of the data set. Clearly, we require a more robust coding technique which is tolerant to changes in the frequency distribution of input data symbols.

A solution to this problem is to use a code table that adapts to the incoming data. Adaptive compression re-sequences the code table each time an input symbol is coded. A running count of the number of occurrences of each symbol is kept and the most frequently occurring symbol is assigned the shortest code. By continuously adapting to the composition of input data and always assigning the shortest code to the most common symbol this method ensures near-optimum compression ratios for any input data.

10.2.2 Data Pre-processing and Entropy Reduction

In addition to entropy coding methods of data compression, many other lossless algorithms exist for reducing the size of a data set. Many of these methods may be used to compress or re-structure source data prior to Huffman coding, to produce increased compression ratios [9]. These algorithms may either seek to reduce the

number of data objects to be coded or to change the frequency distribution of the data to one with a lower entropy.

The success of Huffman coding depends on the frequency of occurrence of different symbols. If the same symbol occurs very often it can be assigned a short code word. Hence if a small number of input symbols occur with a high frequency the achievable compression ratio will be high. In order to produce data sets with a lower entropy it is therefore necessary to produce a source file with a large number of similar data values, as can be deduced from equation (10.2).

10.2.2.1 Run Length Coding

Run length coding attempts to reduce the number of source data objects by identifying blocks of adjacent data samples having the same value and replacing them with shorter codes. A block of identical data objects may be replaced by three values, a special block identifier code, the data value and the number of occurrences of that value in the block.

This run length coding algorithm is demonstrated by the example in Figure 10.2. Each run of data is signified by the identifier, \$, which can be any code not included in the original data set. This identifier is followed by the repeated character, X, and a character count. Since the run length coding scheme requires three data items to represent each run of data, any coding algorithm should not attempt to code blocks which are shorter than three pixels, since this would result in expansion of the data set. The example shown in Figure 10.2 results in a reduction of the data set from nine characters to seven.

Run length coding proves particularly efficient when applied to graphics images produced by computer drawing packages, which consist of large areas of constant intensity, or colour. Similarly, images containing very little detail and large, low contrast areas will benefit significantly from run length compression. In contrast, applying the run length coding algorithm to images containing large amounts of detail will not produce significant compression.

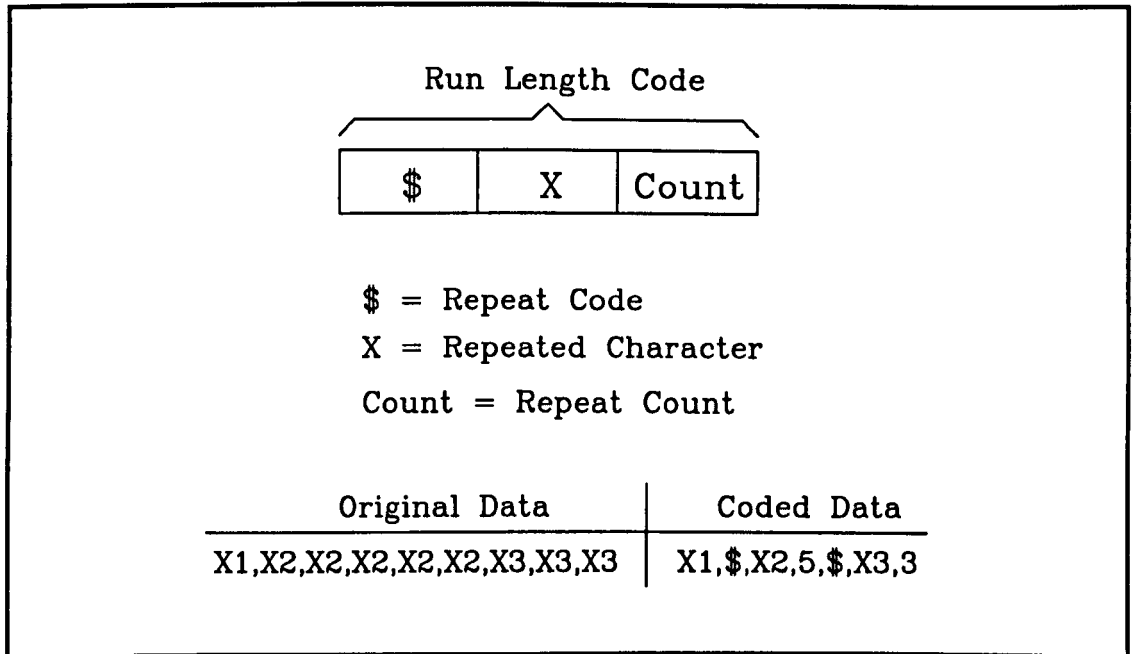


Figure 10.2 Run Length Coding

10.2.2.2 Difference Coding and Predictive Coding

Difference coding attempts to reduce the entropy of a data set by coding only the differences between adjacent pixel values rather than the values themselves. The success of this coding scheme relies on the fact that image data tends to contain large areas of slowly varying intensity. Hence, the difference between adjacent pixels will tend to be low regardless of the absolute value of individual pixels.

When difference coding alone is applied to a data set it will result in an overall expansion of data. For instance, in an image captured at n bits per pixel, each pixel value may vary between zero and $(2^n - 1)$. However, the difference coded data may vary between $\pm(2^n - 1)$, so that $(n + 1)$ bits are required to code each pixel.

The advantage of difference coding becomes apparent when it is applied prior to some form of entropy coding. Since the difference coded image data consists of a small number of frequently occurring symbols, the entropy of the data set will be correspondingly low. Hence, difference coded image data is particularly amenable to entropy coding techniques. A further advantage of this compression scheme is that difference coded data is not particularly sensitive to changes in input data composition, since it will always tend to be distributed around zero. This condition

allows the same Huffman code book to be applied to different images without excessive degradation of compression ratio.

An extension of the difference coding process is predictive coding, which attempts to produce a data set of lower entropy by predicting the next pixel value from the values of pre-ceding pixels. In this way it is only necessary to store error terms between actual and predicted values. Several methods of predicting pixel values can be used, the most simple form of which is zero order prediction, which assumes that adjacent pixel values are equal. Hence, zero order prediction is equivalent to difference coding.

First order prediction involves extrapolating a straight line through pre-ceding data values to predict the value of successive pixels. A least squares fit line is found between preceding pixels, the number of pixels used for this calculation can be adjusted depending on the source data structure. Higher order predictors which attempt to fit source data to polynomial functions can also be used, but there is little evidence to support their effectiveness.

10.2.2.3 Contour Encoding

The coding methods considered so far have all been applied to simple, one dimensional pixel distributions. However, significant compression benefits may be gained from regarding an image as a two dimensional intensity function and using the method of contour encoding.

Contour encoding may be regarded as a two dimensional extension of the run length coding scheme. This technique seeks to identify areas within an image in which all pixels have the same intensity value. The entire area can then be represented by tracing a constant intensity contour around the image plane. All the pixels enclosed by this contour will have the same intensity value and need not be explicitly defined. As with run length coding, the technique of contour encoding is most effective when applied to graphics images, but may yield little benefit when applied to images with a high detail content.

Although contour encoding will generally produce higher compression ratios than those achieved by run length coding, there are several practical barriers to its efficient implementation. A major concern is that the contour coding algorithm requires

simultaneous memory access to all the pixel values within an image frame. The size of image files captured during this project often exceeded the available processor memory, so that contour coding became impractical to use. Even when applied to relatively small images, contour encoding is difficult and slow to implement.

10.3 Lossy Compression Methods

The degree of compression attainable using lossless forms of compression is severely limited by the requirement that all source data must be preserved in the compressed form. For instance, compression ratios of around 3:1 are typical for lossless compression of 8 bit image data, as will be seen in the examples later in this chapter. However, if a certain amount of image degradation can be tolerated then the amount of compression which can be achieved is significantly increased. This class of compression algorithm relies on the discarding of certain components of the image and is referred to as “lossy” compression.

The most simple method of lossy image compression involves the technique of image subsampling and pixel duplication. This technique discards altogether a proportion of the pixels from the original image. When the image is reconstructed these pixels are substituted by duplicate pixels from one of their near neighbours. A refinement of this idea attempts to interpolate between sampled pixels in order to produce a more accurate reconstruction of the original image. These subsampling techniques, if they are used too aggressively, tend to severely compromise image quality and cause a blocky appearance, known as pixelation.

Limitations of the human visual system ensure that certain spatial frequencies within an image will not be discernible under normal viewing conditions [10]. By selectively discarding data which is impossible or difficult to perceive, significant compression can be achieved with minimal degradation of image quality [11]. In order to implement these methods of lossy compression it is first necessary to convert the image representation from the spatial domain to the frequency domain. The most common form of spatial frequency transformation used in image compression involves the Discrete Cosine Transform (DCT).

The most comprehensive specification of this type of lossy image compression algorithm for both colour and monochrome still video images is contained within the JPEG international standard [12]. Compression of RGB colour images using the JPEG algorithm can result in compression ratios of between 20:1 and 30:1 without excessive loss of image quality.

10.4 Tests on Compression Methods

In order to investigate the relative merits of possible data compression algorithms, software was written to implement several of the most promising methods. All methods investigated involved no loss of source data, so that the compressed files could be expanded back to their original form.

The following different types of lossless data compression were implemented in software :

1. Adaptive Huffman Coding
2. Run Length Coding
3. Difference Coding followed by Huffman Coding
4. First Order Predictive Coding followed by Huffman Coding
5. Run Length Coding followed by Huffman Coding
6. Run Length Coding of Differences followed by Huffman Coding

Tests were carried out using each of these methods to assess their relative performance. These tests were carried out early in the project, before images captured by the radiograph digitisation system were available. Instead, image data was acquired using a standard monochrome video camera. These images consisted of 512 by 512 pixel frames captured at 8 bits per pixel.

The results of applying different compression methods to a typical image frame are summarised in Table 10.2. The image used for these tests was a test card consisting of several, highly contrasting intensity steps. The theoretical entropy of the image was calculated using equation (10.2).

$$H = 4.426$$

The theoretical maximum achievable compression ratio for the unprocessed input image could then be calculated using equation (10.3).

$$CR_{\max} = 1.81$$

Compression Method	Compression Ratio
1 Adaptive Huffman Coding	1.73
2 Run Length Coding	1.16
3 Difference Coding + Huffman Coding	2.94
4 First Order Predictive Coding + Huffman Coding	2.77
5 Run Length Coding + Huffman Coding	1.67
6 Run Length Coding of Differences + Huffman Coding	2.89

Table 10.2 Lossless Compression Tests

The first step involved in Huffman coding is the construction of a code table. The structure of this code table depends on the frequency distribution of source data symbols. In order to construct an optimum code table it is necessary to first analyse the input data and then use the Huffman algorithm to assign codes to each symbol based on their frequency of occurrence. Software was developed to analyse a data file and produce an optimised Huffman code table from the results. The software then used this code table to compress the data using the adaptive compression algorithm. The use of adaptive compression eliminated the need for analysis of each individual data file.

It can be seen from Table 10.2 that the Huffman compression method resulted in a compression ratio approaching the entropy limit for the unprocessed pixel data. The adaptive Huffman compression algorithm was applied using a code book containing 50 entries, generated from the analysis of each image data file. The size of the code book could be varied, but increasing it past 50 entries was found to have very little effect on compression ratios, whilst significantly increasing processing times. The use

of an adaptive Huffman code book lessened the effect of changing data composition somewhat. However, the Huffman coding method, when applied to raw pixel data, was still seen to be relatively sensitive to changes in image structure.

Run length coding, when applied by itself was seen to be only moderately successful. The reason for this poor performance is that only blocks of identical data symbols were compressed, while all other data was left unchanged. It was noted that the test card image used for these tests consisted of several areas of near-uniform intensity, which would not generally be present in a real image. It was therefore assumed that run length coding would be less successful when applied to images produced by the radiograph digitisation system.

The use of difference coding prior to Huffman coding had the effect of decreasing the entropy of the data set, and so increasing the amount of compression possible. Assessment of the difference coded data yielded an entropy of 2.61, so that the theoretical maximum compression ratio was increased to 3.06. This figure compares well with the achieved compression ratio of 2.94.

The use of difference coding was also seen to reduce the impact of changing data composition. When using the same code book to compress different data files, high compression ratios were still maintained. This was because, although the frequency of occurrence of individual pixel values varied widely between images, the average difference between adjacent pixels remained very similar.

The use of first order predictive coding was seen to be slightly less effective than difference coding. This suggested that adjacent data values tended to vary slightly around a mean point, rather than increasing and decreasing linearly. This hypothesis was supported by observing the raw image data. It was decided not to attempt second order predictive coding since the form of input data did not appear to be closely modelled by this equation.

The use of run length coding was seen to have a detrimental effect on the compression ratios achieved by Huffman coding. Although run length coding reduced the number of source characters it also increased the entropy of the data set. Run length coding replaces a block of identical data values with three new ones, an identifier code, the

data value and the length of the block. Since run length values tended to vary widely, this introduced many new source values into the code book. The general effect of this was to produce a large code book of low probability characters, and hence high entropy source data.

The combined use of difference coding, run length coding and Huffman coding was seen to be slightly less effective than simple Huffman coding of differences. The inclusion of run length codes again appeared to increase the entropy of the data. Run length coding was also observed to increase sensitivity of the Huffman coding algorithm to changes in data composition.

It was decided, using the results of the above tests, to use Huffman coding of differences between data values as the method for compressing image data files in the radiograph interpretation system.

10.5 Compression of Radiographic Images

Routines were added to the system software environment to implement lossless data compression using a combination of the adaptive Huffman coding and difference coding algorithms. This software included routines for the compression and expansion of image files. A routine was also provided to allow statistical analysis of data files and the construction of an optimum Huffman code table based on this statistical analysis.

Final tests were carried out in order to assess the effect of these compression algorithms on images produced by the film digitisation system. A variety of test images were used, taken from both test cards and industrial radiographs of welded joints. The compression ratio achieved for each image was recorded. Tests on data integrity were also carried out to ensure that no image corruption was introduced during either the compression or expansion processes. This involved comparing each pixel in the original, uncompressed image with pixels from data that had undergone compression and re-expansion. All expanded data produced an exact pixel for pixel reproduction of the original image.

It should be noted that the 12 bit data representation used by the image capture system required two bytes per pixel to implement a standard fixed length coding scheme.

Hence, with no other form of compression, the bit packing algorithm used during Huffman coding would result in a degree of compression. Equation (10.3) shows that this bit packing process would result in a compression ratio of 1.33.

The performance of the data compression algorithm was highly dependent on the structure of the image used. Typical results obtained from the compression of 1 Mbyte image files are summarised in Table 10.3. It was found that compression ratios between three and four to one were typical for lossless compression of images produced by the film digitisation system.

Image Description	Compression Ratio	Compression Time	Expansion Time
Test Card	3.16	19 s	23 s
Welded Pipe	3.97	16 s	21 s
Welded Seam	4.32	15 s	21 s

Table 10.3 Results of Compressing Digitised Images

10.6 Summary

This chapter has discussed methods of lossless data compression which were applied to the digitised images in order to decrease the data storage volume required for image archiving. Due to the need for very high quality images it was decided that no loss of image data could be tolerated during data compression. Consequently, all compression algorithms were 100 % lossless, so that the original pixel data could be exactly reconstructed from its compressed format.

Several forms of lossless image compression were implemented in software and their relative performances have been presented in this chapter. The most successful compression algorithm was found to involve a combination of adaptive Huffman coding and difference coding. This algorithm typically yielded compression ratios of between 3:1 and 4:1, depending on the composition of the original image.

References

- 1 P. A. Wintz & L. C. Wilkins, "Bibliography on Data Compression, Picture Properties and Picture Coding", IEEE Trans. on Information Theory **IT-17** 2, 1971.
- 2 R. Gonzalez & P. Wintz, "Digital Image Processing", Addison-Wesley, 1987, pp 256-329.
- 3 ASME Boiler and Pressure Vessel Code, Section V, "Non-destructive Testing", American Society of Mechanical Engineers, 1992.
- 4 BS7009, "Applications of Real Time Radiography to Weld Inspection", British Standards Institution, 1988.
- 5 "Review and Evaluation of Technology, Equipment, Codes and Standards for Digitisation of Industrial Radiographic Film", U. S. Nuclear Regulatory Commission, Washington D. C. May 1992.
- 6 C. E. Shannon, "A Mathematical Theory of Communication", Bell Systems Tech. Journal **27**, 1948, p. 379.
- 7 J. O. Thomas & L. Turner, "Imagery Data Compression" Oxford Computer Services, 1975, pp 67-101.
- 8 D. A. Huffman, "A Method of Constructing Minimum Redundancy Codes", Proc. IRE **40**, 1952.
- 9 G. Held, "Data Compression- Techniques and Applications, Hardware and Software Considerations", Wiley, 1983.
- 10 F. W. Campbell & L. Maffei, "Contrast and Spatial Frequency", Scientific American **231** 5, 1974, pp 106-114.
- 11 R. A. Quinell, "Image Compression", EDN **38** 2, 1993, pp 62-71.
- 12 G. K. Wallace, "The JPEG Still Picture Compression Standard", Communications of the ACM **34** 4, 1991, pp 31-44.

11. Conclusions and Suggestions for Further Work

11.1 Conclusions

This thesis has described the development of apparatus for the digitisation of industrial radiographic film images. This digitisation process allows considerable savings in physical storage space to be made and increases the ease with which images can be transmitted and reproduced. Digitisation also serves to eliminate problems associated with the degradation of film quality during storage.

As was stated in the introduction to this thesis, the principal objective of this research was to obtain high quality digital images from radiographic film transparencies. This was achieved by constructing a high resolution image acquisition system. The resolution of the digitised images was determined by both the spatial sampling interval and the resolution of the analogue to digital conversion process. The CCD sensor used during this project provided an array of 2048 photosensitive elements spaced at 14 μm intervals. This very high spatial sampling resolution ensured excellent detail reproduction. The output signal from the CCD sensor was then digitised at a resolution of 12 bits/pixel, which quantised intensity data to 4096 different grey scale levels.

The method of image acquisition used during this project involved scanning a linear array photosensor across the face of the illuminated film.. The mechanical scanning apparatus provided accurate repeatable motion to within 1 μm . A uniform intensity light distribution was provided by a scanning laser system, which used a 15 mW laser diode with a wavelength of 675 nm.

The modulation transfer function of the Sony ILX503 sensor was measured to be 0.85 at a spatial frequency of 5 cycles/mm and 0.6 at 10 cycles/mm. This ensured that at least 60% of contrast information was retained when imaging 50 μm wide features. This performance ensured that adequate contrast resolution was maintained at high spatial frequency. These findings were verified by the digitisation of complex test card patterns, which were reproduced with no visible image distortion.

The prototype system operated at a pixel data rate of 625 kHz. This data rate was dictated by the maximum transmission speed of 20 Mbits/s achieved by the transputer links used for data communication. The overall acquisition speed of the system was highly dependent on the exposure time of the sensor. Using an exposure time of 20 ms, a typical 2048 pixel by 1024 pixel frame could be captured in around 70 seconds.

Other work carried out during this project involved post processing of images produced by the film digitisation system. Several image enhancement techniques were implemented including contrast enhancement, noise reduction, image sharpening and edge detection. These algorithms were seen to significantly increase the amount of useful detail visible within the digitised image.

Finally, work was carried out to investigate the lossless compression of image data. All compression algorithms used during this research were 100% lossless, so that no image data was lost during the compression process. The compression algorithm developed during this research resulted in typical compression ratios of up to 4:1.

11.2 Further Work

During assessment of the film digitisation apparatus a number of areas were identified where system improvements could be made, or further work could be carried out to increase the functionality of the existing apparatus.

A problem was identified with the imaging system, due to scratches on the face of the perspex screen used to support radiographic film plates. These scratches were reproduced as undesirable artefacts on the digitised image. The effects of these scratches could not be eliminated by system calibration. Instead, some new type of film holder was required, which would not be susceptible to scratching. Initially a glass screen was considered, but problems with machining the necessary air channels in glass precluded the use of the original design. A new design has been proposed which would use a plain glass screen supported by a steel window arrangement, but time considerations have prevented its construction.

A method was developed to extend the dynamic range of the image acquisition system by automatically adjusting the exposure time of the sensor, depending on the magnitude of the sensor output signal. However, this technique could not be tested

because of a fault developed by the analogue to digital converter. It was anticipated that, once a new ADC was available, this technique of dynamic range extension could be implemented without any major system modifications.

At present a single CCD sensor is used for image capture, which only provides an effective imaging width of around 28 mm. Any wider area must be imaged in a number of separate scans. The speed of image acquisition could be significantly increased by the use of multiple sensor arrays, each controlled by a separate transputer. The parallel operation of the transputer network would allow several sensor arrays to acquire data simultaneously. The use of multiple sensors would also have the advantage of considerably simplifying the mechanical scanning apparatus. The additional data rates achieved by such a system may require some improved data storage media, such as optical disk, to be used, This would also significantly increase the storage capacity of the system.

It is anticipated that the images produced by the radiograph digitisation system may be used as the input to an automatic radiograph inspection system. Such a system would seek to identify and classify relevant artefacts within a radiographic image. These artefacts may represent either legitimate features of a welded joint or defects within the weld metal. A judgement could then be made on the acceptability or otherwise of the weld being imaged, based on the size and type of any defects identified.

Such a system would require several stages of data processing to retrieve the necessary data from an input image. This image processing hierarchy may consist of some or all of the following general stages, image preparation, image enhancement, feature extraction, object recognition and image interpretation. Initial work carried out on the processing of radiographic images suggested that the very fine sampling grid used during image acquisition was liable to complicate the task of feature detection. The high resolution of the images tended to reduce intensity gradients around feature edges, so that legitimate edges were obscured by random noise fluctuations. In particular it was noted that a 3 x 3 Sobel edge detection mask was inadequate for the accurate detection of features within digitised radiographic images. It was anticipated that any effective feature detection algorithm would require some form of noise filtering prior to the application of a large area edge detection mask.

The successful detection of defects from radiographic images depends to a large extent on the size and shape of defect which is being viewed. For instance, large defects such as gas cavities may be readily visible, whereas thin, transverse cracks may be much more difficult to identify. Obviously, the effectiveness of any automatic defect detection system will depend on the quality of the digitised images supplied to it.

Many researchers have already undertaken work in the area of automatic assessment of radiographic images of welded joints and castings [1] [2] [3] [4]. Defects have been successfully detected and classified using several different techniques. As yet these systems are still rather rudimentary and can only be applied to relatively undemanding tasks. However, as this area of research progresses it is anticipated that systems will be developed that can rival the ability of an experienced inspector.

A further extension of these principles would involve the production of an on-line system for the automatic processing and interpretation of high quality radiographic images. Such a system would use some form of real time radiography, rather than conventional film plates, as the first stage of an integrated inspection system. At this time no practical RTR¹ method has been developed which can match the sensitivity of film systems [5]. However, swift advances in RTR technology have been made in recent years so that it is anticipated that on-line radiographic interpretation systems may well emerge as a viable technology in the future.

References

- 1 W. Lin, C. Dinghua & X. Qinghong, "Study on Automatic Detection and Identification of Defects in Welds by Microcomputer", Proceedings of the International Conference on Quality and Reliability in Welding, Hangzhou, China, September 1984, C-28-1-6.
- 2 W. Daum, P. Rose, H. Heidt & J. Bultjes, "Automatic Recognition of Weld Defects in X-ray Inspection", British Journal of NDT, March 1987, pp 79-82.

¹ Real Time Radiography

- 3 A. Kehoe & G. Parker, "An Intelligent Knowledge Based Approach for the Automated Radiographic Inspection of Castings", *NDT & E International* **25** 1, 1992, pp 23-36.
- 4 A. Gayer, A. Saya & A. Shiloh, "Automatic Recognition of Welded Defects in Real Time Radiography", *NDT & E International* **23** 3, 1990, pp 131-136.
- 5 Y. Zhu, D. Babot & G. Peix, "A Quantitative Comparison Between Linear X-ray Sensitive Array and Image Intensifier Systems", *NDT International* **23** 4, August 1990, p 214-220.

Appendix A : Circuit Diagrams

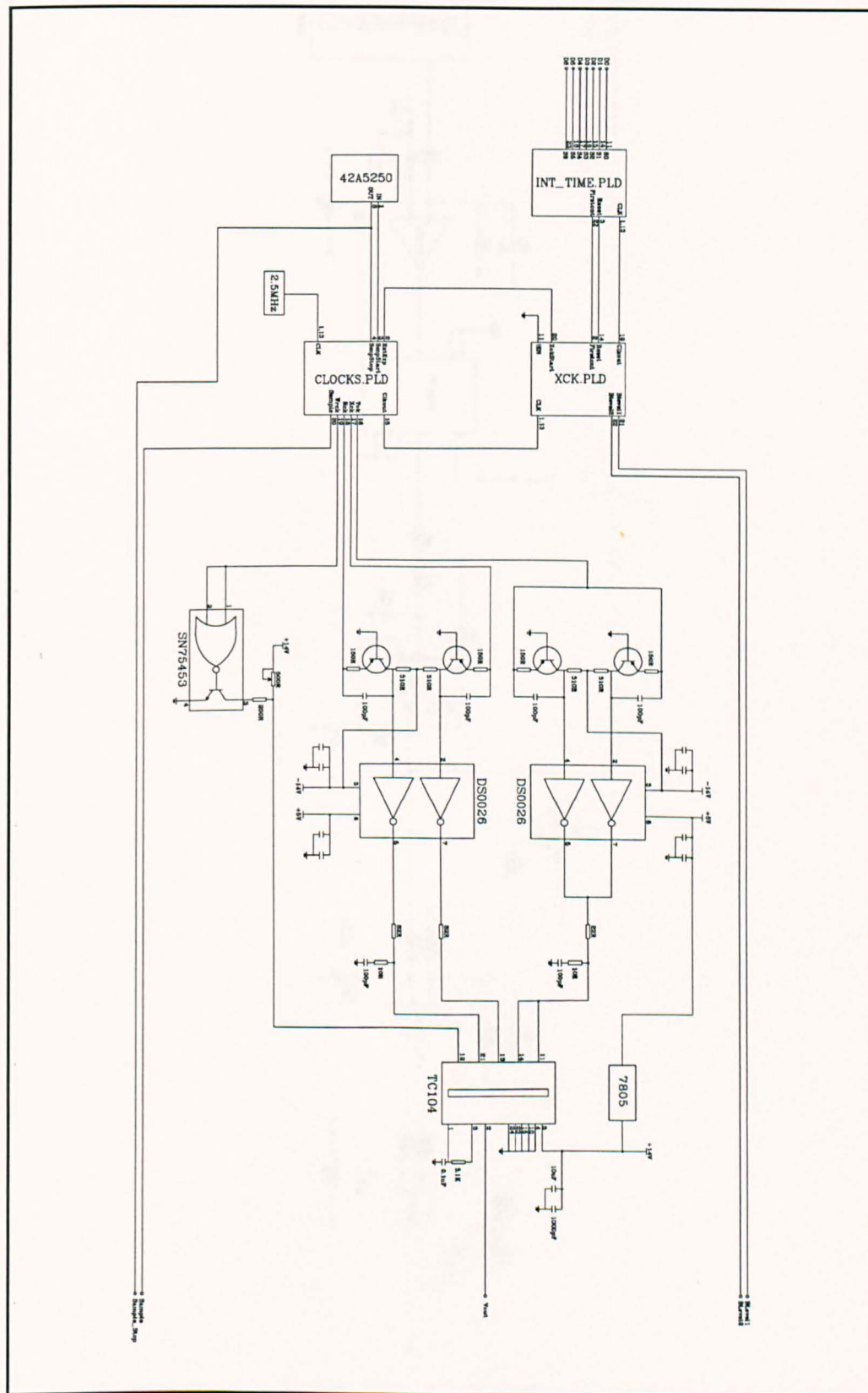


Figure A-1 TC104 Drive Circuit

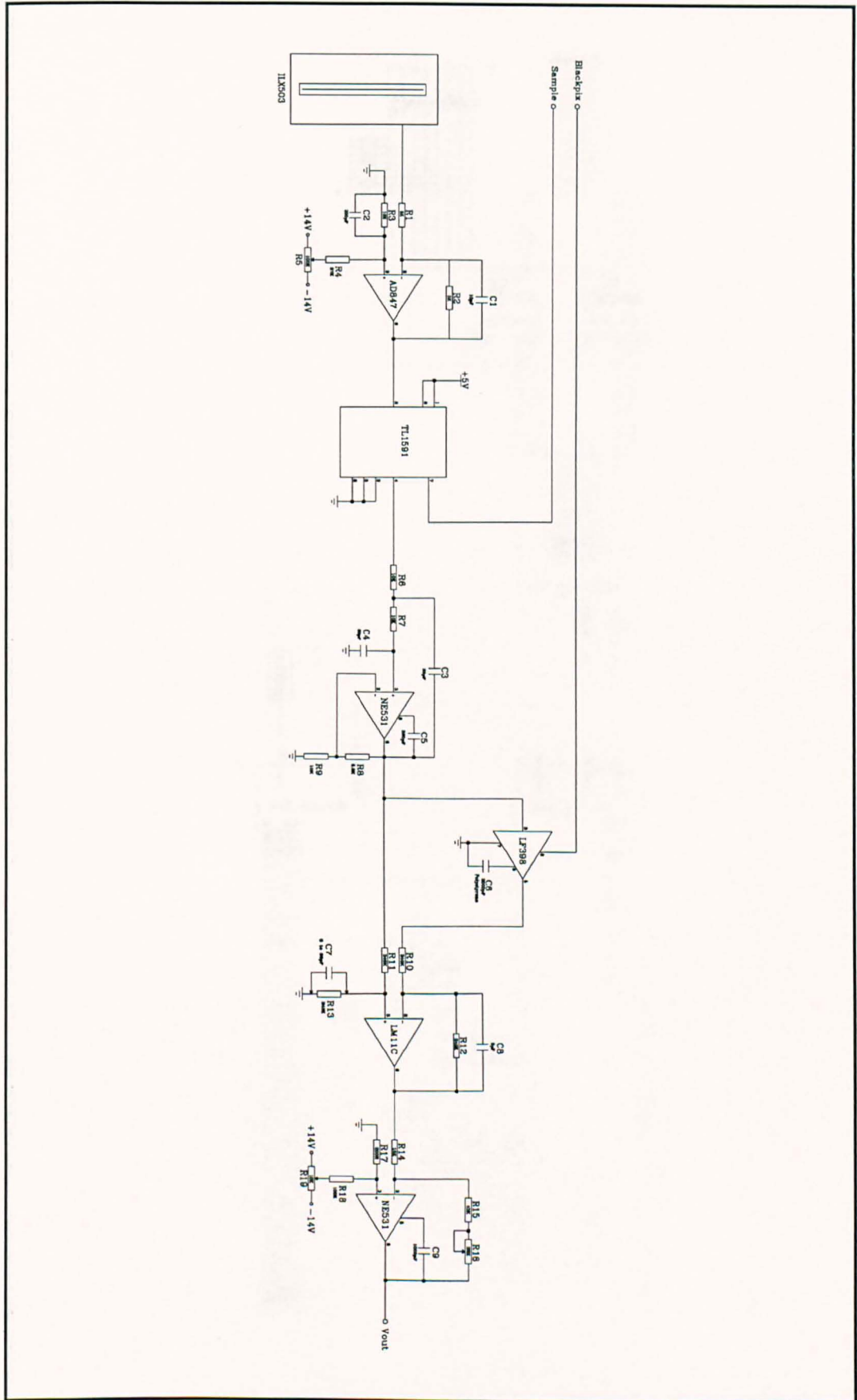


Figure A-2 TC104 Video Signal Processing

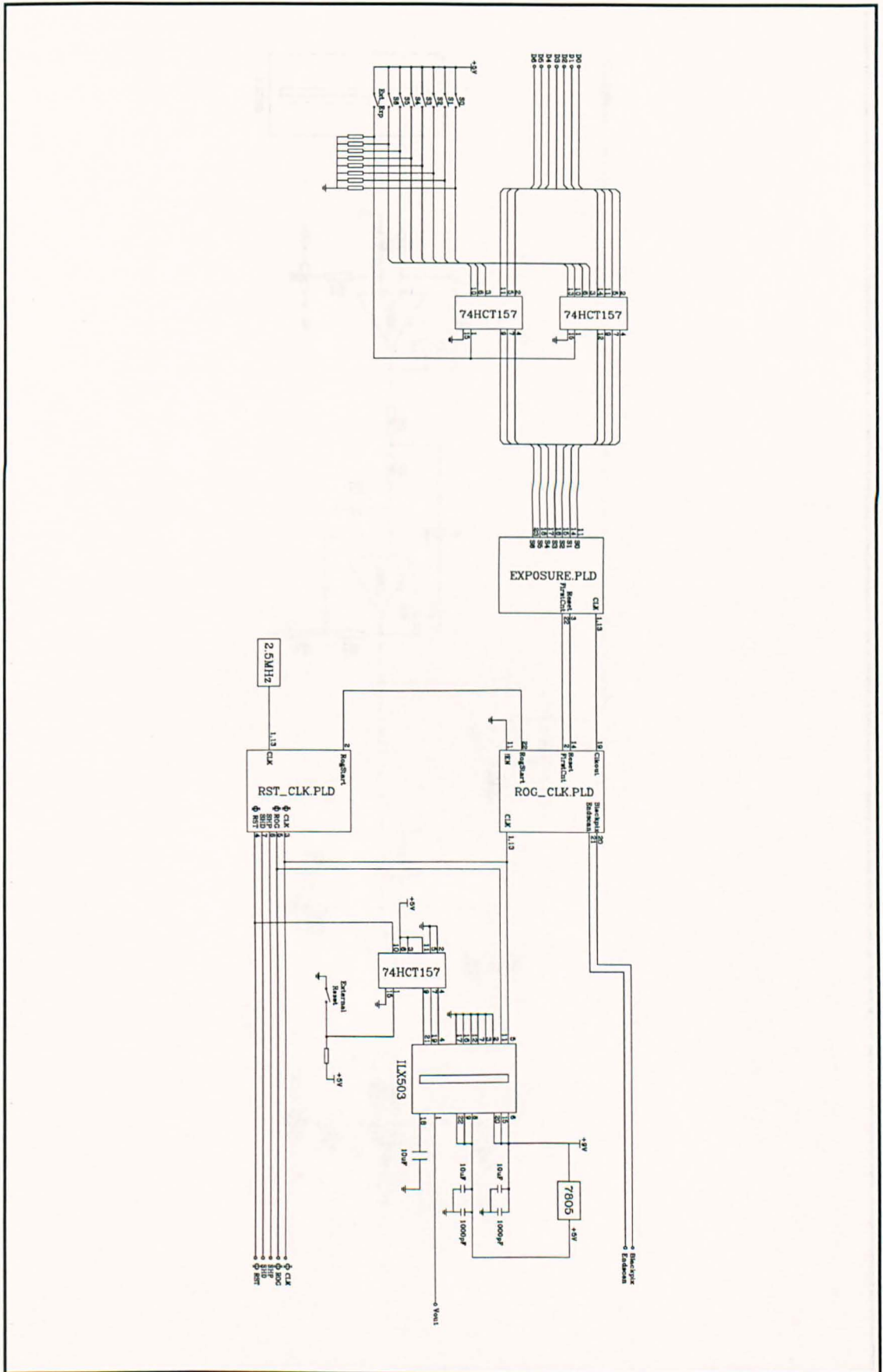


Figure A-3 ILX503 Drive Circuit

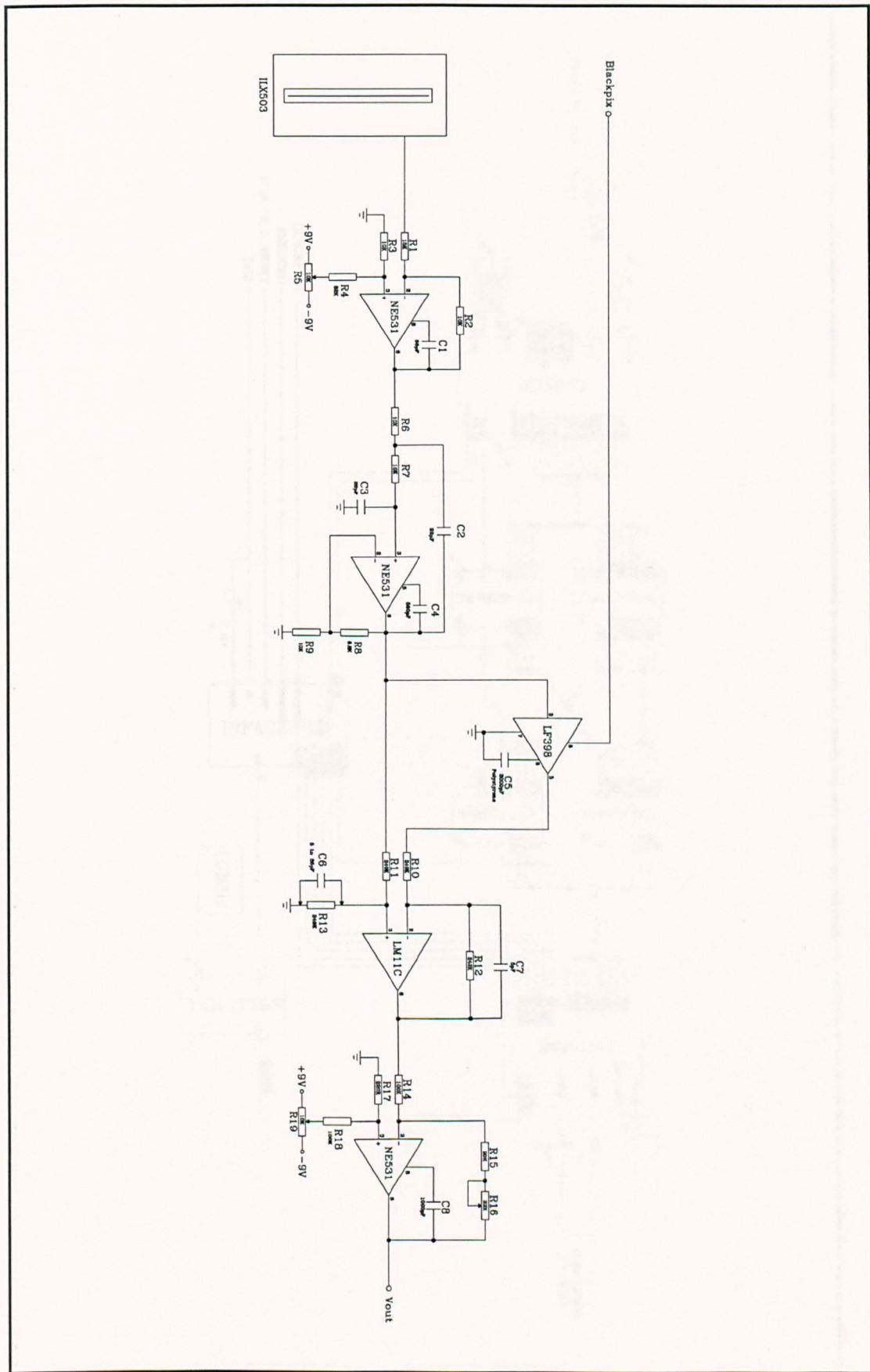


Figure A-4 ILX503 Video Processing Circuit

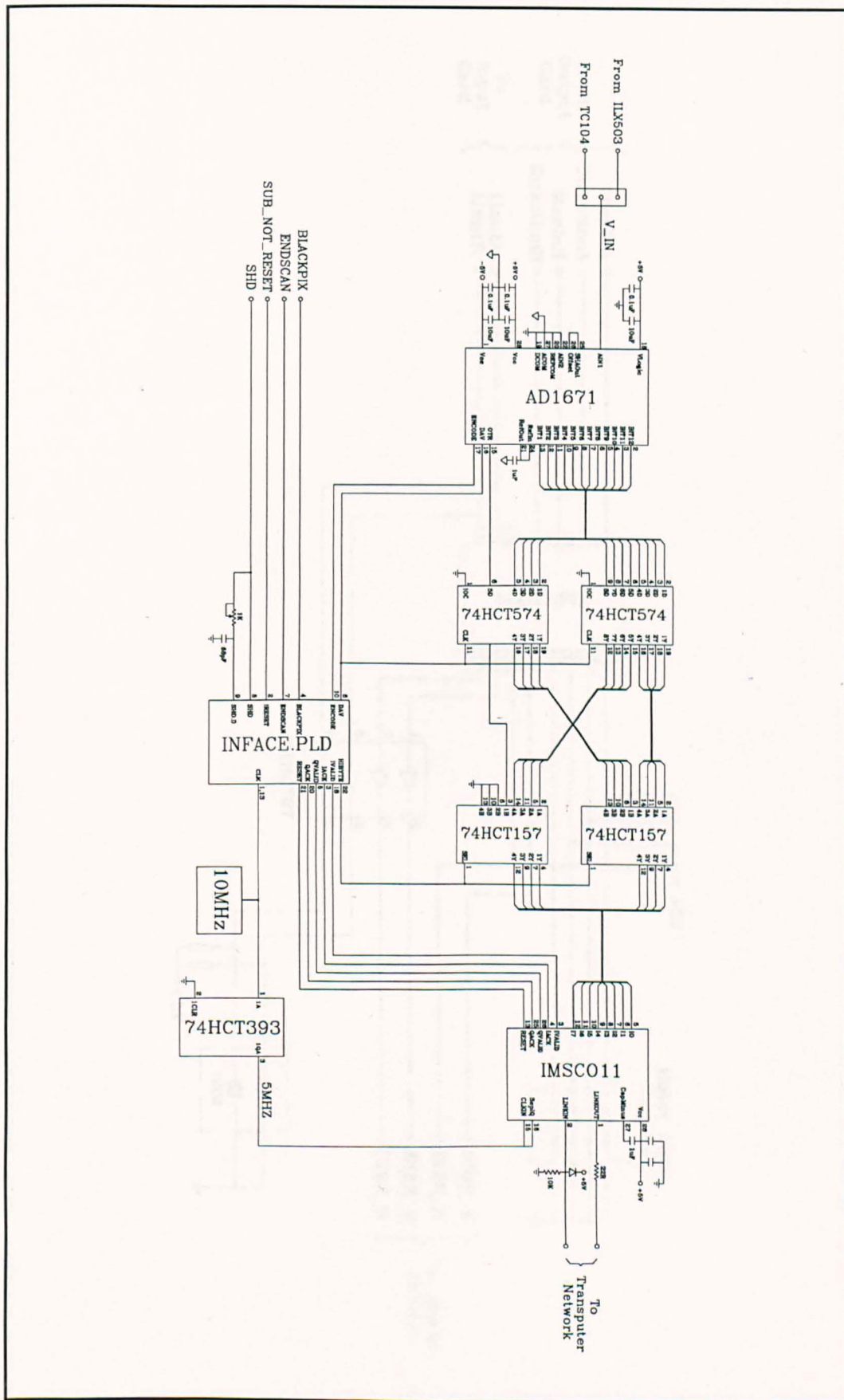


Figure A-5 Interface Circuit

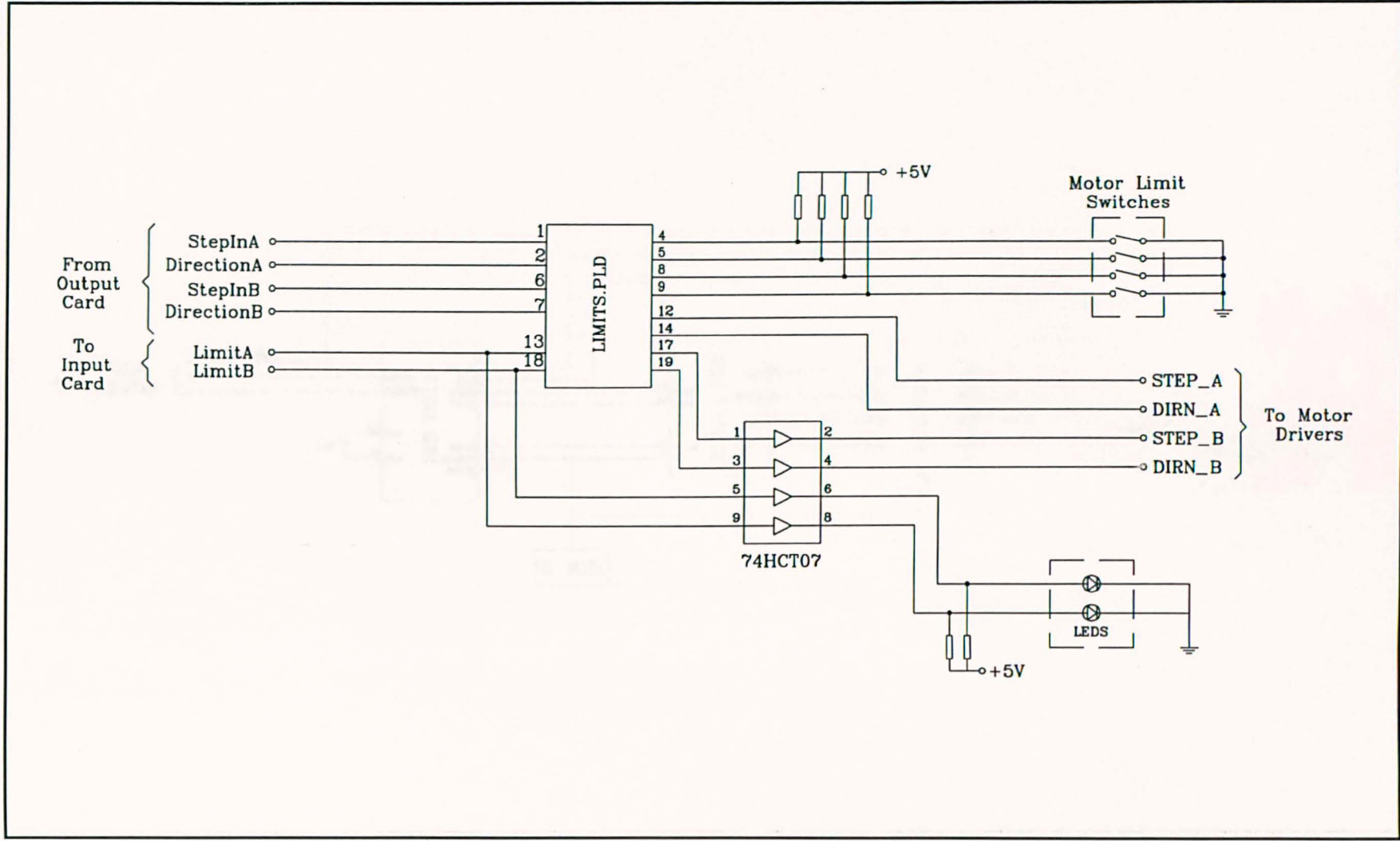


Figure A-6 Stepper Motor Drive Logic

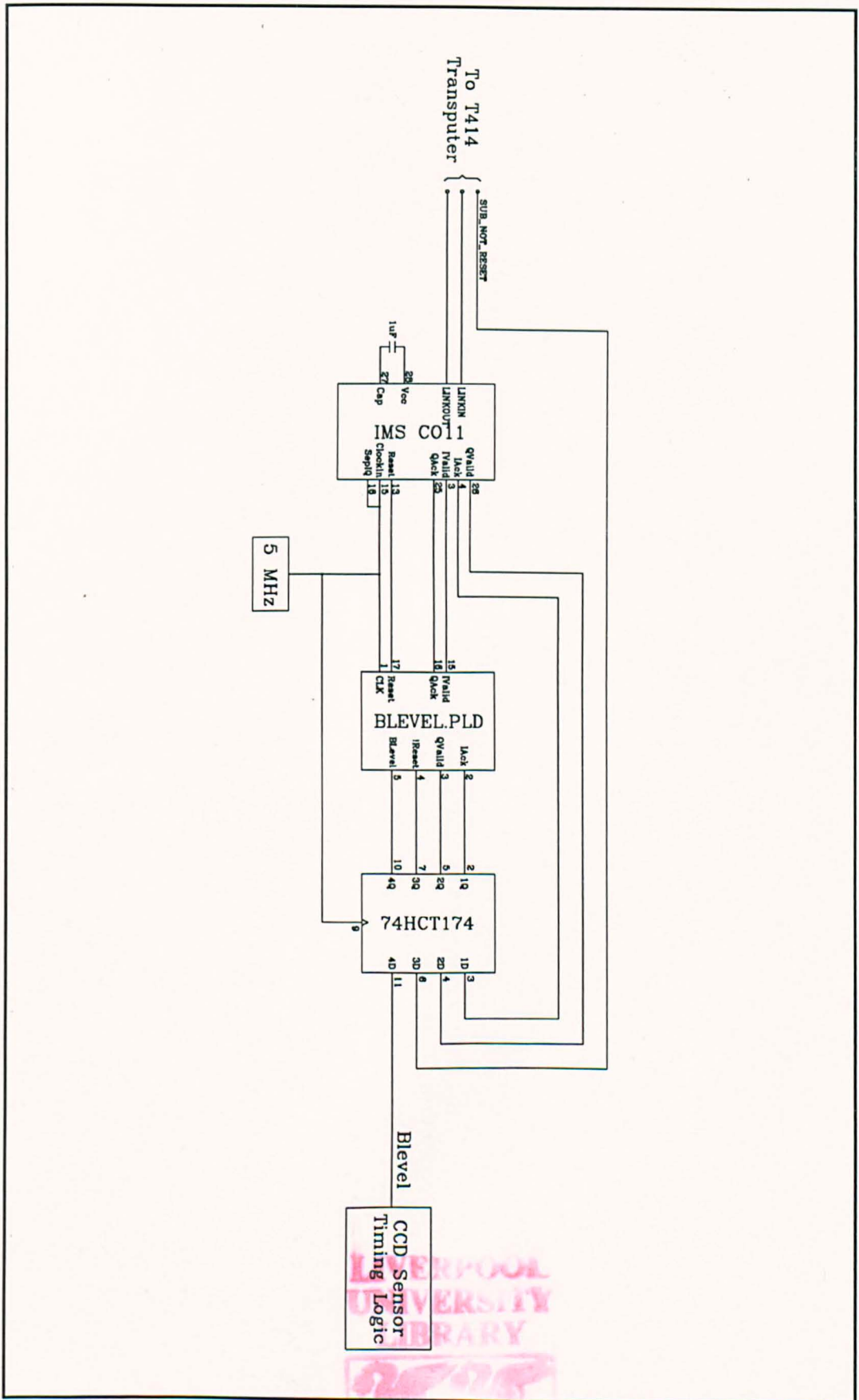


Figure A-7 Laser Scanning Synchronisation Circuit