



University of Tennessee Health Science Center
UTHSC Digital Commons

Theses and Dissertations (ETD)

College of Graduate Health Sciences

6-2000

Qualitative Evaluation of Data Compression in Real-time Ultrasound Imaging

Bijoy J. Sundersingh

University of Tennessee Health Science Center

Follow this and additional works at: <https://dc.uthsc.edu/dissertations>

 Part of the [Equipment and Supplies Commons](#), and the [Investigative Techniques Commons](#)

Recommended Citation

Sundersingh, Bijoy J., "Qualitative Evaluation of Data Compression in Real-time Ultrasound Imaging" (2000). *Theses and Dissertations (ETD)*. Paper 263. <http://dx.doi.org/10.21007/etd.cghs.2000.0308>.

This Thesis is brought to you for free and open access by the College of Graduate Health Sciences at UTHSC Digital Commons. It has been accepted for inclusion in Theses and Dissertations (ETD) by an authorized administrator of UTHSC Digital Commons. For more information, please contact jwelch30@uthsc.edu.

Qualitative Evaluation of Data Compression in Real-time Ultrasound Imaging

Document Type

Thesis

Degree Name

Master of Science (MS)

Program

Biomedical Engineering

Research Advisor

Herbert D. Zeman, Ph.D.

Committee

Donald Emerson, M.D. Stephen Kritchevsky, Ph.D. Lawrence Jordan, Ph.D.

DOI

10.21007/etd.cghs.2000.0308

Qualitative Evaluation of Data Compression in Real-time Ultrasound Imaging

A Thesis
Presented for
The Graduate Studies Council
The University of Tennessee, Memphis

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science
In the Joint Graduate Program in Biomedical Engineering
From The University of Tennessee
and
The University of Memphis

By

Bijoy J. Sundersingh
June, 2000

Copyright ©Bijoy J. Sundersingh, 2000

All rights reserved

DEDICATION

This thesis is dedicated to my parents

Mr. Solomon Sundersingh

and

Mrs. Neela Sundersingh

for their unending love and support in all my endeavors.

ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. Herbert Zeman for giving me the opportunity to work on this project and for his guidance and advice throughout the program. I would like to thank Dr. Donald Emerson who made this research possible and for his interest and support in this project. I would also like to thank my other committee members Dr. Stephen Kritchevsky and Dr. Lawrence Jordan for their active participation and very valuable suggestions on this project. And finally I would like to thank all the radiologists who participated in this project.

ABSTRACT

The purpose of this project was to evaluate qualitatively real-time ultrasound imaging using objective and subjective techniques to determine the minimum bandwidth required for clinical diagnosis of various anatomical and pathological states. In the experimental setup live ultrasound video samples representing the most common clinical examinations were compressed at 128, 256, 384, 768, 1152 and 1536 kbps using a compressor-decompressor (CODEC) adhering to International Telecommunication Union (ITU-T) recommendation H.261. A protocol for qualitative evaluation was developed and subjective and objective testing were performed based on this protocol. Subjective methods comprised of inter-rater reliability tests using kappa statistics and three way Analysis of Variance (ANOVA) using General Linear Models (GLM). Objective testing were performed using histogram analysis and estimation of peak signal to noise ratios.

The kappa scores for all bandwidths greater than 256 kbps indicated good inter-rater reliability and minimum variation in confidence levels. Using the results from GLM and ANOVA we could not establish a trend in degradation of observer confidence with increasing compression ratios. The histogram analysis showed a linear increase in standard deviation values, indicating a linear scatter in pixel intensity, with increasing compression ratios.

Although higher compression levels were evaluated, only video clips with bandwidths greater than 256 kbps displayed satisfactory temporal and spatial resolution, good enough to make clinical diagnosis of various anatomical and pathological states. The evaluations also indicate that compressed real-time ultrasound imagery using H.261 can be transmitted over a T1 or ADSL networks.

Table of Contents

Chapter 1 Background and Introduction	1
1.1 Ultrasound Imaging	1
1.1.1 Introduction	1
1.1.2 Basic Acoustics	1
1.1.2.1 Propagation of Sound	2
1.1.2.2 Acoustic Impedance	2
1.1.2.3 Reflection	2
1.1.2.4 Refraction	3
1.1.2.5 Attenuation	3
1.1.3 Instrumentation in Real-time Ultrasound Imaging	3
1.1.3.1 Transmitter	4
1.1.3.2 Pre-processing	6
1.1.3.3 Scan Converter	7
1.1.3.4 Image Post-processing	8
1.1.4 Artifacts	8
1.1.4.1 Reverberation Artifact	9
1.1.4.2 Mirror Image Artifact	9
1.1.4.3 Refraction Artifact	10
1.1.4.4 Shadowing	10
1.1.5 Real-time Vs Still Ultrasound Imaging	10
1.2 Data Compression	11
1.2.1 Introduction	11
1.2.2 Motivation for Reliable Data Compression in Ultrasound Imagery	12
1.2.3 Information Theory	13
1.2.4 Models	14
1.2.4.1 Physical Models	14
1.2.4.2 Probability Models	15
1.2.4.3 Markov Models	15
1.2.5 Compression Schematic	16
1.2.6 Lossless Compression	17
1.2.6.1 Pre-processing	17
1.2.6.2 Huffman Coding Algorithm	18
1.2.7 Lossy Image Compression	20
1.2.7.1 Discrete Cosine Transform (DCT) Based Coding	21
1.2.7.2 Motion Compensated Prediction	22
1.2.8 H.261 Video Coding Standard	24
1.2.8.1 Motivation for Using H.261	24
1.2.8.2 Common Intermediate Format (CIF)	24

1.2.8.3 H.261 Encoder	25
1.2.8.4 H.261 Decoder	26
1.2.9 Applications of Data Compression in Medical Imaging	27
1.2.9.1 Teleradiology	27
1.2.9.2 Telesonography	29
1.2.9.3 Ultrasound Picture Archiving and Communication Systems	30
1.2.10 Legal Issues and Standards for Medical Image Compression	33
1.3 Research Objectives	34
1.3.1 Introduction	34
1.3.2 Objectives	35
1.3.2.1 Temporal Blurring	35
1.3.2.2 Spatial Blurring	36
1.3.2.3 Increase in Contrast	37
Chapter 2 Methods and Materials	38
2.1 Introduction	38
2.2 Protocols and Guidelines	38
2.3 Experimental Overview	39
2.4 Materials	42
2.4.1 Ultrasound Scanner	42
2.4.2 Compressor-Decompressor Unit (CODEC)	46
2.4.3 Betacam	47
2.4.4 Display System	47
2.4.5 Digital Video Editing Workstation	48
2.5 Subjective Quality Assessment	50
2.5.1 ROC Analysis	50
2.5.2 Design of the Questionnaire	53
2.5.3 Data Collection	55
2.5.4 Inter-rater Reliability Studies	55
2.5.4.1 Index of Reliability- Kappa (κ)	56
2.5.4.2 Kappa Estimation	56
2.5.5 General Linear Models (GLM) and Analysis of Variance (ANOVA)	59
2.5.6 Hypothesis Test Procedure	60
2.6 Objective Quality Assessment	61
2.6.1 Data Collection	62
2.6.2 Histogram Analysis	62
2.6.3 Peak Signal to Noise Ratio (PSNR)	64

Chapter 3 Results and Discussions	67
3.1 Kappa Studies	67
3.2 General Linear Models	70
3.3 Average Confidence	71
3.4 Key for the ANOVA Table and SAS Results	82
3.4.1 Dependent Variable	82
3.4.2 Degrees of Freedom (DF)	83
3.4.3 Sum of Squares (SS)	83
3.4.4 Mean Squares (MS)	84
3.4.5 F-Statistic	84
3.4.6 P-Value	84
3.5 Objective Testing	96
3.6 Clinical Impressions	106
Chapter 4 Future Work	109
4.1 Subjective Testing	109
4.2 Objective Testing	110
Conclusion	112
List of References	113
Vita	117

List of Tables

Table 2-1.	Ultrasound transducer settings for various studies in the experiment	43
Table 2-2.	Truth table containing the comparisons of observations of the images as seen by the trained observer (Radiologist B) and the clinical gold standard (Radiologist A)	56
Table 3-1.	The estimated average values of kappa for different bandwidths across all the studies and raters	68
Table 3-2.	The estimated average values of kappa for different studies across all the bandwidths and raters	69
Table 3-3.	The output of the SAS GLM procedure for 3-way ANOVA with the dependent variable confidence and the independent variables image, observer and bandwidth	86
Table 3-4.	The results of the SAS procedure that models average confidence as a function of the bandwidth across all observers and images	87
Table 3-5.	The results of the SAS procedure that models average confidence as a function of image across all bandwidths and observers	88
Table 3-6.	The results of the SAS procedure that models average confidence as a function of the observer across all bandwidths and images.	89
Table 3-7.	A summary of the results of the SAS GLM procedure modeling each study parameter separately to find the effect of bandwidth on observer confidence	90
Table 3-8.	Mean and standard deviation values from the histograms of the difference images at various compression levels	105

List of Figures

Figure 1-1.	Schematic of a real-time ultrasound scanner	4
Figure 1-2.	Schematics of linear array and phased array sector transducers	5
Figure 1-3.	Schematic diagram of the scan converter	7
Figure 1-4.	A generic compression scheme	16
Figure 1-5.	Block schematic for lossless compression schemes	18
Figure 1-6.	Huffman's tree for an eight gray level image	20
Figure 1-7.	A schematic of DCT based compression schemes	23
Figure 1-8.	Block schematic of the H.261 encoder	26
Figure 1-9.	Block schematic of the H.261 decoder	27
Figure 1-10.	Block diagram of a generic ultrasound PACS module	32
Figure 2-1.	The experimental setup for the subjective and objective testing at the Regional Medical Center, Memphis	41
Figure 2-2.	Sample ROC for 4 compression ratios by the same observer	51
Figure 2-3.	Flow diagram of image handling including the equipment used, format and the software used in each stage.	63
Figure 3-1.	Average confidence plots for the gall bladder, plotted against bandwidth and specific study aspects	72
Figure 3-2.	Average confidence plots for the kidney, plotted against bandwidth and specific study aspects	73
Figure 3-3.	Average confidence plots for the fetal heart, plotted against bandwidth and specific study aspects	74
Figure 3-4.	Average confidence plots for the fetal heart, plotted against bandwidth and other specific study aspects	75
Figure 3-5.	Average confidence plots for the intra-uterine pregnancy, plotted against bandwidth and specific study aspects	76

Figure 3-6.	Average confidence plots for the uterus, plotted against bandwidth and specific study aspects	77
Figure 3-7.	Average confidence plots for the fetus, plotted against bandwidth and specific study aspects	78
Figure 3-8.	Average confidence plots for the liver, plotted against bandwidth and specific study aspects	79
Figure 3-9.	Average confidence plots for the liver, plotted against bandwidth and other specific study aspects	80
Figure 3-10.	Average confidence plots for the female pelvis, plotted against bandwidth and specific study aspects	81
Figure 3-11.	The PSNR values at various compression levels	97
Figure 3-12.	Difference image and the corresponding histogram plot at 128 kbps	100
Figure 3-13.	Difference image and the corresponding histogram plot at 256 kbps	101
Figure 3-14.	Difference image and the corresponding histogram plot at 384 kbps	102
Figure 3-15.	Difference image and the corresponding histogram plot at 768 kbps	103
Figure 3-16.	Difference image and the corresponding histogram plot at 1536 kbps	104

Chapter 1

Background and Introduction

1.1 Ultrasound Imaging

1.1.1 Introduction

Diagnostic ultrasound imaging is based on the detection and display of acoustic energy reflected from various interfaces in the body. The interactions between acoustic energy and acoustic interfaces provide the fundamental information for generating gray scale two-dimensional images of the organs and tissues in the body. The laws of acoustic physics govern the scattering of sound energy by interfaces of materials with different properties.

1.1.2 Basic Acoustics

Sound is produced as a result of mechanical energy travelling through matter as a wave producing alternating compression and rarefaction. Pressure waves are propagated by physical displacement of materials through which the sound is being transmitted. Changes in pressure with time define the basic units of measurement for sound.

The number of complete pressure cycles per second describes the frequency (f) of the sound wave while the distance between corresponding points in adjacent cycles is the wavelength (λ). The frequency of the sound wave is measured in Hertz (cycles/second).

Ultrasound and audible sound essentially possess the same acoustic properties, differing only in frequency. Audible sound is in the range of 20 Hz to 20 kHz, while frequencies greater

than 20 kHz are called ultrasound. The range of frequencies used in diagnostic imaging is in the range of 1 MHz to 15 MHz.

1.1.2.1 Propagation of Sound

In diagnostic ultrasound imaging, short pulses of sound energy are transmitted into the body and are propagated through the tissue. The propagation of sound in the tissue is in the direction of particle movement. The speed of sound in tissue is affected by the physical properties of the tissue [1]. The density of the propagation medium affects propagation velocity.

1.1.2.2 Acoustic Impedance

Images in the scanners are produced by the detection and display of echoes. To produce a reflecting echo, a reflecting interface must be present and the sound should not pass through a totally homogeneous medium. The amount of reflection is determined by the difference in acoustic impedance of the interface materials, which is given by the product of density and the propagation velocity of sound in the travelling medium.

$$Impedance = \frac{Density(kg / m^3)}{Speed(m / s)} \quad (1.1)$$

1.1.2.3 Reflection

Ultrasound is reflected when it strikes an acoustic interface. The size and features of the interface determine the amount of reflection. The amount of energy reflected at the acoustic interface is expressed as a fraction of the incident energy and is called the reflection coefficient (R) [1].

$$\text{Reflection Coefficient(R)} = \frac{\text{Reflected Intensity(W / cm}^2\text{)}}{\text{Incident Intensity(W / cm}^2\text{)}} \quad (1.2)$$

1.1.2.4 Refraction

When sound passes from a tissue with a higher propagation velocity to a tissue with a lower sound velocity, there is a change in direction of the sound wave and this change in direction of the sound wave is given by Snell's law,

$$\frac{\text{Sin}\theta_1}{\text{Sin}\theta_2} = \frac{c_1}{c_2} \quad (1.3)$$

Where θ_1 and θ_2 are the incident and reflected angles and c_1 and c_2 are the respective velocities.

1.1.2.5 Attenuation

Attenuation is a decrease in the intensity of sound waves as they pass through various tissue interfaces. Attenuation is the result of the combined effects of absorption, scattering and reflection. Attenuation is usually measured in absolute units and is represented by decibels. In diagnostic ultrasound imaging, higher frequencies are attenuated more.

1.1.3 Instrumentation in Real-time Ultrasound Imaging

Ultrasound imaging is a method of real-time physical imaging that uses hardcopy images for storage. Figure 1-1 shows a block schematic of an ultrasound scanner. The scanners for real-time imaging have complex instrumentation and despite the complexity, all scanners have basic components performing the same key functions.

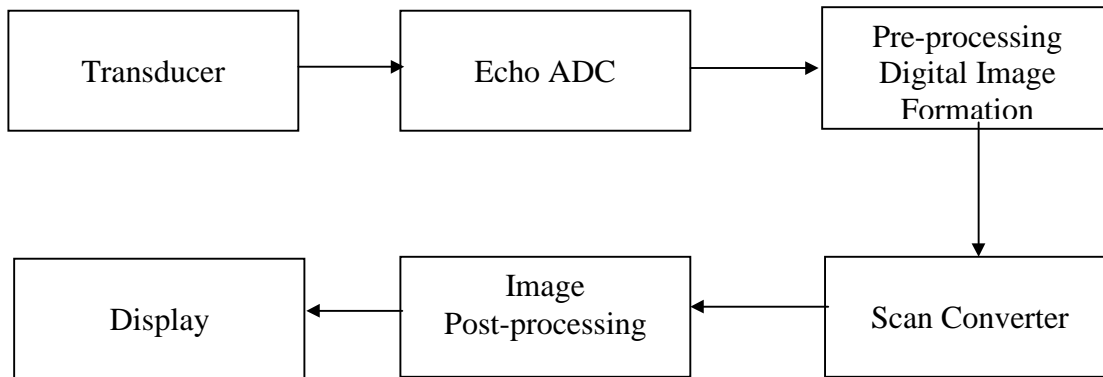


Figure 1-1. Schematic of a real-time ultrasound scanner.

1.1.3.1 Transmitter

The transducer is a part of the transmitting system that converts electrical energy into acoustic energy and is also a part of the receiving system when it converts the returning acoustic energy to electrical signals. During transmission acoustic waves are generated by applying a voltage pulse to a piezo-electric crystal resulting in the creation of a pressure sonic wave. The rate at which the voltage pulses are applied to the transducer is called the Pulse Repetition Frequency (PRF). Typically PRF values range from 0.5 to 2.5 kHz. The frequency of the generated sound wave depends on the PRF and the thickness of the piezo-electric crystal. Types of transducers include linear array and phased array sector transducers.

A linear array transducer is constructed by assembling individual transducer elements into an array or by using a single slab of ceramic which is the size of the full array and by cutting groves into it to produce the separate transducer elements. Each element is typically 1 to 2mm thick. In a linear array transducer, a small number of transducer elements are grouped together to

make up an 'aperture' for a single line of sight. This aperture is stepped along the array one element at a time to produce a rectangular format.

Figure 1-2 shows the schematics of linear array and phased array sector transducers. In a phased array sector transducer, the construction of the transducer uses only the required number of transducer elements to make the aperture. Each element in the transducer can be triggered at different time intervals using the appropriate delay circuits. Typically time delay is applied gradually increasing from one end of the transducer to the other. The delay produces a beam as shown in the figure 1-2. The sound wave front can be steered at any angle to a sector scan depending on the time at which each array element is triggered. This type of sector scan is often referred to as the 'windshield effect' [2]. During the detection, the signal is received by the transducer and echo details such as amplitude and positional data are retained for preprocessing while the noise components are dropped.

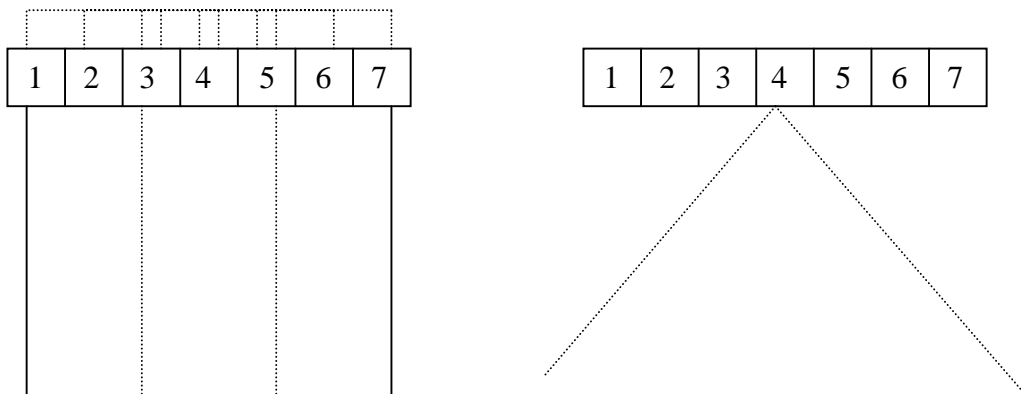


Figure 1-2. Schematics of linear array and phased array sector transducers.

1.1.3.2 Pre-processing

In this stage, all the manipulations of received signal are done to produce the data needed by the scan converter stage. The returning pressure waves do not reach the transmitter simultaneously. However, the next stage (scan converter) requires all the data to be presented together to store the data amplitudes as continuous variations. Therefore, practical operation requires some form of data buffering until all the relevant transmitted frequencies have returned from the reflecting interfaces [3]. Data buffering is achieved through conversion of the time varied signals to quantified digital form. Quantization is a process where the analog signals are assigned specific digital code. In gray scale images, the images are usually quantized by 8-bit analog to digital converters, which assigns 256 shades of gray to the signal. With a limited representation of only 256 gray levels, it is important to allocate available levels to favor those portions of echo amplitude variations having the greatest significance in any given diagnosis. For example in a fetal study, in the measurement of the biparietal diameter (the cross sectional diameter of the head), strong echoes outlining the skull are the most important. Therefore, ideally, in fetal studies most of the quantized levels should be allocated to the stronger returning echoes. In a liver study, attention is more focused on the echoing structures forming the image texture patterns. The returning echoes from the liver are much weaker than the echoes from the other abdominal structures such as the diaphragm. Therefore while studying the liver, more levels of signal representation are applied to the weaker signals. Thus in the pre-processing stage, the user can selectively allocate the levels and amplitude range depending on the type of examination.

1.1.3.3 Scan Converter

The section briefly describes the function of the scan converter. Detailed technical discussions of the scan converter are covered in detail in a publication by Ophir and Mankad [32]. The digital scan converter is used to convert returning ultrasound echoes in the multi-sector scan format to a raster format for displaying in a cathode ray tube. The scan converter comprises of elaborate digital circuitry for the conversion of the format. The scan converter also contains a matrix of digital memory elements where the returning echo and its positional data are stored. A schematic of the digital memory part of the scan converter is shown in figure 1-3. The memory locations are accessed for the writing of the incoming echo amplitudes to the lines of the cathode ray tube from the data stored in each memory location. The reading out to the analog output is done by the digital to analog converter

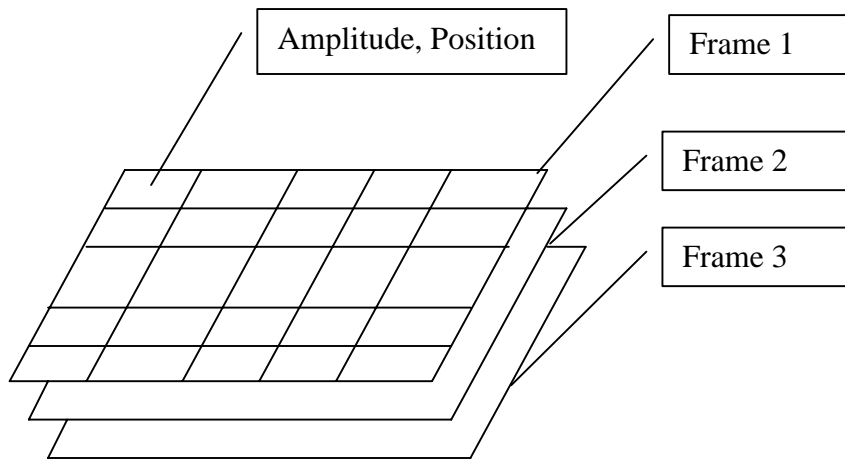


Figure 1-3. Schematic diagram of the scan converter.

The scan converter also contains circuitry which comprises a means of generating X positional address voltage with an inclination proportional to the sine of the angulated scan line being imaged. The circuitry also generates a corresponding Y address linear ramp voltage with an inclination proportional to the cosine of the angulated scan lines being imaged.

1.1.3.4 Image Post-processing

The encoding of the x and y positions of the face of the transducer and the angular orientation of the transducer with respect to the normal of the scanning surface is accomplished by the encoding circuits of the post-processing phase. The assignment of gray levels and processing of the signal to enhance certain image features is also done at this stage. The modified characteristics are those that emphasize outlines and those that maximize the contrast between tissues. The data processing done in this stage does not alter the data stored in the scan converter. Data buffering is done so that other processing such as frame averaging and smoothing which reduces the speckle noise in the returning signals may be accomplished by making a copy of the data from the scan converter.

1.1.4 Artifacts

In imaging, artifacts are defined features in an image, which are not properly indicative of the structure imaged. Some artifacts are undesired and interfere with interpretation whereas others help identify certain structures. In ultrasound imaging, certain artifacts such as shadowing, enhancement and speed error can be very useful, conveying vital information about the physical state, while others can be harmful and hinder interpretation. Artifacts in ultrasound images occur when structures are not real, are missing, are improperly located, or are of improper size and

shape [4]. An artifact can lead to a missed diagnosis or obscure a real abnormality. In this section some of the characteristics of commonly occurring artifacts are discussed.

1.1.4.1 Reverberation Artifact

In reverberation artifacts, the sound bounces back and forth between two tissue interfaces. Some of the sound waves returning to the transducer get trapped in the tissue interface and get reflected again. The reflections between interfaces prolong the time of flight and produce an artifact distal to the surface. The reflection phenomenon often occurs when the sound beam is perpendicular to a strong reflector such as a soft tissue–air interface or the abdominal wall, which lies at greater depths than a significant amount of subcutaneous adipose tissue. Depending on the echogenicity of the tissue, reverberation artifacts can continue. An example of a reverberation artifact in abdominal studies would be the parallel lines in the urinary bladder produced by the ultrasound reverberations between parallel layers of tissue.

1.1.4.2 Mirror Image Artifact

The mirror image artifact is similar to the reverberation artifact, however it is different by the fact that the extra reflection comes from within the organ. Although the extra reflection may be within the line of the sound beam, more commonly the sound is reflected off an angle to another interface so that, like a real mirror, the artifact shows up like a virtual object. For example, a mirror image artifact can cause strange appearances such as the simulation of emphysema or of a lung abscess [5] and an invasion of a transitional cell carcinoma through a bladder wall.

1.1.4.3 Refraction Artifact

Refraction artifacts are seen because of the bending of the sound beam at an oblique interface between the tissues of different acoustic velocities and by the defocusing of the sound beam at these interfaces. Refraction is also caused by the displacement of the distal structure by a proximal refraction of the beam. Refraction artifacts can be seen on the edge of a cyst because the sound beam is refracted more by the proximal surface than the distal surface.

1.1.4.4 Shadowing

The sound pulses that do not reach the deep tissues to produce an echo cause acoustic shadowing. Complete shadowing does not cause confusion and can be more helpful in confirming a gallstone or a renal stone. Focusing plays an important role in the production of acoustic shadows. It is important to use the appropriate transducer and to adjust the focus to the depth of the region being imaged to appreciate the shadow.

1.1.5 Real-time Vs Still Ultrasound Imaging

Ultrasound imaging is a *live*, visual physical examination that uses hard copy images to create a permanent record of the examination [6]. Therefore ultrasound imaging is different from other imaging modalities except angiography, by the fact that a live examination requires the participation of a sonologist for the study. The two types of imagery used in ultrasound imaging are still-images and real-time video images; both modalities having their advantages and disadvantages.

In ultrasound imaging, although there are recommended views of structures, there are no absolute directions and angle requirements and therefore ultrasound imaging is referred to as a

free-form examination. This free form evaluation feature allows the sonologist to design customized views and to decide on vantage points from which to image the structure. An analogy for the relationship between real-time and still imaging is that of a vacation trip and the souvenir photographs [6]. The actual experience is the heart of the trip just as the real-time study is the heart of the ultrasound examination. Just as a photograph taken during the trip to record one's impression of a site cannot match the actual experience of that actual trip, still imaging cannot replace the evaluation possible through real-time imaging.

Another reason why a collection of frozen images can't adequately replace a live ultrasound examination is that a still image leaves out the possibility of retrieving information about the structures beyond the edge of the frame of the frozen image. The retrieval aspect can be understood by referring to the analogy of the vacation trip. The tourist may have taken pictures of an important local landmark but not the new fast-food restaurant next door. When reviewing the still photographs, the person may not recollect the existence of the fast-food restaurant. Therefore in order to utilize best the capabilities of ultrasound imaging, real-time imaging is necessary.

1.2 Data Compression

1.2.1 Introduction

Data compression is the process of representing the information that is present in a compact form. In most cases, the statistical structure of information present in the signal is used to provide data compression. Exploiting the information content in the signal creates these compact representations. Other compression techniques exploit structures existing in the data to provide data reduction.

In image compression, the information content is reduced for storage or transmission. Since images are intended to be perceived by human beings, and their perceptual abilities are limited, data reduction is achieved by removing large amounts of data within an image that is redundant or outside visual perception.

1.2.2 Motivation for Reliable Data Compression in Ultrasound Imagery

Digital representations of ultrasound images are highly beneficial for purposes of archiving, transmission and retrieval. Digital sonography would thus have tremendous utility by permitting random access to the studies and rapid transfer of sonograms to remote sites. However the size of the image files associated with digital sonograms impose significant restrictions for archiving or transmission to remote sites. Typical real-time ultrasound imaging generates about 74 Mbps for gray scale video and 221 Mbps for 24bit-color video. However the telecommunication systems offer only 33.6 kbps and 128 kbps for transmission of signals. T1 lines and T3 lines offer 1.4 Mbps and 44.7 Mbps bandwidths but are very expensive. Ultrasound teleradiology can easily handle still images, however they require very high data transmission rates to enable real-time interaction between a sonographer in a remote location and the radiologist. Asynchronous Transfer Mode (ATM) is used within city limits to perform telediagnosis. However ATMs cannot be used for long distances outside the metropolitan area because of prohibitive transmission costs. Recently newer communications options such as cable modems from RoadRunner, Digital Subscriber Technology (DSL) and Asynchronous Digital Subscriber Line (ADSL) from telephone companies such as Bellsouth provide bandwidth from 1.6 Mbps to 7.1 Mbps and are now available to the end user. DSL is a new modem technology that uses digital coding techniques to increase communication bandwidths to greater than a T1

connection at lower costs. However these data rates still do not suffice for uncompressed real-time video to be transmitted without data compression. Therefore to draw a balance between the transmission rates and the costs, diagnostically acceptable levels of data compression become essential.

Data compression is typically characterized into two groups of algorithms: lossless compression and lossy compression. Lossless compression is usually defined as compression rates of up to 3:1, where no visible degradation of the image is seen because of exact numeric reconstruction of the original image. However lossless compression does not suffice for the purpose of transmission because the compression required for ultrasound imagery is of the order of at least 100:1. At very high data compression rates, the compression is lossy and there will be visibly significant perceptual losses in the images.

Effective modeling of data leads to good coding schemes in data compression, and information theory provides the framework for the development of the compression schemes.

1.2.3 Information Theory

Claude E. Shannon of Bell Labs put a quantitative measure for information by defining a quantity called self-information [18].

Let S be a source that generates symbols $S_1, S_2, S_3, \dots, S_N$; For example if S represents a digital image, then S_i may represent one of its N pixel values.

Let P_i denote the probability of occurrence of one of those random symbols S_i , then the self- information of the symbol is given by,

$$I(S_i) = \log \frac{1}{P_i} \quad (1.4)$$

$I(S_i)$ will be the information received from S_i .

From equation 1.4, it can be deduced that the probability of a symbol is low if the amount of self-information associated with it is high and the probability of the symbol will be high if the amount of associated self-information is low. Self-information is represented in bits when the base of the log is 2, nats when the base of the log is e and hartleys when log base 10 is used.

The average information or the measure of the average number of binary symbols required to code the output of a source is called entropy [18] and is defined as

$$H(S) = \sum_i P_i \log \frac{1}{P_i} \quad (1.5)$$

The value of entropy depends on the structure of the source.

1.2.4 Models

As mentioned earlier, structure in data is an essential part of data compression techniques. Therefore having a good model for the data will be helpful in estimating the entropy of the source thus leading to the design of efficient data compression algorithms. Some of the models are described below.

1.2.4.1 Physical Models

Information is modeled by the physics of the data generation process. For example, in image related applications, knowledge of the physiology of visual perception can be used to

construct a mathematical model. However, usually the physics of data generation is too complicated to develop models in most physiological cases.

1.2.4.2 Probability Models

Given a source, the simplest statistical model is to assume that each generated symbol is independent from the other and is called an '*independent model*'.

For a source S that generates a set of symbols $S_1, S_2, S_3, \dots, S_N$, when the independent probability for each symbol assumption is removed and an equal probability is assigned to each symbols in S a '*probability model*' is obtained. The model is given by

$$P = \{P(S_1), P(S_2), P(S_3), \dots, P(S_N)\} \quad (1.6)$$

$P(S_i)$ is the probability associated with the symbol S_i in the signal. The entropy for the source can be calculated from the model. For example in an 8-bit image, the image can be modeled as the probability of occurrence of each of the 256 gray levels.

1.2.4.3 Markov Models

The Markov model was named after the mathematician A.A. Markov. This model uses different ways of representing dependencies present in the data. Let $\{X_n\}$ be a sequence of observations. This sequence is said to follows a 1st order Markov's model if

$$P(X_n | X_{n-1}) = P(X_n | X_{n-1}, X_{n-2}, \dots) \quad (1.7)$$

$X_n | X_{n-1}$ indicates the conditional probability of occurrence between the symbols. Conditional probability is the probability obtained for a symbol based on probability of occurrence of a previous symbol or a set of previous symbols. In other words, knowledge of the past k symbols is equivalent to knowledge of the next symbol. For example in a binary image, 2

types of pixels are present, the white pixels and the black pixels. If the previous k symbols in this image were known, then the next pixel, whether black or white, in the sequence can be predicted using a Markov model. The prediction is done by defining two states S_w and S_b , where S_w would correspond to the pixel being white and S_b would correspond to the case when the pixel is black. The corresponding probability for these symbols would be $P(S_w)$ and $P(S_b)$.

Using the equation given in 1.7 for each symbol, the entropy of the white pixels in the binary image is given by

$$H(S_w) = -P(b|w) \log P(b|w) - P(w|w) \log P(w|w) \quad (1.8)$$

Where $P(w|w) = 1 - P(b|w)$ and $P(b|w)$ is the conditional probability of black if the probability of white is already known.

1.2.5 Compression Schematic

Figure 1-4 shows block schematic for a generic compression scheme. All compression algorithms have the basic components shown in figure 1-4. The source coder reduces the input rate to a level supported by the transmission and storage media.

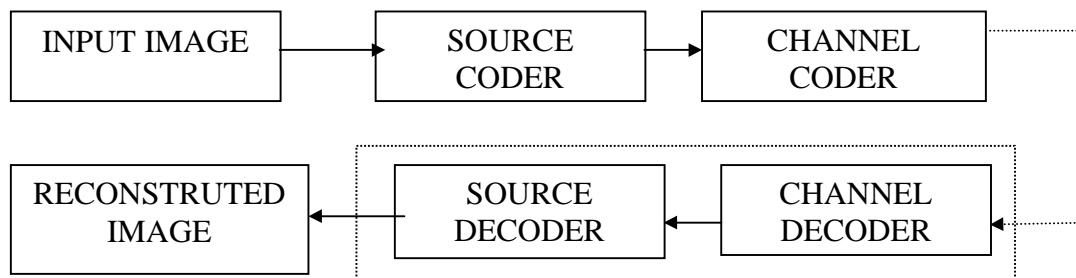


Figure 1-4. A generic compression scheme.

Compression ratio is the term used to characterize the capability of the compression system and is given by

$$C_r = \frac{\text{Source Coder Input Size}}{\text{Source Coder Output Size}} \quad (1.9)$$

The channel coder translates the compressed bit stream to a signal suitable for transmission or storage. In order to reconstruct the input coded data; the decoder reverses the process of channel coding and source coding and outputs the reconstructed image.

1.2.6 Lossless Compression

Lossless compression techniques allow the exact reconstruction of the original data from the compressed data. Hence lossless compression is a reversible process. Given a set of input symbols, the model generates an estimate of the probability distribution of the symbols and this distribution is used to map the symbols into code words.

Effective coding is done by using shorter code words for symbols occurring with a high probability and longer code words for symbols occurring with low probability. Figure 1-5 is a schematic of a lossless compression scheme.

1.2.6.1 Pre-processing

Entropy coding is done on a symbol to symbol basis. However an image of size 256x256 pixels when viewed as a single image is 65,536 units long and to entropy code this segment would be difficult. By considering the gray scale values of the pixels ranging between 0 and 255, the image sequence now reduces to finding a good probability model for 256 symbols. Pre-

processing techniques also include differential coding which exploits the inter-pixel correlation in the image and skews the symbol statistics to better model the data symbols to improve compression ratios.

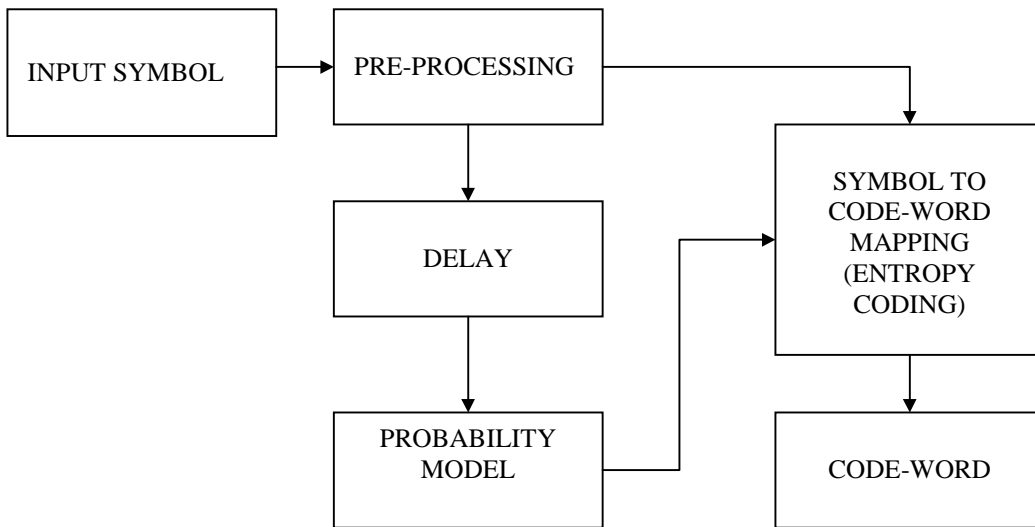


Figure 1-5. Block schematic for lossless compression schemes.

1.2.6.2 Huffman Coding Algorithm

The Huffman's technique of entropy coding was developed as a code construction method for lossless compression by David A. Huffman [19]. Huffman's procedure is based on two observations regarding optimum prefix codes [20]

- a) In an optimum code, symbols occurring more frequently are assigned shorter code words than symbols occurring less frequently.
- b) In an optimum code, the two symbols that occur least frequently will have the same code length.

The Huffman's algorithm consists of the following steps and is explained with an example below

- a) Ordering the symbols according to their probabilities. The frequency of occurrence must be known a priori which is done by estimating a training data set that is representative of the actual data set. If the data set is composed of N distinct symbols S_1, S_2, \dots, S_N and the probabilities of occurrence is P_1, P_2, \dots, P_N then the symbols are rearranged such that $P_1 > P_2 > \dots > P_N$.
- b) Applying a contraction process to the two symbols with the smallest probabilities. Let the two symbols with the smallest probabilities be S_{N-1} and S_N , then these symbols are replaced by one hypothetical symbol H_{N-1} that has the probability of occurrence between P_{N-1} and P_N . The new set of symbols has $N-1$ symbols.
- c) Repeating the previous step until the final set has only one member.

The recursive second step of the process can be considered as the construction of a binary tree, since in each step, two symbols are merged as one.

The steps explained in the Huffman's coding process are best explained by an example.

Consider an image with 8 distinct gray levels (C, F, G, B, E, D, H and A) with frequency of occurrence of (0.16, 0.10, 0.03, 0.18, 0.14, 0.16, 0.02 and 0.21) respectively. The histogram of the image is obtained and rearranged according to the frequency or probability of occurrence of the gray levels. Now the rearranged gray levels are (A, B, C, B, E, F, G, and H) with the frequency of occurrences (0.02, 0.03, 0.10, 0.14, 0.16, 0.16, 0.18 and 0.21) respectively.

Figure 1-6 shows a Huffman tree with two nodes at each level. The two gray levels with the lowest possibility of occurrence, G and H form the first level.

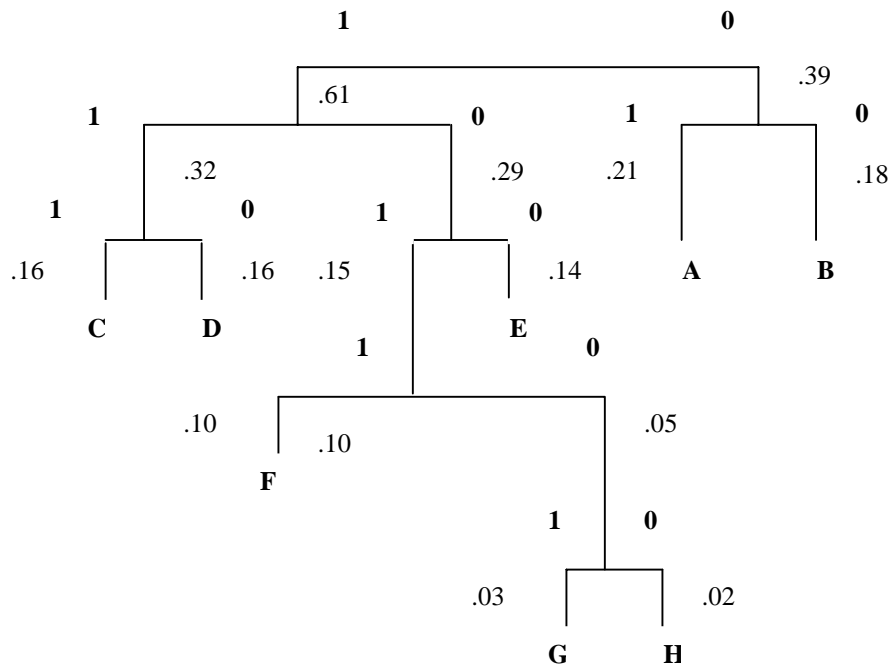


Figure 1-6. Huffman's tree for an eight gray level image.

The probabilities are added. The next gray level (F) is taken and the second level of the tree is formed. The gray level with the lower probability is always put on the right. This process is continued till the entire tree is built. Finally 1's are designated to the left nodes at each level of the tree, and 0's are designated to the right nodes of the tree. The code for each gray level is generated from this tree starting from the top. For example, the code for letter G would be 10101.

1.2.7 Lossy Image Compression

In lossy image compression, the image reconstruction process yields an imperfect reconstruction of the original image during decompression. Lossy image compression techniques

can be applied to both still images and real-time images. Lossy data compression however involves a trade off between the image quality and the bit rate.

Image data tends to have a high degree of spatial redundancy. By exploiting these redundancies, a high degree of compression can be achieved without image degradation being detected by the visual system.

Using lossy compression techniques, images can be compressed as blocks or in its entirety. In block compression methods, the image is subdivided into equal sized blocks and transformations are computed separately for each block. Several of these small blocks can be compressed in parallel and are more efficient than compressing the entire image. Furthermore, reconstruction of these from the transform domain to images in the spatial domain is much easier when the compression is done using the block method.

1.2.7.1 Discrete Cosine Transform (DCT) Based Coding

The DCT based coding is the basis of newer lossy image compression standards for real-time images. In DCT, compression is based on the transformation of a NxN image block from the spatial domain to the frequency domain. The forward DCT [21] is given by

$$y_{kl} = \frac{c(k) \times c(l)}{4} \sum_{i=0}^7 \sum_{j=0}^7 x_{ij} \cos\left(\frac{(2i+1)k\pi}{16}\right) \cos\left(\frac{(2j+1)l\pi}{16}\right) \quad (1.10)$$

Where $k, l = 0, 1, \dots, 7$. The DCT is an orthogonal transform. If the DCT output in the matrix form is given by $\mathbf{Y} = \mathbf{T}\mathbf{X}\mathbf{T}^t$, where \mathbf{X} and \mathbf{Y} are input and output matrices respectively, \mathbf{T} is the transformation matrix and \mathbf{T}^t is the transpose matrix of \mathbf{T} . The inverse transform is given by $\mathbf{X} = \mathbf{T}^t\mathbf{Y}\mathbf{T}$

Equation 1.10 decomposes each input block into a series of waveforms each with a particular spatial frequency. Thus performing a DCT on the signal is the process of finding a corresponding weight y_{kl} for each waveform of frequency components.

The decoder requires the Inverse Discrete Cosine Transform (IDCT) for the reconstruction of the signal. IDCT is given by

$$x_{ij} = \frac{c(k) \times c(l)}{4} \sum_{i=0}^7 \sum_{j=0}^7 y_{kl} \cos\left(\frac{(2i+1)k\pi}{16}\right) \cos\left(\frac{(2j+1)l\pi}{16}\right) \quad (1.11)$$

The advantage of using a DCT based algorithm is that its implementation is image independent thereby requiring no additional processing for different images.

Figure 1-7 shows a block schematic of a generic DCT based compression scheme. The DCT process transforms each of the 8x8 blocks into a set of DCT coefficients Y using equation 1.10. To each of these transformed components, a weight is associated corresponding to the DCT basis waveform. Data compression occurs during the quantization process (a process of representing the signal in the digital form where specific gray levels are assigned to the signal depending on the signal intensity), where some of the weights are discarded depending on the quantization model. Finally the output of the decoder is passed through a run-length coder such as Huffman's coding to give the compressed data stream.

1.2.7.2 Motion Compensated Prediction

In most video sequences, the change between the contents of the image from one frame to the next is negligible. Taking advantage of this temporal redundancy and using the previous frame to predict the current frame produces effective data compression.

The encoding process is done by a technique called *block based motion compensation*. Here the frame being encoded is divided into blocks of size $M \times M$. The previous frame is searched for a $M \times M$ block that closely matches the block being coded. The distance between the two blocks is measured by finding the absolute differences between the corresponding pixels in each block; a motion vector is determined and transmitted to the receiver. Various implementations of algorithms for motion estimation are available and these algorithms vary in the running time required and in computational complexity.

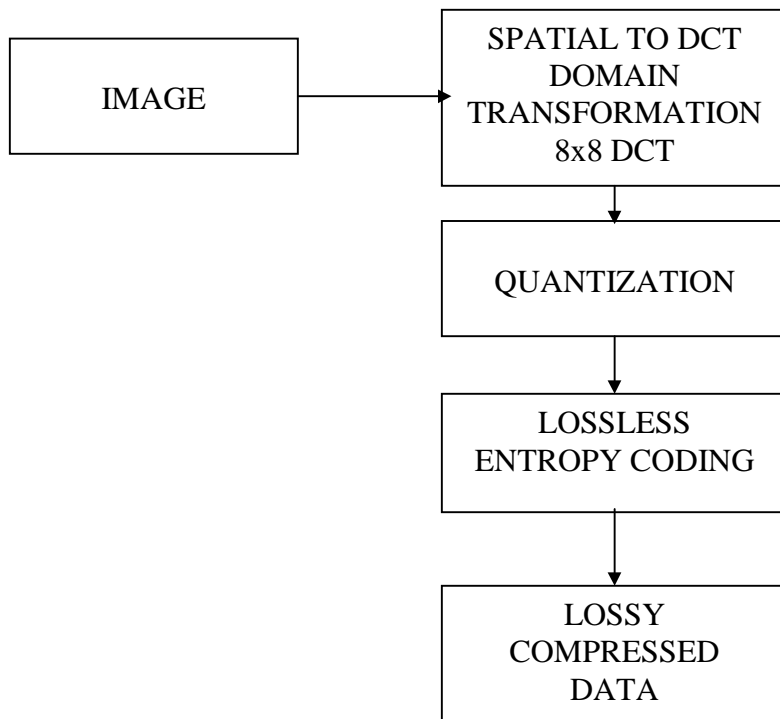


Figure 1-7. A schematic of DCT based compression schemes.

1.2.8 H.261 Video Coding Standard

The H.261 is a video-coding standard published by the International Telecom Union (ITU) in 1993 [22]. This standard is designed for data rates which are multiples of 64 kbps and is also referred to as $P \times 64$ kbps where P varies from 1 to 30. The video compression-decompression unit (CODEC) acquires composite video and audio, digitizes it, and outputs a compressed data stream.

1.2.8.1 Motivation for Using H.261

As discussed earlier, real-time teleconferencing within city limits is feasible by the use of an ATM network and does not require image compression. However as discussed earlier, transmission over long distances necessitates data compression and involves a trade off between the picture quality and cost. Therefore a CODEC capable of generating data rates suitable for T1 or ADSL networks becomes necessary. Since H.261 is a compression scheme designed for lower bit rates, it is ideal for producing rates suitable for T1 or ADSL. The other motivating factor is the availability of the implemented algorithm. A variety of already implemented H.261 algorithms are readily available and can be used for teleconferencing without specialized processing requirements such as Reduced Instruction Set Computers (RISC) and Digital Signal Processing (DSP).

1.2.8.2 Common Intermediate Format (CIF)

The H.261 standard specifies that the input to the CODEC should be of the Common Intermediate Format. This is because of the two different lines and frame rates available for video signals. The National Television System Committee (NTSC) format uses 525 lines per

interlaced picture at 30 frames a second, while other standards such as Phase Alteration by Line (PAL) uses 625 lines at 25 frames per second. A common format called CIF was recommended [22] and both formats were pre-processed before being supplied as inputs to the CODEC.

CIF is in non-interlaced format and is based on 352 pixels/line, 288 non-interlaced lines/frame at 30 frames per second. These numbers are a combination of the other two previous formats and represent half the active lines of the 625/25 signal and the picture rate of the 525/30 NTSC signal. Therefore in order to convert to the CIF format, NTSC performs a line number conversion and other formats perform frame number conversion. The Quarter Common Intermediate Format (QCIF) which is $1/4^{\text{th}}$ the number of pixels /line and $1/4^{\text{th}}$ the lines /frame of the CIF may also be used as inputs to the H.261 encoder.

1.2.8.3 H.261 Encoder

The H.261 coding standard uses a combination of DCT coding and differential coding [21]. The figure 1-8 shows the relationship of the main components in the standard such as frame prediction, DCT, quantization and the Variable Length Coding (VLC) such as Huffman's coding.

The input video frame is translated into a CIF frame and stored in the frame memory. The required signal processing such as noise filtering is done at this stage. In DCT, the frame is converted to 8x8 blocks. Four luminance blocks and 2 color difference blocks are combined to form a macro block. The macro block is either of the inter-frame type or and intra-frame type, which is the differential of the present frame and the previous frame. The DCT coefficients are quantized and coded using a variable length coder such as Huffman's coder and stored in the output buffer. The rate is dynamically controlled by altering the capacity of the buffer.

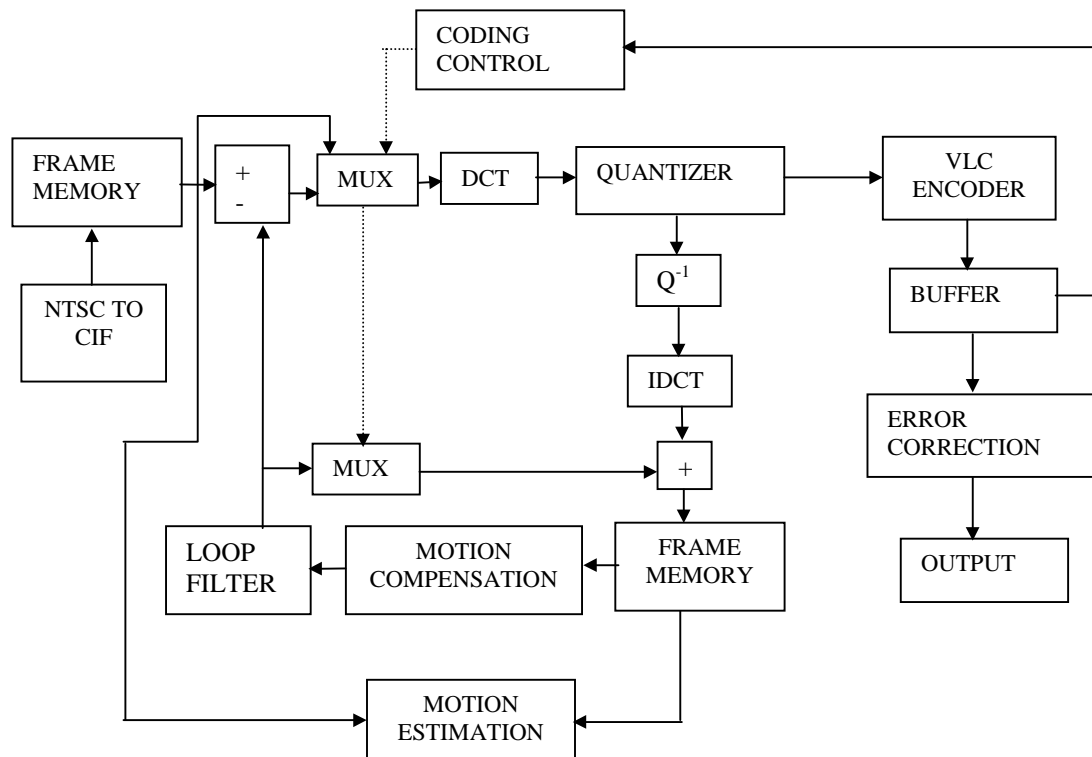


Figure 1-8. Block schematic of the H.261 encoder.

Predictive coding is used and the encoder monitors the quality of the transmitted image by including a feedback loop comprising of inverse quantization and IDCT. The compressed stream is arranged in layers. A picture is the top layer. Each picture is divided into Groups of Blocks (GOB) each of which is 1/12th of the CIF picture frame. Each GOB is divided into 33 macro-blocks and each block consists of six 8x8 blocks.

1.2.8.4 H.261 Decoder

Figure 1-9 is a block schematic of the H.261 decoder. The input data to the decoder is error corrected and buffered. A VLC decoder such as a huffman's decoder is used to reconstruct

the coded data back to the original data. The decoded data is parsed and inverse quantization and IDCT processes of the decoder do the reconstruction of the image.

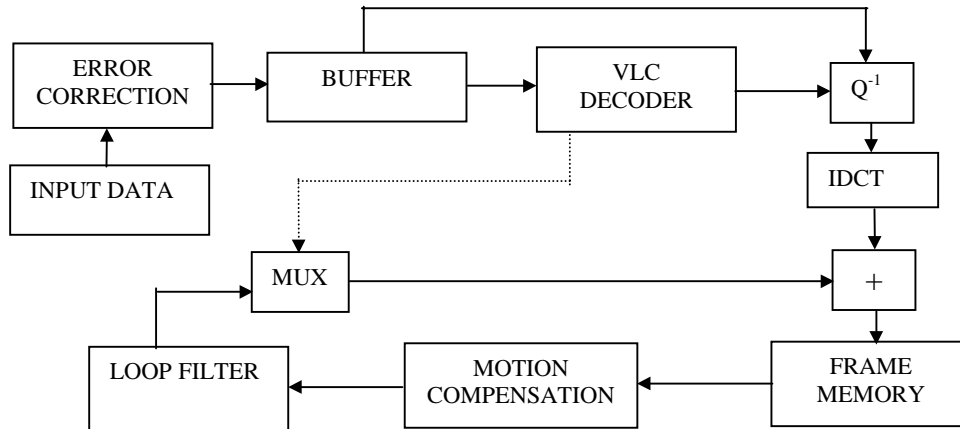


Figure 1-9. Block schematic of the H.261 decoder.

1.2.9 Applications of Data Compression in Medical Imaging

Radiological image files are often large and compressing them can save image storage space and transmission time. The following section describes some of the critical areas of radiology where data compression techniques are essential

1.2.9.1 Teleradiology

Teleradiology is the electronic transfer of radiological images and patient data from one site to another to be read by a radiologist at the receiving site located at a significant distance from the patient. The push for implementation of teleradiology systems by companies to provide efficient services at reduced costs is the result of strong socio-economic factors. Teleradiology units were initially designed to help radiologists manage on-call obligations from home [7] [8];

however they are also used to overcome a variety of other problems caused by the distance between the physician and the patient.

Teleradiology is becoming an extremely important component in radiological practice because of the following reasons. First, teleradiology secures all the images for the radiologist to read so that none of the images will be accidentally lost during transit. Second, teleradiology reduces the cycle time from acquiring the image to the completion of a report. Third, teleradiology facilitates seeking second opinions. For example, a study indicates that teleradiology consultations led to changes in the emergency department physician's initial diagnosis in 27 of 90 cases [9]. Fourth, the reading time of the images is estimated to decrease by 50% [10]. Finally, teleradiology can serve multiple sites by reducing the number of radiologists required to be present at any given site. Teleradiology systems incorporate data formatting for the images using standards such as Digital Imaging and Communications in Medicine (DICOM). The DICOM formatted images are suitable for transmission and for data authoring where the images are combined with patient demographics before transmitting over the network [14].

A teleradiology system without images is easy to implement, A computer gathers the patient information, examination results and diagnostic reports at the referring site and transmits them over a computer network to a remote site to be read. However in reality images have to be transmitted along with the patient demographic information. Digitized medical images consume a lot of bandwidth and the data transfer rates exceed the capacity of hospital computer networks. Bandwidth is a communication term, which is defined as the number of bits received or transmitted over a channel per unit time.

A teleradiology system uses a digitizer to convert a film image to digital format and the digital images are transmitted over a wide area network (WAN). Typically an examination

generates 10 to 20 Megabytes of data and to transmit this information over the WAN in a reasonable amount of time requires very high bandwidths. This high bandwidth requirement necessitates data compression in teleradiology systems. Compression techniques previously mentioned are used in reducing the generated data rates from acquisition equipment to the available bandwidth of the transmitting network. Depending on the available bandwidth, lossless or lossy compression techniques will be used. Teleradiology systems used with lossless compression have proved to be very reliable. A previous study using lossless compression transmitted over a WAN yielded 98% rater agreement on the diagnosis between teleradiology images and the original film images before digitization [11]. Teleradiology involves a trade off between image quality, transmission and costs. By increasing the compression ratios to levels suitable for transmission, the quality of the transmitted images may be compromised by the amount of compression. But by decreasing the compression levels where compressed images can be reconstructed lossless, the time taken for transmitting these images may be enormous. Therefore acceptable levels of transmission time along with levels of compression where the diagnostic integrity of the images is not compromised are required in good teleradiology systems [12].

1.2.9.2 Telesonography

Ultrasound imaging is different from other imaging modalities except angiography by the fact that it requires active participation by the sonologist in the performance of the study. As discussed earlier, real-time ultrasound imaging is a much more effective method of evaluation than still ultrasound imaging. However, real-time telesonography presents greater technical problems than still-imaging telesonography because of the bandwidth required for transmission

of video images. For the implementation of a reliable tele-ultrasonography system, greater bandwidth and compression schemes are required.

In a typical setting the radiologist may be at a central reading station connected to a number of satellite ultrasonography stations. The satellite stations may not have a radiologist present at the time of scanning, however all cases that are done, are electronically transferred to the central reading stations where the sonologist reads and interprets each case. Because an active participation by the sonologist is required in the performance of the study, an interactive tele-ultrasonography system is mandatory. A good telediagnostic ultrasound service would contain the following characteristic [6]

- Involvement of the physician in the examination while the patient is still on the table by viewing images real-time.
- Ease of use of components in the tele-ultrasonography system to overcome physical and psychological barriers to interaction.
- Capability of personal interaction between the patient and physician.
- Direct access to previously conducted examinations
- Rapid communication of diagnostic results to the referring physician.

1.2.9.3 Ultrasound Picture Archiving and Communication Systems

A Picture Archiving and Communication System (PACS) is used in radiology departments to archive images and other patient data in a digital form to enable easy and random access to old records and images of patients. When film images are stored in hospitals, retrieval of these images from over 6 months takes over 0.5 hours to 2 hours [14]. But when the images are stored in digital form in a PACS, retrieving these images can be done faster and in a more

efficient manner. Because of the volume of images stored typically in a PACS, image compression is necessary to speed up the image transmission rates and save on storage requirements. Typically ultrasound imaging generates 221 Mbps for 24-bit color and 74 Mbps for 8-bit gray scale video. At these rates, the PACS will have insufficient storage space in a few weeks time even though they might have started with tera-bytes of free space. These storage constraints in PACS necessitate reliable real-time compression schemes for long term archive of image data.

The ultrasound scanners output NTSC images; the gateway processes the data, indexes the images and adds image headers to the data making them suitable for DICOM conversion. In the DICOM gateway several processes are kept concurrently running. The first daemon (a process in the gateway constantly running) checks for new scans, picks up scans and sends the scans to a second process which converts them to DICOM format. A third process usually compresses the file before sending it across to the PACS acquisition gateway. When the amount of data involved in archiving is very large, the archival system can saturate very fast. For example the echocardiography data stream is close to 221Mbps; therefore effective data compression that does not compromise the image quality is necessary. Figure 1-10 shows a block schematic of an ultrasound PACS.

In the acquisition gateway, several daemons are kept concurrently running. The first daemon checks for DICOM *send* commands, which indicates the arrival of new images from the US gateway. The second daemon checks for the proper DICOM formatting in the images that have arrived at the gateway. The third daemon queues the processed file for archiving. Bottlenecks affecting the operation of a PACS are network contentions where a lot of images are queued through the server, slow responses from workstations and long delays for image retrieval

from the long-term archive. The network bottlenecks can be avoided by improving the system architecture, reconfiguring the networks by segmenting the networks to sub-networks and by using effective compression schemes to streamline the operational procedures. The bottlenecks caused by network contentions can be avoided by using a distributed network archive server that optimizes the access time and reduces the network contention [15,16]. The slow response at the workstation caused by bad database design and insufficient memory [17] can be avoided by upgrading the memory of the workstations and by better database design.

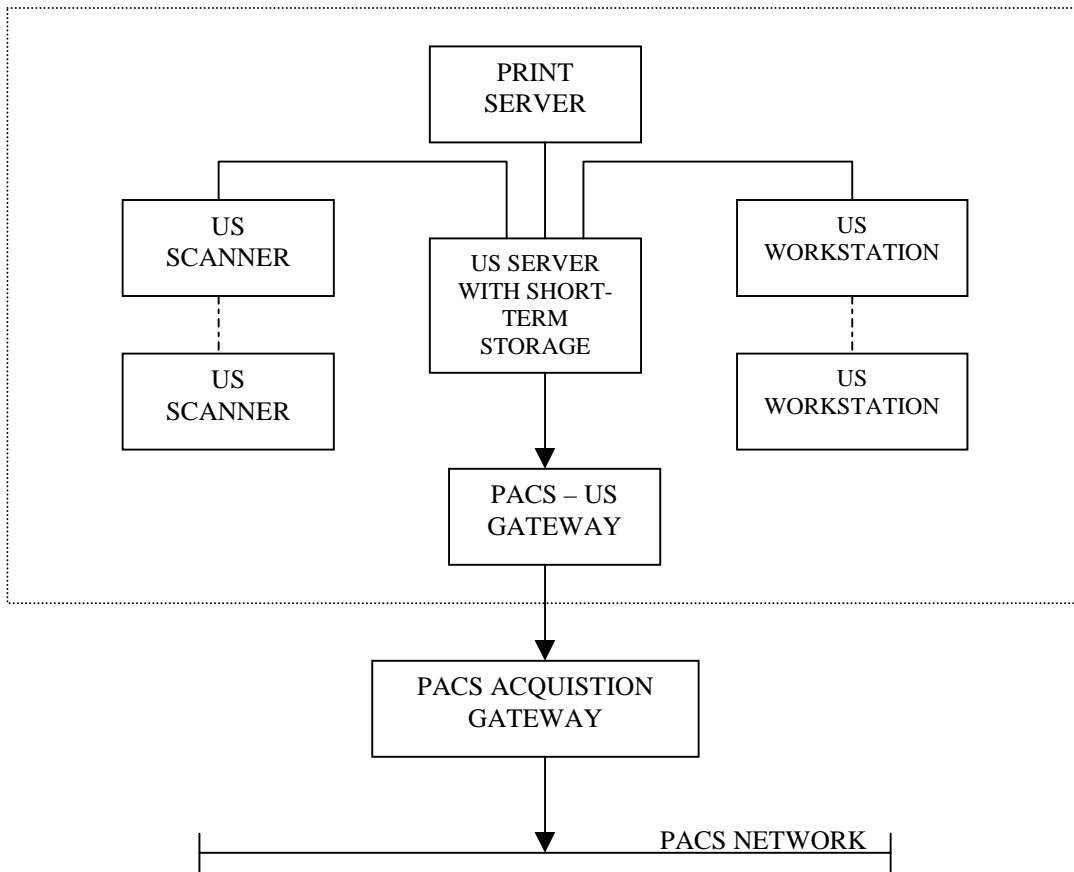


Figure 1-10. Block diagram of a generic ultrasound PACS module.

1.2.10 Legal Issues and Standards for Medical Image Compression

Lossless image compression does not raise any legal complications as the original image is exactly reproduced with modest compression ratios of up to 3:1. Most of the research is therefore focussed on lossy compression techniques for various imaging modalities that discard image data of low diagnostic significance but retain anatomical and physiological information. In previous studies involving the lossy compression of chest radiographs, acceptable levels with no significant reduction in detection ability was 11:1 and for statistically significant degradation of the images, the compression ratio was 125:1 [23]. Studies for magnetic resonance images using multi-spectral image compression showed that compression ratios of 25:1 preserved diagnostic integrity [24]. Results of qualitative testing of data compression using Joint Photographers Expert Group (JPEG) format indicated that lossy compression schemes should have compression ratios less than 5:1 for Quantitative Coronary Arteriography (QCA) [24] and 12:1 for dental images [25] to preserve diagnostic integrity.

Two major standards; ACR-NEMA (American College of Radiology-National Electrical manufacturers Association) and DICOM govern medical image compression. To be clinically acceptable, a compression algorithm requires thorough clinical validation tests, carried out on a large number of images representing all stages of the pathology and involving many clinicians to ensure that the diagnostic accuracy is not compromised. Currently when a misdiagnosis occurs, radiologists testifying for the plaintiff and for the defense will argue the image quality before the jury and let the jury decide whether the image used was of reasonable quality for diagnosis. Therefore manufacturers and researchers are responsible for a compression algorithm which will not sacrifice diagnostic integrity.

1.3 Research Objectives

1.3.1 Introduction

Video compression is required to fit the video stream into the available transmission bandwidth and such compression always involves a tradeoff between the bandwidth and image quality. In lossless compression no visible degradation of the image is seen because of the exact numeric reconstruction of the original image. At compression ratios greater than 3:1, the compression is lossy and there will be visibly significant perceptual losses in the images with increase in compression. However, the use of lossless compression techniques for the purpose of transmission of real-time ultrasound imagery is unrealistic because the compression required for transmission of a digitized NTSC video stream over a T1 line or an DSL is much higher and is of the order of about 100:1.

While evaluating quality in ultrasound imaging, it does not suffice to say that the compressed image “*looks good*” and has a high peak signal to noise ratio. Rather it should be convincingly proved that there is no significant degradation in the image to hinder good clinical diagnosis.

Two important aspects in ultrasound diagnosis are findings and the levels of confidence in the findings. It is not only important to be able to make a diagnosis, but also to be absolutely confident about the diagnosis made. A low level of confidence in diagnosis will inevitably require a second opinion from a different radiologist or the radiologist may have to use a different modality to verify the initial findings. This duplicates efforts and the advantages of data compression are nullified.

1.3.2 Objectives

1. The goal of this project is to qualitatively evaluate compressed real-time ultrasound video imagery to determine the minimum bandwidth necessary for clinical diagnosis of multiple diseases and pathological states. In other words, the goal is to estimate, using objective and subjective techniques, the extent to which the diagnostic information in different anatomical regions and pathological states are affected by data compression.
2. Lossy compression techniques usually entail a tradeoff between spatial resolution and temporal resolution. When a tradeoff is inevitable, it becomes necessary to categorize studies to see in which study temporal resolution should be given more priority over spatial resolution and vice-versa without losing diagnostic integrity.
3. It is our hypothesis that there will be a gradual degradation of observer confidence with increase in compression. It is therefore our aim to subjectively quantify observer confidence using statistical methods and to test for a trend in loss of observer confidence ratings with an increase in compression.
4. In this project objective methods of evaluation were used to assess trends in image degradation and express image degradation quantitatively.
5. This project aims at classifying each study by the observed effects due to compression. Common effects of compression on ultrasound imagery include temporal blurring, spatial blurring and an increase in contrast of the images obtained. These effects are discussed below.

1.3.2.1 Temporal Blurring

Temporal blurring is seen due to the loss of certain frames during a motion sequence of continuous frames. Depending on the compression algorithm used, compression schemes may

cause either a random loss or a loss of a predetermined frame in a motion sequence. Temporal blurring makes a normally continuous video sequence appear jerky and irregular. It also causes a loss in the ability to analyze structures in motion and forces the radiologist to look only at stationary structures. Temporal blurring can be a very serious issue in ultrasound studies that have intrinsic motion and can therefore cause a loss of an important component of what ultrasound imaging can do.

Fetal studies have intrinsic motion and are one of the critical areas where temporal blurring can cause diagnostic problems. In early pregnancy, the anatomical structures of the fetal heart such as valves and walls are not seen. Only the motion of the entire fetal heart is seen. This observation in a fetal study gives vital information about the state of the fetus. Temporal blurring may cause a misdiagnosis and lead to improperly declaring the fetus dead. In late pregnancy, the sonologist should be able to see the internal structures of the heart such as the valves and their movement. Data compression may cause temporal blurring and seriously cripple the visualization of the motion of the internal structures of the heart.

Temporal blurring affects other studies such as liver and kidney scans and hinders the proper visualization of the tissue in motion. In difficult patients, such as obese patients, the scanning usually causes a haze in the obtained image. By moving the transducer, the noise can be averaged out to form a good image. Due to temporal blurring, liver and kidney structures in difficult patients are not properly rendered.

1.3.2.2 Spatial Blurring

Spatial blurring is caused by the removal of pixels within the frame. In ultrasound imaging, spatial blurring causes the picture to become blocky, giving an improper definition of

structures and borders of the organ imaged. In large structures the effects of spatial blurring are not easily noticed. However, spatial blurring causes bigger problems in the detection and diagnosis of small tumors, masses and lesion structures. Technological development in ultrasound imaging for the past 15 years has focussed on improving spatial resolution. The effects on imagery due to spatial blurring give an impression of going back to older technology.

The areas in diagnosis where spatial blurring poses potential problems are detection of small liver lesions, finding stones in the common bile duct, tracing borders of fibroids, tracing fetal structures such as fetal kidneys and the visualization of the skin line behind the neck of the fetus. Other potential areas of increased diagnostic difficulty include defining the characteristics of complex ovarian cysts and cysts filled with blood.

1.3.2.3 Increase in Contrast

Image compression in ultrasound imaging causes an increase in inherent contrast, due to which low-level signals are lost thereby making it difficult to differentiate finer shades of gray. In diagnosis, various shades of gray convey significant information about the texture and parenchyma of solid tissue. By losing certain low-level signals, the ability to diagnose any pathology that may be present in the parenchyma is lost. For example an increase in contrast causes failure to detect debris present in complex cysts and distended tubes.

Chapter 2

Methods and Materials

2.1 Introduction

An approach to establish quality and utility in a specific application is to simulate the application in a carefully designed experiment, gather the necessary data in a way that least interferes with the simulation and analyze the resulting data to assess a specific hypothesis such as “image A is not different from image B” [26]. In our research, objective and subjective methods were used for the qualitative evaluation of data compression in real-time ultrasound imagery. Image quality was subjectively quantified by statistical analysis of radiologists’ scores using inter-rater reliability studies and by Analysis of Variance (ANOVA). Image quality was objectively quantified using histogram analysis and estimation of Peak Signal to Noise Ratios (PSNR).

2.2 Protocols and Guidelines

The protocol should simulate ordinary clinical practice as closely as possible. In particular, the evaluating sonologists should perform in a setting that mimics their daily clinical practice. This criterion includes the room setting, controlled lighting with minimum glare, and other aesthetic constraints.

- Sonologists who do the testing should be board certified and have sufficient expertise in the particular area of evaluation.
- The studies for evaluation should include examples of images containing a wide variety of possible findings, and should exclude very rare conditions.

- The detection task should be clearly defined for the evaluating sonologist.
- Gold standards for the evaluation of equivalence or superiority of images should be defined clearly and should be consistent with the established standards and experimental goals. The gold standards set were
 - Personal: the uncompressed clip is used as the gold standard and is compared with the ratings at other levels of compression.
 - Independent: the agreement of an independent expert panel.
 - Separate: the results of other modalities like CT, MRI, biopsy etc.
- The experimental setup should be carefully designed to eliminate or minimize any sources of bias in the data. For example, studies should contain both easy and difficult cases. The data set used for testing should contain sonograms that are easy and difficult to read.
- The experimental setup should also eliminate or minimize sources of bias in observers such as fatigue and learning effects. Careful design of the experimental protocols should be able to eliminate these sources of observer bias.
- Experimental conditions such as image display, room environment, viewing time and the number of times the observer is allowed to view the study should be specified and controlled.
- The number of subjects for the study should be sufficient to ensure adequate statistical power and size.

2.3 Experimental Overview

Real-time ultrasound images were acquired using an Acuson XP128 computed imaging system (Acuson Inc, Mountain View, CA). Different data/bit rates were produced from the source ultrasound signal by a British Telecom Visual Images (BVTI) Compressor-Decompressor (CODEC) CD-400, which is a compression decompression unit. The CODEC

uses the H.261 compression algorithm recommended by the International Telecommunication Union (ITU-T). The various bit rates produced by the CODEC were 1532, 1152, 768, 512, 384 kbps and uncompressed video.

Uncompressed and compressed real-time ultrasound image segments of the same organ and pathology were recorded onto betacam tapes. Images were recorded from 45 patients and a total of 90 different studies were taped on 14 different betacam tapes. The studies included both normal physiological states and abnormal studies with pathologies. The data set contained images that were representative of the studies done on any given day at an ultrasound department.

Significant segments of anatomical and pathological regions of interest were edited and analyzed using subjective and objective techniques. Two possible methods of editing were considered. The first method was editing significant video segments from the 14 betacam tapes onto a single betacam tape. The second method considered was using digital methods. Both the methods involved degradation of the image quality to a certain extent. Analyzing the advantages and disadvantages of both the methods, digital editing was done to capture significant video segments for analysis. Several experiments were conducted on the digital capturing method to ensure that the captured images were of a quality comparable to that of the original tapes. Segments of anatomical and pathological regions of interest were captured using a MIRO DC-30 plus video capture card (Miro Computer Products Inc.) and edited using Adobe Premiere 5.1.

The experimental system shown in figure 2-1 was set up at the Regional Medical Center, Department of Ultrasound, University of Tennessee, Memphis.

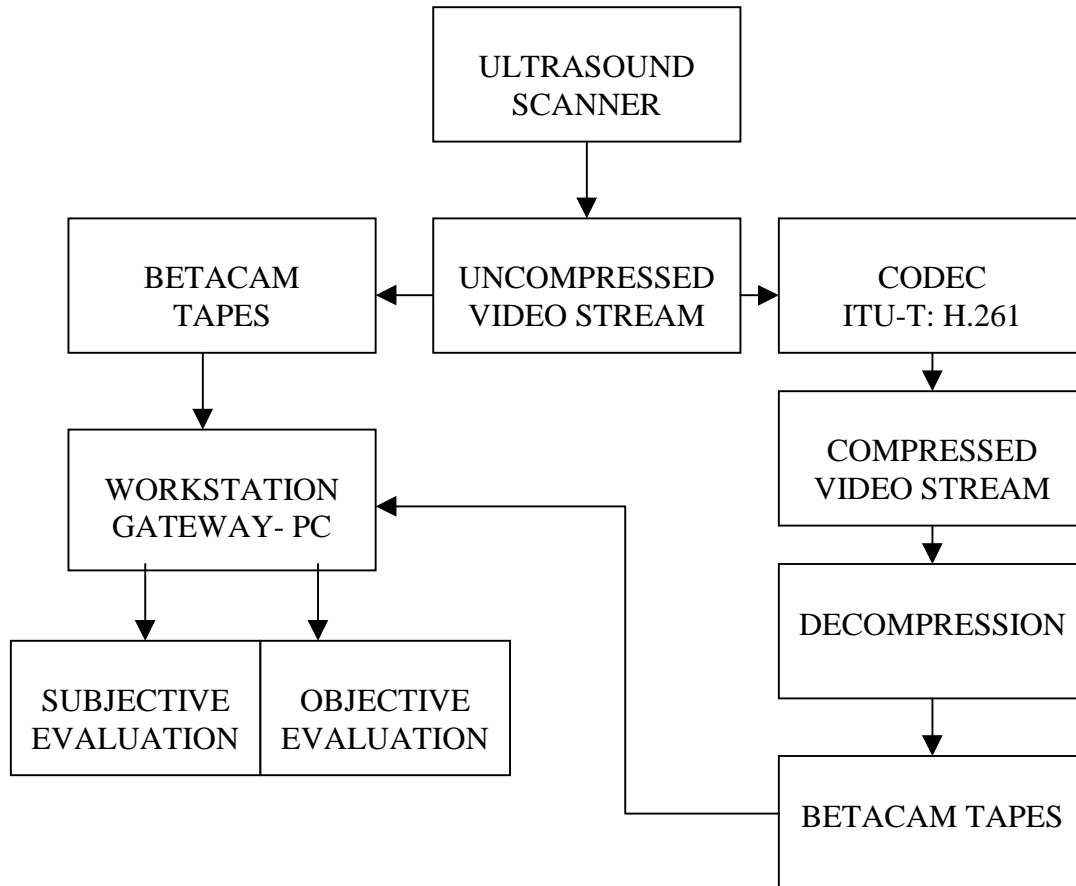


Figure 2-1. The experimental setup for the subjective and objective testing at the Regional Medical Center, Memphis.

2.4 Materials

This section describes the technical specifications of the various components of the system shown in figure 2-1.

2.4.1 Ultrasound Scanner

The ultrasound scanner used for the experiment was an XP 128 computed sonography system from Acuson Inc., Mountain view, CA. This is a 128 channel imaging system with good spatial and contrast resolution. Various transducers from Acuson were used in the experimental setting.

The transducer model numbers from Acuson Inc., are given below. The C366 and C544 transducers (C denotes curved array) are typically used for abdominal and fetal studies fetal studies. The EV519 (EV denotes endovaginal transducers) is used for intra-uterine pregnancy studies.

- C366 High performance curved array with aperture size 66mm
Frequency settings are 3.5MHz and 2.5 MHz
- C544 High performance curved Array with aperture size 44mm
Frequency setting 3.5MHz
- EV519 Vector endovaginal with aperture size 19mm
Frequency setting 3.5 MHz.
- V325 Vector transducer format with frequency setting 3.5MHz.

Table 2-1 shows the transducer settings for each study

Table 2-1. Ultrasound transducer settings for various studies in the experiment.

Bandwidth (kbps)	Study	Transducer Type	Frequency (MHz)	Dynamic Range(dB)	Processing Settings	Gain (dB)
384	G.Bladder	C3	3.5	50	0/3/0	3
384	Kidneys	C3	2.5	50	0/3/0	3
384	Liver	C3	2.5	50	0/3/0	3
384	Uterus	C3	2.5	50	0/3/0	-3

The table 2-1 shows the various transducer settings such as transducer type, frequency, gain settings and processing settings. The transducer types listed in section 2.4.1 and are chosen by the sonologist depending on the study. Processing settings include pre-processing/edge enhancement settings, persistence and post-processing settings such as Signal to Noise Ratio (SNR). For example when the settings are 0/3/0, the first 0 indicates the pre-processing settings and this values is set to zero indicating no edge enhancement. Sometimes the first 0 may be replaced by ‘S’ depending on the scanner. The S indicates that no edge enhancement settings were made. The second value gives the persistence setting, which is the averaging value for the signal over time. The number explains the extent to which the information is held over time before being dropped. Here information from the first frame is held till the 3rd frame appears. This is important in keeping the images real-time. If persistence is not high enough the image appears jerky. The post-processing settings are the SNR values that affect the gray scale. The settings are on a scale of 0 to 5 and 0 indicates that the image is very contrasty. For most studies this setting is set at zero to get a good contrast among various tissue interfaces.

G.Bladder = Gall Bladder, F.Heart = Fetal Heart, IUP= Intra-Uterine Pregnancy and F.Pelvis = Female Pelvis

Table 2-1 (continued)

Bandwidth (kbps)	Study	Transducer Type	Frequency (MHz)	Dynamic Range(dB)	Processing Settings	Gain (dB)
128	F.Heart	C5	5.0	55	0/3/0	-12
256	F.Pelvis	C3	3.5	53	0/3/0	-12
768	Fetus	C3	2.5	53	0/3/0	-12
768	G.Bladder	C3	2.5	50	0/3/0	0
768	Kidneys	C3	3.5	50	0/3/0	3
384	F.Heart	C5	5.0	40	0/3/0	3
768	Uterus	C3	2.5	50	0/3/0	-12
1536	G.Bladder	C3	2.5	50	0/3/0	-3
1536	Kidneys	C3	2.5	50	0/3/0	3
1152	Liver	C3	3.5	50	0/3/0	3
512	F.Heart	C5	5.0	55	0/3/0	7
768	IUP	EV5	5.0	47	S/3/0	-5
1152	F.Heart	C3	5.0	53	0/3/0	-12
1152	Liver	C3	3.5	47	0/3/0	-5
1536	Uterus	C3	2.5	55	0/3/0	-3
1536	IUP	EV5	5.0	50	S/3/0	-15
1152	F.Pelvis	C5	3.5	47	S/3/0	-5
1152	IUP	EV5	5.0	53	S/3/0	-15

Table 2-1 (continued)

Bandwidth kbps	Study	Transducer type	Frequency (MHz)	Dynamic Range(dB)	Processing settings	Gain (dB)
1536	F.Heart	C5	5.0	55	0/3/0	-12
1536	Liver	C3	3.5	50	0/3/0	7
Uncomp	Liver	C3	3.5	50	0/3/0	-5
Uncomp	Liver	C3	3.5	50	0/3/0	7
Uncomp	Liver	C3	2.5	50	0/3/0	7
Uncomp	Liver	C3	3.5	50	0/3/0	-5
Uncomp	Liver	C3	3.5	55	0/3/0	-5
Uncomp	Uterus	C3	3.5	53	0/3/0	-3
Uncomp	F.Pelvis	C3	3.5	50	0/3/0	-10
Uncomp	Fetus	C5	3.5	53	0/3/0	-12

2.4.3 Betacam

The Betacam BVW-35 (Sony Inc.) video recording system was preferred over other NTSC or PAL format video recorders because of its excellent clarity, contrast resolution and high reliability. The betacam takes NTSC or PAL as inputs and converts them to betacam SP format. Some of the features are

- 4 channels of 16 bit/ 48 kHz digital audio.
- Uses ½ inch S-size and L-size tapes
- 525/60, 625/50 switchable environment. The inputs can either be in the NTSC or PAL formats
- 2 outputs
- 3 analog composite outputs and 1 analog component input. In the composite signal, the signal contains both the luminance and the color information as a single waveform. In the component signal, the video signal is made up of separate waveforms for each color red, green and blue. Each color waveform includes a sync signal to make a component signal.
- RS 232 interface. This is a standard for electrical interface used in serial connections. This interface allows the user to control the hardware features such as recorder clock settings on the betacam recorder through software.

2.4.4 Display System

Sony monitor model PVM-1343MD (Sony Inc.) was used for the display of the images. This monitor was chosen because it is the monitor used at the Regional Medical Center for regular viewing. Some of its features include

- Inputs include NTSC 3.58, NTSC 4.43 and PAL

- High resolution trinitron picture tube
- Composite and S video inputs and outputs. S video is a video signal where the luminance and color components are not combined as a single waveform but are treated as separate signals.
- 16 x 9 aspect ratio.

2.4.5 Digital Video Editing Workstation

For reasons previously discussed, the video editing was done digitally. Digital video editing requires highly specialized dedicated video hardware and software on a high-speed personal computer. This video hardware enables a personal computer to handle tasks previously handled only by studios and production firms.

To test the feasibility of digital editing, initial experiments were conducted on a 333 MHz Dell computer with a standard Integrated Device Electronics (IDE) hard drive speed of 5400 rotations/min (RPM). The experiments were done to check the viability of capturing video onto the hard drive directly. The computer was equipped with a Miro DC -30 plus video capture card, along with the video editing software Adobe premier (version 5.1) required for capturing and editing. The card by itself uses custom Motion Joint Photographers Expert Group (MJPEG) hardware during the capture and does not allow capturing uncompressed video. In order to reduce the further degradation of the images during capturing and editing, the compression levels during capturing had to be kept as low as possible. When the data rate was low, the image degradation was more pronounced. In the initial experiments, we found that the speed of this IDE hard drive was insufficient and some of the frames being captured were dropped. This was because the data rate was too high for the hard disk to write even with high compression ratios of about 6:1. Low compression ratios generate high data rates. At compression ratios of 7:1, we

2.5 Subjective Quality Assessment

In image quality evaluation, subjective techniques are used to satisfactorily prove that the degradation in the imagery being evaluated is diagnostically acceptable. To assess the extent of the effects of data compression on the diagnostic information using subjective methods, medical experts in the field are required. It is important to estimate qualitatively how well the structures are detected by a trained observer in the presence of noise or compression. Standard methods for the evaluation of radiological image quality and resolution include linear ranking, paired comparison using inter-rater reliability measurements and free response Receiver Operating Curves (ROC). All these methods compare the ratings from the evaluating radiologists with the gold standards set for the experiments.

2.5.1 ROC Analysis

In ROC analysis, raters with similar expertise in clinical diagnosis review a set of images. For each image, the observers is asked to give a ROC confidence rating on a scale of 1 to 5 representing their impression of the likelihood of presence or absence of a disease. Figure 2-2 shows a sample ROC plot for 4 compression ratios by the same observer. A confidence value of 1 indicates that the disease is not present and a confidence of 5 indicates that the disease was absolutely present. Confidence values 2 and 4 indicate that the disease or finding is probably present and not present respectively. A confidence value of 3 indicates that the rater could not determine with any degree of certainty if the disease was present or absent. Every observer evaluates the same set of images. The rating of each observer is graded based on a clinical gold standard.

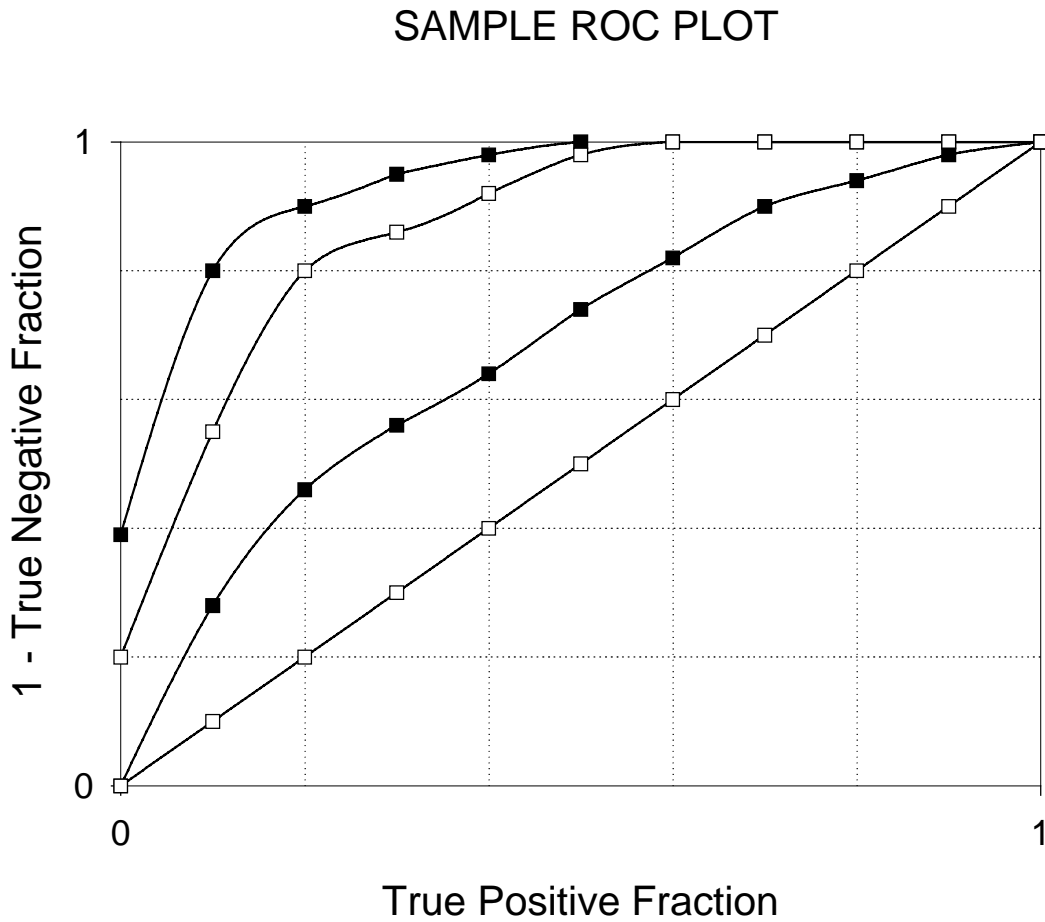


Figure 2-2. Sample ROC for 4 compression ratios by the same observer.

The values are plotted for true positives (when raters could identify the cases with the diseases correctly) and false positives (where raters misidentified, the diseases to be present when they were actually absent)

The graph in figure 2-2 shows the ROC plot for 4 different compression ratios. The plots show the performance of the rater in diagnosing the disease from the compressed images against the gold standard. The area under each curve (A_z) is an index of quantitative measure of the observer's performance on that compression level. Thus while comparing the areas under the

curves A_z for the original image and A_z for the compressed image, if the areas are found to be almost the same, we can conclude that the compressed image is as good as the original uncompressed image. This provides a scale as to how much compression can be accepted for perfect diagnostic capacity. In ROC studies, the data set should possess adequate statistical power. That is the testing should include at least 100 studies of the same anatomical region or pathology.

In an earlier study [26], some of the following reasons were discussed as disadvantages for the use of ROCs in quality evaluation.

- Essentially methods of testing must be established to imitate daily clinical practices in drawing diagnostic conclusions. This is not necessarily true for developing ROCs
- Traditional ROC analysis violates several reasonable guidelines for designing experiments to measure quality and utility in medical imaging because of the statistical assumptions of a Gaussian or Poisson distribution behavior.
- ROC analysis is not suited to study the accuracy of detection and location when a variety of abnormalities are possible. Therefore it does not mimic the real situation in the reading room.
- ROC methodology does not come equipped to distinguish between the various possible “gold standards” such as personal, independent and separate ratings.
- ROCs assume signal independent noise. Digitization of analog ultrasound and lossy compression is not equivalent to an addition of signal independent noise.
- On the statistical side, image data is not well modeled as known signals plus gaussian noise. Hence methods that rely on gaussian assumptions are not valid.

In our experimental setup, images were not specific to one anatomical region and did not contain the same degree of compression as required in ROCs. Therefore we were unable to fashion ROCs with the collected data. In order to design a ROC, the data set should contain the

same anatomical region and incorporate a large number of studies containing positive and negative values to obtain the required sensitivity (true positives) and specificity (True negatives). Also in the qualitative evaluation of data compression in ultrasound imagery, it is imperative to understand that the problem is 3 dimensional and not 2 dimensional as in ROCs. We realized that observer confidence had to be modeled for each specific question as well as each compression level making the model 3-D and far too complex for ROC studies.

2.5.2 Design of the Questionnaire

For the purpose of this project, medical experts had to evaluate a subjective preference questionnaire to assess the quality of compressed ultrasound imagery at different compression rates for several anatomical regions of interest. The questionnaire incorporated key features such as morphological details, size/volume determination, texture and other features whose properties are likely to be altered by data compression. In a blinded and randomized fashion radiologists evaluated the questionnaire for diagnostic and image quality of several normal and pathologic states. The questionnaire was developed and presented to the radiologists using the front-end software Visual Basic 5.0. (Microsoft Inc.). The following were some of the important criteria considered in the design of the questionnaire,

- The detection task had to be clearly defined. In the questionnaire presented to the radiologists, two parameters were considered. First, a binary yes/no type of response to the posed question and second, the levels of confidence on a scale of 1 to 10 (1 indicated no confidence and 10 very high confidence). From initial evaluations it became evident that some radiologists were inherently more confident about their diagnosis than others. Varying scores could skew the results. Therefore in order to set uniform scales, we set guidelines for

the confidence levels to be given to each study. The suggested guidelines to be followed were based on their confidence to read a report based on their impressions. Confidence scores of 8 to 10 would indicate excellent confidence and would be read out as presence or absence of the posed question in the report. A confidence score between four and seven would indicate that the radiologist would normally recommend a secondary evaluation. And confidence scores lower than four would denote unacceptable image quality.

- Rater effects such as learning effects could also skew the statistical results. In order to prevent learning effects, various studies were presented to the radiologists in a somewhat blinded and randomized fashion. For any particular component of a study, the questionnaire was designed in such a way that images with higher compression ratios were evaluated before lower compression ratios and uncompressed images.
- The other important consideration in the design of the questionnaire was the selection of viewing time for each study. Longer viewing time could make it easy for the radiologist but could defeat the purpose of the evaluation. However short viewing spans could make the study very difficult to interpret even without compression. Therefore after a number of initial experiments the viewing time was optimized for each study. For most studies 10 second video segments were used. While evaluating clips of longer sequences, viewers could be strongly influenced by what they saw in the last few seconds of the clip. For example if a more visible degradation occurred in the last part of the clip, then lower confidence scores are given and vice versa. Therefore the viewers were advised to give a weighted temporal score over the entire video segment.
- In order to remove the inherent biases that may be present in the images, patients who are easy to view and difficult patients were both included in the study. For example it may be

more difficult to observe the structures of the liver in an obese patient than in a thin patient. Therefore for certain studies a variety of patients were evaluated for the same levels of compressions.

2.5.3 Data Collection

For this study, nine different board certified practicing sonologists participated in the assessment by evaluating a questionnaire based on their impressions of the images displayed. The data set consisted of 40 video segments incorporating 210 questions involving various studies at different compression levels. The real-time images that were used were scans of the gall bladder, kidneys, liver, fetal heart at 16 weeks, fetus, thyroid and intra-uterine pregnancy. The various compression levels used for the study were 384, 512, 768, 1128 and 1536 kbps.

2.5.4 Inter-rater Reliability Studies

Reliability is defined as the consistency with which a result occurs. In an image evaluation study it is very important to have a high degree of agreement among trained observers to prove that the images are diagnostically acceptable. Types of reliability include internal consistency (a measure of consistency over time), intra-rater reliability and inter-rater reliability. When a person measures the same item twice and the measurements are compared, an index of intra-rater reliability is obtained. When two or more people measure the same item, an index of inter-observer variability called inter-rater reliability is obtained [27]. Alternatively, inter-rater reliability is also defined as the consensus in observation among raters when the content and conditions of the observation are the same [27]. It is also a degree of scoring stability across multiple raters.

2.5.4.1 Index of Reliability – Kappa (κ)

In the experimental setup, radiologists had to interpret an image and indicate the presence or absence of a feature in the study. The feature could be an anatomical region or identification of pathology and the observation is a yes/no outcome. Images were presented to the radiologists and the evaluations were made independently, i.e., one radiologist did not know the results of the other radiologist’s diagnosis. In order to measure the degrees of agreement between the evaluations of two radiologists, a measure of reproducibility for categorical data called the kappa (κ) statistic was used. Kappa is defined as the agreement beyond chance divided by the agreement possible beyond chance [27].

2.5.4.2 Kappa Estimation

The value of kappa (κ) can be estimated from the 2 x 2 truth table generated from responses of two different radiologists. The truth table is shown in table 2-2. In the kappa evaluation, the responses of one radiologist are compared with the another for the same set of images to find a measure of inter-rater reliability.

Table 2-2. Truth table containing the comparisons of observations of the images as seen by the trained observer (Radiologist B) and the clinical gold standard (Radiologist A).

		RADIOLOGIST B		TOTALS B
		POSITIVE	NEGATIVE	
RADIOLOGIST A	POSITIVE	TP (a)	FP (b)	POSITIVE (p_1)
	NEGATIVE	FN (c)	TN (d)	NEGATIVE (q_1)
TOTALS- A		POSITIVE (p_2)	NEGATIVE (q_2)	TOTAL (T)

The responses to the questionnaire from various radiologists can be compared to the clinical gold standard for the experiment and the values obtained can be used to fill the truth table (table 2-2). In our experimental setup, the responses of each radiologist were compared with a gold standard. The gold standard for this study was the observation of a radiologist who knew the case history of the patient. For the truth table, the overall proportion of agreement is given by

$$P_0 = a + d \quad (2.1)$$

If we suppose that the category under study is relatively rare so that the proportion 'd' representing the agreement on absence is likely to be large, then the value of P_0 would be inflated. Therefore an index of specific agreement was proposed [28] and is given by

$$P_s = \frac{2a}{2a + b + c} \quad (2.2)$$

To estimate the agreement by chance, we consider the following steps

Radiologist A identifies a total of p_1 positive scans and radiologist B identifies a total of p_2 positive scans. Total agreement by chance is given by

$$P_T = \left(\frac{p_1}{T} \times \frac{p_2}{T} \right) \times T \quad (2.3)$$

Similarly the negative agreement by chance is given by

$$P_N = \left(\frac{q_1}{T} \times \frac{q_2}{T} \right) \times T \quad (2.4)$$

Total proportion for chance agreement is given by

$$P_C = \frac{P_N + P_T}{T} \quad (2.5)$$

The agreement beyond chance is given by

$$P_0 - P_C \quad (2.6)$$

The potential agreement beyond chance is given by

$$1 - P_C \quad (2.7)$$

Substituting the values obtained in the definition of kappa,

Kappa is obtained as

$$\kappa = \frac{P_0 - P_C}{1 - P_C} \quad (2.8)$$

Alternatively by substituting the value of the indices, kappa is obtained as

$$\kappa = \frac{2(ad - bc)}{p_2q_1 + p_1q_2} \quad (2.9)$$

A value of κ equal to one indicates perfect agreement between raters and κ equal to zero indicates no agreement between raters. For inter-rater reliability in diagnostic image quality measurements, a value of κ greater than 0.75 indicates excellent reliability, values of κ between 0.4 and 0.75 indicates good inter-rater reliability and κ less than 0.4 indicates marginal reproducibility. Interpretations for the scores can also be based on a scale suggested in a previous study [29] with kappa values less than 0.0 as 'poor' agreement, kappa values between 0.0 and

0.20 as '*slight*' agreement, 0.21 to 0.40 as '*fair*' agreement, 0.41 to 0.60 as '*moderate*' agreement, 0.61 to 0.80 as '*substantial*' agreement and kappa values between 0.81 and 1.0 as '*almost perfect*' agreement.

2.5.5 General Linear Models (GLM) and Analysis of Variance (ANOVA)

Observer confidence can be modeled as a function of compression, bandwidth and rater effects using GLM. The GLM procedure routine is a part of the Statistical Analysis Software (SAS), SAS Institute Inc., Cary North Carolina. The SAS models the data using observer confidence as the dependent variable and bandwidth, type of study and compression levels as the independent variables. In the data set collected, the rater variable took nine different values, the compression variable took five different values one for each level of compression and one uncompressed level. The model also included eight different studies and each rater was posed with 210 questions. GLM provides a method to integrate all these various parameters.

Analysis of Variance traditionally abbreviated as ANOVA is used to measure the overall impact of compression on observer confidence (independent variable) taking into consideration the independent factors in the study, namely observer and image. ANOVA provides a method to test a given hypothesis by comparing two or more means in a population accounting for the disparities in the distribution of the independent variables in the hypothesis. The language used in ANOVA derives its roots from agriculture and medicine [30]. Thus different populations are often called "treatments". In this research, the treatments (factors) are considered to be the bandwidth settings, the raters, and the images. Even if none of the treatments has a significant effect on the entire model, there will still be some variation in the data obtained because of the nature of sampling and measurement. Using ANOVA, the variations in the data can be

partitioned about a grand mean and a residual term [31]. The grand mean is the mean value of the confidence scores across all raters and bandwidths. The residual variation is obtained by subtracting the raters confidence scores from the grand mean. By comparing the variation that we attribute to each “treatment” (bandwidth settings, raters and images) with the residual variation, we can decide which factors contribute significantly to the overall variation in the data. Thus ANOVA can be broadly used to analyze the components of variance in confidence across a series of compression levels.

In our experiment we consider the effect of bandwidth, observer and image on observer confidence. Using this method we ask, “ How much variation in observer confidence can be explained by bandwidth? ”. If bandwidth accounts for very little variation in observer confidence (dependent variable), we conclude that bandwidth does not affect the overall observer confidence.

2.5.6 Hypothesis Test Procedure

The GLM procedure of the SAS is used to test the hypothesis of the experiment. Hypothesis testing is a statistical procedure for determining whether to accept or reject a null hypothesis of no difference across group means [30]. For this procedure we first decide on a null hypothesis. In our study, the null hypothesis is that there is no difference in confidence across the different compression levels. To test this hypothesis, we choose a level of significance (α) which is used to accept or reject the null hypothesis. Alpha (α) is the frequency with which the null hypothesis will be rejected when there is no difference in the mean values for the various compression ratios. Traditionally we set α less than or equal to 0.05 which is an error rate of 5%. That is 5% of all analysis when the null hypothesis is true, we will declare it falsely to be

erroneous. The values from ANOVA are used to compare with the value of α to reject or accept the null hypothesis for each bandwidth. Using ANOVA in the GLM we determine the P value for each compression level and each bandwidth to compare with observer confidence. The P value gives the frequency with which we would expect to see a value as extreme or more extreme than the experimental data if the null hypothesis were true. The null hypothesis will be rejected if the P values obtained from the GLM are less than or equal to α and accepted if the P values are greater than α .

2.6 Objective Quality Assessment

Objective methods of quality measurements are used to complement the results obtained from subjective testing methods. Objective forms of testing quantify the quality of a compressed image without the participation of human subjects. These methods include estimation of Peak Signal to Noise Ratios (PSNR), histogram analysis and verification of the rate-distortion function. Data compression has a direct impact on the objective measurements and the results are highly dependent on the content of the image. Results obtained for image quality through subjective methods are not usually predictable and need not necessarily show monotonically increasing confidence scores with decreasing compression. The evaluation scores given by radiologists to the images are based on their perception and it is difficult to predict the scores given by the raters. Some raters may be more inherently confident on their assessment while others may not be confident about their diagnosis. However, objective image quality evaluation gives more predictable results. For example in the study where compression adds distortion to the signal, we can expect increasing distortion with increasing compression ratios. In other words if the source material were stressful, in terms of spatial detail and motion, subjective testing may

not show monotonically increasing confidence scores with decreasing compression ratios because of human factors such as fatigue. The results obtained through objective estimations can be better predicted, as human factors do not play a role in this type of testing.

In objective evaluation, the usual protocol followed is to make a set of measurements of quality on the reference-input signal, make a set of identical measurements on the output signal and compare the two measurements. For the evaluation of ultrasound image compression, the reference signal would be the uncompressed image and the compared signal would be the compressed image. The objective methods used in our study were the estimation of PSNR and histogram analysis.

2.6.1 Data Collection

Figure 2-3 is a flow diagram of image handling to make the images suitable for objective evaluation methods. The flow diagram includes the software and the hardware used in producing the required format suitable for image analysis.

2.6.2 Histogram Analysis

The difference image between the original and the compressed image gives a measurement that compares the quality of the compressed image with that of the original image. The corresponding histogram of the difference image provides a qualitative measurement of the difference between the original and the compressed images. A narrow histogram means that the difference between the two images is small, whereas a broad histogram indicates a large difference. In the difference images, the differences created by the compression algorithm are

evenly distributed over the entire image; the higher the compression ratio the larger the difference.

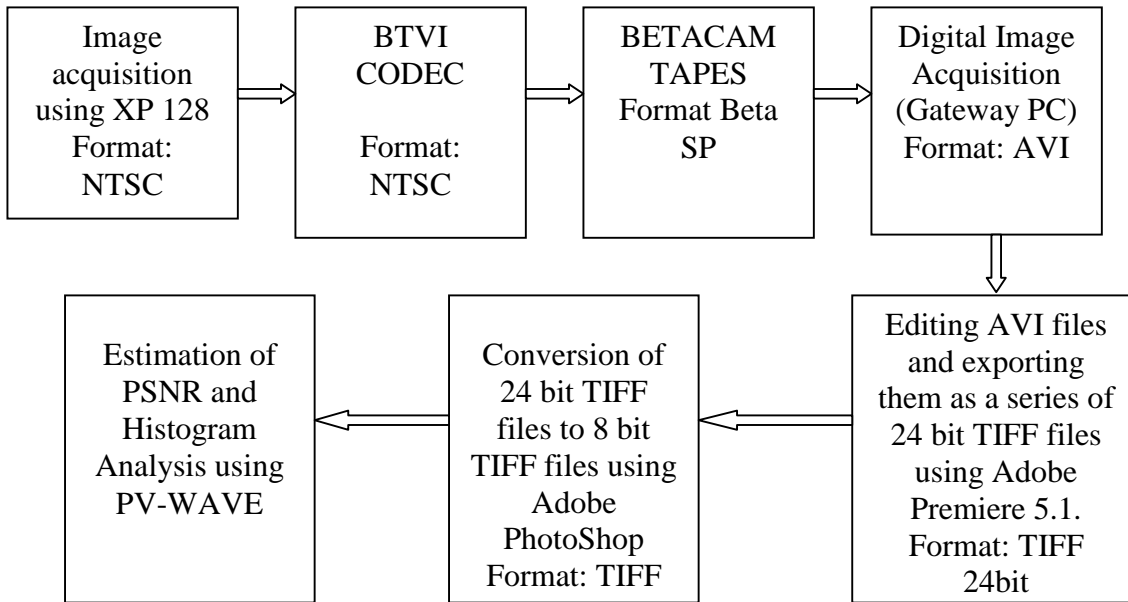


Figure 2-3. Flow diagram of image handling including the equipment used, format and the software used in each stage.

TIFF= Tagged Image File Format, AVI= Audio Video Interleaved

The abscissa for the histogram of the difference image varies between -255 and $+255$ for an 8-bit image. In the experiment, the histogram of the difference image was generated using custom written PV-Wave programs and by using Adobe Photoshop. A cine-loop image of the liver was acquired using an XP-128 scanner and compressed at various compression levels using a BTVI CODEC and taped onto betacam tapes. Images were edited to still format 24 bit TIFFs from the AVI video format using Adobe Premiere. The 24 bit images were imported into Adobe Photoshop and converted to 8 bit TIFFs. The difference image between the compressed image

and uncompressed images were generated and the histograms were plotted for the difference image using PV-Wave routines. The mean, standard deviation and the median values were also estimated using both Photoshop and PV-Wave routines. The plotted histograms were analyzed for various levels of compression

2.6.3 Peak Signal to Noise Ratio (PSNR)

In lossless compression, the reconstructed signal is an exact numeric representation of the original signal; therefore the reconstructed signal is said not to have any distortion. However in lossy compression the reconstructed signal is not the same and there will be a difference between the original signal and the reconstructed signal. This difference is called distortion. To evaluate the effectiveness of the compression scheme, it is important to estimate distortion generated during compression. Ideally we would like to incur minimum distortion while compressing to the lowest possible bit-rate. The distortion measures the closeness of the original and compressed signals and can be evaluated using the PSNR. There is a trade-off between the rate and the amount of distortion with increasing distortion for decreasing bit-rate. The calculation of PSNR involves the estimation of two measures of distortion; the squared mean measure and the absolute difference measure. Images required for the calculation of PSNR were 8-bit TIFFs. The procedure for acquiring the images suitable for the estimation of PSNR was exactly the same as described for the histogram analysis. Custom programs were written in PV-wave for the extraction of PSNRs for various levels of compression.

If $i(x, y)$ is the source image and $j(x, y)$ is the compressed image, then the difference image was estimated using the formula

$$\Delta(x, y) = j(x, y) - i(x, y) \quad (2.10)$$

The mean error Δ_{μ} between the two image was calculated by averaging $\Delta(x, y)$ over all the pixels and was given by

$$\Delta_{\mu} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \Delta(x,y) \quad (2.11)$$

Where M and N were the number of rows and columns in the image. Similarly the scatter in pixel intensity also called the root mean square error for the images was estimated using the formula

$$\Delta_{\sigma} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \Delta^2(x,y) \right]^{\frac{1}{2}} \quad (2.12)$$

The SNR was calculated between the compressed image and the original corresponding frame. The SNR was measured on a logarithmic scale.

$$SNR(db) = 10 \log_{10} \left[\frac{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} j^2(x,y)}{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \Delta^2(x,y)} \right] \quad (2.13)$$

In video compression, signal to noise ratios are expressed in terms of PSNRs. PSNR was obtained by replacing the numerator with the peak possible value of the input signal, which is the maximum decimal value of an 8-bit decimal integer. PSNR was estimated by the formula

$$PSNR(db) = 10 \log_{10} \left[\frac{255}{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \Delta^2(x,y)} \right] \quad (2.14)$$

Chapter 3

Results and Discussions

3.1 Kappa Studies

The responses from nine radiologists to the questionnaire were analyzed using kappa statistics to obtain an index of inter-rater reliability. The responses to the questionnaire were output from visual basic as plain text files and were used to populate the 2x2 truth table (table 2-2). The values of the true positives, true negatives, false positives and false negatives were obtained for each study by comparing each response of the questionnaire to the set clinical gold standard.

In this study, the clinical gold standard was the responses to the questionnaire by an expert panel consisting of two radiologists who knew the clinical outcome and the case history of each patient in the study.

The values from the 2x2 truth table were substituted in equation 2.9 to estimate kappa, which gives a measure of the inter-rater reliability for each radiologist. The value of kappa was estimated across all studies for various bandwidths for each radiologist. The average value of kappa for each bandwidth across all the raters is shown in table 3-1.

The average value of kappa for each study irrespective of bandwidths was also estimated and is shown in table 3-2. The obtained kappa values summarize the reliability measurements between similar observers. A high value of kappa suggests that it is unlikely that different evaluating radiologists with similar experience would give different diagnoses for the same patient.

Table 3-1. The estimated average values of kappa for different bandwidths across all the studies and raters.

Bandwidth	Kappa
128 kbps	0.38
364 kbps	0.69
768 kbps	0.66
1152 kbps	0.59
1536 kbps	0.68
74 Mbps	0.77

Table 3-2. The estimated average values of kappa for different studies across all the bandwidths and raters.

Study	Kappa
Fetal heart at 16 weeks	0.85
Gall bladder	0.90
Intra-uterine pregnancy	0.39
Uterus	0.80
Fetus	0.84
Female Pelvis	0.90
Liver	0.67
Kidneys	0.91

From table 3-1, analysis of inter-rater variability for image quality for different bandwidths across all studies and raters indicated an overall ‘*substantial/good*’ agreement on the basis of the interpretation developed in a previous study [29]. The results in the table also indicated that the values at 128 kbps represented ‘*fair*’ but not sufficient inter-rater reliability. However from bandwidths greater than or equal to 384 kbps to uncompressed, the kappa values did not show significant variation. Lower kappa values at higher compression rates would have indicated greater disagreement between the raters. A monotonic increase in kappa values with increase in bandwidth as initially expected was not observed in the study.

Table 3-2 summarizes the estimated kappa values for various studies across all the bandwidths and raters. The kappa values for most studies were between 0.67 and 0.90 indicating *substantial to almost perfect* inter-rater reliability. The average kappa value for the IUP was significantly lower than the other studies. Further studies need to be done by incorporating a larger sample of patients for the different bandwidths. Fetal heart studies are sometimes considered the clinical gold standard in the evaluation of ultrasound imaging. The average value of kappa for the fetal heart study indicated excellent inter-rater reliability for all the bandwidths.

3.2 General Linear Models

A General Linear Model was used to model the average confidence of the observer as a function of image, bandwidth and observer. As previously mentioned, observer confidence was measured for each study parameter across all the bandwidths in the study. The average value of confidence across all raters and bandwidths for each study parameter was estimated. SAS procedures were used to model the observer confidence in terms of the observer effect,

bandwidth and the image and perform analysis of variance (ANOVA) on these parameters. ANOVA describes the variations associated with each factor and their interactions. SAS procedures estimate the least square means for bandwidth, observer and image independently. SAS is also used to estimate the degree of freedom (DF), sum of squares (SS) and mean square (MS) values for each element in the data table. This model generates the R^2 values for each study, which is a measure of the variance explained by the model. A higher value of R^2 indicates a better explanation of the variances present in the model. The GLM procedure in SAS also generates the F-statistics and corresponding P values. The degrees of freedom in the model represent the number of groups in each factor and the sample size, which affects the sensitivity of the analysis of variance. The DF also gives a measure of the number of groups in each factor. For example, when there were nine observers, the DF was eight. Similarly when the total number of observations in the study was 1449, the DF was 1448.

3.3 Average Confidence

Figures 3-1 to 3-10 show the average confidence across the bandwidths for each study component of the various ultrasound studies. The average confidence for gall bladder studies showed minimal variation across the four different bandwidths. The observations of rater confidence for gall bladder (figure 3-1) indicated a relatively low effect of compression on the ease of diagnosis of gall bladder pathologies. The confidence level variations for the kidney studies were similar to those of the gallbladder studies. The kidney studies also indicated a lower level of confidence in the evaluation of stones and masses as compared to increased echogenicity.

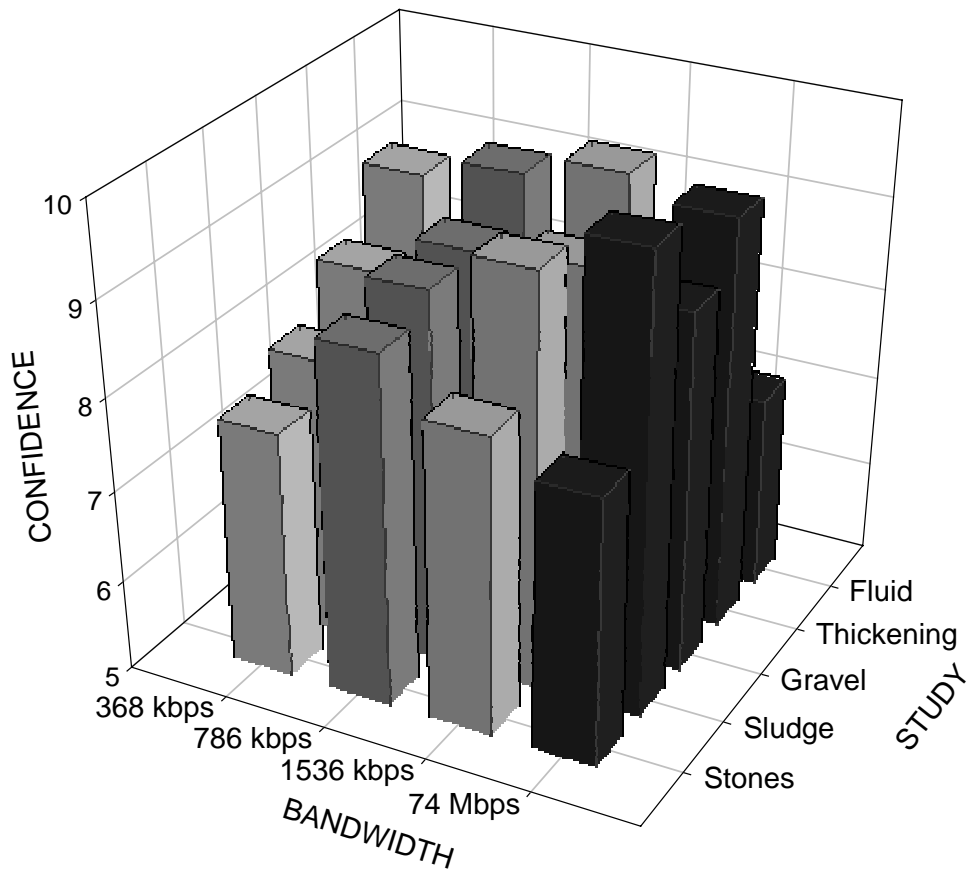


Figure 3-1. Average confidence plots for the gall bladder, plotted against bandwidth and specific study aspects.

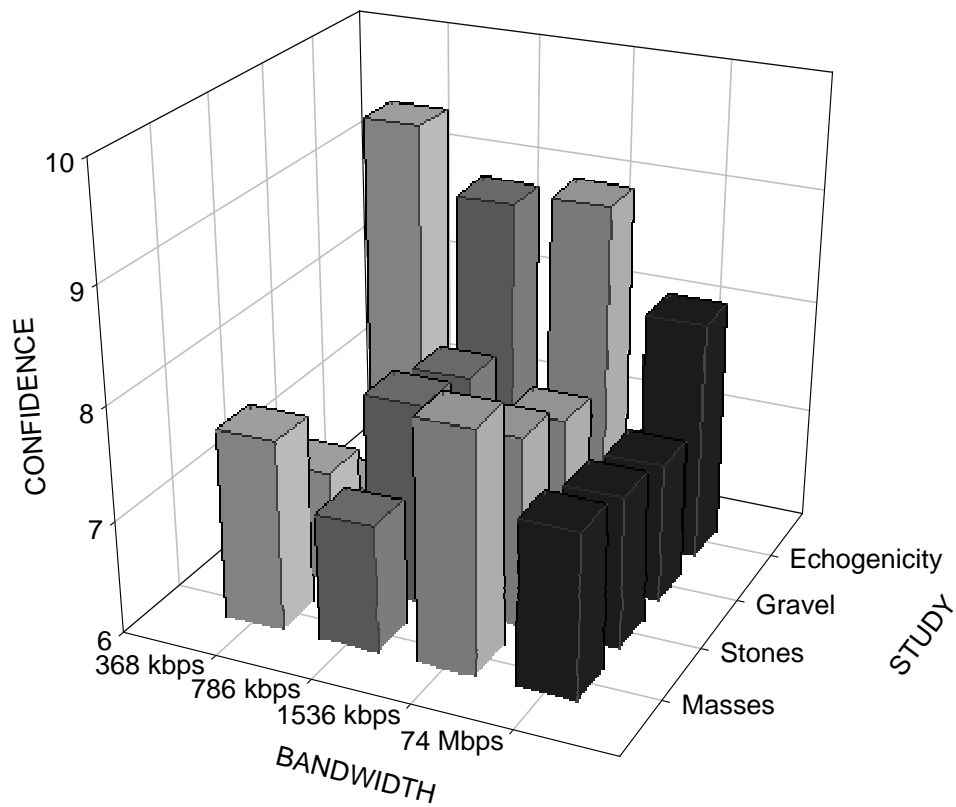


Figure 3-2. Average confidence plots for the kidney, plotted against bandwidth and specific study aspects.

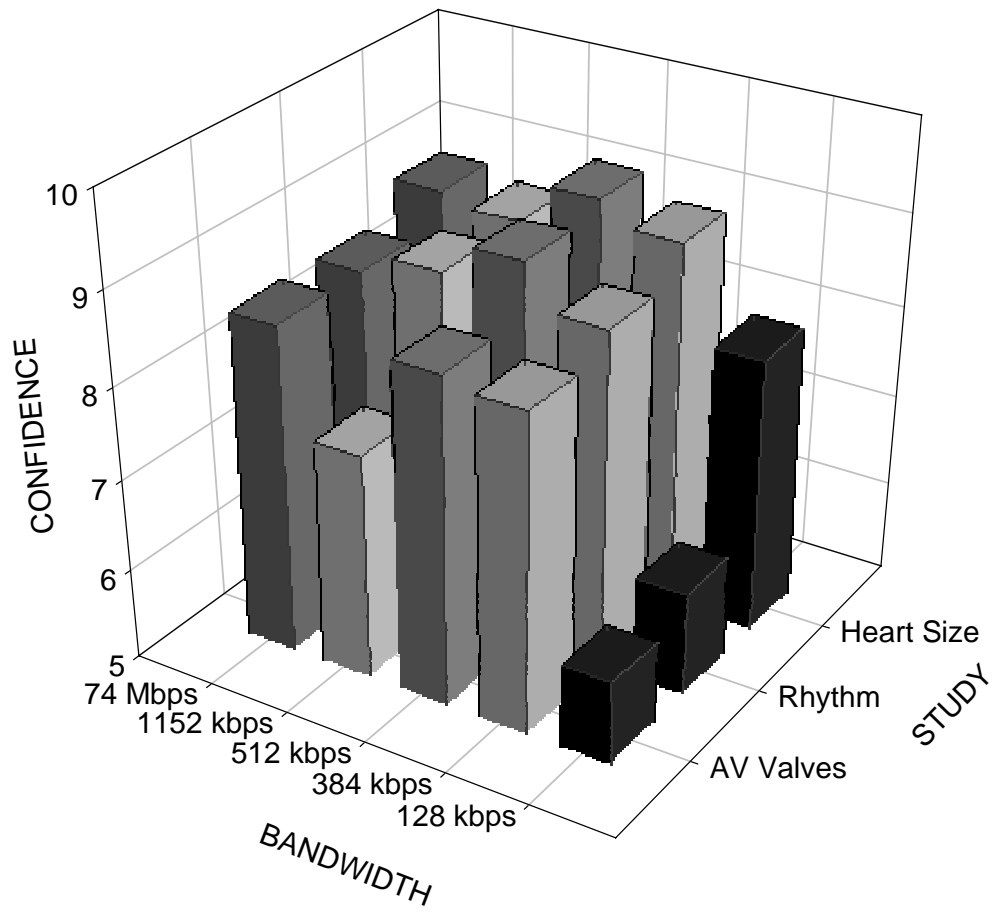


Figure 3-3. Average confidence plots for the fetal heart, plotted against bandwidth and specific study aspects.

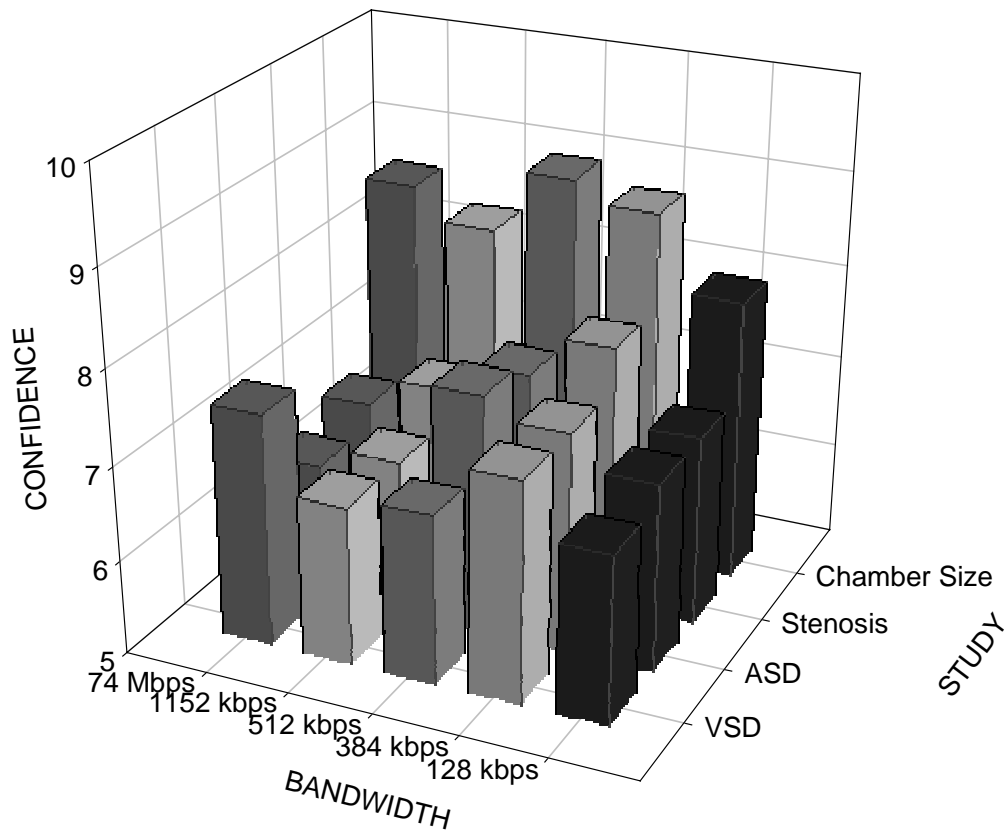


Figure 3-4. Average confidence plots for the fetal heart, plotted against bandwidth and other specific study aspects.

VSD= Ventricular Septal Disease, ASD= Atrial Septal Disease

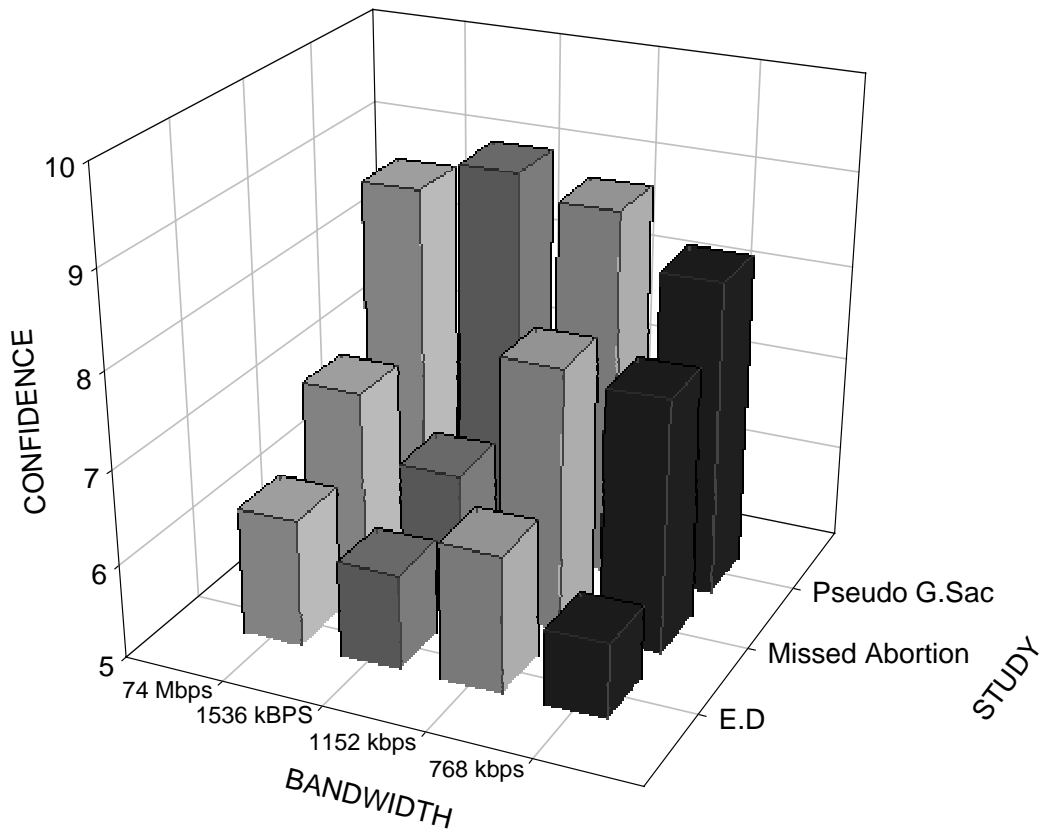


Figure 3-5. Average confidence plots for the intra-uterine pregnancy, plotted against bandwidth and specific study aspects.

E.D = Embryonic Demise, G.Sac = Gestation Sac

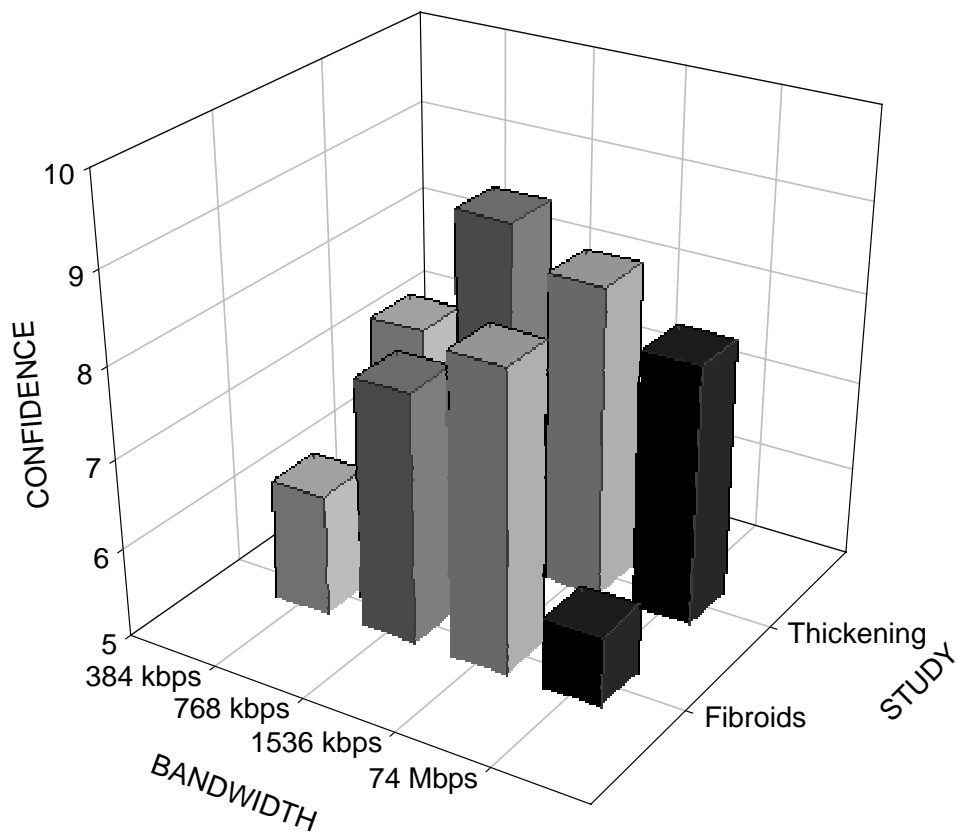


Figure 3-6. Average confidence plots for the uterus, plotted against bandwidth and specific study aspects.

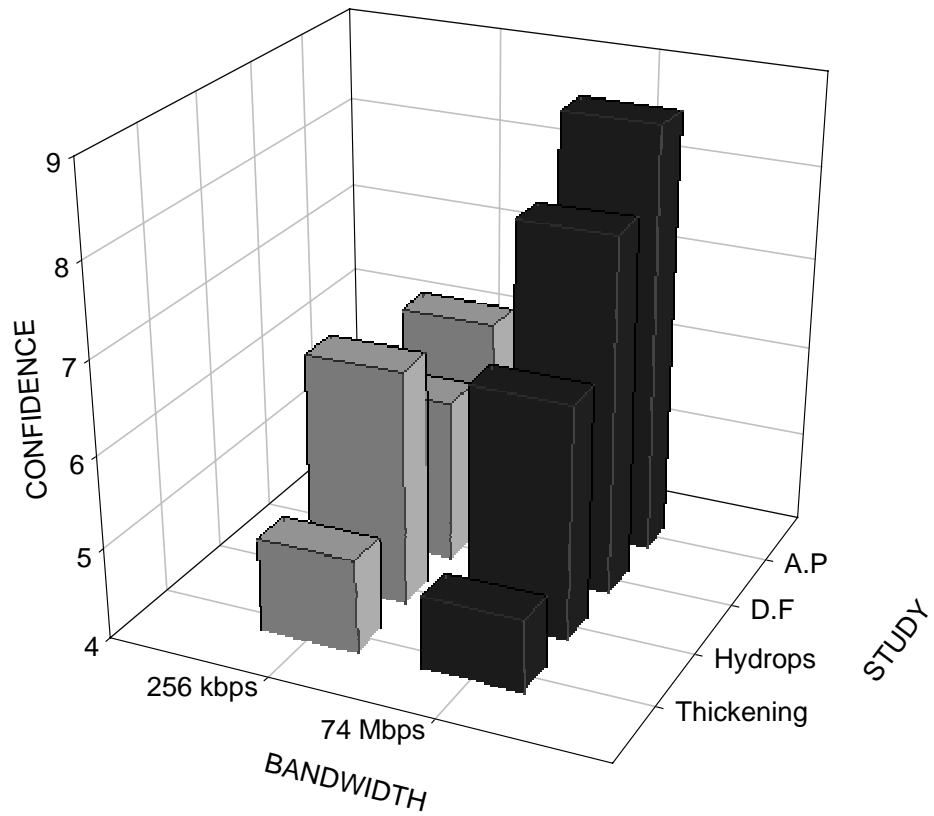


Figure 3-7. Average confidence plots for the fetus, plotted against bandwidth and specific study aspects.

A.P = Abdominal Pregnancy, D.F= Dead Fetus

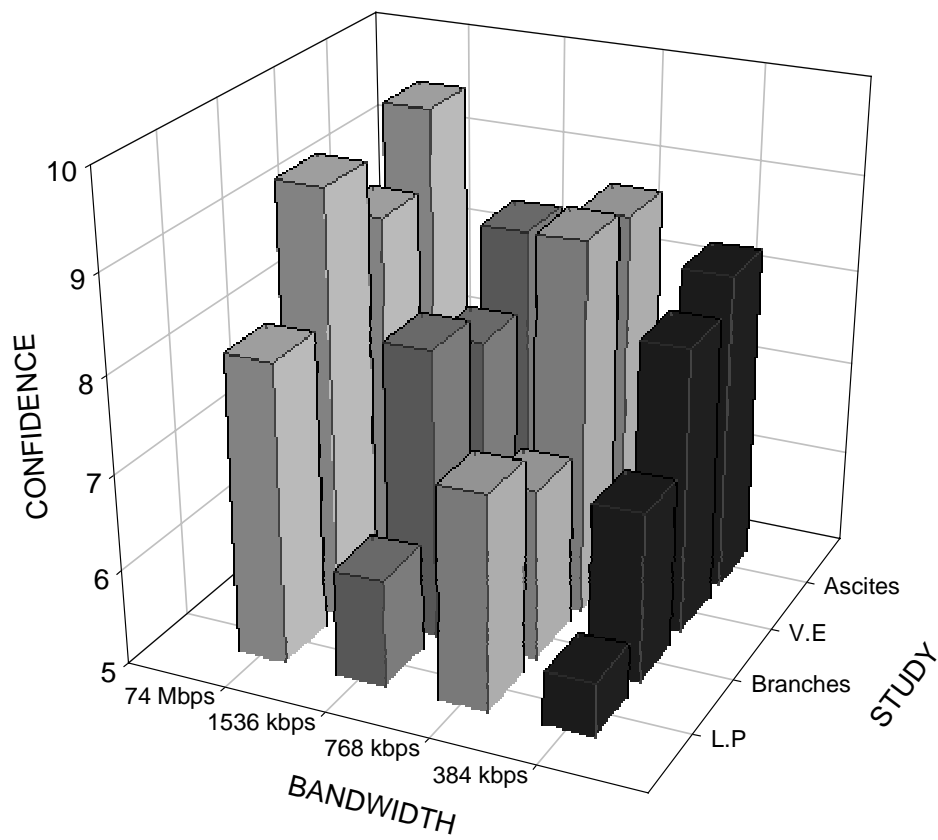


Figure 3-8. Average confidence plots for the liver, plotted against bandwidth and specific study aspects.
 V.E= Veins Engorged, L.P= Liver Parenchyma

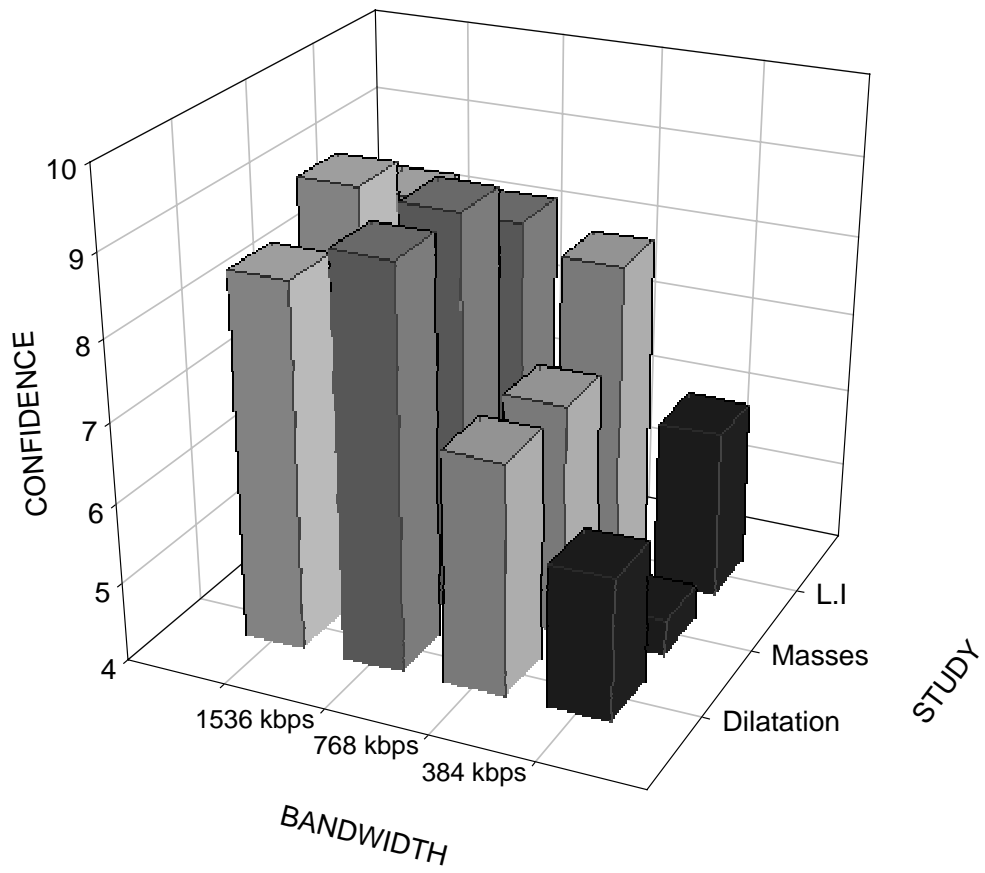


Figure 3-9. Average confidence plots for the liver, plotted against bandwidth and other specific study aspects.
 L.I= Liver Infiltration

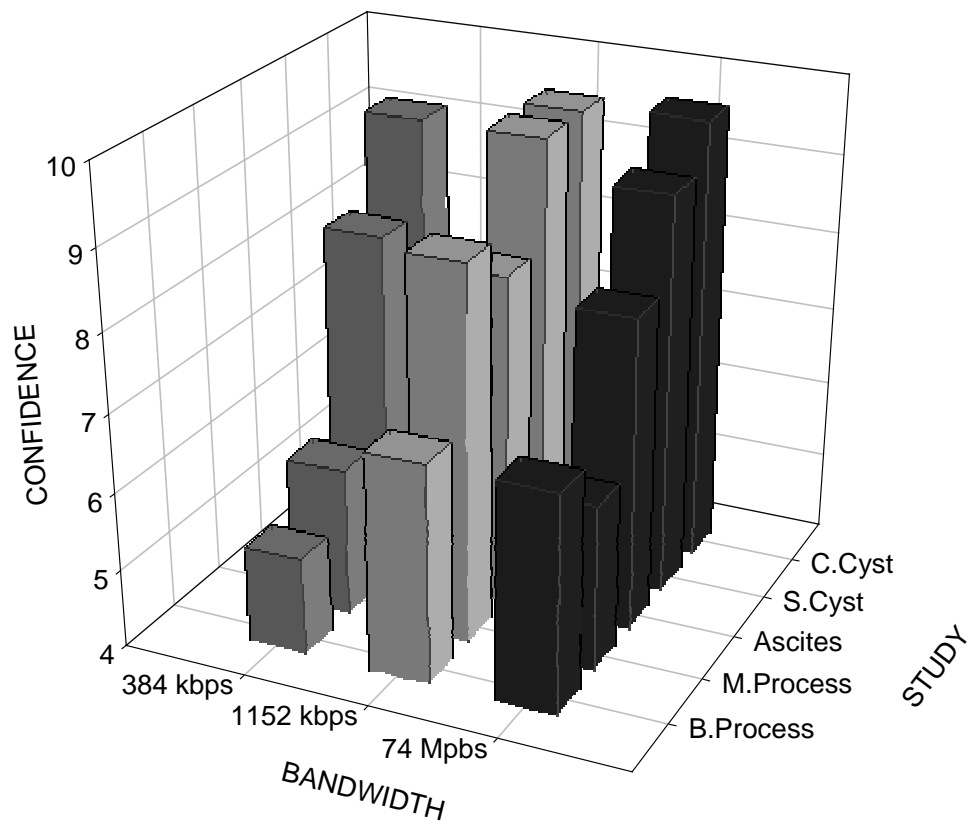


Figure 3-10. Average confidence plots for the female pelvis, plotted against bandwidth and specific study aspects.
 C.Cyst = Complex Cyst, S.Cyst = Simple Cyst, M.Process = Malignant Process and B. Process = Benign Process

The results of the confidence studies for the fetal heart at 16 weeks (figures 3-3 and 3-4) indicated conservative average confidence scores by the evaluating radiologists at 128 kbps making the image diagnostically unacceptable. No significant difference in the diagnostic confidence value was detected for the other bandwidths indicating a low effect of data compression on the image quality in fetal heart studies. Statistically significant information could not be obtained for the fetus and female pelvis studies due to the insufficient variety of bandwidth levels obtained at the acquisition stage of the experiments.

The confidence plots for the intra-uterine pregnancy indicate similar confidence levels at 768 kbps and 1152 kbps and higher confidence at 768 kbps. The confidence scores for the compressed images were less than for the uncompressed images. In these studies we were unable to establish that increasing compression ratios produced a gradual and predictable decrease in the confidence levels.

3.4 Key for the ANOVA Table and SAS Results

3.4.1 Dependent Variable

The dependent variable in the SAS program was set to be the observer confidence scores. The three way ANOVA tests were performed to determine if the values of the dependent variables were affected by different factors such as bandwidth, image or the observer in the model. The ANOVA table uses general linear models in the estimation of the results of the three way measures. The results are calculated for each factor and between factors.

3.4.2 Degrees of Freedom (DF)

The degrees of freedom in this study are a measure of the total number of images, bandwidth settings and observers, which affect the sensitivity of ANOVA. The total DF is the number of independent observations. The DF total is one less than the total number of observations. In our study there were a total of 1449 questions, therefore the DF total is 1448. Factor degrees of freedom are measures of the number of levels available to estimate the mean across each factor. The DF factor is always one less than the actual number in each factor of the ANOVA table. Since there were eight images, nine bandwidths and eight raters, the corresponding DFs were seven, eight and seven respectively.

3.4.3 Sum of Squares (SS)

The sum of squares is a measure of the variability associated with each element in the ANOVA table. In ANOVA as previously mentioned, the observer confidence is partitioned about a grand mean value, which is the mean value of confidence scores for all the bandwidth settings and a residual term. The residual term is obtained by subtracting the raters confidence score from the grand mean.

In the GLM procedure instead of looking at the differences in scores from the grand mean, we look at the sum of the squares of the residual differences. We can separate the total variations into two parts, one due to the different compression levels and the other due to random sampling errors often called as sum of squares for errors or SS Error.

3.4.4 Mean Squares (MS)

The mean square value is the average of the sum of squares. The mean squares provide estimates of the population variances. In other words, the MS value for the each factor is the value obtained when the sum of the squares for the factor is divided by the corresponding DF.

The mean square for each factor is estimated by

$$\text{MS Factor} = \text{SS for each factor} / \text{DF for each factor}$$

For example in our study the MS for the bandwidth factor is 29.93 (SS bandwidth =209.54 divided by the DF bandwidth = 7). By comparing the MS values between factors, we can choose the factor that affects confidence the most

3.4.5 F Statistic

The F-statistic is a measure of the comparison within each factor and between the factors and is estimated by

$$F_{\text{factor}} = \text{MS for the factor} / \text{Error MS}$$

For example in our study, the F value for the entire model was 20.37 (MS model =77.196 divided by MS error =3.79). The F value along with MS and SS indicates which of the 3 independent factors (image, bandwidth and rater) affects the dependent variable. In general, the higher the F value, the greater the contribution to the variation in observer confidence.

3.4.6 P-Value

The P value is the frequency with which we would expect to see a value as extreme or more extreme than the experimental data if the null hypothesis were true. The smaller the P value, the less consistent the data are with the null hypothesis. Traditionally we conclude that

there are significant differences if $P \leq 0.05$, which is an error rate of 5%. For example in our study, while comparing the P values for 128 kbps with any other bandwidth, the values are lower than 0.05. Therefore we can conclude that observer confidence is affected by the images at 128 kbps.

Tables 3-3 to 3-6 are the outputs of the SAS general linear model procedure for the 3 way ANOVA, for the entire data set. Table 3-7 is a summary of the SAS results for average confidence modeling for each study component separately.

Tables 3-3 to 3-6 show the results that were obtained by the SAS GLM procedure modeling the entire data set with the average observer confidence as the dependent variable and bandwidth, image and observers as the independent variables.

The ANOVA table (table 3-3) for the overall model breaks down the sum of squares for the dependent variable average confidence into the portions attributed by the model and portions attributed by error. The portions of error are due to the errors in the sampling of data. In our model, the sum of squares for the entire model accounts for only 23% of the variations in confidence scores and the value of error for the entire model was estimated as 77.19. The mean square values for the errors, which is an estimate of the variance of the true errors, was 3.790

The F tests in ANOVA explain whether the means differ significantly from each other. In other words by comparing the F values in the ANOVA table we can explain which of the 3 dependent factors affects the dependent variable the most. In our study, from the table 3-3, we have the F values of the dependent variables image, bandwidth and observer to be 7.90, 4.35, and 25.29, respectively. From these 3 values we see that the observer variable contributes the most to the variation in observer confidence which is the dependent variable.

Table 3-3. The output of the SAS GLM procedure for 3-way ANOVA with the dependent variable confidence and the independent variables image, observer and bandwidth.

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	1698.32	77.19	20.37	0.0001
Error	1426	5405.10	3.79		
Corrected Total	1448	7103.43			
Image	7	209.54	29.93	7.90	0.0001
Bandwidth	7	115.34	16.47	4.35	0.001
Observer	8	1373.43	171.67	25.29	0.001

Dependent Variable: Confidence

R-Square: 0.23

Confidence Mean: 7.84

Table 3-4. The results of the SAS procedure that models average confidence as a function of the bandwidth across all observers and images.

Bandwidth	Confidence	I/J	1	2	3	4	5	6	7	8
128 kbps	6.66	1	.	0.037	0.0039	0.0004	0.0008	0.0001	0.0001	0.0002
256 kbps	7.50	2	0.037	.	0.9133	0.3461	0.4474	0.0453	0.1586	0.3142
384 kbps	7.54	3	0.0039	0.9133	.	0.2567	0.2703	0.0036	0.0277	0.1476
512 kbps	7.89	4	0.0004	0.3461	0.2567	.	0.6865	0.4474	0.8021	0.7578
768 kbps	7.76	5	0.0008	0.4474	0.2703	0.6865	.	0.0737	0.2558	0.8257
1152 kbps	8.12	6	0.0001	0.0453	0.0036	0.4474	0.0737	.	0.4614	0.0607
1536 kbps	7.97	7	0.0001	0.1586	0.0277	0.8021	0.2558	0.4614	.	0.3294
74000 kbps	7.80	8	0.0002	0.3142	0.1476	0.7578	0.8257	0.0607	0.3294	.

Table 3-5. The results of the SAS procedure that models average confidence as a function of the image across all bandwidths and observers.

Image	Confidence	I/J	1	2	3	4	5	6	7	8
Gall Bladder	8.50	1	.	0.1913	0.1188	0.044	0.0054	0.0001	0.0001	0.0001
Kidneys	8.20	2	0.1913	.	0.6204	0.5095	0.1031	0.0001	0.0001	0.0344
Uterus	8.06	3	0.1188	0.6204	.	0.9688	0.3395	0.0001	0.0084	0.2734
Fetal heart	8.05	4	0.044	0.5095	0.9688	.	0.2363	0.0001	0.0007	0.1482
Female Pelvis	7.76	5	0.0054	0.1031	0.3395	0.2363	.	0.0001	0.0918	0.9316
Fetus	6.53	6	0.0001	0.0001	0.0001	0.0001	0.0001	.	0.0102	0.0001
IUP	7.34	7	0.0001	0.0001	0.0084	0.0007	0.0918	0.0102	.	0.0138
Liver	7.78	8	0.0001	0.0344	0.2734	0.1482	0.9316	0.0001	0.0138	.

Table 3-6. The results of the SAS procedure that models average confidence as a function of the observer across all bandwidths and images.

Observer	Confidence	I/J	1	2	3	4	5	6	7	8	9
1	7.97	1	.	0.0002	0.0001	0.0001	0.0001	0.0004	0.0014	0.0001	0.7099
2	7.16	2	0.0002	.	0.0005	0.5866	0.0001	0.0001	0.0001	0.0001	0.0001
3	6.41	3	0.0001	0.0005	.	0.0036	0.0001	0.0001	0.0001	0.103	0.0001
4	7.04	4	0.0001	0.5866	0.0036	.	0.0001	0.0001	0.0001	0.0001	0.0001
5	8.82	5	0.0001	0.0001	0.0001	0.0001	.	0.7099	0.4744	0.0001	0.0004
6	8.74	6	0.0004	0.0001	0.0001	0.0001	0.7099	.	0.7313	0.0001	0.0015
7	8.67	7	0.0014	0.0001	0.0001	0.0001	0.4744	0.7313	.	0.0001	0.0047
8	6.06	8	0.0001	0.0001	0.103	0.0001	0.0001	0.0001	0.0001	.	0.0001
9	8.05	9	0.7099	0.0001	0.0001	0.0001	0.0004	0.0015	0.0047	0.0001	.

Table 3-7. A summary of the results of the SAS GLM procedure modeling each study parameter separately to find the effect of bandwidth on observer confidence.

Study	R² Value	Confidence	Bandwidth F- Value	Bandwidth P-Value	Observer F-Value	Observer P-Value
Ascites	0.46	8.56	0.12	0.9878	6.33	0.001
ASD	0.48	6.98	0.65	0.6342	3.47	0.0054
AV Valves	0.74	7.91	6.59	0.0006	8.14	0.001
Benign Cysts	0.63	6.37	1.42	0.2698	3.06	0.0273
Cardiac Rhythm	0.61	8.26	5.70	0.0014	3.17	0.093
Chamber Size	0.83	8.57	1.65	0.1858	18.50	0.0001
Complex Cyst	0.64	9.56	0.57	0.5758	3.43	0.0173

Table 3-7 (continued)

Study	R² Value	Confidence	Bandwidth F- Value	Bandwidth P-Value	Observer F-Value	Observer P-Value
Embryonic Demise	0.71	6.22	0.50	0.6846	7.38	0.0001
Endometrial Thickening	0.82	8.52	1.23	0.3191	13.59	0.0001
Fatty Infiltration	0.36	7.62	1.46	0.2366	1.55	0.1791
Fibroids	0.66	7.44	5.20	0.0065	3.87	0.0047
Flow in Myometrium	0.58	7.72	1.39	0.2708	3.64	0.0066
Gravel	0.32	7.08	0.07	0.9772	1.38	0.2567
Heart size	0.76	8.78	2.11	0.1019	11.50	0.0001
Hepatic Veins	0.52	8.67	0.51	0.5301	4.03	0.0021
Intra-hepatic Dilatation	0.54	7.47	2.19	0.0927	3.67	0.0039
Malignant Cyst	0.45	5.63	0.19	0.8316	1.59	0.2056

Table 3-7 (continued)

Study	R² Value	Confidence	Bandwidth F- Value	Bandwidth P-Value	Observer F-Value	Observer P-Value
Polycystic Masses	0.73	6.95	5.52	0.0017	7.79	0.0001
Missed Abortion	0.53	7.19	1.39	0.2708	2.92	0.0199
Cyst	0.59	7.53	0.75	0.5316	4.00	0.0039
Increased Echogenicity	0.53	9.14	0.25	0.8594	3.26	0.0118
Normal IUP	0.41	7.78	0.31	0.8196	1.96	0.0965
Normal Parenchyma	0.57	7.40	3.45	0.0187	3.66	0.0039
Mitral / Tricuspid Stenosis	0.68	7.40	1.40	0.2551	7.95	0.0001
Nuchal Thickening	0.70	5.28	0.11	0.7539	2.35	0.1247
Pericholycystic Fluid	0.56	9.60	1.18	0.3377	3.444	0.0089
Portal Veins	0.44	8.00	2.46	0.0657	1.90	0.0939

Table 3-7 (continued)

Study	R² Value	Confidence	Bandwidth F- Value	Bandwidth P-Value	Observer F-Value	Observer P-Value
Simple Cysts	0.54	9.23	0.79	0.04719	2.81	0.0877
Sludge	0.65	5.56	6.98	0.0015	3.05	0.0163
Stones	0.35	7.82	0.38	0.7711	3.88	0.001
Thickening	0.51	9.03	0.20	0.8984	3.03	0.0167
Vascular	0.42	7.61	1.35	0.2828	1.66	0.1594
VSD	0.70	7.31	1.74	0.3426	8.69	0.001

The sample mean of confidence for the entire model irrespective of the bandwidth or the image was estimated as 7.84.

The R^2 value, which measures the variation in the average confidence that can be accounted for by the model, was also estimated to be 23.9% indicating that 76.1% of the variation in confidence levels is unaccounted for by the model. Thus, the results obtained by modeling the entire data irrespective of the image and bandwidth produced an insufficient explanation for the variations in average confidence. However, while comparing the F values of individual elements in the model such as bandwidth, observer and image, the results indicate that the observer mean squares account for 78.71% of 23.9% of the total variations in the model while the mean square values for the image and the bandwidth factors account for only 13.72% and 7.5% respectively. In general, the higher the F value, the stronger the relationship is between the variables (image, bandwidth and observer) and the average confidence. Therefore by observing the F values and the P values in the model, we can conclude that by knowing the observer beforehand, the average confidence can be better predicted than by knowing the image or the bandwidth. Hence, most of the variation in confidence in the model can be attributed to the raters and not the different images or bandwidths.

Table 3-4 shows individual estimations of average confidence for the model parameter bandwidth while keeping the other parameters, image and observers, constant. The results from table 3-4 indicated that the average confidence least square means at 128 kbps were significantly lower than the least square mean values of average confidence at other bandwidths. The least square means for average confidence at all the other bandwidths were similar and indicated negligible variation. The P values in the table gives a pair wise comparison of any 2 bandwidths. These P values are the values estimated to be the differences seen by chance alone and is used to

reject the notion that the difference in the confidence scores are seen by chance alone. The smaller values of $Pr > |T|$ (student's t value for testing the null hypothesis) for the confidence at 128 kbps indicated that the mean confidence at this bandwidth is significantly lower than at all other bandwidths. From table 3-4, the pair-wise comparisons for all the other bandwidths indicate P values greater than 5%. Therefore we accept the null hypothesis that there is no difference in average confidence at bandwidths greater than 256 kbps. Table 3-4 conclusively proves that average confidence only at 128 kbps varies significantly from the other levels.

Table 3-5 indicated that while comparing the effects of the different images on the observer confidence, the average confidence scores dropped significantly for the study of the fetus at 16 weeks. Table 3-6 shows the effects of individual observers on the average confidence. The results from this table indicated that two of the nine raters (raters 3 and 8) were inherently conservative with their scores

Table 3-7 is a summary of the results of the GLM procedure obtained by modeling average confidence for each study aspect. Since the R^2 values for the entire model accounted for only 23% of the variations, each study component was modeled separately by using the multivariate analysis of variance in GLM.

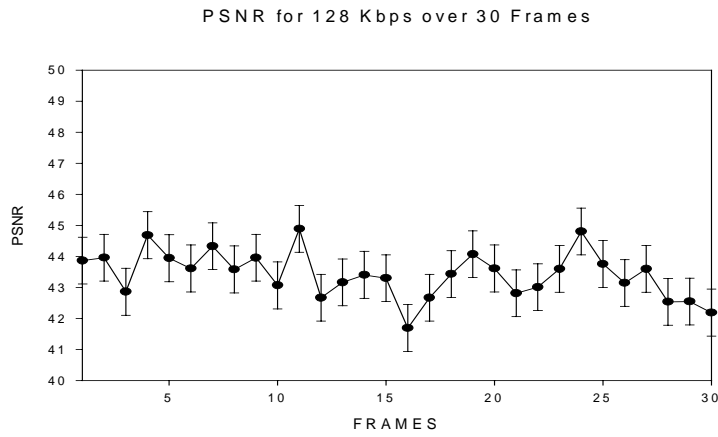
The R^2 values for most studies indicated that by modeling the individual study components, the variations in the studies are well accounted for. The average confidence scores for the studies in the model varied from a low of 5.22 on a scale of 10 for nuchal thickening and the highest confidence scores were 9.59 for pericholycystic fluid detection. These scores were obtained by modeling average confidence as the dependent variable keeping the bandwidth, raters as the independent variables for each study component. In the model, while observing the P values and F values for bandwidth and observers in different studies, it was found that in most

cases the F values for bandwidth were lower and in some cases significantly lower than the F values of observers. This again indicates that in most cases, that the consequence of variations in bandwidth on the average confidence is lower or in some cases significantly lower than the effect of variations in observer. Also, by modeling for individual components, the differences in F values ratios were significantly lower than by modeling for the entire data set.

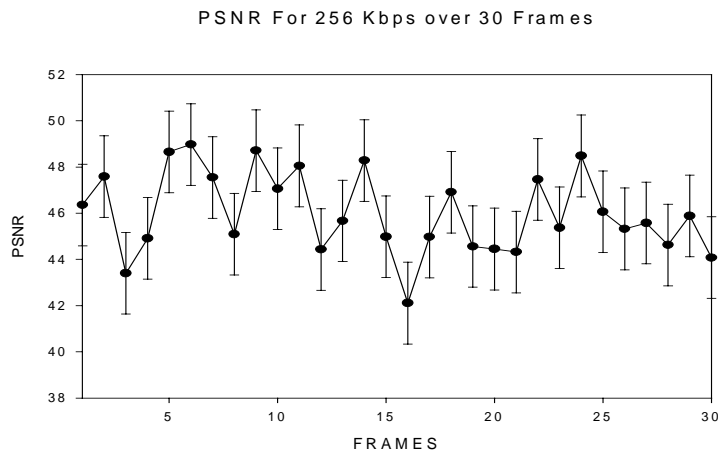
From table 3-7 by observing the P values of bandwidth we can also deduce that for most studies we can accept the null hypothesis that the average confidence does not significantly vary across the bandwidths. However the P values of bandwidth on studies such as AV valves, cardiac rhythm, fibroids, normal parenchyma of the liver, portal vein branches, sludge and simple cysts, indicate that we would reject the null hypothesis. This implies that in fetal heart studies and liver studies, bandwidth does have a significant effect on the observer confidence. The fact that bandwidth overall did not predict confidence does not rule there being one bandwidth with either higher or lower confidence.

3.5 Objective Testing

Figures 3-11(a) to 3-11(f) are the estimated PSNRs at each compression level for 30 frames with the standard errors in the plots indicating the standard deviations. Figures 3-12 to 3-16 are the difference images and the corresponding histograms obtained from the difference images at various compression levels. Table 3-8 summarizes the mean and standard deviation values obtained for the histogram plots at various levels of compression. The results for the objective testing were obtained using the methods described in the flow diagram for image handling in figure 2-3.

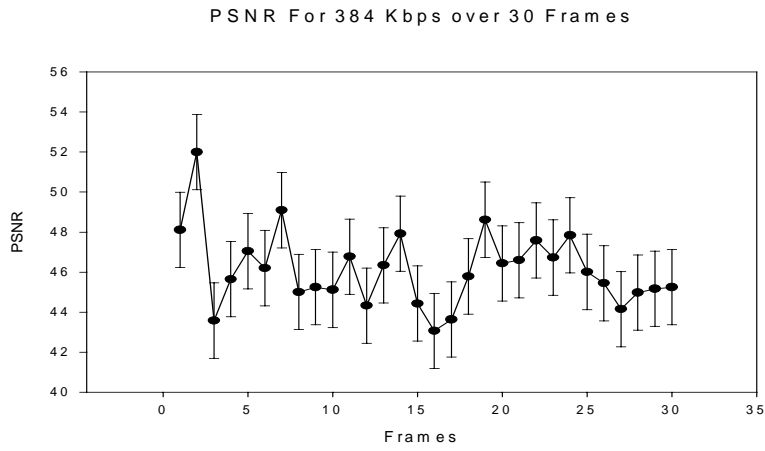


(a)

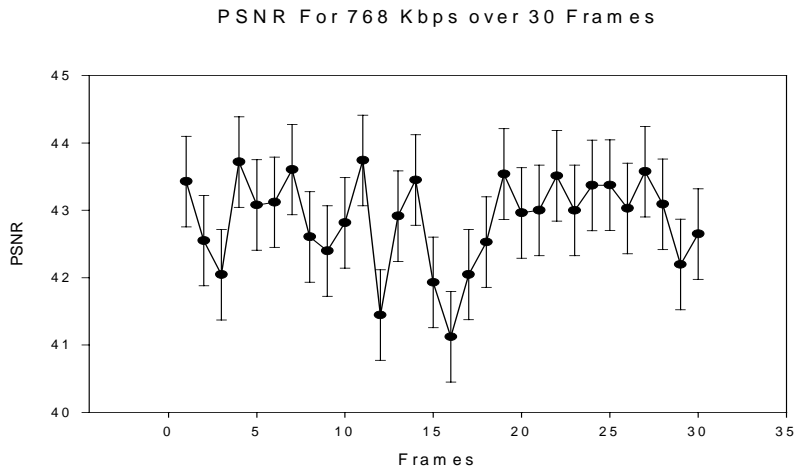


(b)

Figure 3-11. The PSNR values at various compression levels.
 (a) The PSNR values for 256 kbps for the same 30 frames. (b) The PSNR values for 256 kbps for the same 30 frames

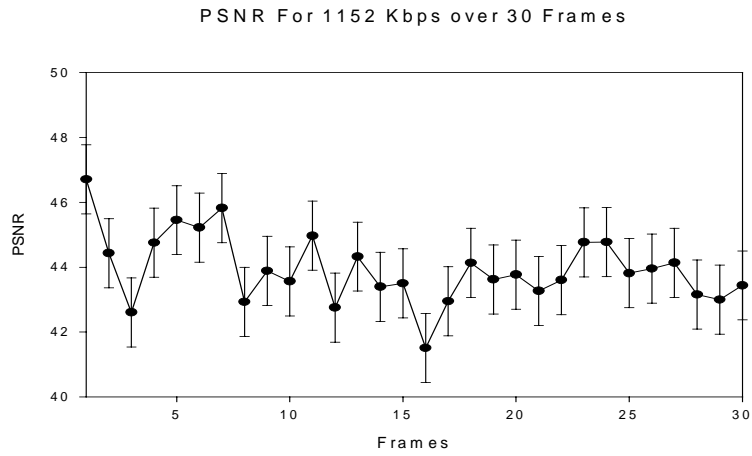


(c)

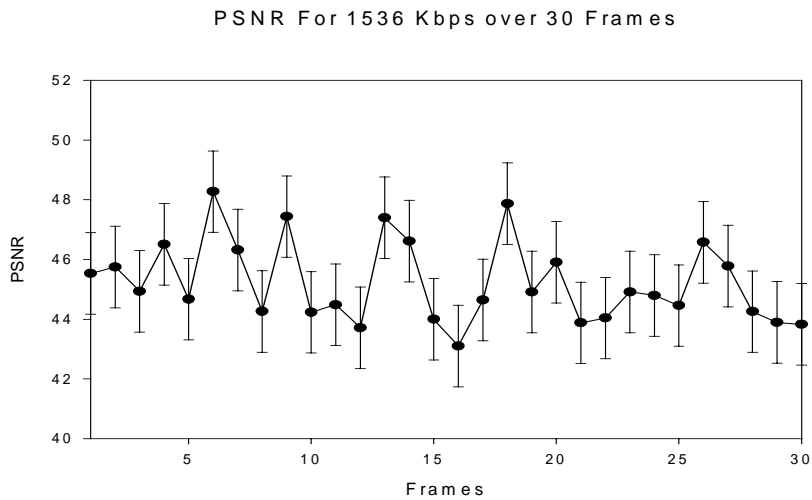


(d)

Figure 3-11. The PSNR values at various compression levels.
 (c) The PSNR values for 384 kbps for the same 30 frames. (d) The PSNR values for 768 kbps for the same 30 frames

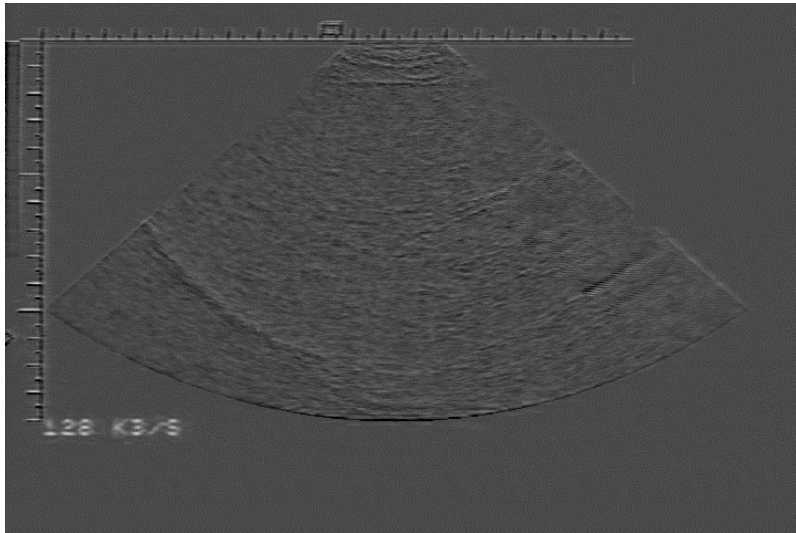


(e)



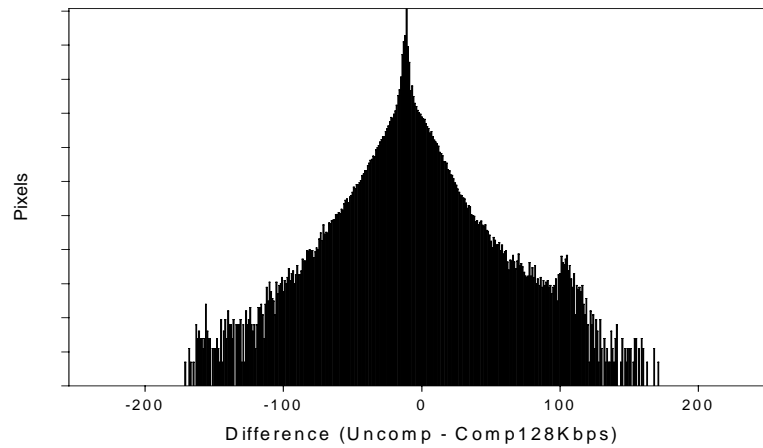
(f)

Figure 3-11. The PSNR values at various compression levels.
 (e) The PSNR values for 1152 kbps for the same 30 frames. (f) The PSNR values for 1536 kbps for the same 30 frames



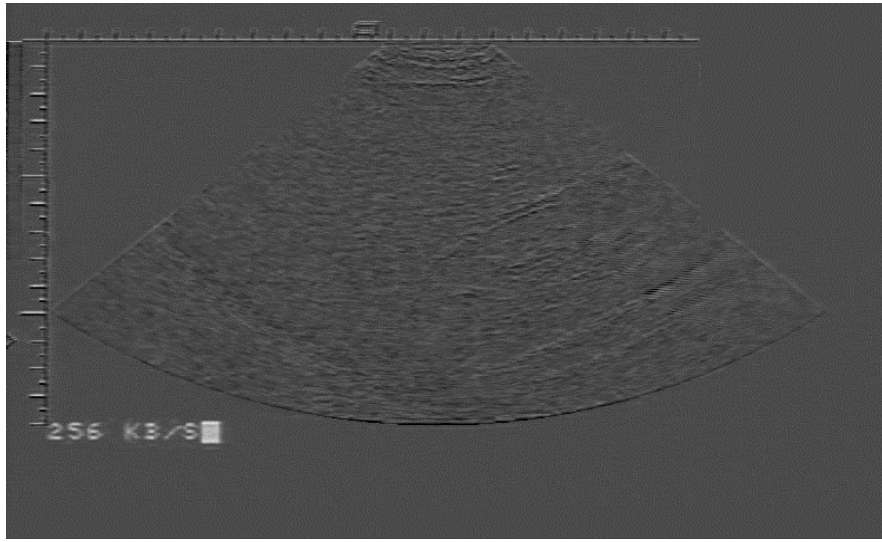
(a)

Histogram of the Difference Image at 128 Kbps - Log Scale



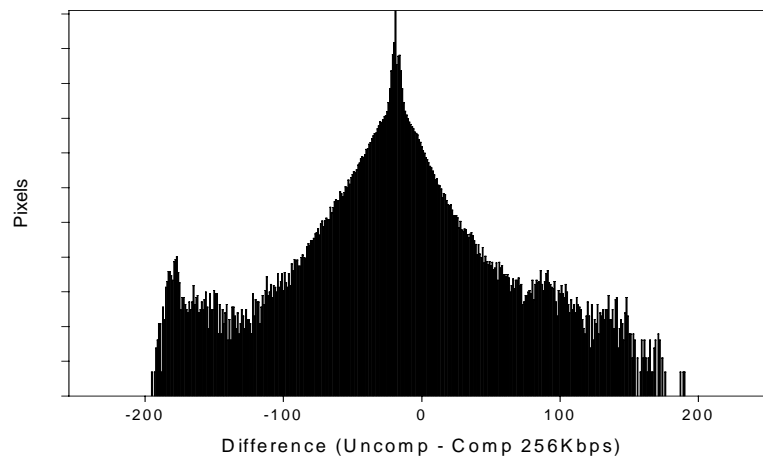
(b)

Figure 3-12. Difference image and the corresponding histogram plot at 128 kbps.
(a) Difference image at 128 kbps (b) Histogram of the difference image



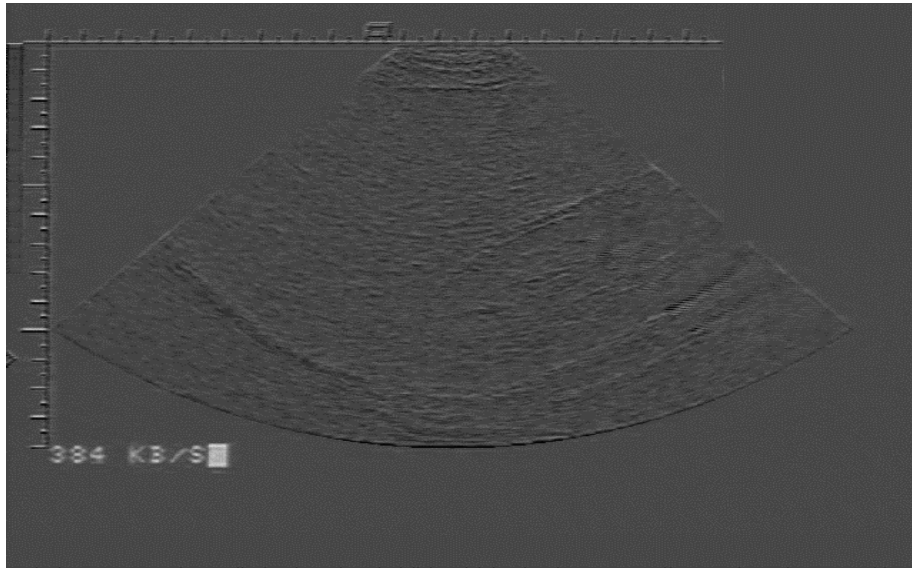
(a)

Histogram of the Difference Image at 256Kbps -Log Scale



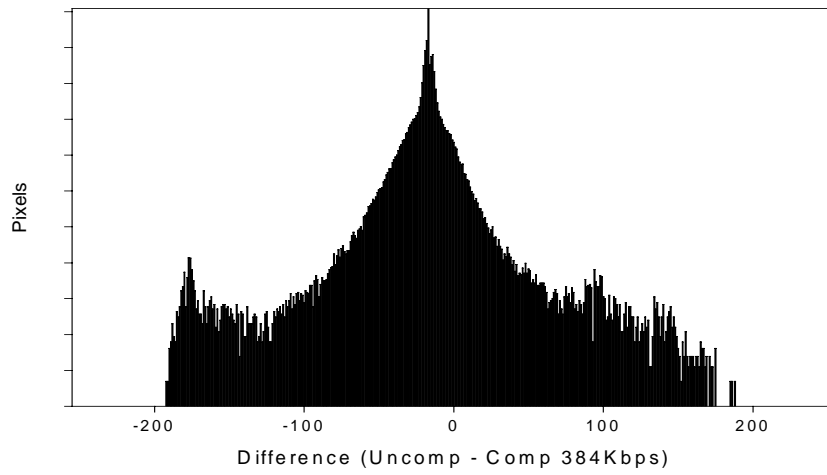
(b)

Figure 3-13. Difference image and the corresponding histogram plot at 256 kbps.
(a) Difference image at 256 kbps (b) Histogram of the difference image



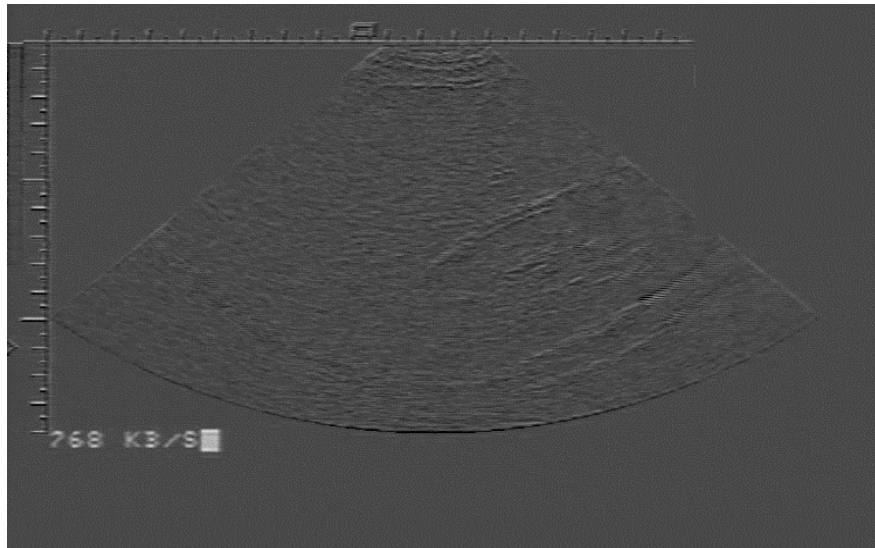
(a)

Histogram Of the Difference Image at 384Kbps- Log Scale



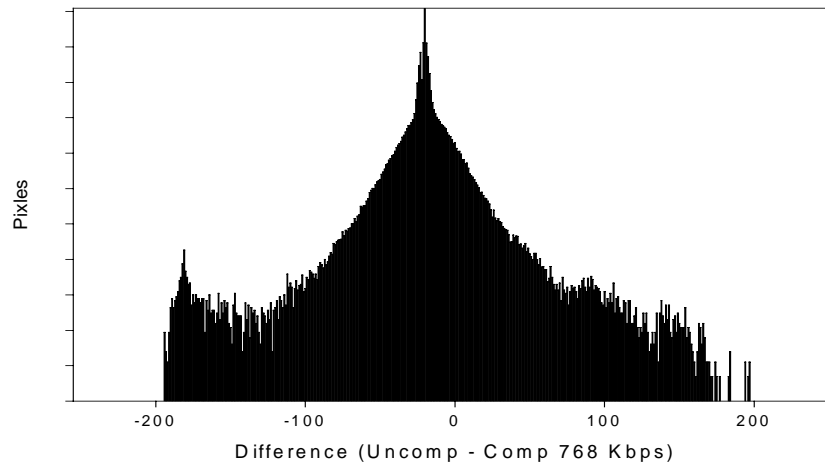
(b)

Figure 3-14. Difference image and the corresponding histogram plot at 384 kbps. (a) Difference image at 384 kbps (b) Histogram of the difference image



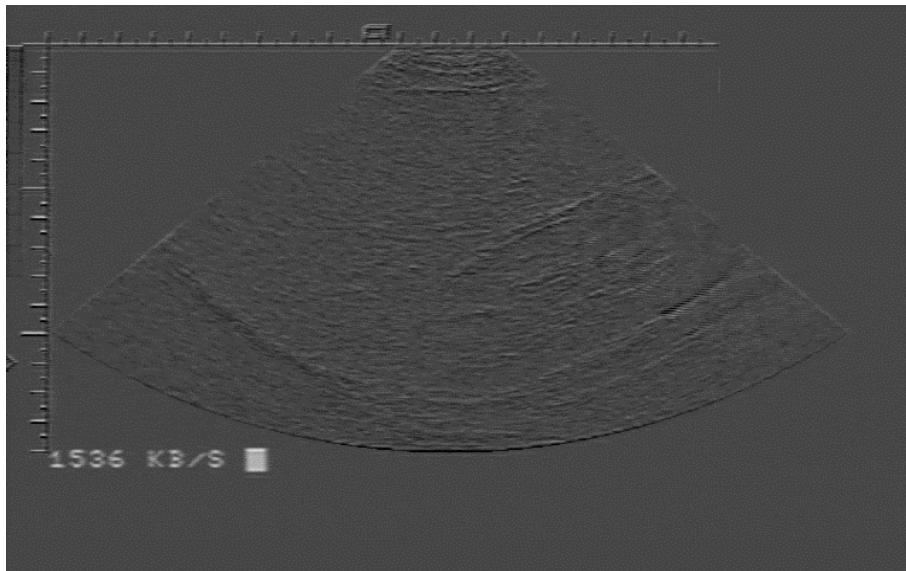
(a)

Histogram of the Difference Image at 768 Kbps - Log Scale



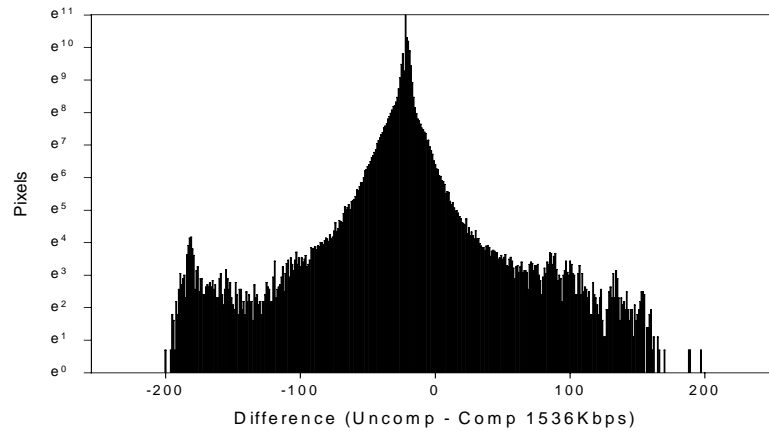
(b)

Figure 3-15. Difference image and the corresponding histogram plot at 768 kbps. (a) Difference image at 768 kbps (b) Histogram of the difference image



(a)

Histogram of the Difference Image at 1536 Kbps- Log Scale



(b)

Figure 3-16. Difference image and the corresponding histogram plot at 1536 kbps.
(a) Difference image at 1536 kbps (b) Histogram of the difference image

Table 3-8. Mean and standard deviation values from the histograms of the difference images at various compression levels.

Compression Levels	Mean Values	Standard Deviation
128 kbps	116.22	9.13
256 kbps	109.73	9.12
384 kbps	108.50	8.87
768 kbps	108.07	8.84
1152 kbps	106.65	8.88
1536 kbps	105.52	8.83

While comparing the uncompressed and corresponding compressed images at various bandwidths, we found that there was a significant change in the gray scale range between them. A reason for the change in gray scale could be the electronics present in the hardware of the CODEC. Although the reason for the shift in gray scale is not clear at this juncture, there was a visibly significant difference in the gray scale ranges between the compressed and uncompressed images, which affected the objective analysis.

In the objective testing, since the gray scale ranges were not carefully matched before performing the PSNR estimations and the histogram analysis, the difference images obtained showed only the changes in the gray scale between the images and not any differences due to compression. Therefore the entire objective testing was insensitive to the effects of compression. Ideally if the gray scale ranges were carefully matched, we could expect a monotonic increase in PSNR values with decreasing compression ratios. Due to the changes in gray scale in the data collected we could not establish an increasing trend in PSNR values with decreasing compression ratios.

At this juncture of the research, we are unable to tell if the observed changes in gray scale are uniform over all the frames and compression ratios. Extensive correction of the data needs to be done before the objective methods can be actually used.

3.6 Clinical Impressions

At 368 kbps, the abdominal studies of kidneys, gall bladder and liver showed similar image degradation. The structures in the images were blurry causing difficulties in distinguishing the various shades of gray in the image. The images also were coarse and appeared to be grainy, posing difficulties in the diagnosis. The grainy image formations affected mostly the diagnosis of the liver parenchyma. At this compression level, the artifacts for the abdominal studies were

more accentuated causing mild shadowing in the liver due to some loss of signal. The kidneys and gall bladder studies at 768 kbps showed progressively increasing confidence scores among the raters. There was an improvement in spatial and temporal resolution at this bandwidth. The images of gall bladder and kidney at 1536 kbps were of diagnostic quality but the effects of compression were still clearly visible. Although the images displayed blurring artifacts and increased contrast in the parenchyma, the images were diagnostically acceptable. The liver studies at 1536 kbps displayed heterogeneous pattern and echogenic masses. The vessels were well defined in the liver and the anatomy was clearly seen even in the presence of noise. For the same compression level in an obese patient study, the visualization of the internal structures was difficult. The intrahepatic biliary dilatations were also not visible.

The studies at 384 kbps for the uterus and female pelvis indicated poor image quality leading to difficulty in clinical evaluation of the pathological states. The real areas (as opposed to artifacts such as shadowing) in the images were not clear and it was difficult to differentiate the tissues. The fibroids in the uterus were not clearly visible at this compression level.

The fetal heart studies at 128 kbps were severely affected by compression. This compression induced image degradation led to false negatives in the studies and to the misconception of some of the raters that the fetus was dead when the fetus was healthy. There was also severe loss in temporal resolution making detection of the heart rate and rhythm difficult. The images displayed improper high contrast and this excessive contrast lead to loss in information about the septal walls. From the diagnostic scores and clinical impressions, it is clear that a fetal heart study at 128 kbps cannot be used for clinical diagnosis. The study of the fetus at 256 kbps also presented temporal resolution discrepancies and spatial blur. The fetal hear studies at 384 kbps showed better temporal resolution and spatial resolution. However the increased

contrast was still persistent at this compression ratio making the visualization of tricuspid valves difficult. The studies for fetal heart at 1152 kbps showed excellent quality images in terms of temporal resolution yielding better visualization and diagnosis of heart rhythm. Although the raters reported a loss in spatial resolution to a certain degree, the image was still diagnostically acceptable.

The IUP images at 768 kbps showed improper artifacts. The raters had difficulty in the determination of the heartbeat in the embryo. The radiologists also reported that motion of the sac appeared jerky and irregular. The texture of the tissue also appeared coarser in the compressed image. The confidence and kappa scores across all of the raters for the IUP indicated the images were unacceptable. Further studies needs to be done to prove the significantly lower kappa scores observed only in IUP studies.

Chapter 4

Future Work

4.1 Subjective Testing

The results obtained indicated good inter-rater reliability. However, another important aspect in this study is the measurement of intra-rater reliability. Intra-rater variability checks for the consistency of confidence values by the same observer in the second and third reading of the same image. By measuring the rate of disagreement between one reading and the next, an estimate of intra-rater variability can be made. The same data set should be used in the evaluation after an interval of time. If the results obtained are similar, the inter-rater reliability measurement values will yield more meaningful interpretations.

The values of kappa for certain studies such as the intrauterine pregnancy indicated low inter-rater reliability values. Further validation of the obtained results should be done by incorporating more images at various stages in the study.

In this research the images were evaluated primarily for diagnostic quality. Our initial expectation for the research was that there would be a considerable loss in observer confidence for these highly compressed video segments. However, on evaluation, we concluded that the diagnostic ability of sonologists was not hampered by the poor quality of the images. The first impressions from the raters suggested that the image quality was degraded and was not aesthetically acceptable. However, through this study we could not show that this poor aesthetic quality affected diagnostic ability. The fact that the bottom line diagnosis was not affected has to be verified by other concomitant methods such as an ROC study.

Receiver Operating Characteristic (ROC) analysis is one of the dominant techniques used for evaluating the suitability of radiological techniques. In order to fashion ROCs, the same study

should be evaluated for detection of anatomy and pathologies at various levels of compression. The ROC technique requires a large data set of true positives and true negatives for evaluation. A description of the ROC technique is given in section 2.5.1. We suggest that the ROC be done for fetal studies, as they are often considered the most difficult for the clinical diagnosis of pathological states. Separate studies to evaluate the image quality could be incorporated into the ROC study by asking the raters to evaluate the images for image quality on a scale of 1 to 10 and estimating the mean values for each level of compression.

4.2 Objective Testing

The results obtained from the objective testing methods, such as histogram analysis and PSNR estimations, were inaccurate and unacceptable because of the changes in the gray scale range between the uncompressed and compressed images. The changes in the gray scale range were introduced by the hardware of the CODEC. The first step to correcting the change in gray scale would be to find out if the gray scale shift is uniform over a whole video frame. If the transformation is found to be uniform across each frame, careful matching of the gray scale range between the compressed and uncompressed image frames has to be done to make the images suitable for objective testing. One possible technique for the gray scale matching would be to use a transformation function that represents the shift in gray scale between the compressed and uncompressed image frames. This transformation can be calculated by performing a least squares fit between the two image frames to derive a polynomial equation. This polynomial function would represent the transformation and could be applied to a compressed image frame to match the gray scale range of the corresponding uncompressed image frame.

Another subjective technique, which can be used for objective evaluation, is the verification of the rate-distortion function for motion compensated intra-frame coding techniques such as H.261. The rate-distortion function defines the lowest limit for achievable compression rate (R) for a specified distortion (D). By specifying the value of distortion a theoretical value of rate can be estimated from the rate-distortion function. The resulting $R(D)$ can be used to determine if a desired compression ratio suitable for diagnostic purpose is achievable in practice.

Conclusion

The qualitative evaluation of data compression in real-time ultrasound imaging for multiple anatomical and pathological states using the H.261 algorithm was discussed. A protocol for evaluation was developed and subjective and objective methods were used for testing the quality of the image.

The results indicated good intra-rater reliability and confidence scores and indicated that compressed video stream with bandwidths up to 1536 kbps to fit the T1 and ADSL networks and could be transmitted with little or no degradation in diagnostic confidence. However alternate methods of evaluation such as ROC studies would have to be completed to verify the results obtained. These studies should use the fetal studies as the clinical gold standard as they require very good spatial and temporal resolution for clinical diagnosis.

Newer improved compression algorithms such as wavelets and MPEG 2 could be tested to ascertain if they are a more viable choice to H.261

List of References

List of References

1. Rumack CM. *Diagnostic ultrasound*. Mosby Publications, 1998.
2. Chevenak FA. *Ultrasound in obstetrics and gynecology*. Little Brown and Company, 1993.
3. Sanders RC. *The principles and practice of ultrasonography in obstetrics and gynecology*. Appleton-Century Raven, 1995.
4. Kremkau FW. *Diagnostic ultrasound: principles, instruments and exercises*. WB Saunders, 1989.
5. McGahan JP. *Diagnostic ultrasound: a logical approach*. Lippincott-Raven, 1998.
6. Emerson DS, Felker RE. *Remote real-time ultrasound interactive teleradiology: putting it into practice*. J Ambulatory Care Management, vol. 18, pp.20-34, 1995.
7. Batnitzky S. *Teleradiology: an assessment*. Radiology, vol. 177, pp.11-17, 1990.
8. Spicer KM. *Hospital-to-home teleradiology*. Imaging Economics, May, pp.6-7, 1997.
9. Lee JK, Renner JB, Saunders BF, Stamford PP, Bickford TR, Johnston RE, Hsaio HS, Philips ML. *Effects of real-time teleradiology on the practice of the emergency department physician in a rural setting: initial experience*. Acad. Radiology, vol. 5, pp. 533-538, 1998.
10. Bieze J. *Tech assessment adapts to image communication*. Diagnostic Imaging, September, pp. 71-75, 1995.
11. Goldberg MA. *New high resolution teleradiology system: prospective study of diagnostic accuracy in 685 transmitted clinical cases*. Radiology, vol. 186, pp.429-434, 1993.
12. Datkins DR. *Cost, bandwidth, volume: telemedicine by numbers*. Telemedicine, July, vol. 23, pp.29-30, 1996.
13. Berger SB, Cepelwicz BB. *Medical-legal issues in teleradiology*. Am J Roentgenology, vol. 166, pp. 505-510, 1996.
14. Huang HK. *PACS: basic principles and applications*. Wiley-Liss, 1999.
15. Honeyman JC, Huda W, Palmer CK, Staab EV. *Picture archiving and communication system bandwidth and storage requirements*. J. Digital Imaging, vol. 9, pp. 60-66, 1996.

16. Stewart BK, Lou SL, Wong WK, Huang HK. *An ultra-fast network for communication of radiologic images*. Am J Roentgenology, vol. 156, pp.835-839, 1991.
17. Honeyman JC, Frost MM, Huda W, Ott M, Stabb EV. *Picture archiving and communication systems*, Curr. Probl. Diagn. Radiology, vol. 23, pp. 101-158, 1994.
18. Shannon CE. *A mathematical theory of communications*. Bell Systems Technical Journal, pp. 379-423, 1948.
19. Huffman DA. *A method for constructing minimum redundancy codes*. Proceedings of the IRE, vol. 40, pp. 1098-1101, 1951.
20. Sayood K. *Introduction to data compression*. Morgan Kaufmann Publishers, 1996.
21. Bhaskaran V. *Image and video compression standard: algorithms and architectures*. Kluwer Publishers, 1997.
22. International Telecom Union. *Line transmission of non-telephone signals. Video CODEC for audio-visual services at P x 64 kbps*. ITU-T Recommendation H.261, March, 1993.
23. Cook LT, Insana MF, McFadden MA, Hall TJ, Cox GG. *Contrast detail analysis of image degradation due to lossy compression*. Med. Phys., vol. 22, pp. 715-721, 1995.
24. Cahill PT, Vullo T, Hu JH, Wang Y, Manzo R, Weingarten K, Markisz JA. *Radiologist evaluation of a multi-spectral image compression algorithm for magnetic resonance images*. J Digit. Imaging, vol. 11, pp. 126-136, 1998.
25. Wenzel A, Gotfredsen E, Borg E, Grondahl HG. *Impact of lossy image compression on the accuracy of caries detection in digital images taken with a storage phosphor system*. Oral Surg. Oral Pathol. Oral Radiolog., vol. 81, pp. 351-355, 1996.
26. Permuter SM. *Image quality in lossy compressed digital mammograms*. Signal Processing, Special Section on Medical Image Compression, vol. 59, pp. 189-210, 1997.
27. Saunders BD. *Basic and clinical biostatistics*. Appleton-Lange, 1994.
28. Fleiss JL. *Statistical methods for rates and proportions*. John Wiley and Sons, 1981.
29. Landis JR. *The measurement of observer agreement for categorical data*. Biometrics, vol. 33, pp. 159-174, 1977.
30. Allen AO. *Probability statistics and queuing theory*. Second Edition, Academic Press, 1990.
31. Williams B. *Biostatistics concepts and applications for biologists*. Chapman and Hall Publishers, 1993.

32. Ophir J, Maklad NF. *Digital scan converters in diagnostic ultrasound imaging*. IEEE Proceedings, vol.67, pp. 654-664, 1976.

Vita

Bijoy J. Sundersingh was born in TamilNadu, India on December 2nd 1974. He graduated from Madras Christian College Higher Secondary School in April 1992. The following August he entered Manipal Institute of Technology, India and received the degree of Bachelor of Engineering in Biomedical Engineering in 1996. In the fall of 1997 he entered The University of Tennessee, Memphis for the Master of Science program. He is currently employed as a software design engineer for Cardiovascular Computer Systems, L-3 Communications Corporation.