2. MEDICAL IMAGE QUALITY

The goal in the production of medical images is to acquire and display pictures of the human anatomy in a manner such that the diagnostic information contained in that representation may be optimally transferred to the trained reader. The concept of image quality pertains to how well that goal is accomplished. However, such a broad definition resists a quantitative determination. This chapter examines this elusive notion of image quality and presents some approaches to its measurement. The chapter concludes with a description of measures based on models of human vision, the approach ultimately pursued in this research.

Medical images can be essential tools in determining a medical diagnosis and treatment. In the medical image acquisition and display system, there are many parameters specific to the various imaging modalities that have an impact on the observer's ability to make that diagnosis accurately. In the development and evaluation of imaging systems, one always wishes to adjust the relevant system parameters to optimize observer performance.[†] Likewise, as image acquisition, processing, and display devices and methods are developed, an assessment regarding the efficacy of the potential innovations is often desired. Both of these kinds of determinations are based upon an assessment of the "quality" of the images that those systems produce.

2.1 Image Quality

It is readily accepted that image quality possesses a functional definition: the determination of quality should be conducted by measuring the performance of an observer for a diagnostic task. The outcome from the interpretation of a medical image is a clinical diagnosis, and the accuracy of that inference about the status of the human anatomy is imperative for successful patient treatment. Therefore, the ultimate measure of the quality of an image is the accuracy with which a trained observer determines a diagnosis. Determining that diagnosis may consist of one or more very specific visual tasks. So measuring image quality must be performed with respect to a particular visual task of interest, and achieving superior image quality is then a matter of optimizing accuracy for that task.

The figure of merit in expressing the quality of an image is frequently a signal-to-noise ratio. This measure provides an indication of the strength of the signal or effect enabled by the image manipulation of interest. For the task of detecting an object in an image, the signal-to-noise ratio provides a measure of the discriminability of the object from its background and might be expected to change as, for instance, different contrast enhancement methods were applied to the image. Similarly, for the task of estimating the volume of an object, a signal-to-noise ratio can provide a measure of the magnitude of the effect adjusted by the variability in the observed estimates. The signal-to-noise ratio as an image quality outcome is an informative indication of the extent and consistency with which the diagnostic truth is determined.

Unfortunately, all too often medical images are evaluated by preference studies in which an albeit trained observer chooses a superior image based upon his/her aesthetic opinion. Because such judgments inevitably fail to determine the goodness of an image with respect to a particular task, basic subjective preference studies are often not predictive of performance in the clinical setting.[1,2] In rating their preference for different image manipulations, observers may be influenced by aspects of the images that are not related to their ability to actually utilize the images in the diagnosis, particularly for images that are somehow different from those encountered in standard clinical practice. Similarly, an objective evaluation by humans of physical quantities of the imaging system, such as spatial resolution or contrast, that might be performed with test images ("phantoms") containing geometric objects,[3,4] are often insufficient for predicting diagnostic performance. Different phantom assessment

---

[†] Sometimes the way that parameters of the imaging system are adjusted is dictated in part by considerations of patient safety. There are, for instance, trade-offs between the radiographic contrast achieved and the radiation dose delivered to the patient. These issues however are not factors for image quality *per se* but considerations that contribute to the overall development of the system.

[1] H.L. Kundel, "Perception and Representation of Medical Images," SPIE: Image Processing 1898 (1993): 2-12.

[2] C.R. Furman, D. Gur, B. Good, H. Rockette, L.A. Cooperstein, J.H. Feist, "Storage Phosphor Radiographs vs Conventional Films: Interpreters' Perceptions of Diagnostic Quality," AJR 150 (1988): 1011-1014.

[3] G.W. Eklund, G. Cardenosa, W. Parsons, "Assessing Adequacy of Mammographic Image Quality," Radiology 190 (1994): 297-307.

[4] L.E. Reinstein, L. Alquist, H.I. Amois, B. Lagueux, "Quantitative Evaluation of a Portal Film Contrast Enhancement Technique," Medical Physics 14, no. 3 (1987): 309-313.

tasks have been shown to exhibit poor correlation with each other and high inter-observer variability.[5] More importantly, the visual tasks performed in diagnosis may be qualitatively different from those required in rating images of phantoms. The methods utilizing human observers for evaluating medical image quality that will be discussed in the later sections of this chapter are much more rigorous and insistent upon the measurement of performance for an important clinical task.

Because human observer experiments measuring performance for a specific task require so much time, computed methods have been pursued as attractive substitutes for evaluating quality. Figures of merit formulated from physical characteristics of the image, including but not limited to the modulation transfer function,[6,7] a signal-to-noise ratio of image intensities,[8] the information spectrum,[9] as well as indices derived from these,[10] are easily computed from the image and attempt to capture the basic physical constituents of low-level human vision. Some of them may even incorporate known characteristics of the contrast or resolution response of the visual system.[11] However, the measurements, or combinations of them, characterizing these methods often lack an intuitive, and moreover statistical, relationship to human visual interpretation for all but the simplest tasks.[12,13]

It is important for computed measures to incorporate some understanding of the tasks required by the evaluation. Nuclear medicine scans, for example, possess notoriously poor signal-to-noise ratio and resolution, yet they are unparalleled in providing a visual representation for certain interpretive tasks related to physiological function. Noise and resolution calculations would fail to capture the usefulness of the nuclear medicine modality for such purposes. The computed methods described in later sections attempt to incorporate an emphasis on determination with respect to a defined task.

2.2 Tasks

The tasks that might be performed with a medical image are subsumed by two categories: classification and estimation. Classification is the task of assigning an observation to two or more non-overlapping categories. In particular, detection, or the assignment to two classes, is often performed in medical image viewing. Radiologists continually determine the presence or absence of anomalies such as tumors or mammographic calcifications, or classify abnormalities with dichotomous labels such as normal or abnormal, and benign or malignant. Sometimes these tasks are labeled as discrimination tasks: an observer might determine whether the shape or size of an object or anatomical feature deviates sufficiently from its normal manifestation to warrant an alternative diagnosis.

Alternatively, interpretation may involve quantitative determinations whose outcomes span a continuous scale. These estimation tasks include judgments about absolute or relative intensities of, within, or between anatomical structures in the image. Also, observers may estimate the shape or interrelationship of objects; they may be required to characterize the curvature, width, or area of objects, or estimate the distance between two structures. Determinations of the amount of radioactive tracer in the organs of a nuclear medicine image, or the amount of blood ejected by the heart revealed by frames from an angiographic sequence, are typical estimates.

Methods for evaluating quality with respect to classification and estimation are described in the following sections, with emphasis on the approaches for estimation. Classification tasks, because of their abundance in image viewing and their relative ease of methodological analysis, have been almost exclusively examined in image quality measurements. However, it is frequently the case that estimation tasks are integral in the utilization of many imaging modalities. It is in fact the case that the imaging modalities investigated in this research may

---

[5] C.B. Caldwell, E.K. Fishell, R.A. Jong, W.J. Weiser, M.J. Yaffe, “Evaluation of Mammographic Image Quality: Pilot Study Comparing Five Methods,” AJR 159 (1992): 295-301.

[6] C.R. Carlson, R.W. Cohen, “A Simple Psycho-physical Model for Predicting the Visibility of Displayed Information,” Proceedings of the Society of Information Display 21 (1980): 229-246.

[7] P.G.J. Barten, “Evaluation of Subjective Image Quality With the Square-root Integral Method,” Journal of the Optical Society of America A 7, no.10 (1990): 2024-2031.

[8] R.M. Nishikawa, M.J. Yaffe, “Signal-to-noise Properties of Mammographic Film-screen Systems,” Medical Physics 12 (1985): 32-9.

[9] J.C. Dainty, R. Shaw, Image Science (London: Academic Press, 1974).

[10] L. Desponds, C. Depeursinge, M. Grecescu, C. Hessler, A. Samiri, J.F. Valley, “Image Quality Index (IQI) for Screen-film Mammography,” Physics in Medicine and Biology 36, no. 1 (1991): 19-33.

[11] J.A. Saghri, P.S. Cheatham, A. Habibi, “Image Quality Measure Based on a Human Visual System Model,” Optical Engineering 28, no. 7 (1989): 813-818.

[12] L.D. Loo, K. Doi, C.E. Metz, “A Comparison of Physical Image Quality Indices and Observer Performance in the Radiographic Detection of Nylon Beads,” Physics in Medicine and Biology 29, no. 7 (1984): 837-856.

[13] E. Buhr, C. Herrmann, D. Hoeschen, “Correlation Between Physical Image Quality Parameters and Visually Perceptible Image Quality in X-ray Diagnosis,” Journal of Photographic Science 41, no. 3 (1993): 90-92.

arguably be optimized with respect to the estimation tasks that were chosen for study. As Chapters 4 and 5 will demonstrate, the estimation tasks of vessel constriction characterization or interobject distance measurement that are explored in this research are fundamental to the consultations for which the angiography and portal imaging systems are designed.

2.3 Image Quality for Classification

A rigorous and universally-applied framework for conducting human observer experiments for the determination of image quality is the psychophysical methodology of ROC (receiver operating characteristic) analysis.[14] By measuring sensitivity and specificity as the decision criterion adopted by the observer varies, the methodology generates the ROC curve and accompanying signal-to-noise ratio. The curve fully characterizes observer response behavior for the binary decision of clinical detection tasks, and as such summary statistics of the curve may be used as measures in evaluating different image manipulations for their effect on that kind of task.

Alternatively, an n-alternative forced choice paradigm for the detection of simulated lesions or other abnormalities embedded in real anatomical backgrounds has been proposed.[15] Using this methodology, a psychometric function is obtained that plots percent correct versus structure contrast. Detection performance is described by the threshold and slope parameters derived from a probit description of the function. This approach has been used extensively in the evaluation of several contrast enhancement algorithms of potential efficacy in mammography.

Computed approaches to evaluating quality for classification tasks are based upon statistical discrimination decisions. The ideal observer,[16,17] a mathematical construct that utilizes complete statistical knowledge of a task, makes a decision about a binary variable by comparing a computed likelihood ratio with an established threshold. This approach results in performance that is optimal, because of the knowledge of the properties of the signal and background. For that reason, the results produced by the ideal observer are often poorly correlated with human performance in complex imaging tasks where such assumptions are inappropriate. The Hotelling observer[18] and channelized Hotelling observer[19] are attempts to incorporate some of the limitations of the human observer so that they might correlate more closely with human performance. The Hotelling model produces a test statistic that is a linear approximation to the nonlinear ideal observer test statistic when assumptions of object and background variability are relaxed. Results from a Hotelling model application have been demonstrated to correlate well for detection in nuclear medicine images.[20]

2.4 Image Quality for Estimation

The statistical methods of analysis of variance (ANOVA) and multiple regression commonly underlie the experimental designs and their accompanying analyses for image quality studies of estimation tasks. These methods estimate the means and variances of the responses for an estimation task for two or more experimental conditions in order to make statistical judgments about the effect of those conditions on the performance of the task. The methods might be used to determine whether observers performed a task significantly better with imaging system A as opposed to system B. Alternatively, one might wish to make statistical conclusions about the effect on observer performance of adjusting system or processing parameter C along several levels.

Both techniques, ANOVA and regression, may be approached from the generalized statistical framework of the General Linear Model (GLM).[21] A particular experimental observation is said to arise from the sum of both intentionally-manipulated and random effects. The GLM expresses the observation as a linear combination of the controlled factors plus a term that reflects all other randomly-varying, error terms. An alternative to this

---

[14] C.E. Metz, "Basic Principles of ROC Analysis," Seminars in Nuclear Medicine 8, no. 4 (1978): 283-298.

[15] D.T. Puff, E.D. Pisano, K.E. Muller, R.E. Johnston, B.M. Hemminger, C.A. Burbeck, R. McLelland, S.M. Pizer, "A Method for Determination of Optimal Image Enhancement for the Detection of Mammographic Abnormalities," Journal of Digital Imaging 7, no. 4 (1994): 161-171.

[16] D.B. Green, J.A. Swets, Signal Detection Theory and Psychophysics (New York: Wiley & Sons, 1966).

[17] A.E. Burgess, R.F. Wagner, R.J. Jennings, H.B. Barlow, "Efficiency of Human Visual Discrimination," Science 214 (1982): 93-94.

[18] H.H. Barrett, J. Yao, J.P. Rolland, K.J. Myers, "Model Observers for Assessment of Image Quality," Proceedings of National Academy of Science 90 (1993): 9758-9765.

[19] J. Yao, H.H. Barrett, "Predicting Human Performance by a Channelized Hotelling Observer Model," SPIE: Mathematical Methods in Medical Imaging 1768 (1992): 161-168.

[20] R.D. Fiete, H.H. Barrett, W.E. Smith, K.J. Myers, "The Hotelling Trace Criterion and its Correlation with Human Observer Performance," Journal of Optical Society of America A 4 (1987): 945-953.

[21] S.E. Maxwell, H.D. Delaney, Designing Experiments and Analyzing Data (Belmont, CA: Wadsworth, 1990).

"full model" description, called the "restricted model," can be developed as well that deletes some of the parameters from the full model. The restricted model embodies the null hypothesis, and statistical decisions are made by studying the increase in the error in representing the data due to adopting the simpler model. For instance, in deciding whether imaging systems A and B differ significantly, the null hypothesis might be that only a single parameter is required in the model to reflect that the data for the two systems are the same. The test then is whether that restricted model is as adequate a representation of the data as is the full model with its *separate* means for the scores for the two systems. Any increase in error caused by adopting the restricted model may be related in terms of a "p" value, a probability estimate that the difference arose by chance. The null hypothesis may be rejected when the p-value is less than a predetermined significance level.

ANOVA and multiple regression then are special cases of the GLM that attempt to develop linear models that accurately describe the data. ANOVA may be applied to experimental designs with a categorical variable, such as those for which there exist several discrete settings, or levels, of system parameter C, together with a continuous response, such as a distance or area measurement. Multiple regression must be applied when one or more of the independent variables in the study exist on a continuous scale.

The GLM analysis rests upon several important assumptions, and the failure to meet those criteria compromise its validity. Specifically, the assumptions are that 1) the scores on the dependent variable are normally distributed, 2) the population variances for the scores from all groups are identical, and 3) the scores are statistically independent. For several reasons, in experiments conducted with human subjects, the design is often "within-subjects," and two or more observations are collected from each subject. For example, an observer might view and generate responses for images from both systems A and B. First, in such "repeated measures" designs, trained subjects, who are difficult to recruit, may contribute many scores and in effect reduce the number of observers needed as participants. Secondly, error variance is reduced in such cases, because the variability resulting from individual differences is reduced. However, because subjects serve in different conditions, repeated measures causes the errors in the different conditions to be correlated. This occurrence violates one of the basic assumptions required for ANOVA, and thus requires the specialized techniques of univariate or multivariate analysis.[22]

Finally, within-subjects designs carry other considerations. The order of presentation of the conditions in a study is crucial: training effects, due to practice with images or other stimuli over the course of the experiment, and conversely fatigue effects, are inevitable in any human observer study. Similarly, "carryover" effects can occur, causing performance on later trials to be influenced by initial presentations. Thus the order of presentation must be carefully counterbalanced or randomized both within and between subjects.[23] The number of observers required to conclusively (with a desired level of probability) reach a statistical decision may be suggested by "power" calculations.[24]

Very few computed methods for image quality assessment with respect to an estimation task have been proposed. The ideal observer model is rarely applicable to complex, multiparameter estimation tasks since the estimate inevitably may not be expressed as a *linear* combination of the parameters and the pixel values. Consequently, estimation methods for determining a single quantity within a region of interest (ROI), which prove to be linear, have been developed. Alternatively, non-linear maximum likelihood estimators have been explored.

Barrett[25,†] outlines three linear ROI estimators: the integration of a quantity within an ROI, the optimal linear unbiased, or Gauss-Markov, estimator, and the Wiener estimator. In each case the methods seek an estimate, θ, that is a (possibly weighted) integral over a region of interest with template $\mathbf{w}$, within the original object $\mathbf{f}$ to be imaged:

$$\theta = \mathbf{w}^{\mathrm{t}}\mathbf{f} \qquad 2.1$$

The simple ROI estimation method integrates the acquired image values ($\hat{\mathbf{f}}$) in the region:

$$\theta_{\mathrm{ROI}} = \mathbf{w}^{\mathrm{t}}\hat{\mathbf{f}} \qquad 2.2$$

---

[22] D.F. Morrison, Multivariate Statistical Methods, 2nd ed. (New York: McGraw-Hill, 1976).

[23] K.E. Muller, C.N. Barton, V.A. Benignus, "Recommendations for Appropriate Statistical Practice in Toxicology," Neurotoxicology 5 (1984): 113-126.

[24] K.E. Muller, V.A. Benignus, "Increasing Scientific Power With Statistical Power," Neurotoxicology and Teratology 14 (1992): 211-219.

[25] H.H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," Journal of Optical Society of America A 7, no. 7 (1990): 1266-1278.

† The discussion and equations that follow are directly from the reference.

This estimator, although simple, is biased. However, that bias and mean-square error (a measure of variance that takes into account the bias) can be calculated. A figure of merit, the signal-to-noise ratio (SNR), for this estimator is shown in Equation 2.3.

$$[\mathrm{SNR}_{\mathrm{ROI}}]^2 = \frac{\mathrm{tr}[\mathbf{FW}]}{\mathrm{tr}[\mathbf{BW}] + \mathrm{tr}[\mathbf{K_m W}]} \qquad 2.3$$

In Equation 2.3, $\mathbf{F}$ is a measure of initial object strength given by $\langle \mathbf{ff}^{\mathrm{t}} \rangle_{\mathbf{f}}$, an average over $\mathbf{f}$. $\mathbf{W}$ is defined as $\mathbf{ww}^{\mathrm{t}}$. $\mathbf{B}$ is an average bias matrix, $\langle \mathbf{bb}^{\mathrm{t}} \rangle_{\mathbf{f}}$, where $\mathbf{b} = \langle \hat{\mathbf{f}} - \mathbf{f} \rangle_{\mathbf{n}|\mathbf{f}}$ is the bias, or the conditional average of the error in the estimate of object $\mathbf{f}$, $\hat{\mathbf{f}}$, over all realizations of the noise, $\mathbf{n}$, for a given object. $\mathbf{K_m}$ is a covariance matrix ($\mathbf{OK_n O}^{\mathrm{t}}$) obtained from the noise covariance matrix, $\mathbf{K_n} = \langle \mathbf{K}_{\mathbf{n}|\mathbf{f}} \rangle_{\mathbf{f}}$, and the operator $\mathbf{O}$ that produces the estimate $\hat{\mathbf{f}} = \mathbf{OHf} + \mathbf{On}$, where $\mathbf{H}$ is the imaging system operator.

The Gauss-Markov estimator is a linear unbiased estimator that produces minimum variance for its estimate for a particular appearance of the object $\mathbf{f}$. Equation 2.4 shows an expression for the Gauss-Markov estimate.

$$\theta_{\mathrm{GM}} = \mathbf{w}^t \left[\mathbf{K_m}^{-1/2}\mathbf{A}\right]^+ \mathbf{K_m}^{-1/2}\hat{\mathbf{f}} \qquad 2.4$$

The figure of merit in this case is

$$[\mathrm{SNR}_{\mathrm{GM}}]^2 = \frac{\mathrm{tr}[\mathbf{FW}]}{\mathrm{tr}\left[\left(\mathbf{A}^{\mathrm{t}}\mathbf{K_m}^{-1}\mathbf{A}\right)^+ \mathbf{W}\right]} \qquad 2.5$$

In the expressions above, $\mathbf{A} = \mathbf{OH}$, the "+" indicates the Moore-Penrose matrix pseudoinverse, and all other arguments were defined previously.

Finally, the Wiener estimator minimizes the mean-square error and is optimal when the object and noise statistics are known to be Gaussian-distributed. The Wiener estimate for that case is shown in Equation 2.6.

$$\hat{\theta}_{\mathrm{WE}} = \mathbf{w}^{\mathrm{t}}\left(\mathbf{A}^{\mathrm{t}}\mathbf{K_m}^{-1}\mathbf{A} + \mathbf{K_f}^{-1}\right)^{-1}\mathbf{A}^{\mathrm{t}}\mathbf{K_m}^{-1}\left(\hat{\mathbf{f}} - \bar{\hat{\mathbf{f}}}\right) + \mathbf{w}^{\mathrm{t}}\bar{\mathbf{f}} \qquad 2.6$$

The corresponding figure of merit is

$$[\mathrm{SNR}_{\mathrm{WE}}]^2 = \frac{\mathrm{tr}[\mathbf{FW}]}{\mathrm{tr}\left[\left(\mathbf{A}^{\mathrm{t}}\mathbf{K_m}^{-1}\mathbf{A} + \mathbf{K_f}^{-1}\right)^{-1}\mathbf{W}\right]} \qquad 2.7$$

The application of the Wiener estimator has been explored[26] in evaluating a simulated nuclear medicine coded-aperture imaging system. Estimates of simulated radioactive pharmaceutical uptake within lesions in the region of the liver were computed as noise and lesion size and brightness were varied. As an indication of the efficacy of that estimator, its figures-of-merit for a range of potential aperture configurations were compared with those computed from a second evaluator, the Hotelling metric, based upon the very different task of tumor discrimination. The high correlation between the proposed estimator and the Hotelling criterion, an approach already demonstrated in one instance to predict human performance, was a promising result.

Several researchers have proposed non-linear estimators that produce measurements based on a maximum-likelihood criterion.[27] This approach chooses as an estimate, among all the possibilities, that which maximizes the probability of obtaining the observed outcome. It consists of maximizing a likelihood function, a

---

[26] W.E. Smith, H.H. Barrett, "Linear Estimation Theory Applied to the Evaluation of A Priori Information and System Optimization in Coded-Aperture Imaging," Journal of Optical Society of America A 5, no. 3 (1988): 315-330.

[27] J.E. Freund, R.E. Walpole, Mathematical Statistics, 3rd ed. (Englewood Cliffs, NJ: Prentice-Hall, 1980).

joint probability distribution of the contributing variables at the observed sample point, developed with respect to the estimated parameter. Maximum-likelihood estimates have a number of attractive properties; in particular, they are known to be unbiased.

Müller, *et.al*.,[28] have explored a maximum-likelihood model for multiparameter estimation in nuclear medicine imaging. The authors examine the imaging of a disc of variable size, location, and gamma activity on a circular background. The maximum likelihood estimate for the disc parameters is that which possesses the maximum statistical probability for representing the true parameters. The authors develop an expression for the likelihood of the agreement between the model and observed image data that includes terms for the modulation transfer function (MTF) and noise-power spectrum (NPS) of the imaging system under investigation. That parameter vector, across the many candidates generated in a Monte Carlo simulation, that maximizes the likelihood expression is chosen as the multiparameter estimate. The accuracy and precision of the method has been studied in the presence of simulated white and tomographic noise: activity estimates for the maximum-likelihood model were significantly improved over ROI estimates, but the precision was worse. The maximum likelihood estimator serves as a potential alternative to linear estimators when the task complexity is such that the relationship is non-linear.

These computed methods for estimation require that the imaging system properties, such as the resolution, be known or approximated. A model for, or covariance matrix representing, the noise must inevitably be determined. Yet the human visual system cannot know these physical properties of the system in any mathematical sense; it simply operates upon the greyscale distribution of the visual scene. In fact, there is little theoretical reasoning to suggest that any of these estimators in any way reflect the operation of the visual system. While it is possible that the predictions of these models might correlate with human results in some situations, without a more immediate representation of visual properties and limitations it seems unlikely that they could consistently prevail in complex and diverse imaging scenarios. This dissertation suggests an approach whereby the properties of the visual system are explicitly embodied in the model. The claim is that only with such a modeling process can an image quality measure be applied in a global manner with predictive results.

2.5 Methods Incorporating Visual Mechanisms

Each of the approaches described in the previous sections has marked disadvantages and limitations for determining image quality. When one is conducting human observer experiments, in order to obtain a complete and statistically powerful characterization of observer response behavior, many trials are required of trained, expensive participants. Recruiting and collecting data from them can be a major undertaking. Furthermore, the results of such studies have limited applicability, as they are not generalizable to systems or pathologies not included in the particular experiment. The computed methods that have been developed in an attempt to supplant observer studies have limitations as well. Typically, the models rely on assumptions about, or knowledge of, the signal and noise in the images in order to compute a result. Many of the complex tasks in medical image interpretation do not satisfy those assumptions.

Perhaps the most theoretically promising approach to computing image quality is to measure task performance for a model of human vision. The model might incorporate known and hypothesized neurophysiological and cognitive operations to compute some of the tasks of visual perception of the image. Under such a framework, computations for medical image perception tasks of all varieties might be performed using the visual mechanisms posited by the chosen visual model. The accuracy of the outcome computed from the performance may be utilized as the measure of the quality of each of the images.

This approach possesses the obvious attribute of avoiding lengthy human observer experiments. Also, an accurate and thorough visual model that possesses hypothesized mechanisms for a wide range of visual tasks performed in medical image interpretation would enable global application of this image quality approach to many medical imaging tasks and systems. Most importantly, as it theoretically performs the tasks with the abilities and limitations of the human observer, this approach might be expected to predict most accurately the assessment of image quality determined by the human.

On the other hand, any image quality measure that relies on the current level of understanding of the human visual system will admittedly be limited in its predictability and applicability. There simply do not exist complete and proven models of visual perception that would be certain to succeed here. Yet recent efforts in visual psychophysics and computer vision have produced testable and usable principles. A goal of the research in this dissertation is to demonstrate, by the utilization of a promising visual model and the production of informative results, the potential for an approach that incorporates these visual fundamentals.

---

[28] S.P. Müller, M.F. Kijewski, S.C. Moore, B.L. Holman, "Maximum-Likelihood Estimation: A Mathematical Model for Quantitation in Nuclear Medicine," Journal of Nuclear Medicine 31 (1990): 1693-1701.

That a model of human vision might be applied in assessing image quality has been proposed and investigated in previous dissertation research at this university. Zimmerman[29] first computed a measure of image quality in comparing simulated lung mass detection for two image processing methods. Cromartie[30] has applied a version of the core model similar to that described in Chapter 3 in calculating image quality estimates for the parameters of a contrast enhancement algorithm, sharpened AHE (SHAHE). The task in that research involved the estimation of the distance from a radiation treatment field edge to a simulated blob. While Cromartie has demonstrated the power of this approach for computing image quality within a massive parameter space, it remains to demonstrate its usefulness by proving that the predictions of such a model-based image quality approach correlate with human assessment.

This research has as its purpose precisely such an investigation. Capitalizing on the previously-demonstrated potential of this kind of an approach, this work seeks psychophysical verification. The promising claims of a vision model-based image quality method may only be realized by demonstrating that its predictions correlate with the assessments of human observers. The remainder of this dissertation is devoted to exposition of the components of the comparison of the model with human observers for two medical image estimation tasks. The following chapter describes the visual model that was studied.

The process of image quality evaluation is crucial to accurate interpretation of medical images. This discussion has attempted to emphasize that such measurements must be performed rigorously and with confidence in their predictability. Additionally, it was suggested that it would be convenient to employ the computer to produce those image quality estimates. The methods proposed in this work attempt to combine these theoretical and practical attributes.

---

[29]J.B. Zimmerman, “The Effectiveness of Adaptive Contrast Enhancement,” Ph.D. dissertation, (University of North Carolina-Chapel Hill, 1985).

[30]R. Cromartie, “Structure-Sensitive Contrast Enhancement: Development and Evaluation,” Ph.D. dissertation, (University of North Carolina-Chapel Hill, 1995).