Spring 2012

Texture analysis of PET scans as a tool for image quality assessment

Puneet Madaan
University of Iowa

TEXTURE ANALYSIS OF PET SCANS AS A TOOL FOR IMAGE QUALITY ASSESSMENT

by

Puneet Madaan

A thesis submitted in partial fulfillment
of the requirements for the Master of
Science degree in Biomedical Engineering
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor John J. Sunderland

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

MASTER'S THESIS

This is to certify that the Master's thesis of

Puneet Madaan

has been approved by the Examining Committee
for the thesis requirement for the Master of Science
degree in Biomedical Engineering at the May 2013 graduation.

Thesis Committee: _____
John J. Sunderland, Thesis Supervisor

_____
Joseph M. Reinhardt

_____
Vincent A. Magnotta

To my faculty members who gave me flexibility and support throughout the course of this thesis

ACKNOWLEDGMENTS

In presenting this work of research I would first and above all like to express my sincere gratitude to my Advisor, Dr. John J. Sunderland, for guiding and teaching me through this field. His experience and method to convey ideas encouraged and motivated me to progress in the project. I would also like to thank him for ensuring my financial support throughout the course of research and making things flexible enough for me to pursue internship and a job along with my thesis work.

I would also like to thank Dr. Joseph M. Reinhardt for his support, guidance and flexibility during my classwork, thesis and internship. It is impossible to forget the contribution of all the faculty and staff at the University of Iowa who made my intellectual journey enjoyable and enriching.

I would also like to thank my lab members Levent Sensoy who helped me in collecting data necessary for my research and Sucheta Mohapatra in guiding me through programs in MATLAB. My deepest thanks to Biomedical Engineering department, the University of Iowa who helped me with every possible thing in which they could.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER I

INTRODUCTION

1.1 Motivation

In today's world of biomedical technology, advanced imaging techniques have emerged as indispensable tools in the diagnosis and treatment of cancer, heart disease, and brain disorders. Imaging is also increasingly used within the context of clinical trials as a means to determine non-invasively whether a particular therapeutic intervention (drug, surgery, or other therapy) is having the intended effect.

Cancer is a major public health problem in the United States and many other parts of the world. One in 4 deaths in the United States is due to cancer. According to the American Cancer Society a total of 1,638,910 new cancer cases and 577,190 deaths from cancer are projected to occur in the United States in 2012, corresponding to more than 1,500 deaths per day [1]. Cancers of the lung and bronchus, prostate, and colon in men and cancers of the lung and bronchus, breast, and colon in women continue to be the most common causes of cancer death.

Clinical assessment of progression or regression of tumors has been significantly improved by measuring the effect of therapies on patients using the variety of imaging techniques. Our current technological sophistication enables us to measure several image-based quantitative indices associated with tumors - size, radiologic density, and metabolic activity to name a few. These measurements have not only led to important insights into the mechanisms of cancer, but also provided critical and timely diagnostic information allowing more individualized treatments of their disease.

Positron emission tomography (PET) is one of the imaging techniques that are used in making this assessment. PET generates three-dimensional images of the biochemistry of body, helping us to identify and characterize diseases like cancer in their early stages, identifying its the location and severity of spread, often eliminating the need

for biopsy. PET stands out from other imaging techniques because it allows the visualization of physiological functions of the body (and disease) with their associated biological and chemical processes, whereas other diagnostic imaging techniques like Computed Tomography (CT), Magnetic Resonance Imaging (MRI) etc. predominantly visualize image anatomy.

1.2 Quantitative PET Imaging

PET is a quantitative imaging technique that measures the concentration of the radiopharmaceutical in the body, non-invasively. Depending upon the specific PET radiopharmaceutical used, PET can assess glucose metabolic rate, DNA proliferation rate or even quantitative blood flow. Identifying hot spots over background is the primary means of identifying a malignant tumor in the body with PET scans. In a neurological context, a patient suffering from Alzheimer's disease imaged with PET/CT using a labeled glucose analog would show a glucose metabolic defect in these areas as compared to normal subjects. Alternatively, PET imaging can be performed with a tracer that binds to the amyloid proteins that abnormally form on affected neurons in Alzheimer's disease. Quantitative measurement of the amyloid burden in patients is proving to be the earliest measure of onset of disease yet discovered. In either case, the variation of intensity values on the PET scan of a patient and its deviation from a normal distribution helps a physician to determine if the patient is suffering from an illness or not.

The quantitative aspect of PET is important as it helps in both intra- and inter-patient comparison for monitoring response to therapy as well as monitoring disease progression between successive patient scans. For example when using PET in clinical trials to measure response to therapy, a good outcome might be identified by either a decrease in the intensity value of tumor or a decrease in the volume of tumor before and

after the treatment. A negative outcome would be indicated by an increase in the tumor's intensity value after treatment or an increase in its volume indicating disease progression. Ideally PET would serve as a predictor of improvement in a patient's condition. Quantitation also helps in inter-patient comparisons between diseased and normal subjects or in comparison between two or more populations in order to derive valuable data including disease progression and response to therapy.

To optimally use PET in a multi center clinical trial environment, the biases and variation between the quantitative values obtained from multiple sites under controlled conditions should be minimized. Unfortunately absolute quantitation varies significantly with many factors including make and model of the scanner, scanner calibration, scan acquisition parameters, scatter correction, attenuation correction, patient movement, and even normal physiological variations within the patient. Standards and protocols are being developed to minimize these variations.

1.3 Qualitative PET Imaging

Besides the quantitative aspects of PET imaging, there is a qualitative aspect that impacts both clinical and research use. Often in an oncologic clinical trial, disease progression is defined as either an increase in the size of the tumor, or increase in activity on a PET scan. But disease progression can also occur if additional tumors are identified in the PET scan. Identification of new lesions is a critical clinical finding. However, newly identified tumors are often quite small and difficult to detect; so subtle are some of these that it is possible they will be visible on one PET scanner system, but not visible on another lesser quality scanner. Or a lesion may be clearly visible on a PET scanner using one reconstruction technique, but not when using a different reconstruction approach. These qualitative differences manifest themselves largely as differences in noise properties of the images, or what might be termed image "texture". Excessive smoothing

with post-processing filters makes images more pleasing to read, but results in a reduction of detectability of smaller lesions – with significant clinical and research implications, while reconstruction techniques intent upon higher resolution often result in significant amounts of high frequency noise, often simulating the appearance of actual lesions when they, in fact, don't exist. There is a spectrum of image quality, with associated lesion detection capability between these extremes. The bottom line is that image quality matters, but any meaningful quantitation of "image quality" has been elusive. More specifically, the ability to tell whether one image is diagnostically equivalent to another has neither been successfully achieved nor even well investigated.

One of the techniques that can be employed to differentiate the way one image looks from another is to quantify the quality in terms of the reconstructed image "texture". Texture is defined as a surface property or surface characteristic of an image, which depicts an interrelationship between neighboring image pixels. In general the mathematically described texture of an image is complex, and can vary with the material, surface pattern, reflectance from the surface, and vary substantially with the noise properties of the image. Because of the wide variety of potential textures, it is impossible to fully define texture of an image using a single mathematical parameter. Instead the field has created multiple mathematically explicit texture parameters, each describing different relationships with neighboring pixels on local or global image scale. Examples of texture parameters include measures of "contrast", "entropy", "correlation", "homogeneity", and others – each having a specific matrix-based mathematical function describing its explicit calculation.

In the context of multi-center clinical trials, where new therapies or drugs are being tested, regulatory bodies like the US Food and Drug Administration (FDA) demand highly controlled studies. FDA has been fundamentally uncomfortable for many years with the variable performance levels of different make and model scanners, and the

increased uncertainty this brings to clinical trials. FDA would prefer that all PET scanners had identical imaging properties, at least within the context of an individual trial. This concept holds not only the quantitative aspect of the PET image, but also the detectability factor which determines if something is present or not.

PET imaging is an inherently quantitative technique. Physicists have been striving for years for ever more accurate corrections and reconstruction approaches to optimize the quantitative fidelity of the image data. In parallel, significant effort has gone into trying to improve the clinical image quality of PET images through various iterative and statistically based reconstruction optimization strategies. However, to date, there has been little to no effort to quantitatively describe and parameterize the qualitative aspect of the PET image. This area is functionally uninvestigated.

In our study we acquire PET image data of two common PET phantoms using two PET/CT scanners of different vintage, with significantly different imaging characteristics. The aim of our study is to exhaustively interrogate the reconstruction space of both scanners to see if we can:

1. Find reconstruction parameters for each of the two scanners that results in identical quantitative characteristics as defined by their respective recovery coefficient curves, and
2. Identify those texture metrics calculated from various reconstructions from each of the two PET scanners that, when taken alone and together, will define when two images are of the same diagnostic image quality.

CHAPTER II

BACKGROUND AND SIGNIFICANCE

2.1 PET as an Imaging technique

Nuclear imaging is a diagnostic imaging technique in which a patient is injected with trace quantities of radioactively tagged molecules (radiopharmaceuticals) and scanned by equipment specifically designed to detect radioactive emissions to generate images of their distribution. Radiopharmaceuticals are compounds specifically designed and synthesized to probe a particular biochemical or physiological function. Often, but not always, the radiopharmaceutical is a nearly authentic biomolecule that has one of its native atoms replaced by a radioactive element.

There are two general categories of nuclear imaging equipment. First are those designed to image radiopharmaceuticals that decay by standard single photon emissions. Gamma cameras (also known as Anger cameras) and Single Photon Emission Computed Tomography (SPECT) scanners fall into this category. Second, is the area of Positron Emission Tomography (PET) where the radiopharmaceuticals are biomolecules that are labeled with radioactive atoms that decay by positron-emission. PET uses the annihilation radiation produced by positron-emissions from unstable, radioactive positron-emitting nuclei to create an image of the distribution of the radiotracer.

A positron is a subatomic particle having the same mass as electron but bearing a positive charge and a spin of ½. When a low energy positron and low energy electron collide, annihilation occurs with the release of two or more gamma ray photons.

The most widely used PET radiopharmaceutical is [^{18}F] 2-Fluoro-2-deoxy-D–glucose (FDG), a radioactive analog of the sugar molecule, glucose. In this molecule, a radioactive, positron-emitting F-18 atom substitutes for a hydroxyl group in the 2-position on the molecule. Small quantities of this molecule are synthesized and dissolved in an isotonic saline solution and is injected into patients so that physicians can image

regional glucose metabolism in the body.

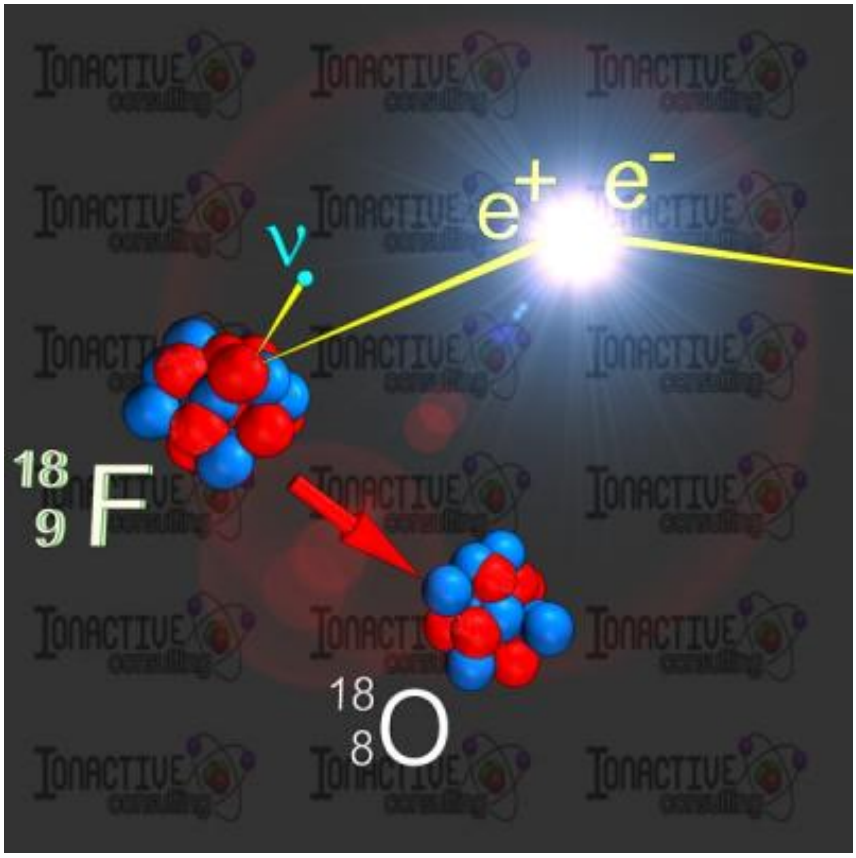


Figure 1: Decay of positron emitting ^{18}F atom

Source: Scienceblogs.com
http://scienceblogs.com/startswithabang/2009/12/03/lhc-smash/

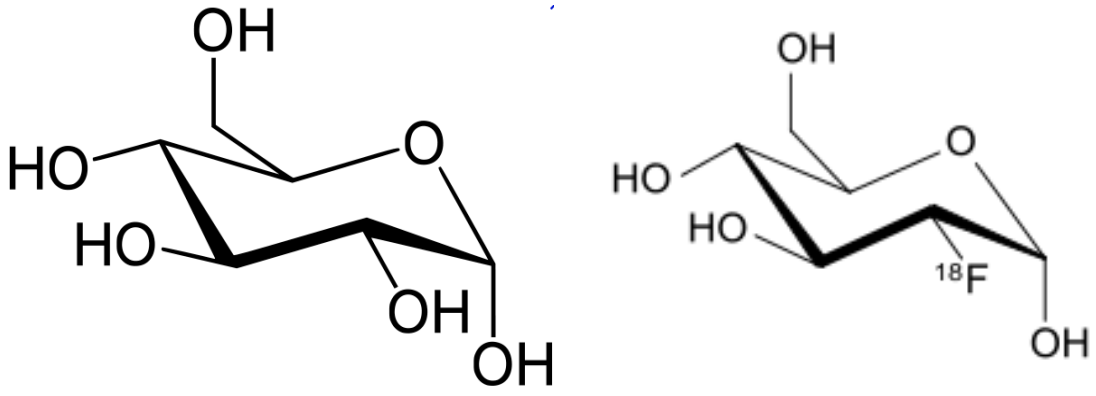


Figure 2: Glucose molecule (left) and FDG (right)

The FDG is transported across the cell membrane through the same facilitated transport mechanism used by naturally occurring glucose. When cells are cancerous, they are generally replicating at a faster pace than the normal body cells and are utilizing a much less efficient glucose metabolism pathway and hence utilize much more glucose than normal functioning cells. As a result, the cancer cells consume and concentrate more radioactive sugar. The presence of higher activity in cancer cells distinguishes then from the healthy cells when the patient is imaged with FDG and a PET scanner.

The FDG PET/CT scan exposes the patient to radiation doses comparable with other tomographic imaging techniques, like CT. It is widely described as the most sensitive and specific way to locate cancerous tissues in the body and monitor its progression with treatment.

PET radiopharmaceuticals, like FDG, are generally biologically active molecules specifically designed to image and measure biochemical processes in vivo. The radioactive nucleus of the positron-emitting isotope (in this case [F-18]) attached to the radiopharmaceutical undergoes positron decay, hence emitting positrons. These

positrons slowdown over a course of travel on the order of 1 mm before colliding with oppositely charged electrons, at which point they physically annihilate, releasing two 511 keV gamma ray photons travelling at 180 degrees from one another

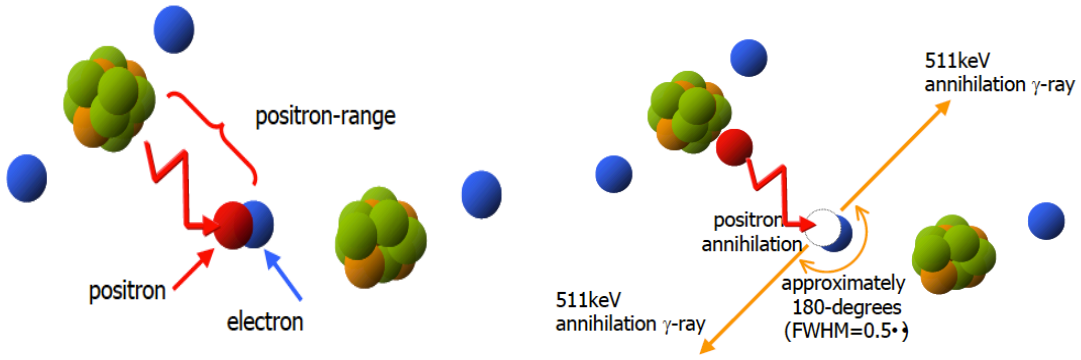


Figure 3. Positron emission and subsequent annihilation followed by creation of two collinear 511 keV photons.

Source: CYRIC Tohoku University
http://kakuigaku.cyric.tohoku.ac.jp/PETAnalysis/PET_ABC_v2-1.pdf

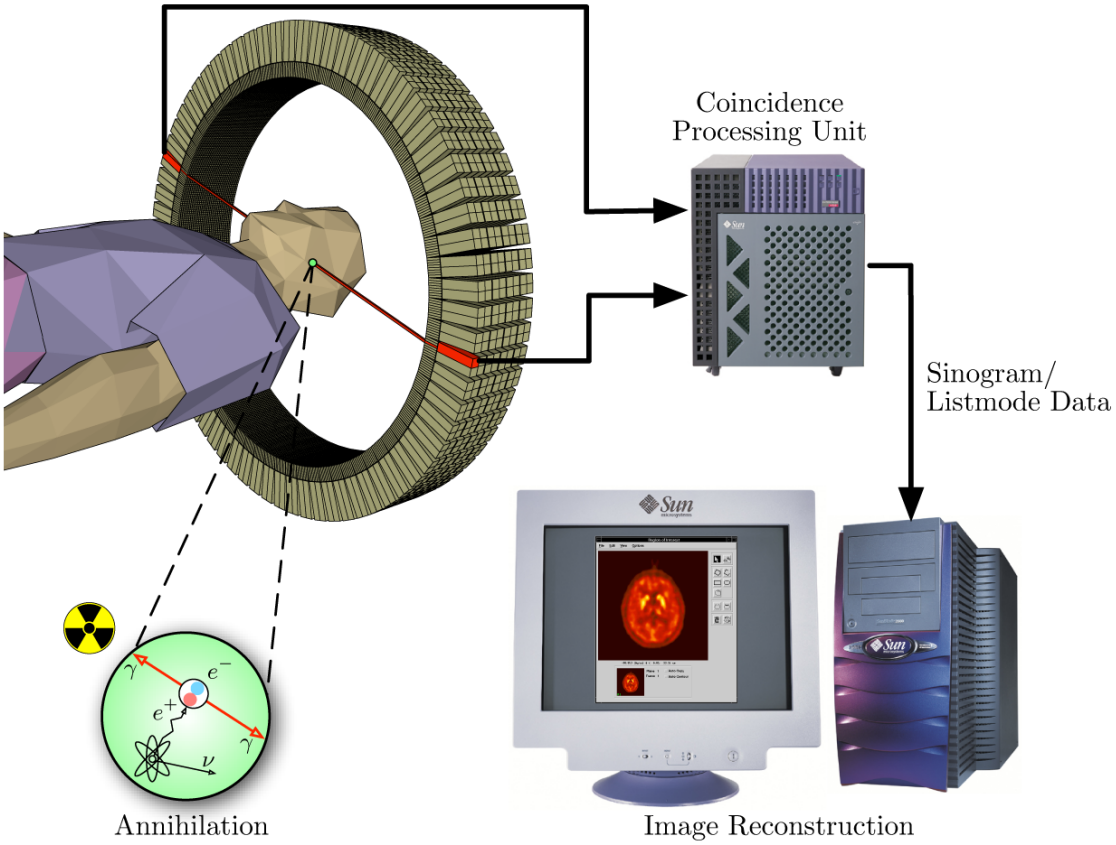


Figure 4: Patient inside a PET scanner

Source: Wikipedia
http://en.wikipedia.org/wiki/File:PET-schema.png

Figure [4], above, depicts a patient being scanned in the PET scanner. The PET detector ring is comprised of an array of thousands of scintillation crystals coupled to hundreds of photomultiplier tubes (PMTs) arranged in an annular ring around the patient. The circular ring of the detectors is designed for optimal simultaneous detection efficiency for the 511 keV annihilation radiations emanating from the positron emitting radiotracer.

The scintillation detectors surrounding the patient detect these gamma ray photons in coincidence. A scintillator is a material which when struck by an incoming high-energy photon absorbs its energy and re-emits that energy in the form of a higher number of lower energy (usually visible light) photons. To efficiently stop the high-energy 511 KeV gamma ray photons, scintillators must be made of high electron density material that is translucent to the wavelength of light they emit. LSO (Lutetium Oxyorthosilicate $Lu_{2(1-x)}Ce_{2x}(SiO_4)$ is very commonly used because of its high density of 7.4g/cc, and fast photo fluorescent decay time, and relatively high scintillation efficiency. The high density, and associated electron density, allows for a smaller detector size and therefore potential for higher resolution. LSO has a high refractive index of 1.82 permitting higher amount of light to pass through total internal reflection to the photomultiplier tube. It has a short decay time of approximately 40ns, which allows the PET scanner to accurately determine the simultaneity of two emitted photons as well as detect and process high count rates.

The lower energy photons emitted from scintillator crystal then travel to the photomultiplier tube.

A photomultiplier tube, optically coupled to the scintillator crystal, is a vacuum tube that releases electrons when visible light photons strike the photocathode, the entry point of a photomultiplier tube. The photocathode is a photosensitive surface that releases electrons when struck by light photons. The electrons are accelerated by "dynodes". Each successive dynode is at higher positive potential than the previous one, resulting in multiplication of electrons by secondary electron emission at each successive channel. The current amplification or gain of photomultiplier tube depends on the number of dynode stages. At the end of the dynode chain is an anode where the electrons are collected. At this stage a current pulse is sent to coincidence processing unit.

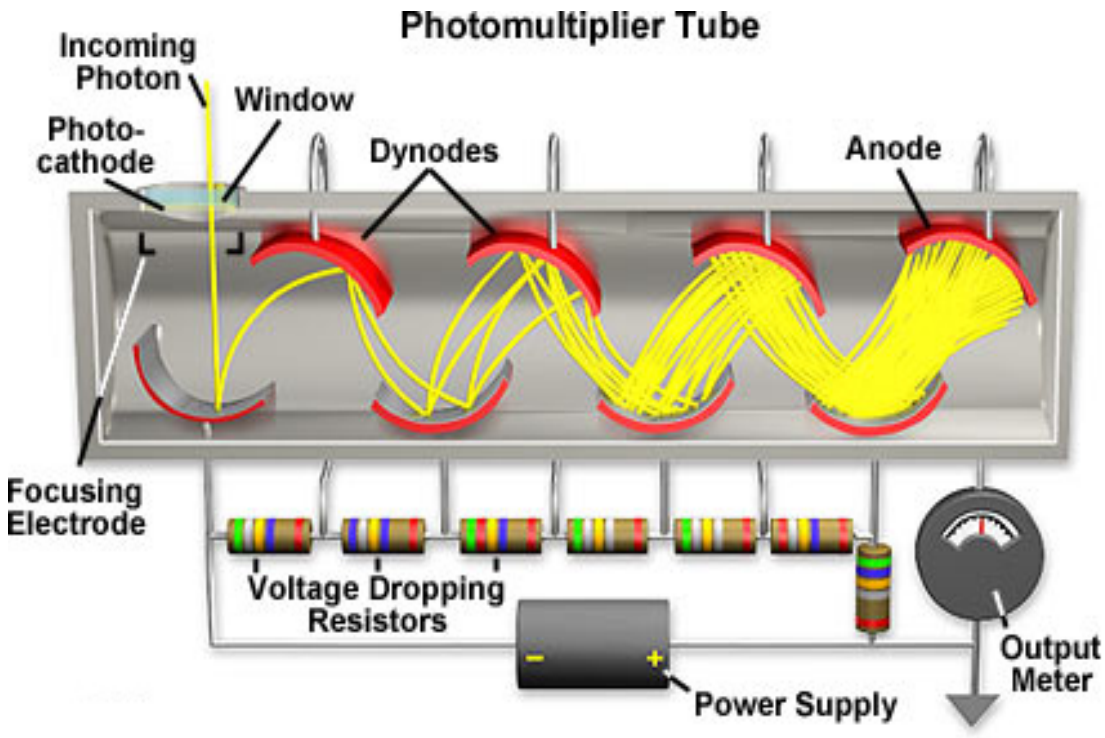


Figure 5: Photomultiplier Tube

Source: Hamamatsu Learning Center
http://learn.hamamatsu.com/articles/photomultipliers.html

The current pulse from the photomultiplier tube is amplified using a low noise pre amplifier. Passing the signal through a variable gain amplifier to compensate the variability in sensitivity of different photomultiplier tubes in the system follows it. The signal then travels to the pulse height analyzer, which performs energy discrimination and in parallel undergoes position and timing processing. If the pulse received was 511

keV, a logic pulse is generated and is sent to a coincidence circuit. Each detector generates a timing pulse when it registers an incident photon. These pulses are combined in coincidence circuitry. If the two pulses occurred within a specific time frame of approximately 5-6 ns they are deemed to be coincident and considered as a valid signal. Photons that do not arrive in this time window are ignored. Millions of such events are detected by the PET scanner system. By combining these annihilation events using a tomographic reconstruction algorithm, a map of the radiopharmaceutical distribution can be calculated.

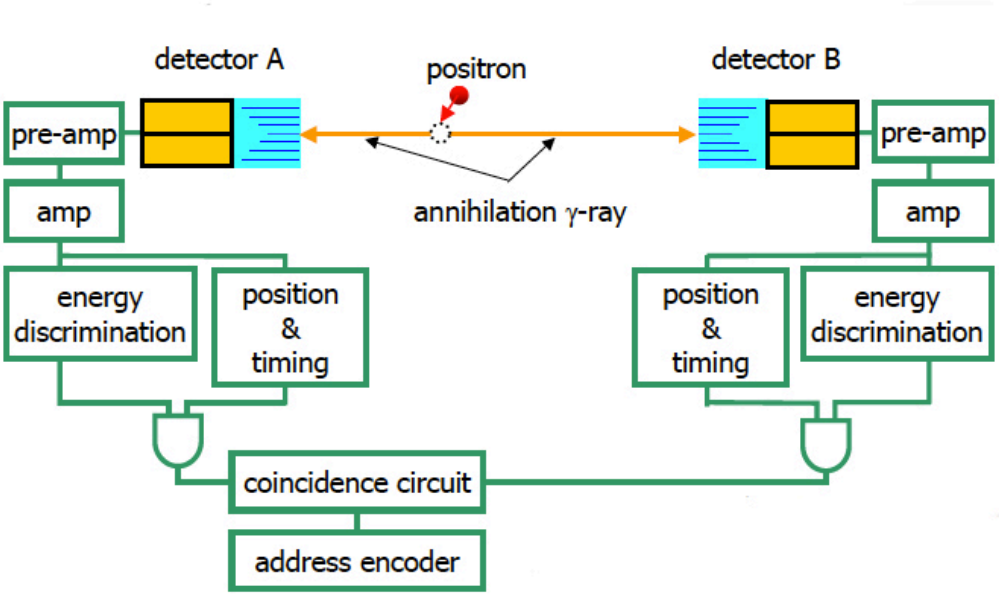


Figure 6: Schematic diagram of a PET scanner

Source: CYRIC Tohoku University
http://kakuigaku.cyric.tohoku.ac.jp/PETAnalysis/PET_ABC_v2-1.pdf

2.1.1 Image formation/reconstruction in PET imaging

The pair of detected photons defines a line on which the annihilation event took place. From an image reconstruction standpoint, this is called the line of response (LOR). A coincidence event is assigned to a line of response (LOR) joining the two relevant detectors. This temporal coincidence of the two co-linear photons results in "electronic collimation" and eliminates the need for a bulky lead or tungsten physical collimator. Electronic collimation results in better sensitivity and uniformity of the point source response function than from systems using a physical collimator.

The reconstruction software then processes the collected coincidence data and reconstructs the image that depicts the localization and concentration of the radionuclide. Various algorithms can be used to reconstruct the data into a quantitative mapping of the radionuclides. The two most common algorithms used for reconstructing PET images are: the filtered back projection algorithm and the ordered subset expectation-maximization iterative algorithm.

2.1.2 The Reconstruction Process: The Sinogram

A sinogram is the 2-D matrix representation of the projection data used to reconstruct a tomographic image. Each row of the sinogram is a one-dimensional histogram of detected annihilation coincident photon events that occurred through the object at a particular angle as illustrated below in Figure [7]. A sinogram is not a representation of the acquired image. It is an organization of the acquisition data from which the tomographic images are reconstructed using one of the reconstruction algorithms discussed below.

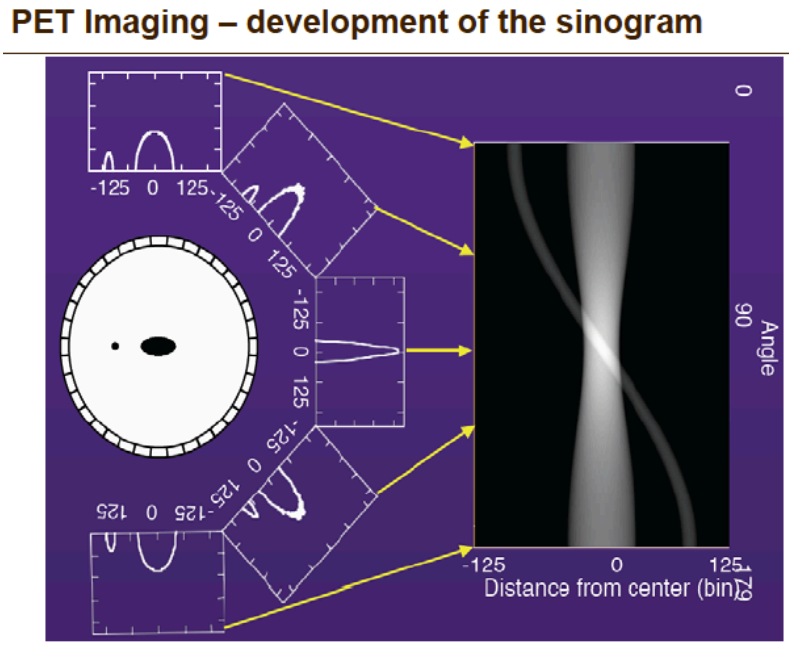


Figure 7: Development of a Sinogram

Source: http://engineering.dartmouth.edu/courses/engs167/20%20PET%20Imaging.pdf

2.1.3 Filtered Back Projection Reconstruction

One of the methods to convert the data organized in the sinograms back to an image is the filtered back projection technique. In filtered back projection the one dimension projections are convolved with one dimension filters before back projecting in order to eliminate the blurring effect of symmetric 1/R point spread function associated with the simple back projection technique.

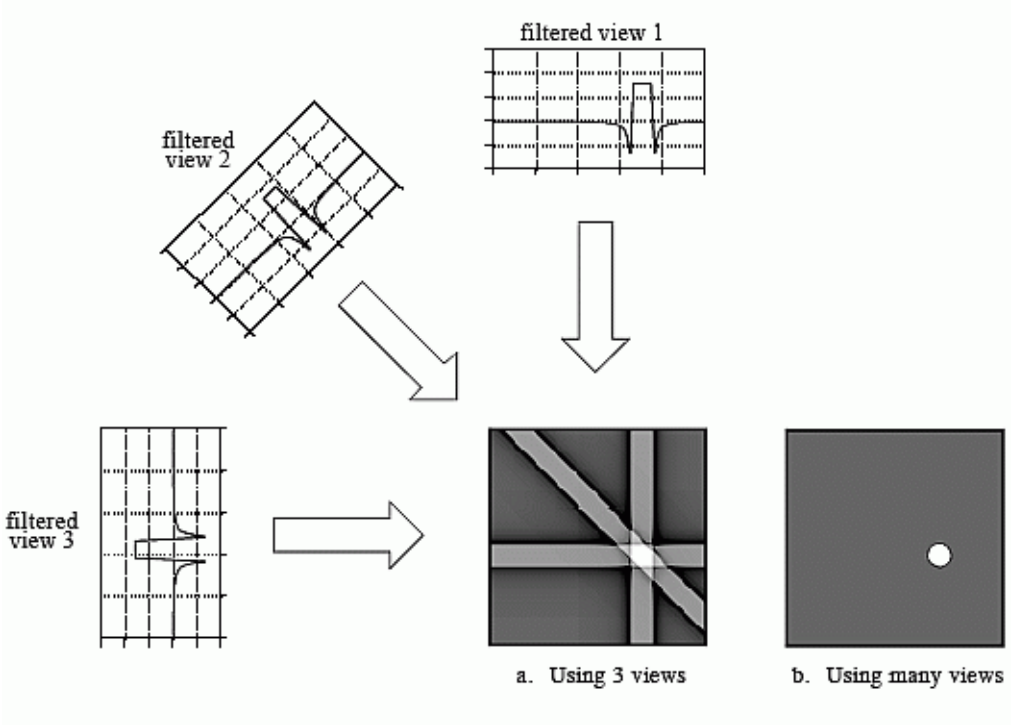


Figure 8: Filtered Back Projection

Source: http://www.dspguide.com/ch25/5.htm

The filtered back projection algorithm is fast, and is not computationally expensive. However it has particularly poor noise properties when used with data that has limited statistics, as is the case in PET scans. The statistical noise in PET data is functionally white noise from a spatial frequency standpoint – it is invariant in magnitude across all spatial frequencies. However during filtered back projection reconstructions using a ramp filter, high frequencies components of the image are multiplied by

proportionally larger numbers, resulting in a reconstructed image with a lot of high frequency noise. Another drawback to FBP reconstructions is that they are unable to handle complicating factors such as scatter, which is necessary for optimal PET reconstruction.

2.1.4 Introduction to Iterative reconstructions

The expectation-maximization reconstruction approach uses an iterative method, which attempts to match the image most consistent with the actual projection data. It usually makes an initial guess at the distribution of the one radionuclide based on a preliminary filtered back projection reconstruction. It then forward projects the data to generate a new set of sinogram data that it statistically compares to the original measured sinogram data. Modifications are then made to the new sinogram data to better match the actual original projection data by compensating for discrepancies. The new sinogram data set is then reconstructed into a new estimated image. This new image is subsequently forward projected to create a third sinogram set, which is compared to the original projection data. The process is iterated and reiterated until an accurate model is found. The iterative reconstruction method is explained in greater detail in Section 3.3.

2.2 Texture as a property of an image

Texture refers to the surface arrangement or surface characteristics of an object. It could be a local or global, repeating or non-repeating surface pattern, which helps us in differentiating various objects from their surroundings. It is a key property of an image that makes one image look different from another.

In computer vision texture is referred to as the spatial inter-relationship between neighboring image pixels. These relationships describe the changes in intensity value or gray tones of the neighboring pixels.

2.3 Image texture analysis

Image texture analysis refers to the quantification of the image texture in terms of local and global qualities like, but not limited to roughness, smoothness, silky, bumpy, uniformity, density, coarseness, energy, homogeneity, entropy, contrast, correlation, regularity, linearity, directionality and phase. These qualities are used to define the texture of an image as a spatial inter-relationship between the neighboring pixels.

Texture of an image represents its physical surface structure, which could vary with material, surface pattern, reflectance from the surface etc. Hence it is not possible to define the texture with a single parameter. For example, Entropy is a measure of randomness. A surface with low entropy value will display less variation in overall intensity value of pixels as compared to a surface with high entropy value. Similarly in a smooth surface, the range of values in the neighborhood around a pixel will be a smaller than a rough texture, where the range of adjacent pixels will be larger.

Computer vision texture analysis is a well-developed mature field of study, and has been applied in many disciplines including the chemical industry, remote sensing, image analysis, photography, and in automated inspections. Only recently has there been efforts to use texture analysis in medical image processing.

2.4 Previous works in texture analysis

H. Kaizer [2] introduced the concept of texture in digital images in the paper “A quantification of textures in aerial photographs” in 1955 which was further elaborated by Robert M. Haralick [3] in his paper “ Textural features for Image classification” in 1973 in which he introduces formulae for calculating various texture matrices defining different properties of images. Since then texture analysis of images had been a topic of interest among researchers in various fields.

Daeyoun Kim [4] explained in his paper the use of texture analysis in the field of chemical engineering for determination of steel quality by real time capturing images of its surfaces. Using Wavelet packet transformation, his method first converts single resolution images taken from product surface to multi resolution sub-images or a summation of waveforms with various frequencies, in order to make it more comparable to human visual system. Wavelet transform retains spatial information along with the frequency information of the signal and texture parameters were extracted from the image. The method was designed for automated surface quality determination system.

Textural features have also been applied for classification or categorization of pictorial data as suggested by Haralick [3]. In this work, he classified the pictorial data on a pixel basis by identifying the agricultural crop category from satellite imagery. The various data sets used in Haralick's study included photomicrographs of sandstone, aerial photographic data and data set derived from satellite imagery. Using textural analysis Haralick could accurately identify/categorize them using textural features on these data sets with an accuracy of 80-90 percent.

There have been recent efforts to apply texture analysis to medical images. In fact there have been several hundred published articles in recent medical literature that have "Texture Analysis" in the title. Most of this work has been performed in CT, and to a lesser extent MRI, with some preliminary work in PET. Texture analysis has been used to identify lung tumors with increased tumor aggression, in a study by Balaji [5]. He demonstrates a textural feature extraction method for recognition of these areas. The relationship between texture features in computed tomography (CT) images of non-small cell lung cancer (NSCLC) and tumor glucose metabolism (PET) and stage has been studied in a study performed on 18 patients suffering from NSCLC. Texture parameters of the CT scans of the tumors were compared with tumor FDG uptake. Before finding the textural parameters a Gaussian filter was applied to the images to separate them on the scale of fine to coarse textures. The result of the study concluded that the coarse texture

features correlated well with tumor SUV whereas fine texture features correlated with the tumor stage i.e. stage I, II, III or IV as grouped by conventional CT criteria for tumor size. This study suggests that the textural features can assist in providing prognostic data and response to therapy for NSCLC.

In the field of PET images, texture analysis has been used by Tixier [6] to predict the response to therapy in Esophageal cancer. Patient with esophageal cancer were treated and their PET scans were taken before and after the treatment to detect the response to radio-chemotherapy. Patients were grouped to be non-responders, partial responders or complete responders as per their response to disease. As the cells in a cancer tumor divide and replicate at a faster pace (high metabolism), they have a higher glucose uptake and hence higher SUV value. Textural features from the tumors were extracted and compared against the response to therapy as seen by changes in SUV value of the tumors after 2 weeks from the start of therapy. Tixier concluded that differentiating the patients with tumor texture analysis showed a higher sensitivity than any SUV measurement and is a better way to detect the patient's response to therapy in Esophageal cancer.

Huan Yu [7] has done another study in a combined multimodality area in "Automated segmentation using texture analysis of PET/CT images in Head and Neck cancer" to define radiation targets in the body. Due to high contrast between a normal and a cancerous tumor, it is easy to locate a tumor; but it lacks a well-defined boundary and shape. The boundary and shape of a tumor depends on the reconstruction parameters and is highly operator dependent. The target delineations made by an automated segmentation method in head and neck cancer using texture classification turned out to be more accurate than the target delineation done using PET data alone.

2.5 Significances and innovation of the work

Texture analysis in PET, and for that matter, medical imaging in general, has only been used in an attempt at tissue characterization. To date it has not been used as a tool for image quality assessment. FDA ideally desires identical quantitative and qualitative data from different sites using different systems for multi-center clinical trials.

Our study is the first attempt at quantification and harmonization of image quality data using texture metrics. In this preliminary study we will use five texture metrics, each describing different texture properties for our image quality assessment study. These texture matrices use different mathematical relations among the pixels within an image. Central to four of these matrices is a preliminary calculation of a so-called co-occurrence matrix (also called the grey level co-occurrence matrix (GLCM)). The GLCM is defined in the next section in detail.

CHAPTER III

METHODOLOGY

Qualitative harmonization in the context of PET imaging, and for the purposes of this work, is defined as the ability to produce PET reconstructions for different make and model scanners which would result in similarity of qualitative features like lesion size, shape, contrast, noise properties and lesion detectability. This analysis will be performed through the application of texture analysis using several common texture metrics. Background information on this is described below in Sections 3.4.1-3.4.4.

3.1 Experiment design

3.1.1 Phantom Specifications

A phantom, within the context of medical imaging, is a controlled, well-characterized test object that is imaged by the imaging modality of interest for performing scanner quality control and assessing various aspects of both image quantitation and image quality. A phantom is generally a physical object whose physical geometry and material content are designed to challenge various aspects of the imaging system, including spatial resolution, contrast, and uniformity, to name a few. It can be a series of geometric shapes of various sizes, or it can imitate the physical appearance or characteristics of an organ or the human body in its entirety.

Before a scanner can be used in a clinical trial, it almost always must meet certain performance criterion defined by a phantom imaging exercise. The phantoms we chose for our study are used for onsite qualification of the scanners and clinical oncology trials.

The two phantoms used in our study are described below:

1) NEMA NU-2 Image Quality PET phantom

2) Clinical Trials Network Oncology PET phantom

3.1.1.1 NEMA NU-2 Phantom

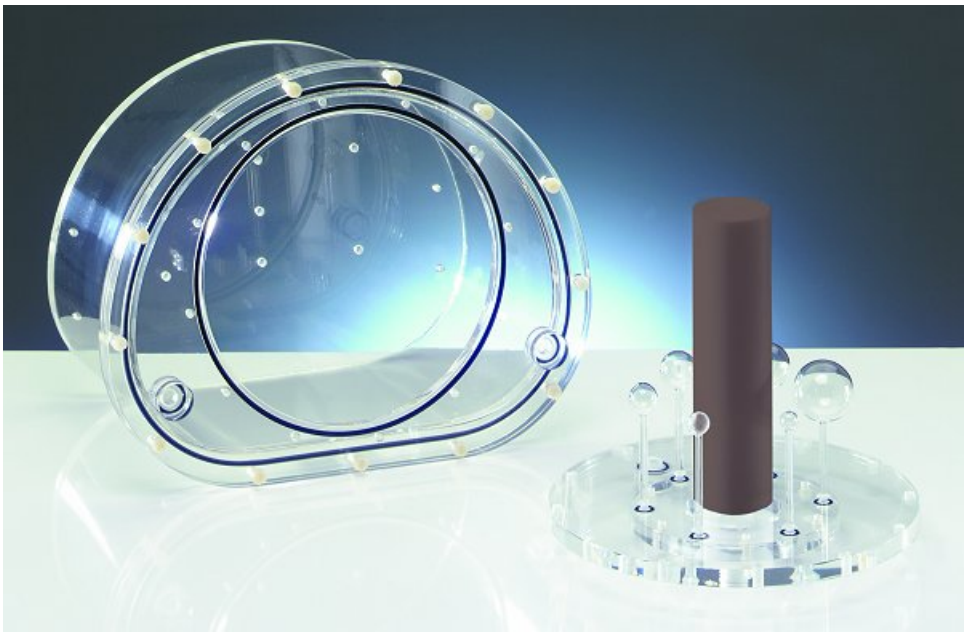


Figure 9: NEMA NU-2 Phantom

Source: http://www.ptw.de/pet_emission_phantom_nema.html

NEMA NU-2 is a phantom developed by National Electrical Manufacturer Association. It is mainly used for quality control purposes of a PET scanner. It is designed to assess the quantitative performance of the scanner. It is used by European Union for the onsite qualification of scanners.

The phantom is an acrylic vessel having an asymmetric shape, which imitates the human body from thorax to abdomen. It consists of a hollow Lucite cylinder in the center and 6 hollow Lucite spheres on the periphery arranged in an annular ring about the axis

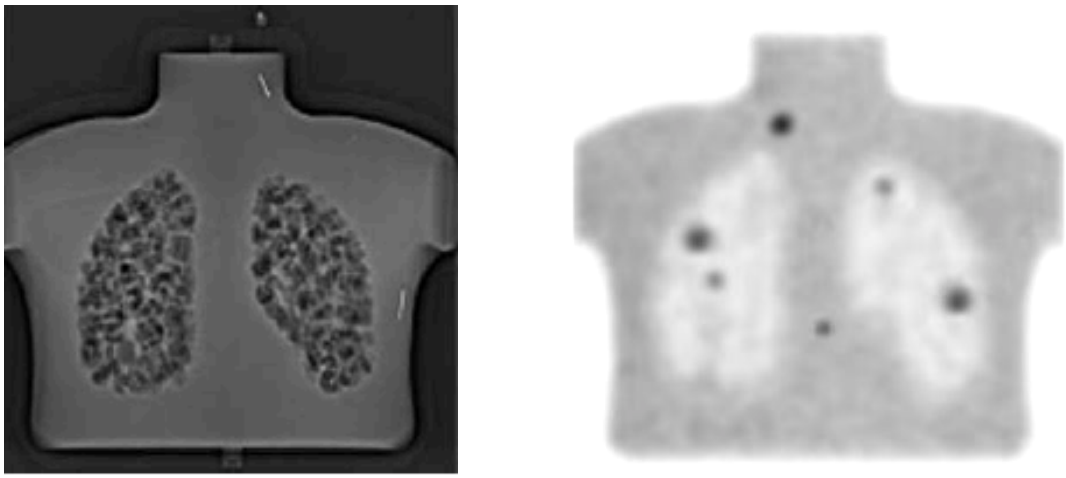
of the phantom. The inner active diameters of the spheres are 10 mm, 13 mm, 17 mm, 22 mm, 28 mm and 37 mm respectively. Outer diameter for the Lucite cylinder is 50 mm. The wall thickness of the spheres is 1mm. Activity can be filled or emptied in any part of the phantom using their respective fill ports.

3.1.1.2 CTN Phantom

Figure 10: PET Chest Tumor Simulator

Source: Society of Nuclear Medicine
http://interactive.snm.org/index.cfm?PageID=10641

The Clinical Trials Network clinical simulator oncology phantom is a clinically realistic phantom imitating a human with tumors of various sizes in various parts of the thorax. It was developed by the Society of Nuclear Medicine and is used in qualitative and quantitative evaluation of images and also as a standard in site validation in the US. It consists of two separate fill-ports, one to fill the lesions within the phantom with activity and the other for filling the background. The CTN phantom tests the scanner's quantitative capabilities and accuracy, and also assesses lesion detectability.



CT scan of CTN phantom PET scan of CTN phantom

Figure 11: Image depicting CT scan/ PET scan of CTN phantom

Source: Society of Nuclear Medicine
http://interactive.snm.org/docs/CTN/CTN_Fact_Sheet.pdf

3.1.2 Scanner Specifications

3.1.2.1 Siemens Biograph 40

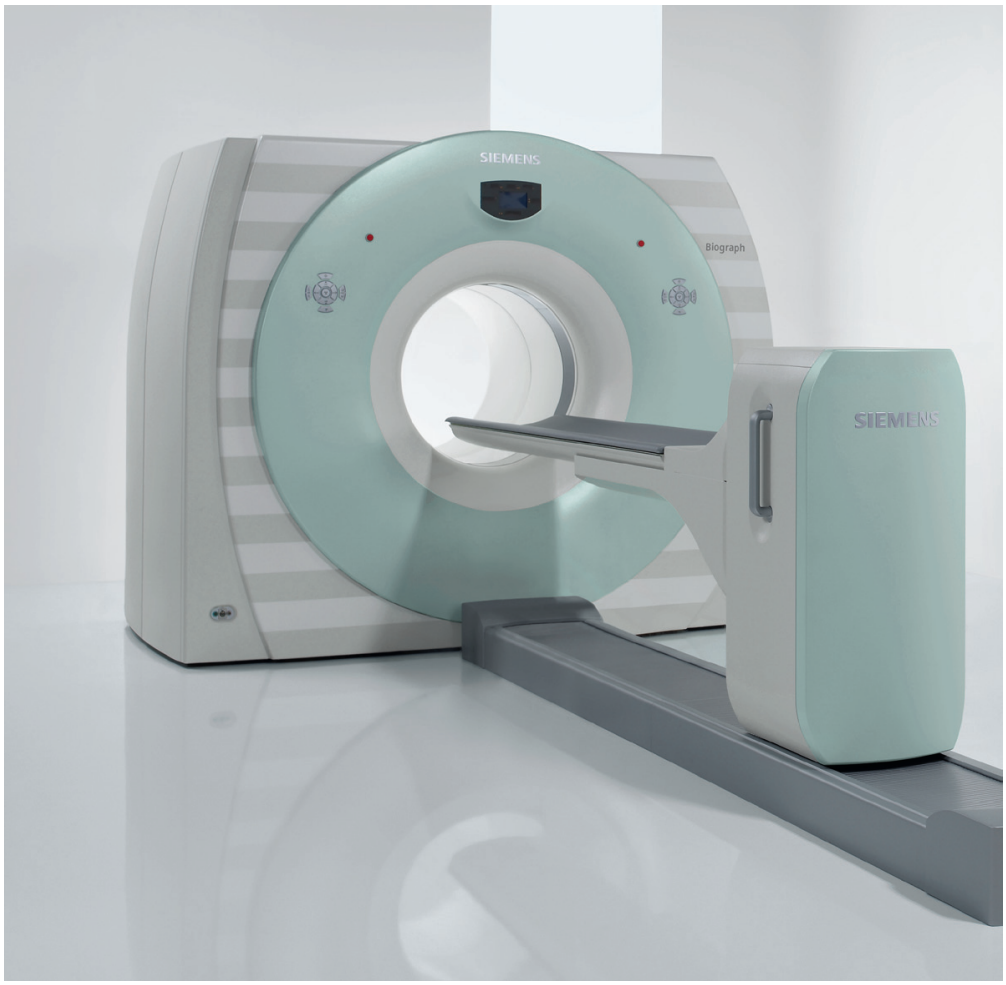


Figure 12: Siemens Biograph 40 Scanner

Source: www.siemens.com

The Siemens Biograph 40 is a whole body PET/CT scanner used for general clinical and research PET imaging, including oncologic, neurologic and cardiac applications. It incorporates PET and CT detectors in its assembly and is capable of 3D image data acquisition. It consists of 144 detector blocks and constructs 81 trans-axial PET image slices from single acquisition covering an axial field of view of approximately 16 cm. The scanner generates 40 CT slices per acquisition, which can be used for both attenuation correction purposes and anatomic correlation of PET images. The coincidence window i.e. the time frame within which an incident photon pair will be considered as a true event is 4.5ns. It provides user defined image reconstruction parameters, which can be altered to achieve the desired result after acquisition has finished. It is capable of list mode acquisition of data, the details of which will be described in Chapter 3.2.

3.1.2.2 Siemens Biograph Duo

Similar to Biograph 40, Siemens Biograph Duo is also a whole body PET/CT scanner. It also incorporates both PET and CT detectors in its assembly and can acquire images data in 3D. It consists of 64 detector blocks and constructs 47 PET image slices from an acquisition. It produces 2 CT slices per acquisition that can be used for attenuation correction purposes as well as anatomic co-registration. The coincidence window for Biograph Duo is 12ns. It also provides various user defined parameters but is not capable of list mode data acquisition.

Number	Specification	Bio Duo	Bio 40 Value
1	Detector material	ECAT multi- Lutetium Oxyorthosilicate	Lutetium Oxyorthosilicate
2	Crystal dimensions	6.29x 6.29 x 20mm	4.0 x 4.0 x 20 mm
3	Crystals per detector block	64	169
4	Number of detector blocks	64	144
5	Ring Diameter	824 mm	842 mm
6	Trans-axial FOV	600 mm	605 mm
7	Axial FOV	162 mm	162 mm
8	Number of image planes	47	81
9	Plane spacing	3.375 mm	2.025 mm
10	Trans-axial FWHM resolution	4.2 mm	6.3 mm
11	Axial FWHM resolution	4.5 mm	4.7 mm
12	Max. No. Of CT slices	2	40
13	Sensitivity	4.4 cps/kBq	4.2 cps/kBq
14	Coincidence Window	12 ns	4.5 ns

Table 1: Comparing Siemens Biograph 40 and Siemens Biograph Duo Specifications

The NEMA NU-2 and CTN phantoms were scanned using the abovementioned scanners using the following procedures. Fluorine F-18 FDG was the radioactive tracer used to fill the phantoms. The phantoms were filled according to instructions defined by the NEMA NU-2 and CTN protocols respectively. The NEMA NU-2 phantom spheres were filled with a solution with a 1.75uCi/ml concentration and the background with 0.44uCi/ml of activity. In the CTN phantom 0.948mCi mixed with 1000ml of sterile water/isopropanol solution was used to fill the lesions and 2.343mCi mixed with 9840ml of water is used for the background. A reconstruction grid with approximately equal pixel dimensions were selected for the two scanners in order to keep the X, Y dimensions of the output images of the two scanners approximately the same. Although X and Y pixel dimensions are functionally software selectable in the reconstruction process, the Z dimension is defined by crystal size. The Biograph Duo therefore has thicker slices than the Biograph 40. The following table describes the image matrix dimensions obtained from the two scanners using similar parameters and using same phantom.

Number	Property	Bio Duo	Bio 40
1	Slice Thickness	3.375 mm	2.025 mm
2	Number of Slices	47	81
3	Rows	256	256
4	Columns	256	256
5	Pixel Spacing	1.768 mm	1.781 mm

Table 2: Comparing Siemens Biograph 40 and Siemens Biograph Duo output Image Properties

3.2 Acquisition methodology

The two phantoms previously described in Section 3.1.1 were imaged so as to gather controlled data sets using the more geometrically simple NEMA NU-2 phantom, and then more clinically realistic, and complex CTN phantom. The goal of the study is to investigate which of the common texture matrices have relevance to determining equivalence of "Image Quality" as it relates to noise properties and lesion detectability in PET.

A total of 4 imaging studies were preformed with two phantoms imaged on the two PET/CT systems:

Phantom	Biograph 40	Biograph Duo
NEMA NU2 Imaging Quality Phantom	30 minute scan divided into 6 statistically identical scans	6 Statistically identical scans with equivalent statistics to the Biograph 40
CTN Oncology Phantom	30 minute scan divided into 6 statistically identical scans	6 Statistically identical scans with equivalent statistics to the Biograph 40

Table 3: Matrix of studies performed with the phantoms using two scanners

PET data can be acquired using two general techniques

3.2.1. Frame mode

Frame mode refers to collection of PET image "frames" sequentially over discrete preset amounts of time. In frame mode the acquisition of data continues until a preset time has elapsed. Immediately upon completion of that scan a new, independent acquisition begins for the next preset duration. In this way a sequence of image sets can be acquired.

3.2.2. List mode

In list mode acquisition the detector pairs associated with each annihilation detection event are stored along with a time stamp, in real time. Areal the detected events are stored as a list of events in order of their occurrence combined with time markers. For image generation, data is replayed from storage to create sinograms associated with user specified time intervals. In list mode we can change the frame duration as per our requirement. The major drawbacks of list mode are its higher computational requirements and more memory usage as compared to frame mode. But it provides added advantages of flexibility of data manipulation of frames. So List mode gives us the ability to reprocess the data after it has been acquired in a number of ways.

3.2.3 Data Acquisition - Biograph 40

A 30-minute data acquisition sequences was performed on each of the two phantoms on the Biograph 40 PET/CT. The list mode data was acquired so that progressively longer duration sinograms could be constructed to allow for the creation of six independent sinogram sets having approximately equal counts so all of the images would be of roughly the same statistical quality. The data collection in PET is a stochastic process and there is understandably some randomness associated with it.

The 30-minute list mode scans of Biograph 40 were later subdivided into six intervals to generate 6 different statistically identical acquisition data sets. Because the F-18 was continually decaying, the length of each interval needed to be successively a little longer than the previous one to maintain statistical equality between scans. This procedure was performed for both the NEMA NU-2 and CTN phantom acquisitions on the Biograph 40. The amount of activity injected in the phantoms was in consistence with the quantities defined in NEMA-NU2 and CTN imaging protocols as described in Chapter 3.1.2. For each phantom, six statistically similar data sets (sinograms) were generated. We generated the 6 functionally identical data sets to be able to assess the variance associated with the stochastically based uncertainty of the images. From these six sinogram sets, we generated a number of different reconstructed images sets with different resolution and noise properties using various reconstruction parameters sets.

We collected data for 30 minutes with the intention of dividing this data into 6 statistically identical data sets. Because the F-18 is continually decaying with a half-life of 110 minutes, each frame needed to be slightly longer than the previous scan to account for the radioactive decay.

To determine the scanning intervals for each of the six scans with comparable statistics we integrated under the decay curve defined below in equation (1) to achieve 6 equal areas under the decay curve (Figure [13]) from t=0-30 minutes,

$A(t) = A(0)\, e^{-lt}$ …….. (1)

Where:

A(0)= Activity of the sample at time t=0.

A(t)= Activity of the sample at time t.

e^{-lt} = Decay Factor. Fraction of radioactive atoms remaining after time t

l = Decay constant. Probability of atom undergoing decay per unit time

t = time interval

$l = \ln(2)/t_{1/2}$

$t_{1/2}$ = 109.8 min for Fuorine-18

The total number of decays that occur between any 2 time points t_i and $t_{i=1}$ is equal to the area under the activity curve between any two of the 6 adjacent time points, as shown in Figure [13].

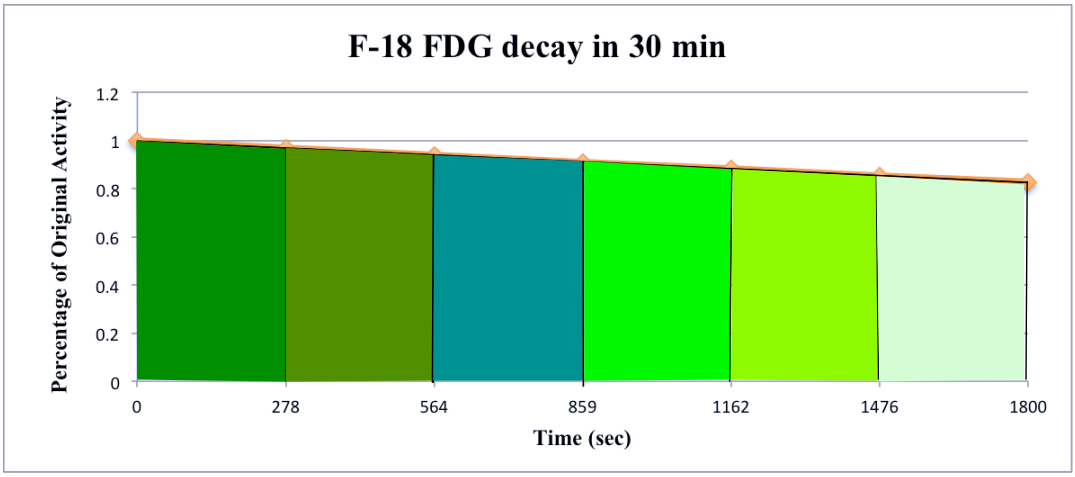


Figure 13: Graph depicting area under the curve for 6 time frames in 30min separated by equal number of counts using F-18 FDG

For the Siemens Biograph 40, when using List mode data and calculating the intervals in order to obtain six data sets with equal number counts from a 30 min scan, we get the following results as shown in table below.

Frame Number	Frame Interval (sec)	Frame Duration
1	0-278	278
2	279-564	286
3	565-859	295
4	860-1162	303
5	1163-1476	314
6	1477-1800	324

Table 4: Siemens Biograph 40, time intervals for six separate frames having equal number of counts

3.2.4 Data Acquisition - Biograph Duo

The Siemens Biograph Duo does not allow list mode data collection. In this case, six separate scans were taken for the NEMA NU-2 and CTN phantoms respectively. Because the scans were on Biograph Duo, which requires several minutes of preparation between successive scans, each individual acquisition was separated by additional time intervals. Hence, the duration of successive scans were increased in accordance with the increasing activity of the radioactive tracer to keep the number of counts equal for all frames. The increasing time interval between acquisitions of two scans varies, as it is affected by factors sometimes outside the control of the operator. The actual scan durations for the Duo acquisitions are shown in Table [5].

Using the scan durations defined in Table [5], we performed six separate scan acquisitions to achieve an equal number of counts for the NEMA NU-2 phantom for the Siemens Biograph Duo. The time gap between successive scans is the time delay caused by the scan preparation before starting a new acquisition.

Frame Number	Frame Interval (min)	Frame Duration (sec)
1	0-5	300
2	10-15.33	319
3	20-25.65	339
4	30-36	360
5	40-46.45	387
6	50-56.9	414

Table 5: Siemens Biograph Duo, time intervals for six separate frames having equal number of counts using NEMA NU-2 phantom

Similarly running six separate scans using Siemens Biograph Duo we collect six frames of CTN oncology phantom. The time intervals from these scans are shown in the Table [6] below.

Frame Number	Frame Interval (min)	Frame Duration (sec)
1	0-5	300
2	23-28.8	348
3	36-42.3	378
4	83-94.3	678
5	99-111.5	750
6	115-128.9	834

Table 6: Siemens Biograph Duo, time intervals for six separate frames having equal number of counts using CTN phantom.

Unlike planar imaging where individual events are stored in a (x, y) matrix, the PET system stores its coincidence events in the form of a sinogram where a sinogram is the 2-D representation of the projection data in (r, q) coordinates where r is the distance of Line of Response (LOR) from the center of scan field and q is the angle between LOR and vertical field axis. The LOR is the line between the two detectors where an event occurred. Each sinogram in PET represents the data acquired for one slice across all angles. PET data is acquired directly in the sinogram space, which acts like a histogram for LOR in (r, q) coordinates. Each LOR constitutes a single element in the sinogram matrix. Whenever the PET scanner detects a coincident event, its LOR is identified and the count is incremented in sinogram space for that particular event.

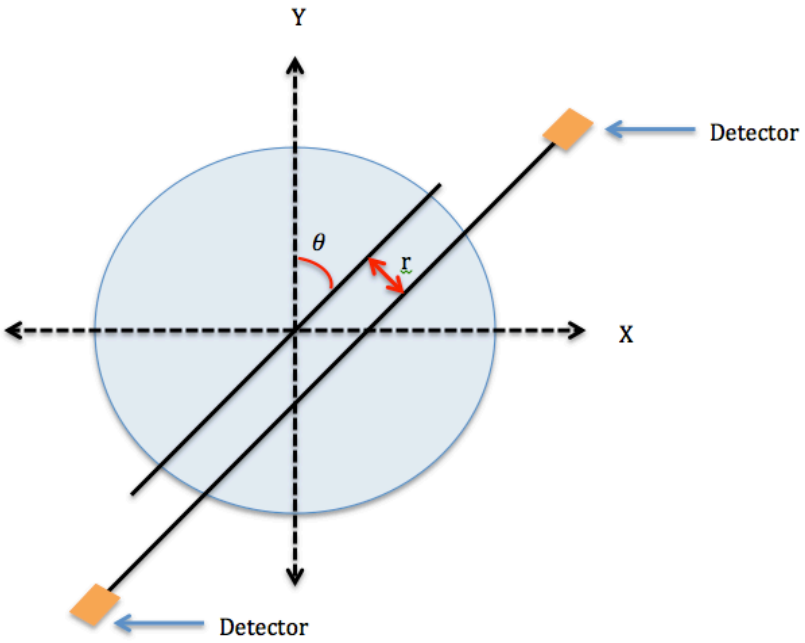


Figure 14: Defining the LOR in terms of q and r.

3.3 Reconstruction methodologies

3.3.1 Iterative Reconstructions

PET images have historically been reconstructed using Filtered Back Projection. As explained in Section 2.1.3, in FBP a 2D Fourier transform is applied to the projections data from the sinograms obtained from the scanner, which are then multiplied with a high frequency filter, usually a ramp filter or a ramp filter variant, in frequency space. Obtaining the inverse Fourier transform of this product and adding them yields the final reconstructed image. Applying the ramp filter is the mathematically correct operation, and would be optimal given noise free data. In the case of Computed Tomography the number of counts are large and FBP can be adequately implemented to reconstruct the image. However, the ramp filter is suboptimal in the PET case because the stochastic noise from the count limited PET scans result in frequency independent white noise whose high frequency components will be amplified by the ramp filter. The problem of high frequency noise amplification is primarily handled by rolling off the ramp filter at higher frequencies. Hanning and Hamm filters are an example of this. However, increasing computing power in modern-day PET systems have allowed for so-called iterative reconstructions.

Because the PET data set is not an ideal data set (it has noise), there is never a unique solution to the reconstruction problem, but a number of possible solutions. We can approach an optimal solution by applying an iterative algorithm, which provides the best fit with measured data. The Iterative reconstruction methods are well suited for these applications.

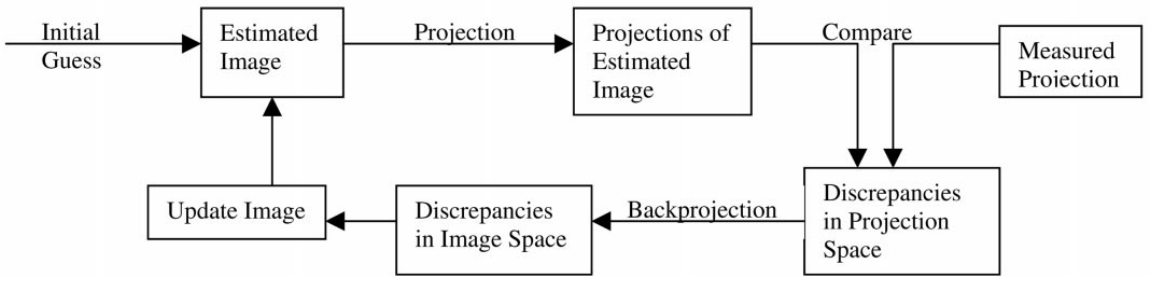


Figure 15: Flow chart of iterative image reconstruction scheme

Source: http://www.guillemet.org/irene/coursem/simpletomo.pdf

The iterative techniques generally work by assuming an initial estimate of the activity distribution or a forward projection of the image, which serves as an estimated object. It could be any image e.g. an estimate of projection that we assume the detectors to measure from the object, an image with all pixel values 0 or an average pixel value image. However an accurate initial estimate of the object itself is recommended as it can drastically reduce the number of steps required to reach an optimal solution.

The patient is scanned and actual raw data is obtained serving as the measured projection data set. The estimated and the measured projections are compared to find the discrepancy between the two. The difference between the two images is used to construct a corrected image, which is back projected and multiplicative or additive corrections are applied to it. The updated estimate becomes the starting point for next iteration and the process is repeated until the image begins to converge and a desired result is achieved. This iterative reconstruction process is common to all the iteration reconstruction algorithms.

3.3.2 Ordered Subset Expectation Maximization (OSEM)

OSEM is the most commonly used reconstruction method for PET. In the ordered subset or block-iterative method the projections are divided into subsets or smaller images, which are chosen in a specific order. We divide the image into subsets so as to make the algorithm fast and maximize its likelihood to achieve an optimal solution. Processing a subset of projection is called a sub-iteration.

Expectation-maximization is an iterative method, which attempts to match the most consistent model with the data. It initially makes assumptions about the distribution of the activity by using an initial model based on filtered back projection for the same object during the first iteration and uses the result from previous iteration in subsequent steps and then compares it to the measured data. Maximum likelihood i.e. closeness to the actual distribution, is achieved by using a multiplicative approach in which the previous estimate is multiplied by the back projection of the ratio of measured over estimated projections. The data is iterated and reiterated until an accurate model is found. The method proceeds by iterating and updating the successive subsets of the projection data until all subsets are processed. The likelihood increases with the number of iterations, to some extent after which it starts getting noisy.

The scans for this work were acquired with a Siemens Biograph 40 and a Siemens Biograph Duo. They were reconstructed using an array of different OSEM reconstruction parameters. The software system provides some flexibility in defining the parameters like iterations, subsets, full width half maximum filter size, matrix size, and pixel size. In our experiment scan duration was also a preliminary variable, as we wanted to assess the impact of statistical noise of image quality and texture, although this is technically not a reconstruction parameter.

Depending upon the choice of reconstruction parameters we can end up with drastically different looking images and drastically different quantitative characteristics

from the same initial data set. The parameters such as iteration, subsets and full width half maximum (FWHM) of a Gaussian smoothing filter effect the image in various ways, which is explained later in the chapter.

3.3.3 Iteration

As discussed above, OSEM is an iterative reconstruction approach in which a mathematical function is applied to an input, which itself is the output of a previous iteration operation using same function. Iteration can be applied repeatedly until the desired result is obtained. In clinical image reconstructions for PET imaging, the number of iterations is a user specified number ranging from 2-20. If the reconstruction is over-iterated, the image tends to become noisier without enhancement of accuracy of the reconstruction.

Iterative reconstructions typically include the correction factors like attenuation correction and scatter in their iterative loop. The major drawback of iterative methods over filtered back projection is it that it is more computationally expensive and time consuming in processing the data. The other drawback is that an image can be over-iterated resulting in image divergence and noise enhancement. In an OSEM reconstruction the user is required to specify the number of iterations the reconstruction will perform. To a point, more iteration will bring you closer to the true distribution.

3.3.4 Subsets

The other user-specified variable in OSEM implementation is the number of subsets. The use of subsets considerably decreases the reconstruction time and makes it feasible to apply OSEM in daily clinical routine. Using the subset method, we divide the projections into subsets or smaller images. Many images cover a large area, while the actual area being studied can be a small portion of the entire image. Dividing the images

into subsets saves both disk space and processing time. For example, if we acquired 64 projections from a patient, we can subdivide it into 8 subsets with each subset containing 8 images. The subsets should contain equal number of projections to ensure convergence of algorithm. The algorithm is individually applied to all the subsets with each subset undergoing equal number of iterations respectively. An iteration of OSEM is complete when it has passed through all the subsets. Using subsets increases the convergence time by a factor approximately equal to the number of subsets.

3.3.5 Gaussian Filtration

Image resolution for a PET imaging system is determined by the full width at half maximum (FWHM) of a Gaussian fit of a point-spread function (PSF) - the reconstructed image of a small (<1mm) point of radioactivity. Standardized NEMA specification requires the reconstruction to be performed with a simple FBP reconstruction with ramp filtration.

It is virtually universal in the field, whether using FBP, or an iterative approach, to apply a post-reconstruction Gaussian function with a defined FWHM for smoothing of the image to make the image easier to interpret, clinically. It acts like a filter and smoothens the image by suppression of high frequency signal and enhancement of low frequency signal. Smoothing is applied as a function of original data and the data points surrounding it. Application of a smoothing filter impacts quantitation as well as image quality by increasing the amount of Partial volume averaging (PVE) by blurring the boundary of structures

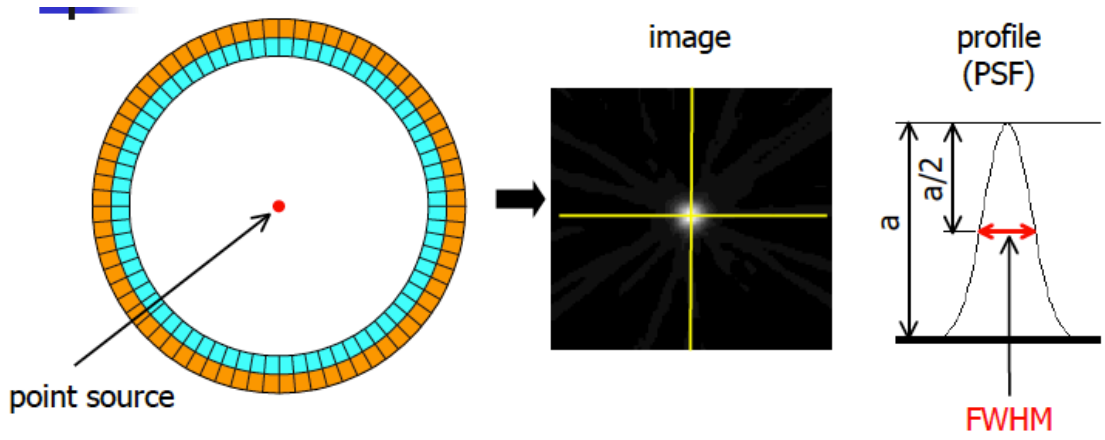


Figure 16: FWHM Description

Source: Image from CYRIC Tohoku University
http://kakuigaku.cyric.tohoku.ac.jp/PETAnalysis/PET_ABC_v2-1.pdf

3.4 Qualitative texture analysis methodology

3.4.1 Gray-Level Co-occurrence Matrix (GLCM)

The GLCM is a representation of how frequent a pixel combination occurs in an image. It is a matrix defined over an image characterizing spatial inter-relationship between pixels at a specific distance and orientation. It is normalized to make the sum of all occurrences equal to 1 to facilitate joint probability distribution calculations. Different GLCMs representing various relations between pixels can be created depending on the specified spatial relationships. A large number of features can be derived from the resulting matrices. GLCMs are a critical intermediate calculation to help with texture analysis.

3.4.2 Creating a GLCM

The figure below represents a grey level image created by a 6X6 matrix. Different shades of grey represent the various pixel intensity values, 1 depicting the darkest or black and 8 depicting white.

5	4	3	2	1	2
6	8	3	2	1	4
1	1	3	2	4	1
3	2	1	1	1	1
4	2	1	1	1	3
3	3	5	4	2	6

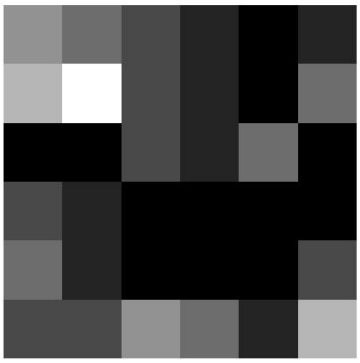


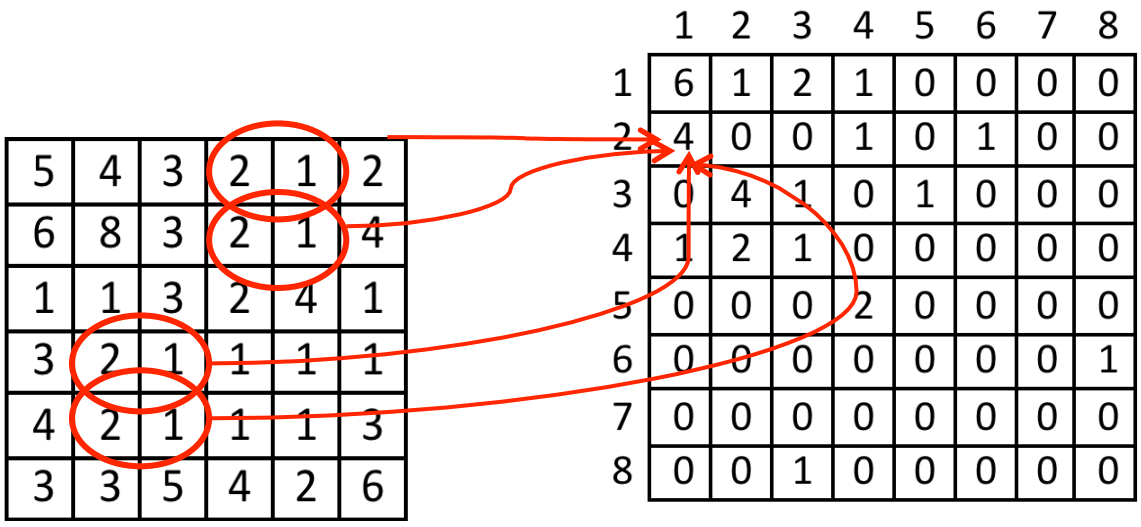
Figure 17: An image represented as integers and grey levels

To construct a GLCM for an image we need to create a matrix from the image that histograms particular spatial relationship among pixels by computing the frequency of occurrence.

As an example we will compute the GLCM for the 6X6 image described above. We will calculate the frequency of occurrence of a pixel of interest i with a specific intensity value to its neighboring pixel j at a distance of 1 pixel in the direction of 0 degrees. As shown in Figure [18], there are 6 occurrences of pixel (1, 1) being adjacent

and horizontally aligned to each other. GLCM being a histogram for the co-occurring value in an image, the matrix cell value at (1, 1) in the GLCM will therefore be 6. Similarly there are 4 occurrences of values (2, 1) being in a similar relation which will yield a value 4 at cell (2, 1) in the GLCM (Figure [18]). The entire GLCM is created from the image matrix in this fashion. Elements in the GLCM matrix thus represent the frequency of occurrence or the number of instances when a pixel i is present in a spatial relationship with pixel j.

The size of the GLCM is dependent on the number of grey levels or the intensity levels present in an image. In the example explained above, size of resulting GLCM is 8X8 due to the 8 intensity levels present in the image.



Creating GLCM Resulting GLCM

Figure 18: Creating GLCM matrix

6	1	2	1	0	0	0	0
4	0	0	1	0	1	0	0
0	4	1	0	1	0	0	0
1	2	1	0	0	0	0	0
0	0	0	2	0	0	0	0
0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0

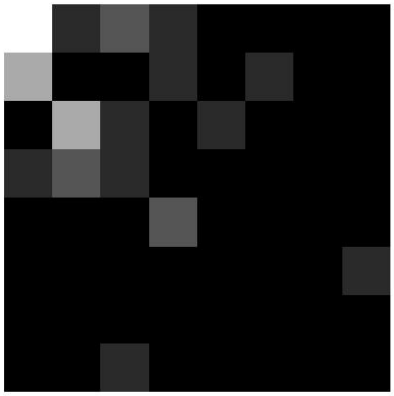


Figure 19: Resulting GLCM as integers and grey levels

3.4.2.1 GLCM in MATLAB

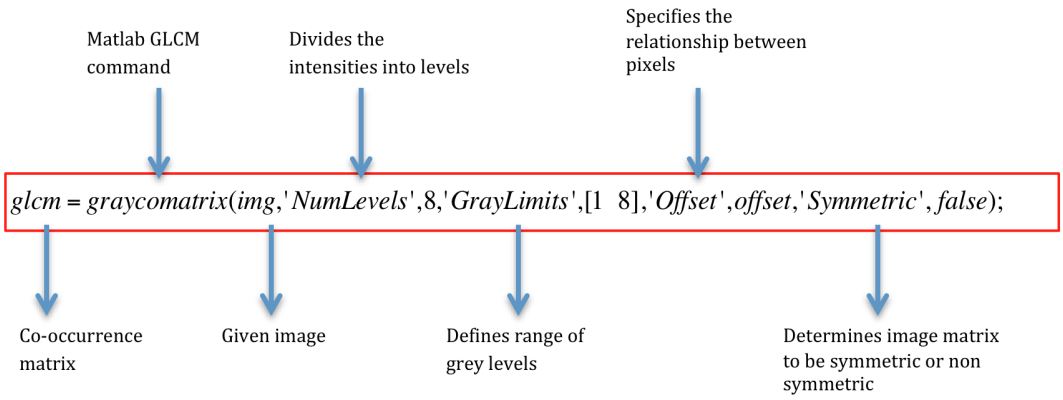
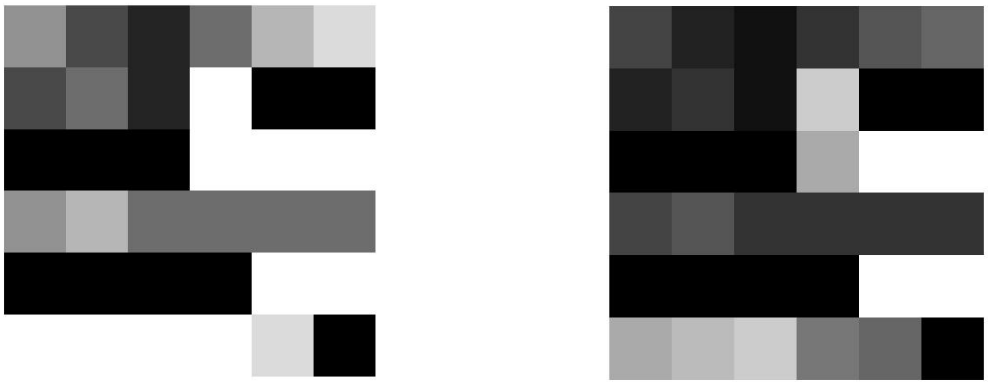


Figure 20: GLCM in MATLAB

MATLAB provides a special function ‘graycomatrix’ to generate the GLCM from a given image. It also provides a number of parameters, associated with the command, which can be altered in order to create a GLCM with user defined spatial inter relationships among pixels. The various parameters that govern a GLCM in MATLAB are:

3.4.2.2 Number of Grey Levels

NumLevels is the parameter that specifies the number of grey levels in which an image will be scaled. It determines the size of resultant co-occurrence matrix. We can specify any number of grey levels. For example, if the NumLevels parameter in the GLCM is 16, graycomatrix will scale the values in image to be integers between 1 and 16.



8 grey levels 16 grey levels

Figure 21: Image depicting 8/16 grey levels

Figure [21] is the grey level description of an example image in different scales with NumLevels 8 and 16 respectively. The figure clearly shows the difference between the two variations with different scales used.

3.4.2.3 Grey Limits

This is a vector with two elements [low, high] that decides how the intensity values in an image will be linearly scaled into gray levels. Gray scale values that are less than or equal to low are scaled to 1 whereas the values greater than or equal to high are scaled to NumLevels.

3.4.2.4 Offset

Offset is the parameter that defines the relationship between the pixel of interest and its neighbor at varying distance and direction. It is not limited to a single direction or distance. In MATLAB, an array of integers [row offset, col_offset] specifies the offset. 1 pixel offset would consider a pixel of interest and its immediate neighbor in a specified direction. Similarly a pixel offset of 2 would consider the pixels at 1 pixel distance apart in a specified direction. Row offset is the number of rows between the pixel of interest and its neighbor whereas the column offset is the number of columns between the pixel of interest and its neighbor. The relation between the pixel of interest and its neighbor could be in 8 possible directions i.e. N, S, E, W, and the four diagonals in between.

The following example creates an offset specifying four directions and 4 distances for each direction.

```
Offsets = [0 1; 0 2; 0 3; 0 4;
          -1 1; -2 2; -3 3; -4 4;
          -1 0; -2 0; -3 0; -4 0;
          -1 -1; -2 -2; -3 -3; -4 -4];
```

The Figure [22] illustrates the spatial relationships of pixels, where D represents the distance from the pixel of interest.

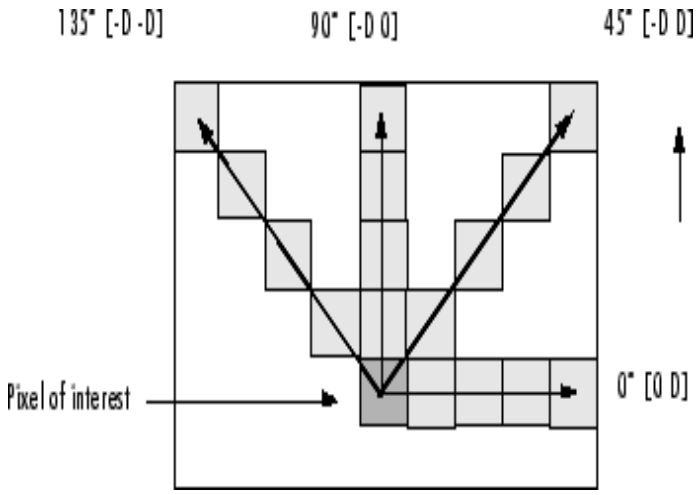


Figure 22: Defining Offset

SOURCE: MATLAB help topics

3.4.2.5 Symmetric

A symmetric matrix is one in which same values occur on the opposite sides of diagonal. It is a GLCM parameter that decided if the ordering of the values in pixel pairs would be considered or not. It is a Boolean with two possible values of 'true' and 'false'. For example, during the calculation to find the number of times the value 1 is adjacent to a value 2, if 'Symmetric' parameter is set to true, a GLCM will count both 1,2 and 2,1 pairings whereas if the 'Symmetric' is set to false, GLCM will only count 1,2 or 2,1, depending on the value of 'offset'.

3.4.2.6 Creating a GLCM using MATLAB

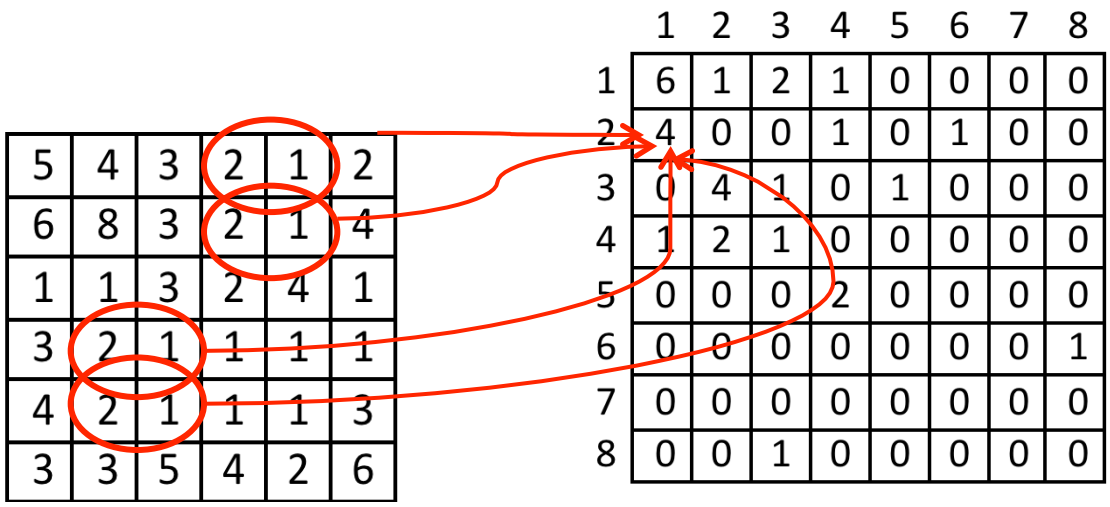
Reiterating the GLCM example in terms of MATLAB commands, we have the task of computing the GLCM for the 6X6 image. We will calculate the frequency of occurrence of a pixel of interest i with a specific intensity value to its neighboring pixel j at a distance of 1 pixel in the direction of 0 degrees. To create this GLCM in MATLAB we have to use the MATLAB command 'greycomatrix'.

With the value of offset defined as

```
offset = [0  1]
```

The following command is used to find the value of GLCM

```
glcm = graycomatrix(img,'NumLevels',8,'GrayLimits',[1  8],'Offset',offset,'Symmetric',false);
```



Creating GLCM Resulting GLCM

Figure 23: Creating GLCM matrix

In the abovementioned command, the value for parameter NumLevels is 8 i.e. the image will be scaled to 8 grey levels when calculating GLCM. The value for parameter GreyLimits is [1 8] which will set the lowest grey value to 1 and the highest grey value to 8 i.e. any value below 1 will be scaled to one and any value above 8 will be considered as 8. The offset is defined as [0 1] that means the pixel of interest or reference pixel i is at a spatial distance of one from the pixel j in the horizontal direction or 0 degrees. The symmetric parameter is set to 'false' hence the ordering of pixel will be taken into account.

3.4.3 Deriving Statistics from a GLCM

The GLCM is a tool that we use to help us more easily calculate meaningful texture metrics as described below. Graycoprops is a function defined in MATLAB that uses the GLCM to calculate other texture metrics. There are four texture metrics generated by the command, all used in this work: Contrast, Homogeneity, Correlation and Energy

```
Stats = graycoprops(glcm1)
```

The abovementioned command generates these metrics from the GLCM for an image.

For example

The following command will yield the contrast value of the GLCM

```
mean(Stats.Contrast)
```

Similarly we can find the correlation value of GLCM

```
mean(Stats.Correlation)
```

Energy Value

```
mean(Stats.Energy)
```

Homogeneity value for the image

$$mean(Stats.Homogeneity)$$

The physical and mathematical meaning of these parameters is described below.

3.4.4 Texture Properties

Texture properties can isolate and quantitate different image pixel spatial relations. For this study we have considered the following five texture properties. [8]

3.4.4.1 Contrast

Contrast is the local change in brightness and is defined as the ratio between average brightness of an object and the background.

$$= \frac{Io - Ib}{Ib}$$

I_o = Average brightness of Object

I_b = Average brightness of background

In mathematical terms

$$Contrast = \sum_{n=0}^{Ng-1} n^2 \left(\sum_{n} p(i,j) \right) = \sum_{i,j=0}^{Ng-1} |i-j|^2 p(i,j)$$

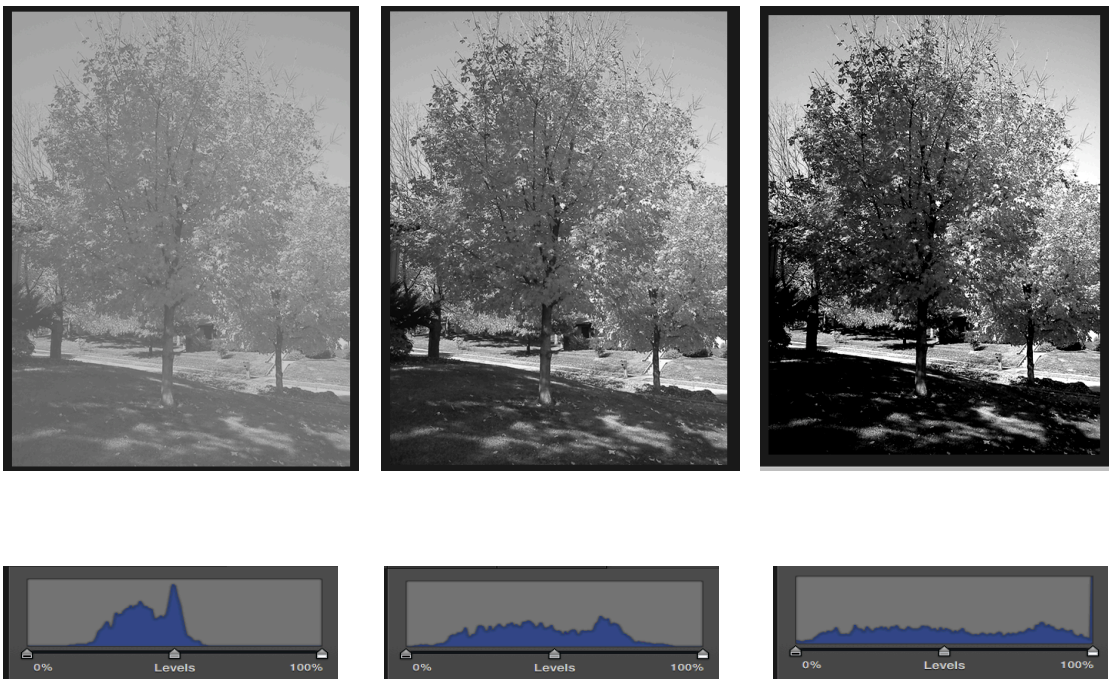
$$p(i,j) = (i,j)^{th} \text{ Entry in a normalized GLCM matrix} = \frac{P(i,j)}{R}$$

R= Normalization factor.

Ng = Number of distinct gray levels in the quantized image.

$n = |i - j|$

The image in Figure [24] shows an example when we move from low contrast to high contrast.



(a) Low Contrast C=334.2777

(b) Normal Contrast C=420.57

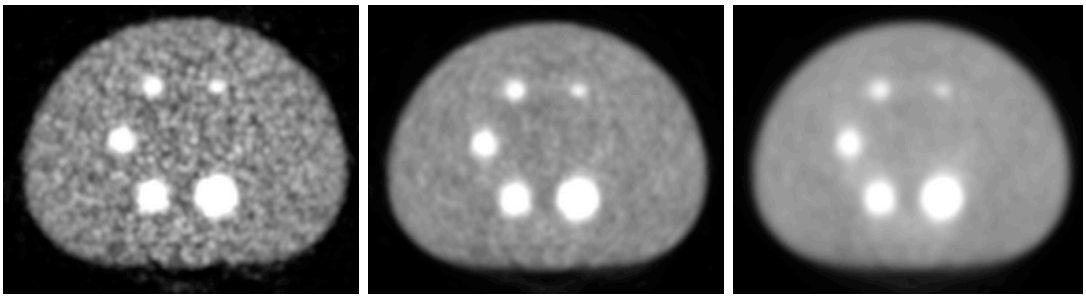
(c) High Contrast C=798.42

Figure 24: Images depicting different Contrast Levels

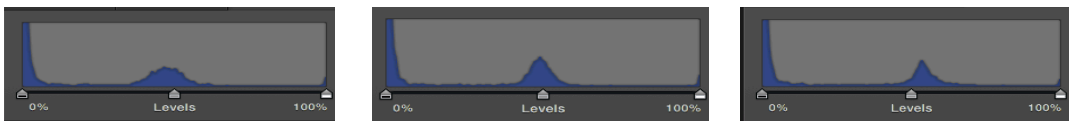
As the value of contrast increases so does the range of pixel values being utilized in the image. The contrast value sums the product of the individual values of the normalized GLCM matrix by the square of the difference between i and j. That is, on diagonal elements of the GLCM matrix are multiplied by zero (i-j) and therefore, de-emphasized. Off diagonal elements where (i-j) is large (which corresponds to pixels with large differences between them in proximity to one another), generate proportionally larger contributions to the calculated contrast value due to the

$$= \sum_{i,j=0}^{Ng-1} |i-j|^2 p(i,j)$$

Functional nature of the definition of contrast



NEMA-6I-16S-5FWHM NEMA-4I-4S-5FWHM NEMA-2I-4S-7FWHM



Contrast max =902.65 Contrast med =724.82 Contrast min =470.42

Figure 25: Phantom scans depicting different Contrast Levels

The histograms reflect the range of pixels in the images above. The image with the broad histogram depicts a wider range of pixels and hence the highest contrast, whereas image with a narrow histogram depicts smaller range of pixels and hence less contrast. Therefore, images with pixel values with large differences in close proximity to one another (a more granular image) will generate higher contrast values while smoother images that, by definition, will only have gradually increasing or decreasing pixel values will generate lower contrast values.

Mathematical Description

Let P be a test image,

Test =

0	1	2	3
2	2	2	2
1	0	1	0
3	2	1	2

With an offset value of [0 1] and Boolean symmetric being 'True', we are looking for the pixel adjacent combinations that are one pixel space apart in East and West directions.

The resulting normalized GLCM is

p(i, j) =

0	0.166666667	0	0
0.166666667	0	0.125	0
0	0.125	0.25	0.083333333
0	0	0.083333333	0

To find Contrast we need to multiple p (i, j) with $|i-j|^2$ where i and j are the respective row and column numbers.

For a 4X4 matrix the value of $|i-j|^2$ or the weights used are

0	1	4	9
1	0	1	4
4	1	0	1
9	4	1	0

Multiplying the GLCM matrix p(i, j) with the scalar $|i-j|^2$ and summing the resulting values will yield the value of contrast.

$$Contrast = \sum^{Ng-1} |i-j|^2 p(i,j)$$

$$=0*(0-0)^2+0.1667*(0-1)^2+0*(0-2)^2+0*(0-3)^2+0.1667*(1-0)^2+0*(1-1)^2+$$
$$0.125*(1-2)^2+0*(1-3)^2+0*(2-0)^2+0.125*(2-1)^2+0.25*(2-2)^2+0.0833*$$
$$(2-3)^2+0*(3-0)^2+0*(3-1)^2+0.0833*(3-2)^2+0*(3-3)^2$$
$$=0.7500$$

The diagonal elements of the normalized GLCM represent pixel pairs with no grey level difference. The higher probabilities in these elements will represent an image with identical neighboring pixels and hence less contrast. The value of contrast increases with the distance from the diagonal.

3.4.4.2 Energy

Returns the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment. It is a measure of uniformity, which describes the structural variation in images. Images with low structural variations are more uniform and have high-energy values.

$$Energy = \sum_{i,j=0}^{Ng-1} \{p(i,j)\}^2$$

where p(i,j) is the GLCM.

The range of energy is 0 to 1.

Energy for a constant image = 1

High value of energy occurs when the image is very orderly with minimal variations.

The orderliness of an image describes how regular its pixel values are. Orderliness of an image is described in the following example.

Image 1 =

1	1	1
1	1	1
1	1	1

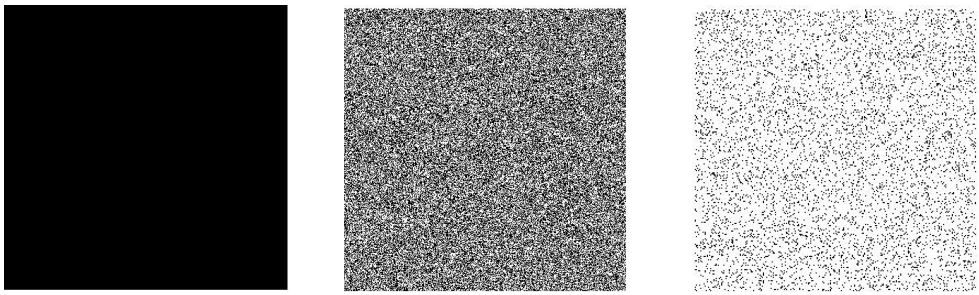
Image 2 =

1	2	3
1	2	3
1	2	3

Image 3 =

3	5	2
4	0	4
2	1	3

In the images above, Image 1 is more orderly as each pair of values are same. In Image 2 a similar arrangement of pixels occur many times, for example 2 is next to 1 three times, 3 is next to 2 three times, etc. Image 2 is less orderly than Image 1. Image 3 is least orderly as the combinations occur less often and image appears more random. Image 1 will return a higher energy value than Image 2 and Image 3.



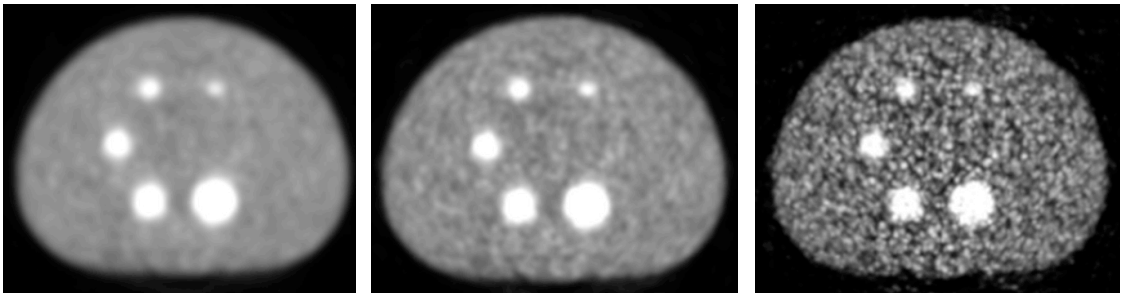
Energy=1 Energy= 0.25 Energy =0.01

Energy =16E-04 Energy =9.45E-05

Figure 26: Images depicting different Energy Levels

In Figure [27], the image on the left has low structural variations and is smoother in comparison with other two images. The energy value for this image is the highest. The image in the middle has higher structural variations, which reduces its energy. The image

on far right has highest structural variations and is least smooth and hence has the minimum energy value.



NEMA-4I-4S-7FWHM Energy max = 6.02E-04

NEMA-3I-8S-5FWHM Energy med = 1.950E-03

NEMA-6I-16S-3FWHM Energy min = 3.70E-03

Figure 27: Phantom scans depicting different Energy Levels

Using the same test image

Test =

0	1	2	3
2	2	2	2
1	0	1	0
3	2	1	2

With an offset value of [0 1] and Boolean symmetric being 'True', we are looking for the pixel adjacent combinations that are one pixel space apart in East and West directions.

The resulting normalized GLCM is

p(i, j) =

0	0.166666667	0	0
0.166666667	0	0.125	0
0	0.125	0.25	0.083333333
0	0	0.083333333	0

$p(I, j)^2$ =

0	0.027777778	0	0
0.027777778	0	0.015625	0
0	0.015625	0.0625	0.006944444
0	0	0.006944444	0

Using the formula to calculate Energy

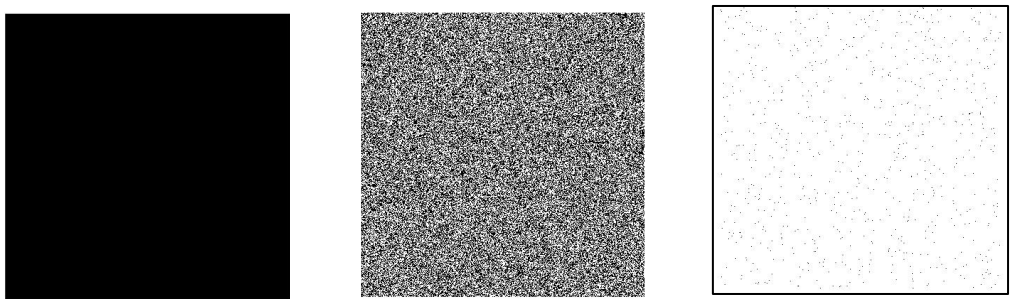
$$Energy = \sum_{i,j=0}^{Ng-1} \{p(i,j)\}^2$$

=0.0278 + 0.0278 + 0.0156 + 0.0156 + 0.0625 + 0.0069 + 0.0069

= 0.1632

3.4.4.3 Homogeneity

Also known as Inverse Difference moment it measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal or the dynamic range of pixels in an image. It has maximum value when all elements in the image are same.



Homogeneity =1 Homogeneity =0.5 Homogeneity =0.369

Homogeneity =0.3044 Homogeneity =0.2584 Homogeneity =0.2373

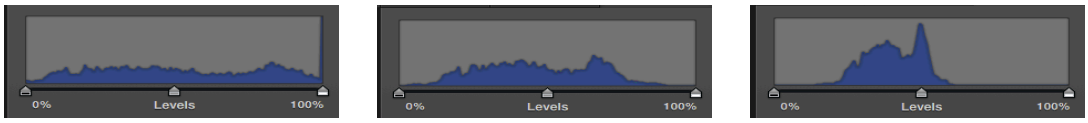
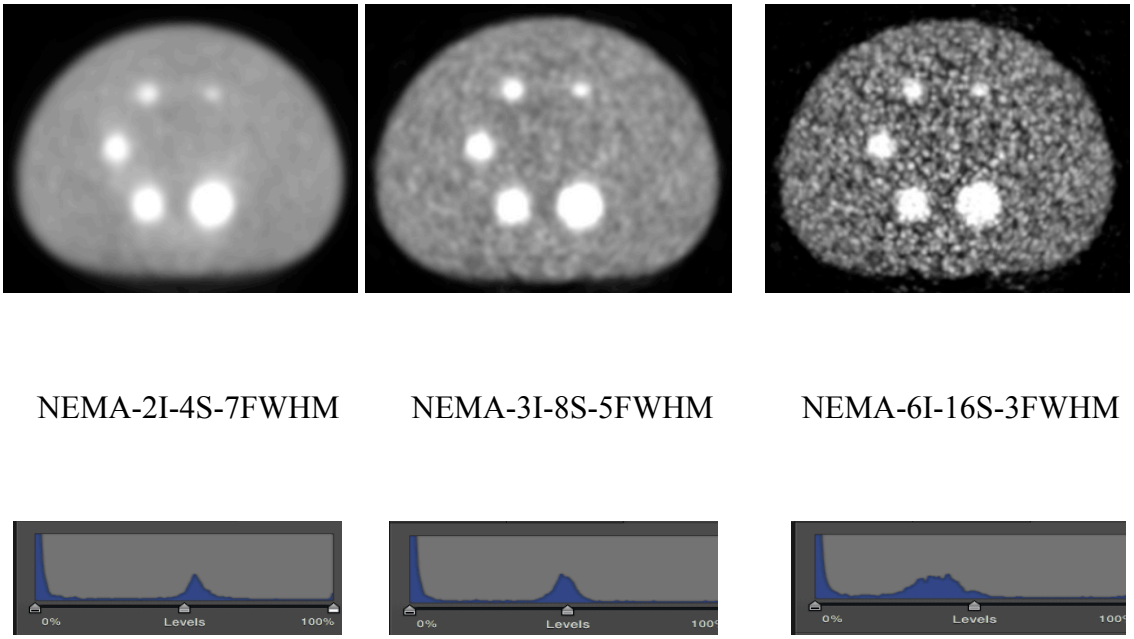


Figure 28: Images depicting different Homogeneity Levels

GLCM contrast and homogeneity are strongly, but inversely, correlated in terms of equivalent distribution in the pixel pair population. Its value ranges from 0 to 1.

$$Homogeneity = \sum_{i,j=0}^{Ng-1} \frac{p(i,j)}{1+(i-j)^2}$$

The Homogeneity value is 1 for a GLCM with diagonals equal to 1.



Homogeneity max=3.36E-01 Homogeneity med=2.48E-01 Homogeneity min=1.52E-01

Figure 29: Phantom scans depicting different Homogeneity Levels

A homogeneous image contains fewer gray levels, giving a GLCM only a few

values of p(i, j) and hence a close distribution. The image on the far left has lesser gray levels and hence a higher homogeneity. Image on the far right has more grey levels and a lesser homogeneity.

Using the same test image

Test =

0	1	2	3
2	2	2	2
1	0	1	0
3	2	1	2

Normalized GLCM for the test image is calculated as

p(i, j) =

0	0.166666667	0	0
0.166666667	0	0.125	0
0	0.125	0.25	0.083333333
0	0	0.083333333	0

To calculate Homogeneity the factor or weight by which the normalized GLCM is multiplied is

$$\frac{1}{1+(i-j)^2}$$

Homogeneity Weights =

1	0.5	0.2	0.1
0.5	1	0.5	0.2
0.2	0.5	1	0.5
0.1	0.2	0.5	1

Using the mathematical relation

$$Homogeneity = \sum_{i,j=0}^{Ng-1} \frac{p(i,j)}{1+(i-j)^2}$$

= 0.167*0.5 + 0.167*0.5 + 0.125*0.5 + 0.125* 0.5 + 0.25*1 + 0.0833*0.5 + 0.0833*0.5

= 0.6250

3.4.4.4 Correlation

It returns the value of how correlated an image pixel is to its neighbor over the entire image. It is a measure of linear dependency of grey levels to its neighboring pixels.

The neighboring pixels are defined by the parameter offset, which was explained in the last section. Offset is a user-defined parameter, which is determined in consideration with the image size. If the image size is large we can use a larger offset and will have larger neighboring pixels with respect to the reference pixel. The range of correlation value is from -1 to 1. Correlation is not defined for a constant image.

$$Correlation = \sum_{i,j=0}^{Ng-1} p(i,j) \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \right]$$

To calculate correlation we need to find GLCM mean and GLCM variance

GLCM mean is represented by the equations

$$\mu_i = \sum^{Ng-1} i * p(i,j) \text{ and } \mu_j = \sum^{Ng-1} j * p(i,j)$$

For our test image $\mu_i = \sum^{Ng-1} i * p(i,j)$

= 0*0.1667 + 1*0.1667 + 1*0.1250 + 2*0.1250 + 2*0.25 + 2*0.0833 + 3*0.0833

$= 1.4583$

GLCM variance is defined as

$$\sigma_i = \sum_{i,j=0}^{Ng-1} p(i,j)(i-\mu_i)^2 \text{ and } \sigma_j = \sum_{i,j=0}^{Ng-1} p(i,j)(i-\mu_j)^2$$

For out test image variance $\sigma_i = \sum^{Ng-1} p(i,j)(i-\mu_i)^2$

$=0.1667*(0-1.45832)^2 + 0.1667*(1-1.45832)^2 + 1.25*(1-1.45832)^2 + 0.125*$

$(2-1.45832)^2 + 0.25*(2-1.45832)^2 + 0.0833*(2-1.45832)^2 + 0.0833*(3-1.45832)^2$

$=0.7465$

Using the mathematical relation for correlation

$$Correlation = \sum_{i,j=0}^{Ng-1} p(i,j)\left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i\sigma_j}\right]$$

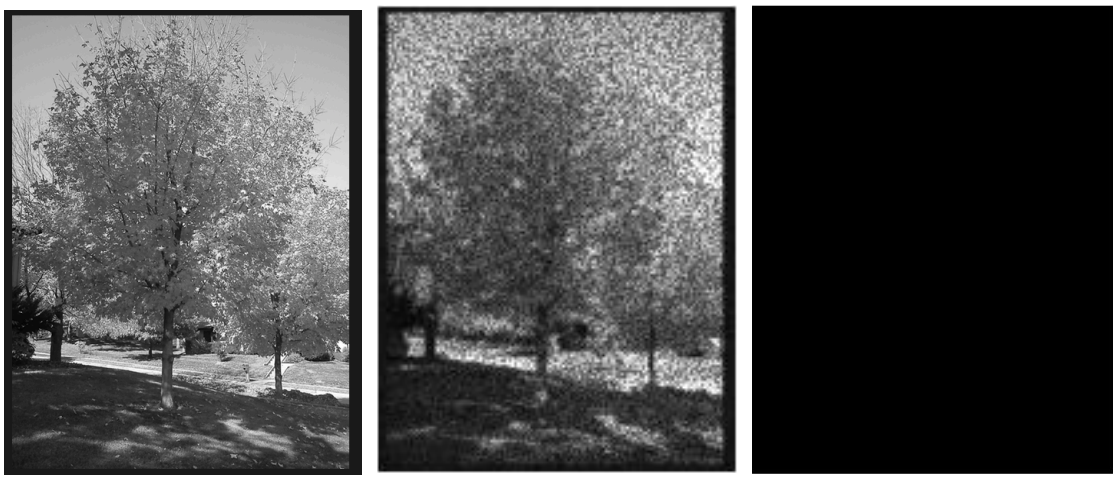
$$= \left[\frac{0.1667(0-1.4583)(1-1.45683)}{0.7465}\right] + \left[\frac{0.1667(1-1.4583)(0-1.45683)}{0.7465}\right] +$$

$$\left[\frac{0.125(1-1.4583)(2-1.45683)}{0.7465}\right] + \left[\frac{0.125(2-1.4583)(1-1.45683)}{0.7465}\right] +$$

$$\left[\frac{0.25(2-1.4583)(2-1.4583)}{0.7465}\right] + \left[\frac{0.0833(2-1.4583)(3-1.4583)}{0.7465}\right] +$$

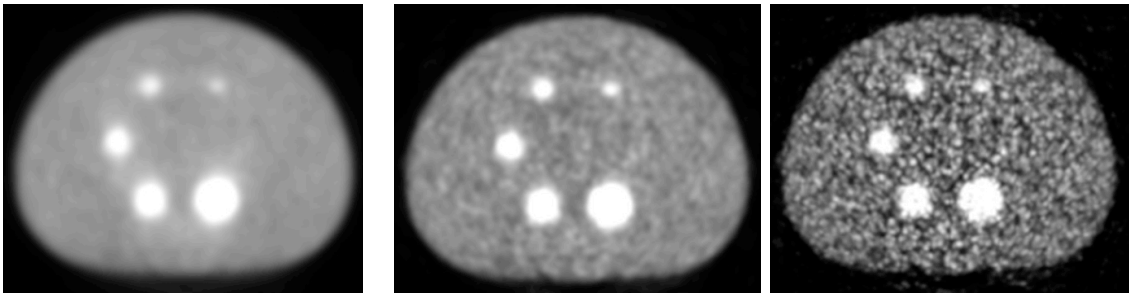
$$\left[\frac{0.0833(3-1.4583)(2-1.4583)}{0.7465}\right]$$

$=0.4988$



Correlation =0.9310 Correlation =0.7540 Correlation =Not Defined

Figure 30: Images depicting different Correlation Levels



NEMA-2I-4S-7FWHM NEMA-3I-8S-5FWHM NEMA-6I-16S-3FWHM

Correlation max = 9.00E-01 Correlation med = 8.53E-01 Correlation min = 7.45E-01

Figure 31: Phantom scans depicting different Correlation Levels

In the image on the far left, the pixel range is smaller and the GLCM values are closely distributed, and hence the image is highly correlated. In images on the right, the pixel range is broader reducing the correlation.

3.4.4.5 Entropy of grayscale image

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. The entropy of the random variable X is the sum, over all possible outcomes k of X, of the product of the probability of outcome x_k with the logarithm of the inverse of the probability of x_k. $\log_2(1/p(x_k))$. It is also called the surprisal of the outcome x_k.

In MATLAB entropy is defined in the following way

E = entropy (I)

Where E is a scalar value representing the entropy of grayscale image I.

I can be a multidimensional image. If I has more than two dimensions then the entropy function treats it as a multidimensional gray scale image and not as an RGB image.

The mathematical representation for Entropy is

$$Entropy = -\sum^{Ng-1} p(i,j)\ln(p(i,j))$$

Using the same test image from the previous examples, we will find the Entropy value of the image

Test =

0	1	2	3
2	2	2	2
1	0	1	0
3	2	1	2

Normalized GLCM for the test image is calculated as

p(i, j) =

0	0.166666667	0	0
0.166666667	0	0.125	0
0	0.125	0.25	0.083333333
0	0	0.083333333	0

ln(p(I, j) =

0	-1.7917	0	0
-1.7917	0	-2.0794	0
0	-2.0794	-1.386	-2.4849
0	0	-2.4849	0

Using the mathematical relation for entropy

$$Entropy = -\sum^{Ng-1} p(i,j)\ln(p(i,j))$$

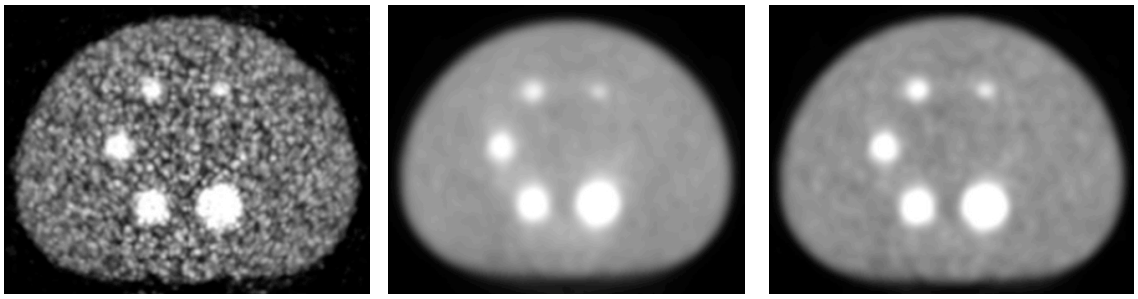
=0.1667*(1.7917) + 0.1667*(1.7917) + 0.125*(2.0714) + 0.125*(2.0714) + 0.25*(1.386) + 0.0833*(2.489) + 0.0833*(2.489)

=1.8761

The following example describes two different images with different entropy values. Entropy being the measure of randomness has a higher value in an image with more intensity values than an image with less intensity values.

Entropy =1 Entropy =7.34 Entropy =7.64

Figure 32: Images depicting different Entropy Levels



NEMA-6I-16S-3FWHM NEMA-2I-4S-7FWHM NEMA-4I-4S-7FWHM

Entropy max = 6.31 Entropy med = 6.05 Entropy min = 5.85

Figure 33: Phantom scans depicting different Entropy levels

The value of the entropy decreases with the increase in orderliness of the image. Disorderly and unstructured image that contains more information will have higher entropy. Image on the left is highly disordered and contains more information and hence has higher entropy. Images on the right are more orderly and structured and hence have lesser entropy values. An image that is perfectly flat will have zero entropy.

3.5 Quantitative harmonization in PET

As mentioned in Chapter 2.5, FDA is struggling for identical quantitative and qualitative data from different sites using different systems for multi center clinical trials. The reason for using 2 different scanners in this study was to generate harmonized PET reconstructions that would be identical in both quantitative and qualitative respects. Quantitative harmonization would be the ability to produce identical PET measurements between different make and model scanners by modifying their specific OSEM reconstruction parameters. This would require identification of PET/CT scanner make-and-model-specific OSEM reconstruction parameters to generate identical recovery coefficient (RC) curves.

The quantitative performance of a scanner is measured by recovery coefficient curves for the objects of different sizes. When small structures are scanned using a PET scanner their quantitative measurements are influenced by the background activity in surrounding structures. Background activity spills into the small structures, and activity from the small structure spills into the background. This combination of effects results in systematic and predictable bias in measurements. A metric called the "recovery coefficient" (RC) is a factor used to correct the bias from the partial volume effect induced in smaller structures.

There are many quantitative strategies to measure activity in a volume of interest in a PET scan. Using the maximum pixel value in a volume of interest is the simplest

and most common approach. But using a single pixel as the sole data point to represent a large volume makes the measurement highly sensitive to pixel noise. To diminish noise sensitivity, a larger volume of interest (VOI) can be used in lieu of the maximum value. The more pixels averaged within the VOI, the less the noise, but the more bias towards lower values.

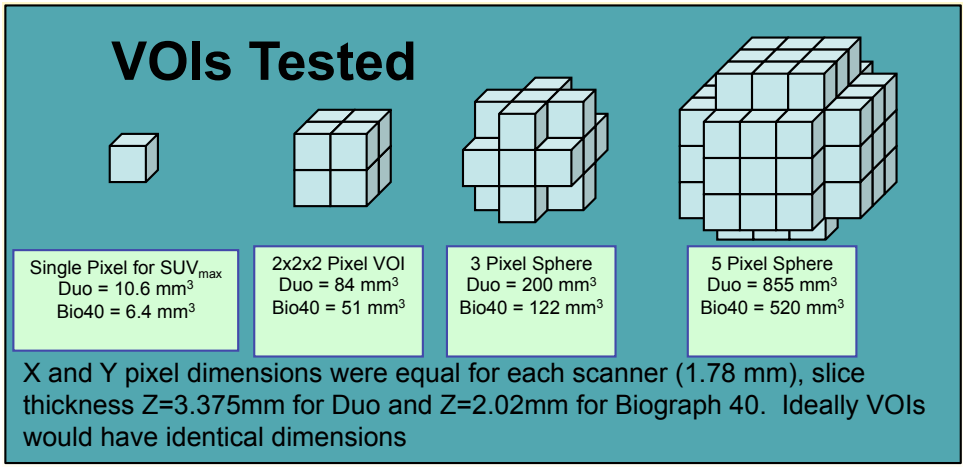


Figure 34: Volume of Interest used to measure Recovery coefficients

Various volumes, objects or measurements can be used to measure the quantitative performance of a scanner. Those used in our study are a 2X2X2 pixel Volume of Interest (VOI) cube, 3-pixel diameter VOI sphere, and 5-pixel diameter VOI sphere. The X and Y dimensions for Biograph 40 and Biograph Duo are same but they have different slice thickness, which result into different VOI sizes. Z=2.02 mm for

Biograph 40 and Z= 3.375 mm for Biograph Duo. As shown in the Figure [34] above, the volumes obtained using a 2X2X2 pixel VOI for Bio 40 and Bio Duo are 51 mm^3 and 84 mm^3 respectively and the volumes obtained using a 5X5X5 pixel VOI for Bio 40 and Bio Duo are 520 mm^3 and 855 mm^3 respectively.

In our sample calculations and plots below we have used the SUVmax value of the lesions as their activity values with respect to the reference activity. Actual results in the following chapter will also show results from the several VOIs defined above.

Changing the values of iterations, subsets, and Gaussian filters impact the quantitation in the reconstructed images, and therefore the shape of the recovery coefficient curves. These parameters have their individual and combined effects on the RC curves, which are discussed in detail in the section below.

Figure [35], below, depicts a PET scan of NEMA NU-2 phantom with 6 spheres and a central cylinder insert. The activity concentration in the spheres is filled as per NEMA phantom protocols discussed in Chapter 3.1.2. The cylinder being greater in diameter and covering the entirety of the axial length of the phantom is filled with the same stock solution as the spheres and is considered to be the reference concentration as it is large enough to be devoid of partial volume effect. All sphere concentration measurements are calculated with respect to the cylinder. In this particular scan, there is no background activity. The plot in Figure [35] illustrates radioactive concentration as measured by Siemens Biograph 40 scanner for a 30 min scan of the phantom. The smallest sphere has the highest partial volume effect, measuring in at approximately 32% of reference value. This percentage increases with the size of the sphere. The largest sphere depicts an activity of 98%, which is very close to the reference value.

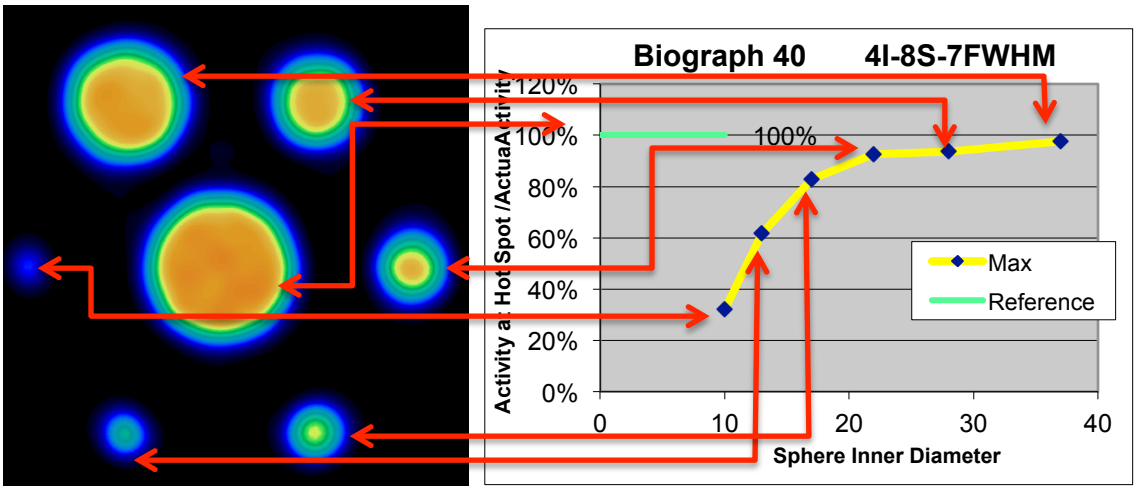


Figure 35: Distribution of Activity in a NEMA NU-2 phantom scanned by Siemens Biograph 40

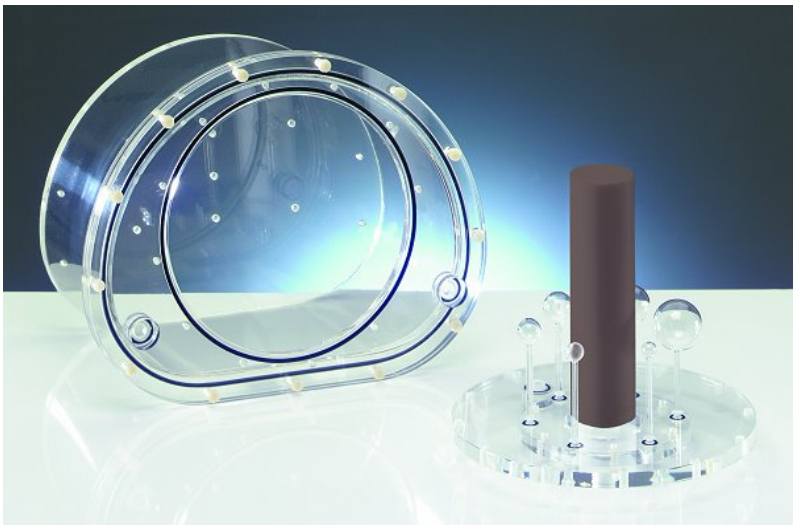


Figure 36: NEMA NU-2 phantom

Source: http://www.ptw.de/pet_emission_phantom_nema.html

3.5.1 Quantitative Trends in Reconstructions on the Biograph 40 with Increasing Iterations

In the following study NEMA NU-2 phantom undergoes a 30 min scan. Injected dose used is as specified by the NEMA protocols. Keeping the number of subsets and FWHM constant, image was reconstructed for different values of iterations. For the data shown in Figure [37], the number of subsets was kept constant at 4 and FWHM of the Gaussian filter at was fixed at 3mm while we varied the number of iterations in the OSEM reconstruction. In the Figure [37] we show the recovery coefficient curves while we vary the iterations in the reconstruction loop to 2, 3 and 8.

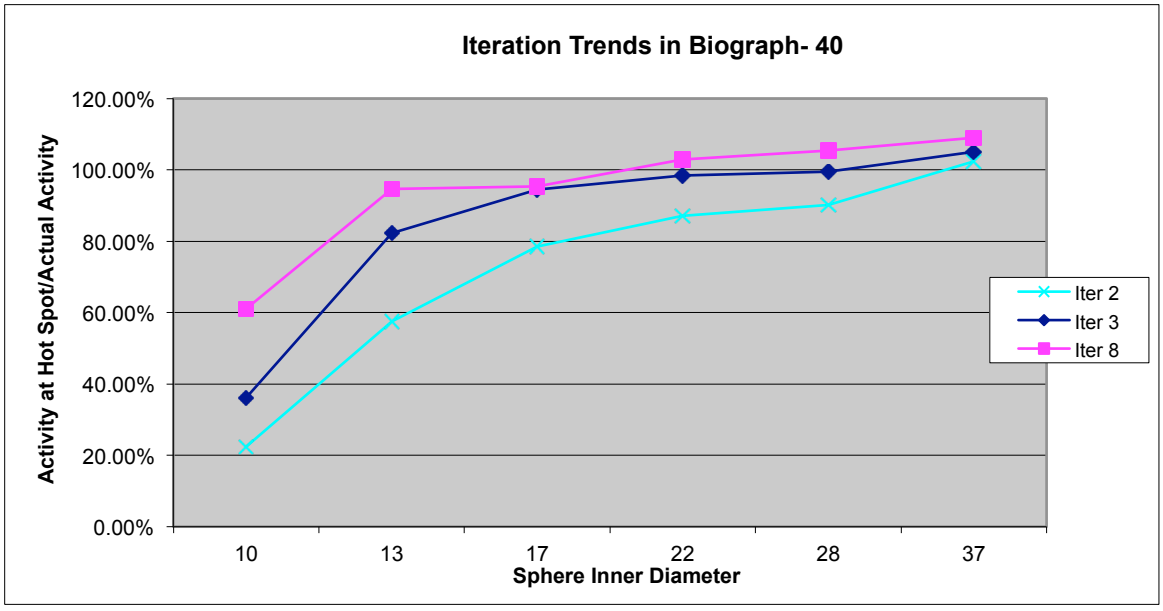
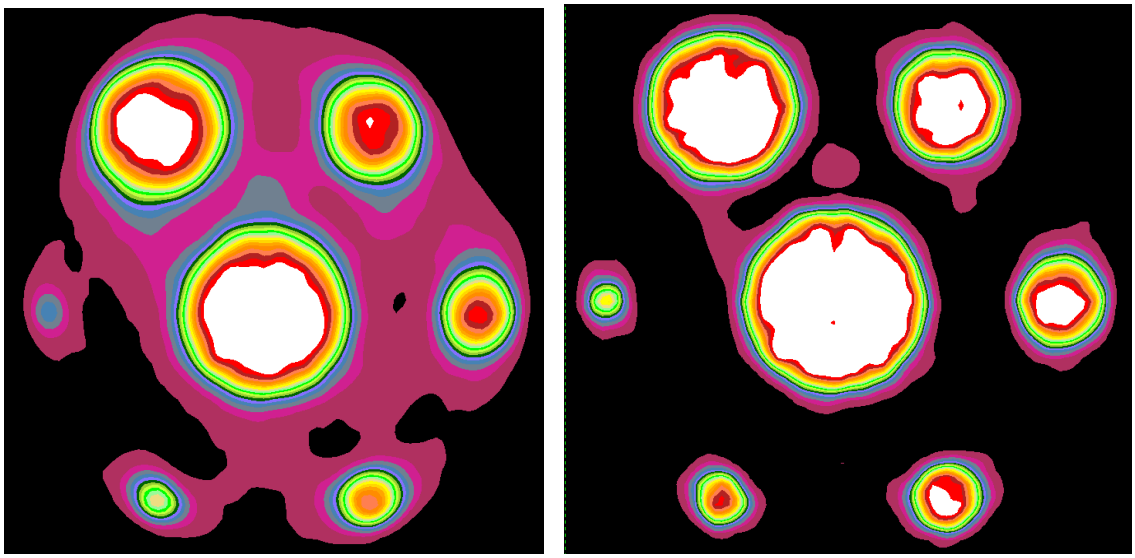


Figure 37: Trends in recovery coefficient curves with increasing iterations

Increasing the number of Iterations alone allows the reconstruction to converge upon more accurate values. This is particularly true for smaller objects, which typically converge more slowly, which is consistent with the literature on iterative OSEM reconstructions. Figure [38] illustrates a comparison between a 2 iteration image with an 8 iteration image demonstrating the increase in the amount of noise, but also the enhanced convergence with increasing iterations.



2I-4S-3FWHM **8I-4S-3FWHM**

Figure 38: Effect of changing number of iterations during image reconstruction for the Biograph 40

NOTE: Voxel intensity in both images range from 0-6000Bq/ml

3.5.2 Quantitative Trends in Reconstructions with Increasing Subsets

In the following study NEMA NU-2 phantom undergoes a 30 min scan. Injected dose used is as specified by the NEMA protocols. Keeping the number of iterations and FWHM constant, image was reconstructed for different values of subset. The value of iteration was kept constant at 3 and FWHM at 3. Values of subsets used were 4, 8, 16 and 21.

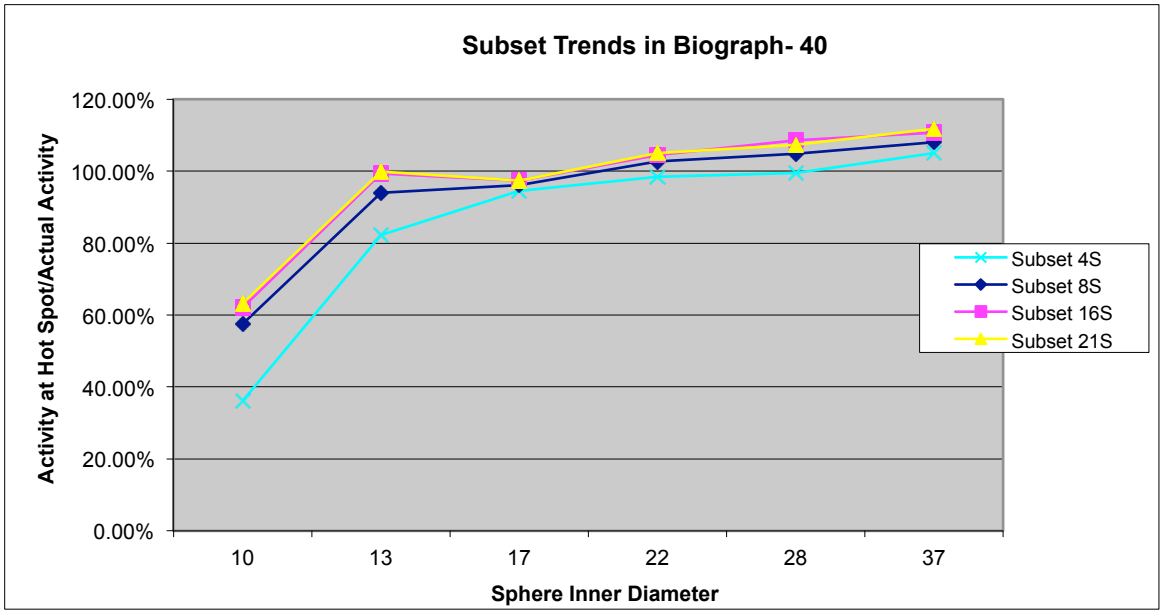


Figure 39: Trends in Biograph 40 with increasing subsets

Increasing the number of subsets alone shows an increase in the value of the measured activity from the reconstruction.

A similar study is performed using the Biograph Duo PET/CT. The NEMA NU-2 phantom undergoes a 30 min scan. Injected dose used is as specified by the NEMA protocols. Keeping the number of iterations and FWHM constant, the image was reconstructed for several subset values. The number of iterations was kept constant at 2 and the FWHM of the Gaussian filter was constant at 5mm. The numbers of subsets used were 8, 16 and 32

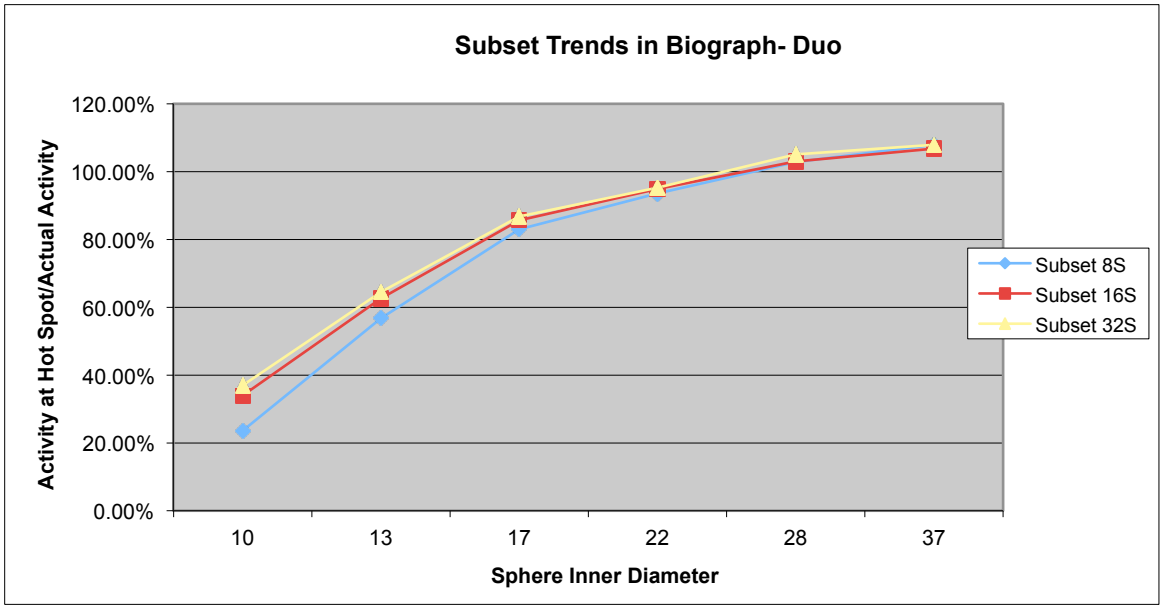
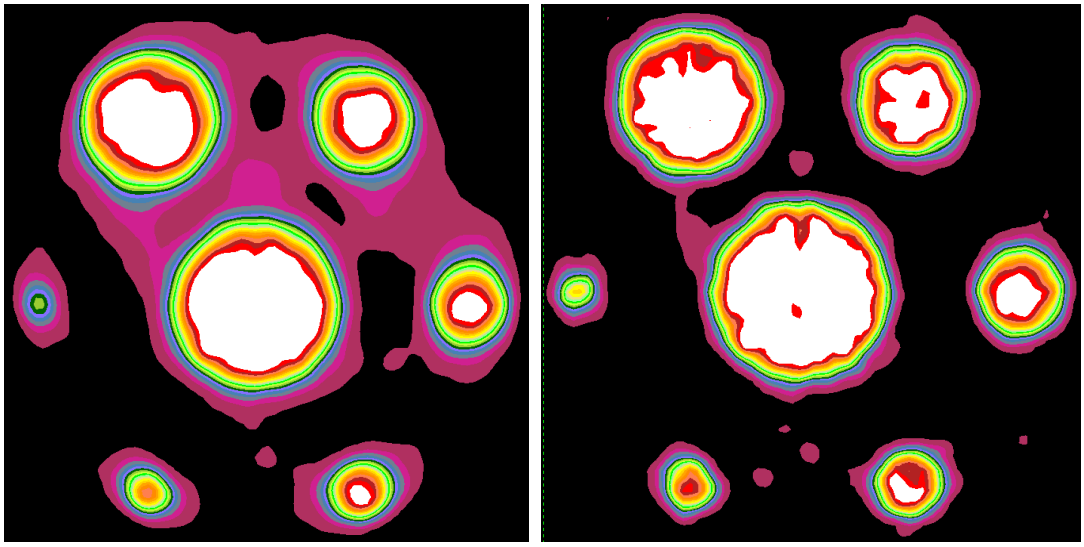


Figure 40: Trends in Biograph Duo with increasing subsets

Increasing the number of subsets alone shows an increase in the value of activity after reconstruction, but little change was observed by increasing the number of subsets above 32, with the 2 iterations used in this dataset.



3I-4S-3FWHM 3I-21S-3FWHM

Figure 41: Effect of changing number of Subsets during Image Reconstruction in Biograph 40

NOTE: Voxel Intensity of the Images range from 0-6000Bq/ml

Using four subsets with the 3-iteration reconstruction clearly does not allow the iterative reconstruction to fully converge, as demonstrated by the overly smoothed image

on the left in Figure [41]. Increasing the number of subsets enhances convergence significantly as shown on the right.

3.5.3 Quantitative Trends in Reconstructions with increasing FWHM

In the following study NEMA NU-2 phantom undergoes a 30 min scan using Biograph 40. Injected dose used is as specified by the NEMA protocols. Keeping the number of iterations and subsets constant, the image was reconstructed for different values of a Gaussian FWHM. The value of iteration was kept constant at 4 and subset at 8. Values of FWHM used were 2mm, 4mm, 5mm and 7mm.

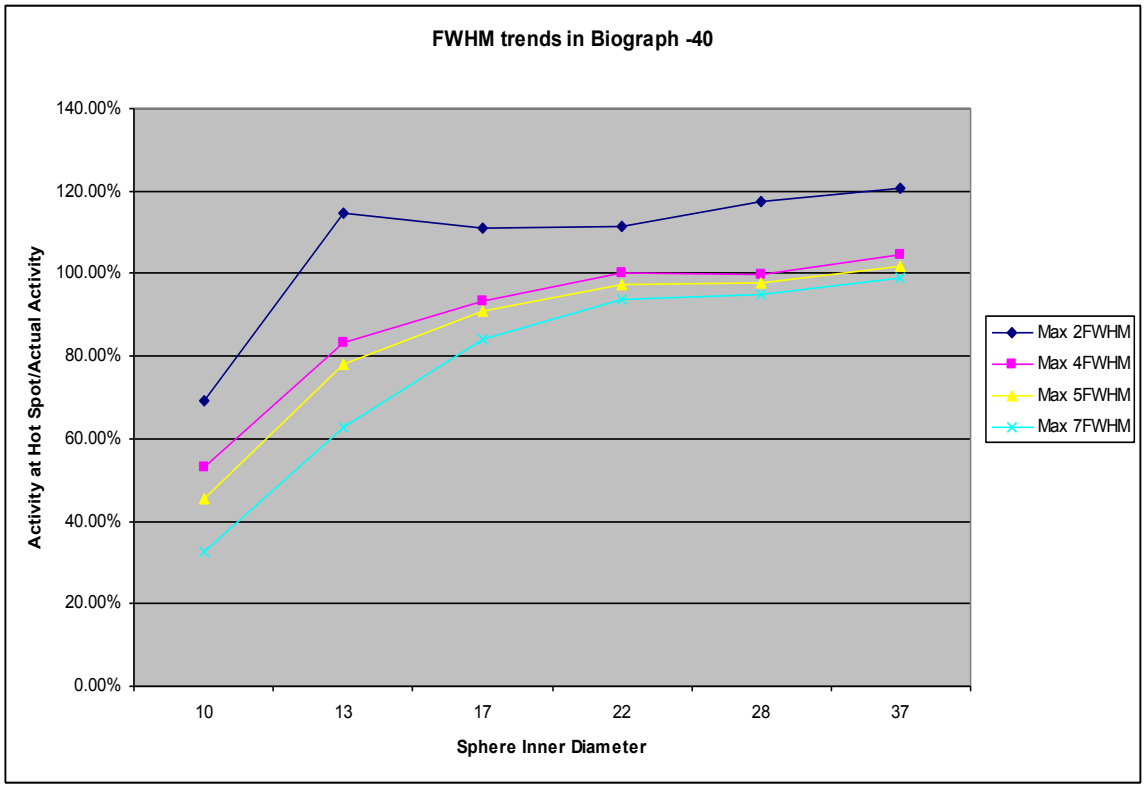


Figure 42: Trends in Biograph 40 with increasing FWHM

Increasing the value of FWHM used in reconstruction shows the expected decrease in the value of measured concentration after reconstruction. Also, as expected, the image becomes smoother with the increasing FWHM of the convolving filter.

A similar study was performed using NEMA NU-2 phantom using Siemens Biograph Duo. In the sample data presented below, the value of the number of iterations was kept constant at 2 with a constant 8 subsets. The FWHM of the convolving filter was 3mm, 5mm, 7mm and 8mm. The Siemens Biograph Duo demonstrated the expected similar response to increase in FWHM value.

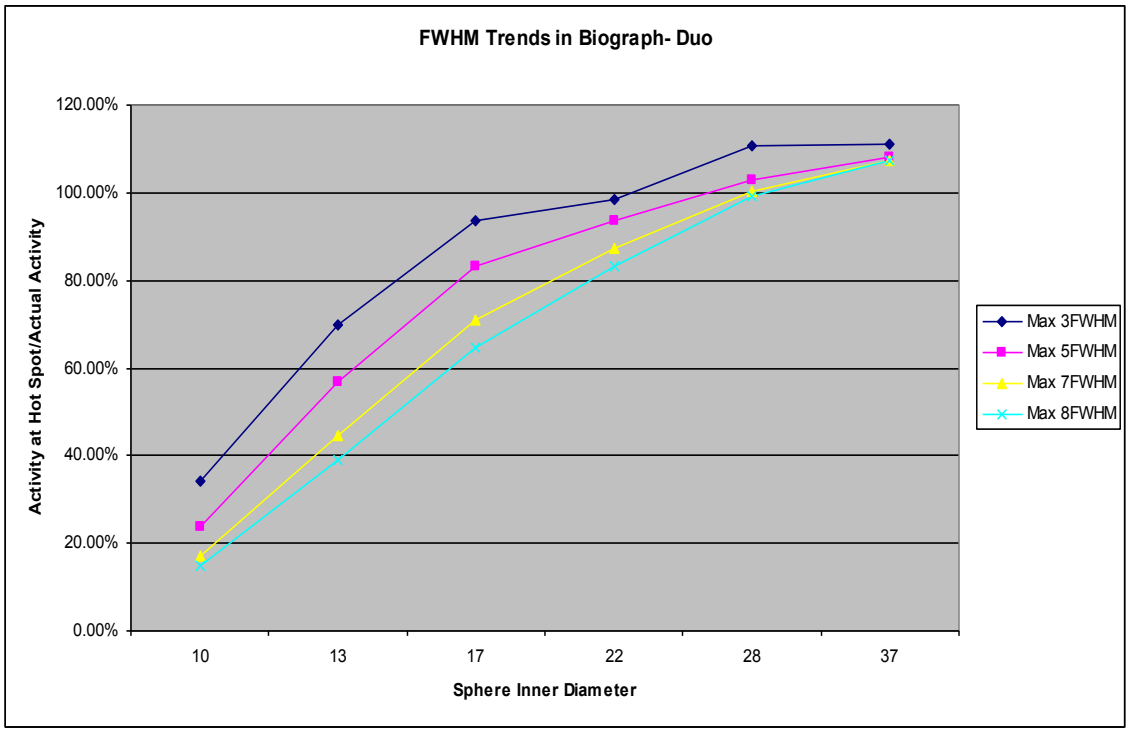
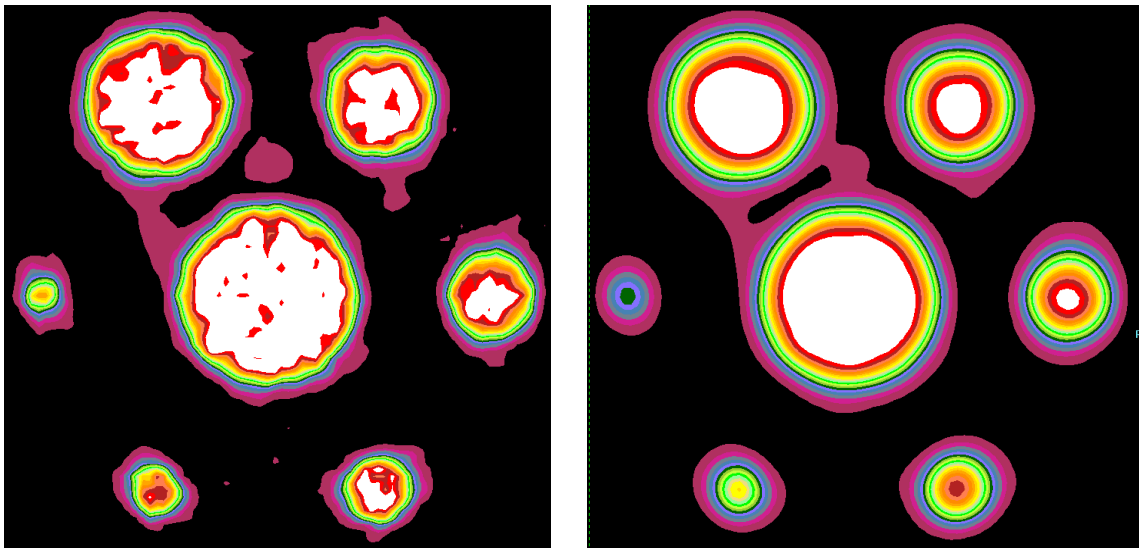


Figure 43: Trends in Biograph Duo with increasing FWHM



4I-8S-2FWHM 4I-8S-7FWHM

Figure 44: Effect of changing FWHM during Image Reconstructions in Bio 40

NOTE: Voxel Intensity of the Images range from 0-6000Bq/ml

CHAPTER IV

RESULTS AND CONCLUSIONS

We have two separate results

1) Quantitative Harmonization Results

2) Qualitative or Texture Analysis Results

4.1 Quantitative harmonization results

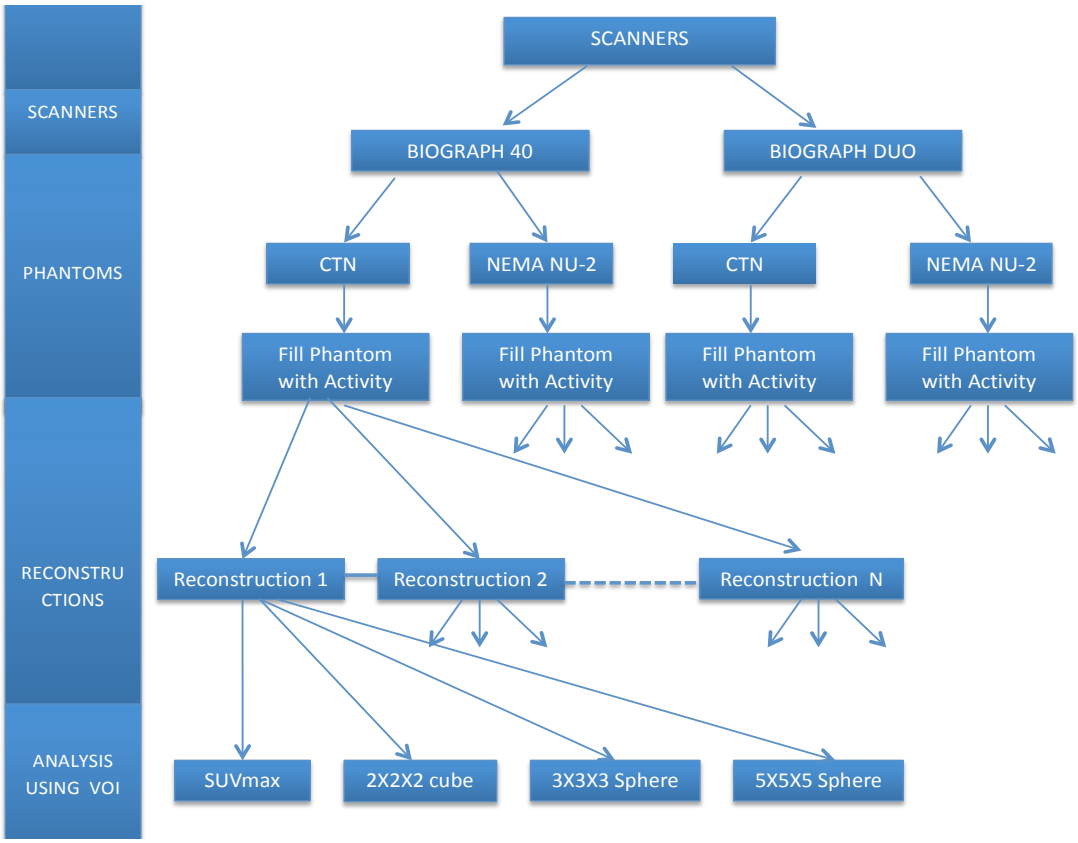


Figure 45: Stat Analysis tree describing an overview of the quantitative analysis

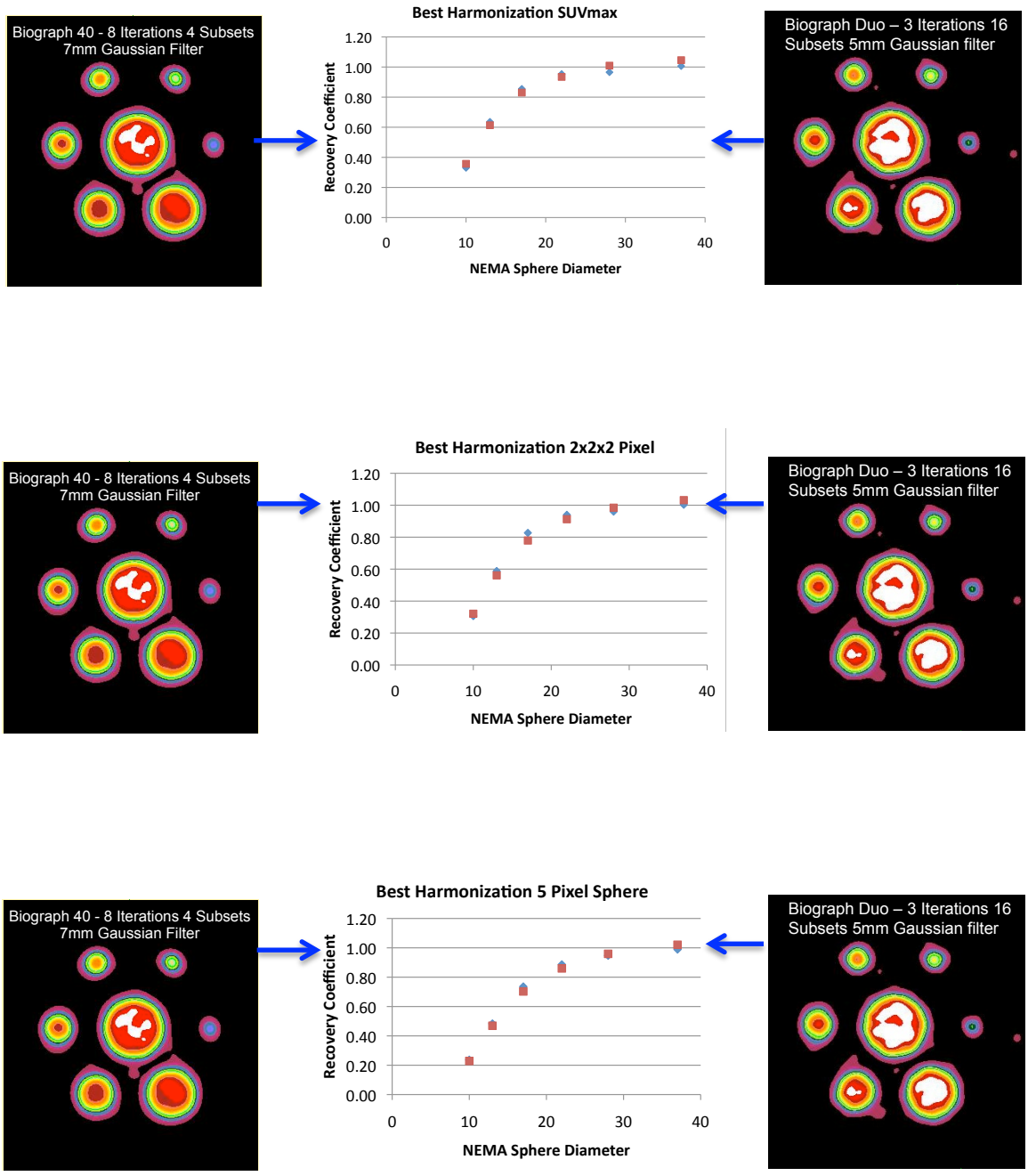


Figure 46: Quantitatively harmonized reconstructions with identical RC curves using NEMA NU-2 phantom with Biograph 40 and Biograph Duo

Quantitative harmonization between Biograph 40 and Biograph Duo was achieved by identifying scanner specific reconstruction parameters that optimally aligned their respective RC curves. For this purpose, we have used the VOIs as described in Chapter 3.5.4 for a comparable RC curve generation. For each of the VOI definitions, quantitatively harmonized reconstructions were defined as that pair of Biograph 40 and Biograph Duo reconstructions that resulted in Recovery Coefficient curves with minimized sum of the square differences.

Figure [46], describes the Biograph 40 (8 iterations, 4 subsets, 7mm Gaussian) reconstruction matched with Biograph Duo (3 iterations, 16 subsets, 5mm Gaussian) reconstruction when using SUVmax, 2X2X2 voxel VOI and 5X5X5 voxel VOI. In all cases, the optimally harmonized reconstructions included the lowest resolution reconstruction of the higher performing scanner (Biograph 40 - 8 iterations, 4 subsets, 7mm Gaussian).

The harmonization zones for three of the VOI metrics for Biograph 40 and Biograph Duo are plotted to determine an area of RC overlap described by the shaded region in Figure [47]. We get a large harmonization region using SUVmax and 3-pixel diameter spherical VOI whereas the harmonization region when using the larger 5-pixel diameter spherical VOI is very limited. It demonstrates that in some cases even the lower performing range of a higher performing scanner may exceed the RC performance of a lower resolution scanner under any reconstruction conditions. Plots describing RC curves with different VOIs, when using different reconstruction parameters with Biograph 40 and Biograph Duo are shown in Appendix A.

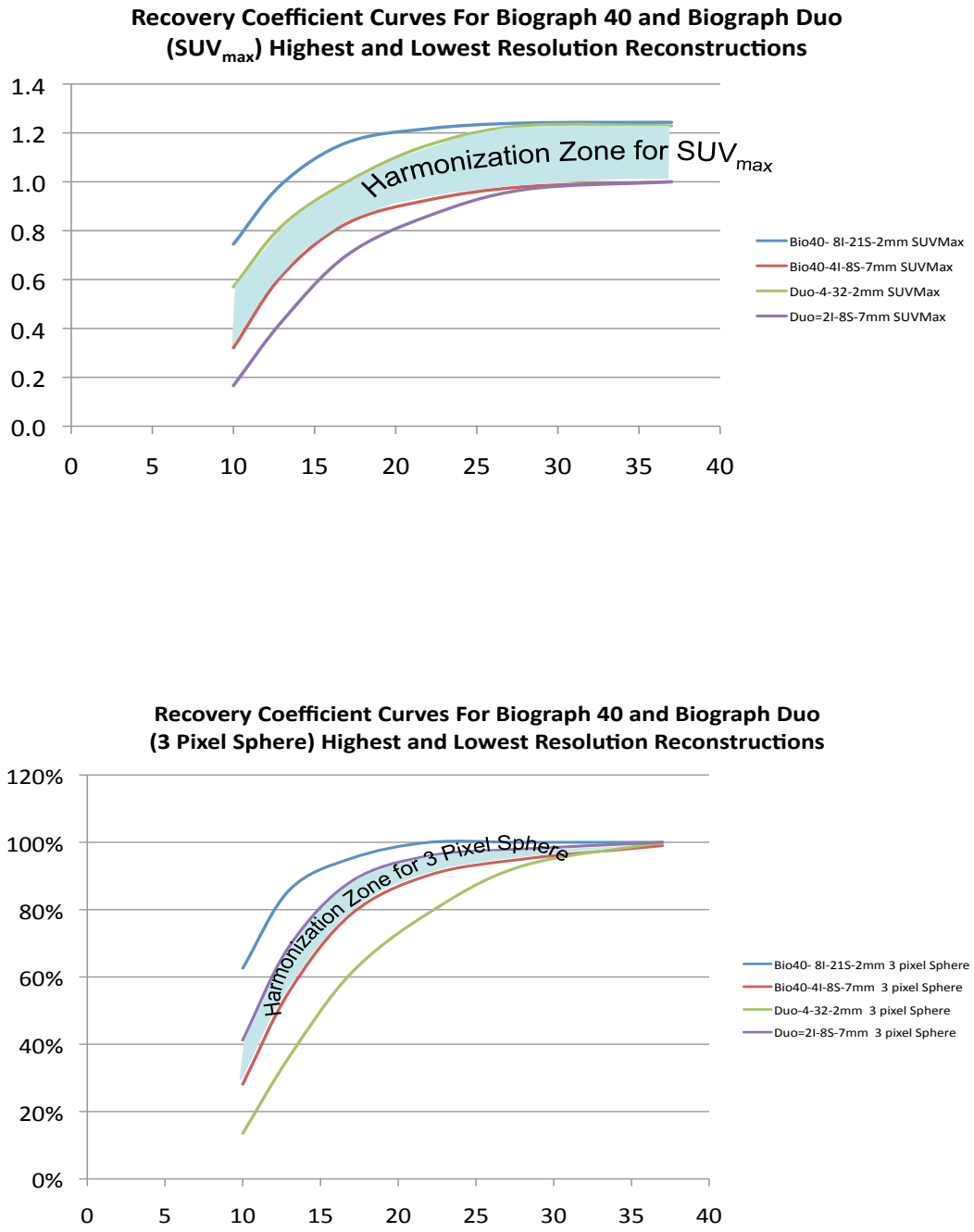
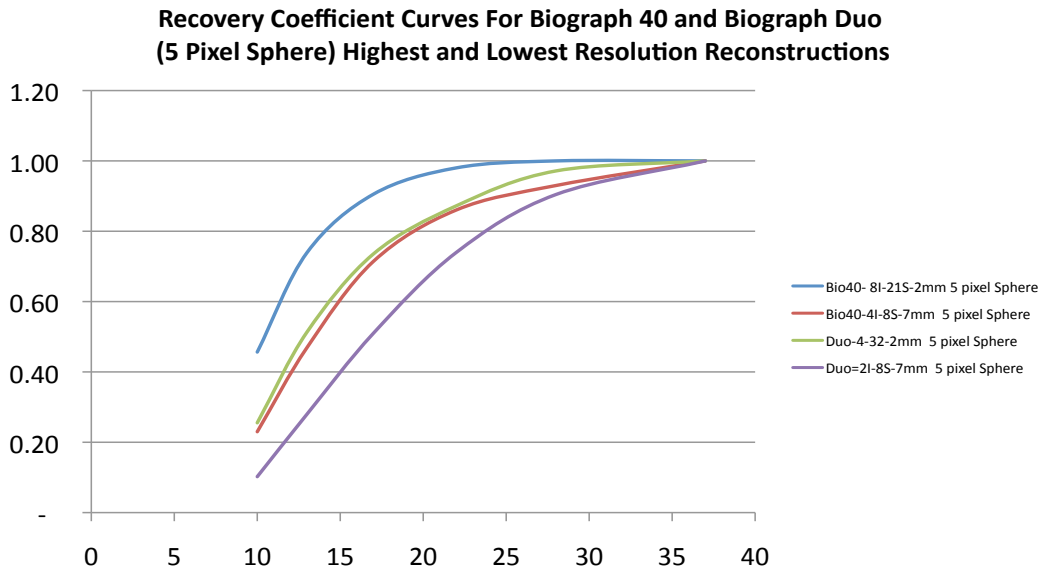


Figure 47: Harmonized RC curves for Biograph 40 and Biograph Duo for SUVmax, 3-pixel and 5-pixel VOI

Figure 47 – Continued



4.2 Qualitative or Texture analysis results

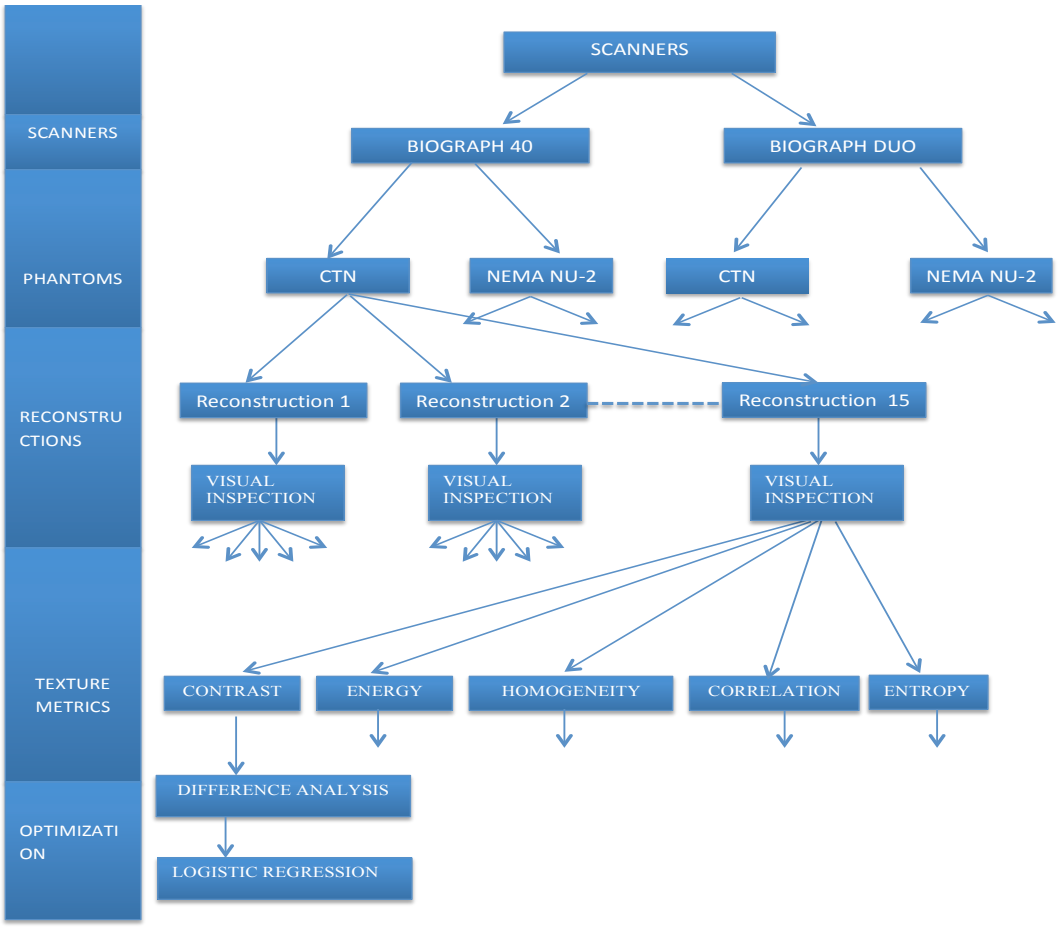


Figure 48: Stat Analysis tree describing an overview of the qualitative analysis

As mentioned in Chapter [3], to calculate texture parameters for a particular reconstruction a rectangular portion of the slice encompassing the lesions and a limited

amount of background was selected rather than the entire image. This selection resulted in texture parameter values not being biased due to an inordinately large background component. The images below illustrate the regions considered while calculating texture parameters.

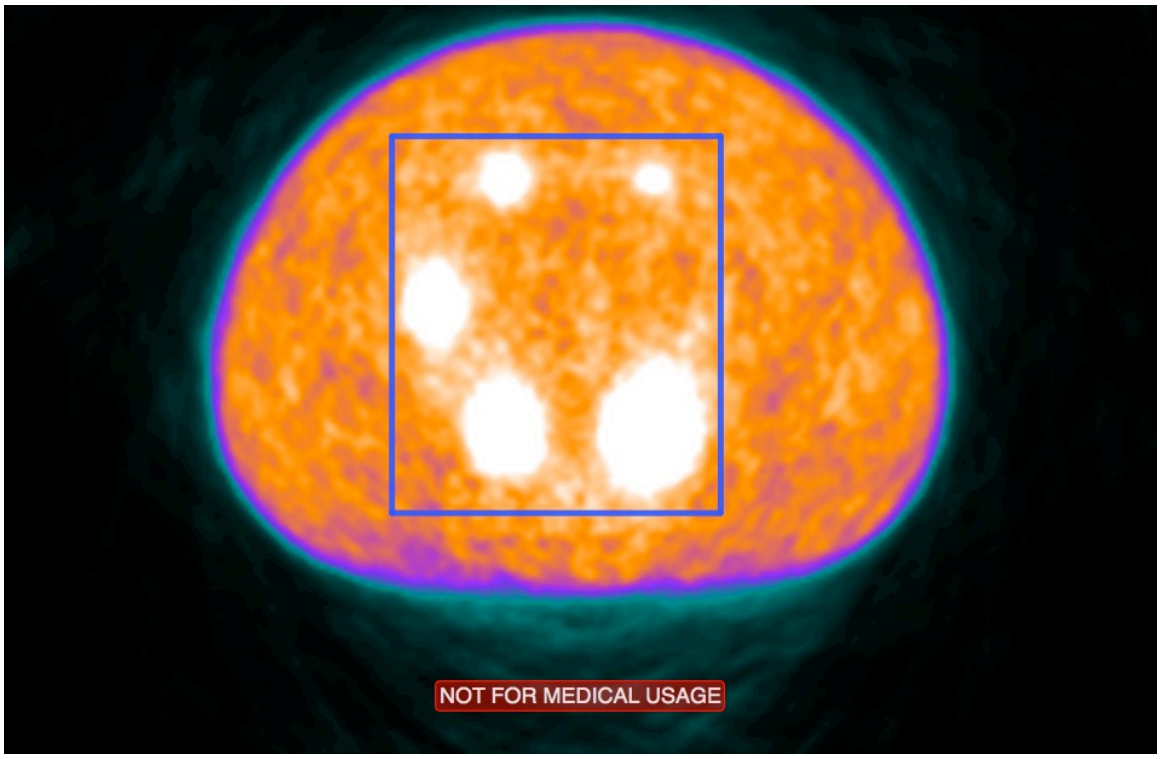


Figure 49: Region selected in NEMA NU-2 phantom for the purpose of calculating texture parameters.

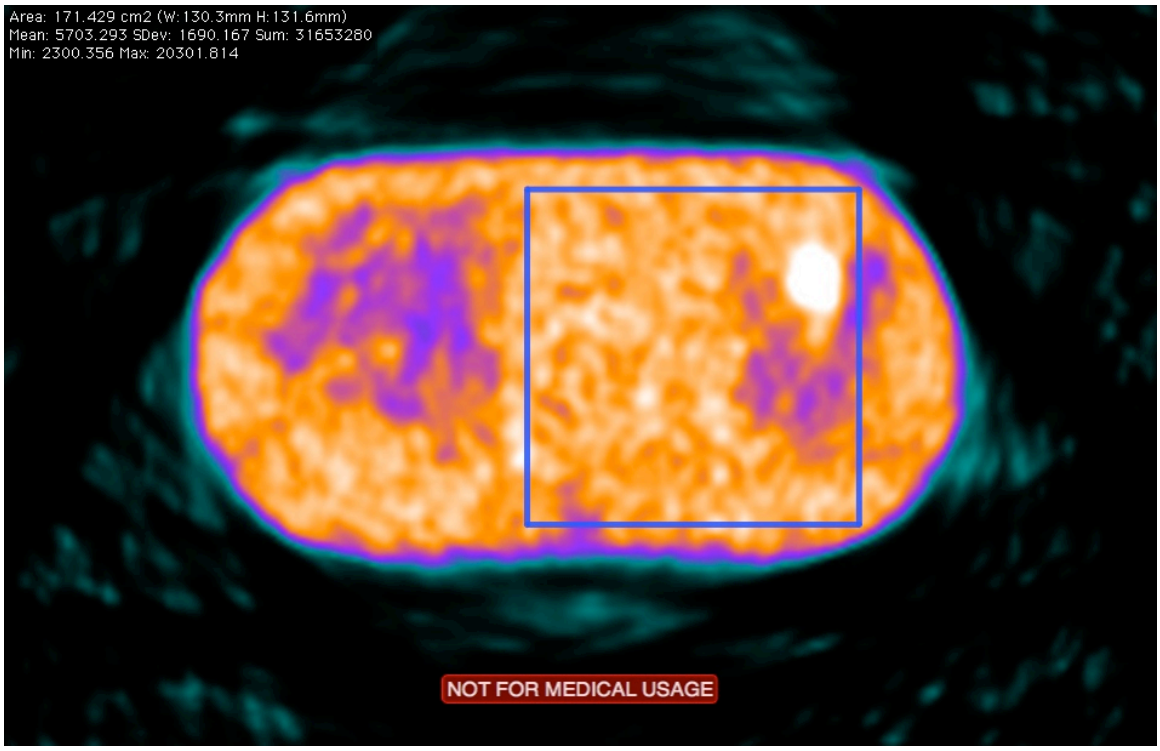


Figure 50: Region selected in CTN phantom for the purpose of calculating texture parameters

In the NEMA NU-2 phantom shown in Figure [49], the rectangular portion of image is selected to achieve a balance between the background and features inside the phantom so that the relationships of pixels in tumor and background have weighting of the same magnitude.

The CTN phantom as shown in Figure [50], is a structurally complex phantom closer to an actual patient image. For this phantom we extracted a representative region that included a uniform background simulating uniform muscle uptake, a tumor, and structurally non-homogenous lung field.

Specific Aim

The specific aim of this part of the study is to attempt to define similarities/differences in diagnostic image quality between two scanners based upon five commonly used texture metrics, namely: contrast, energy, homogeneity, correlation and entropy.

To accomplish this aim we followed a two-step methodology:

4.2.1 Image Comparison through Visual Inspection

In this first phase of the project a human observer was tasked with visually comparing an image with other images created with different reconstruction parameters and assigning discrete levels of similarity. The comparison was performed using the following scale.

SCALE	DESCRIPTION
1	The images are remarkably different and could never be mistaken as even nearly identical.
2	The images are clearly different from one another, but not remarkably so.
3	The images have the same general look, but there are easily observable differences.
4	The images have small observable differences.
5	The images are virtually identical.

Table 7: Scale used to assign comparison ranks to image-pairs.

Two independent observers compared all image-pairs and assigned the ranks according to Table [7]. In case of discrepancies, a third adjudicator made the final decision. This comparison was performed for both inter-scanner and intra-scanner image combinations for the NEMA NU-2 and CTN phantoms. Image sets that rank 4 or 5 on the scale described above were considered to be visually alike enough to categorize them as being functionally of similar image quality. Image set comparisons with ranks of 1, 2 and 3 were considered to be visually different enough to exclude from subsequent analysis.

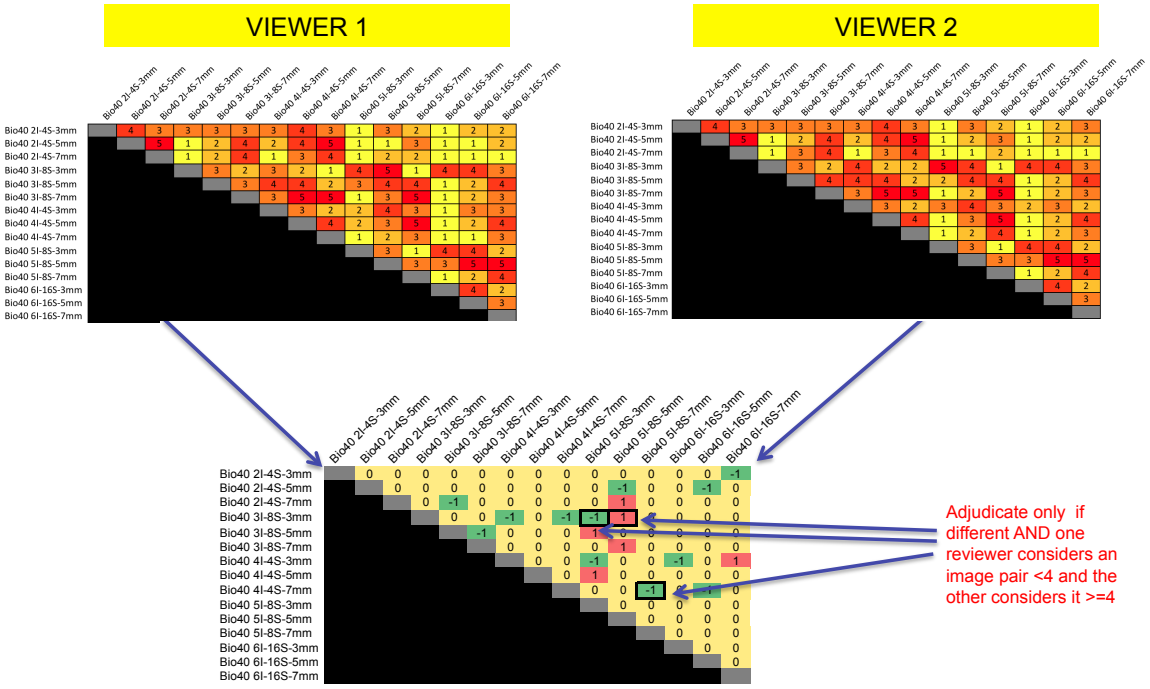


Figure 51: Visual Inspection Testing Process example describing Biograph 40 Vs Biograph 40 comparison with NEMA NU-2 phantom

Visual comparison is performed between the groups listed in Table [8] below. Each group consists of 15 images processed using different reconstruction parameters.

Groups	Group 1	Vs	Group 2
1	NEMA NU-2 phantom with Biograph 40	Vs	NEMA NU-2 phantom with Biograph 40
2	NEMA NU-2 phantom with Biograph Duo	Vs	NEMA NU-2 phantom with Biograph Duo
3	NEMA NU-2 phantom with Biograph 40	Vs	NEMA NU-2 phantom with Biograph Duo
4	CTN phantom with Biograph 40	Vs	CTN phantom with Biograph 40
5	CTN phantom with Biograph Duo	Vs	CTN phantom with Biograph Duo
6	CTN phantom with Biograph 40	Vs	CTN phantom with Biograph Duo

Table 8: Reconstructed image sets that are compared to determine Similar looking images. Each group consists of 15 images processed using different reconstruction parameters

The choice of reconstruction parameters used greatly affects the appearance of its background and structures within. To demonstrate the impact of different reconstruction parameters, several examples of the NEMA NU-2 phantom images using different reconstruction parameters are shown below:

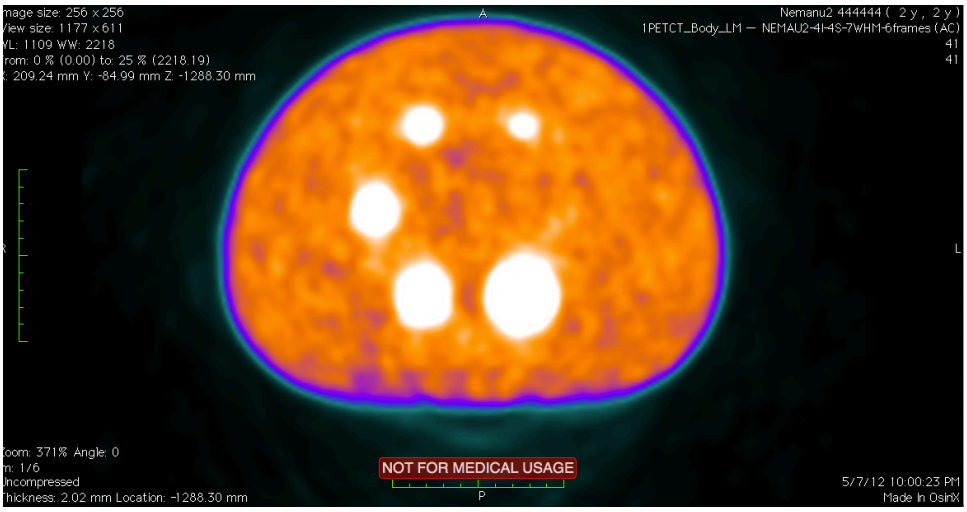


Figure 52: Example image with low noise frequency, low noise amplitude, circular lesion geometry, low edge distortion. NEMA NU-2 (4 Iterations, 4 Subsets, 7mm filter)

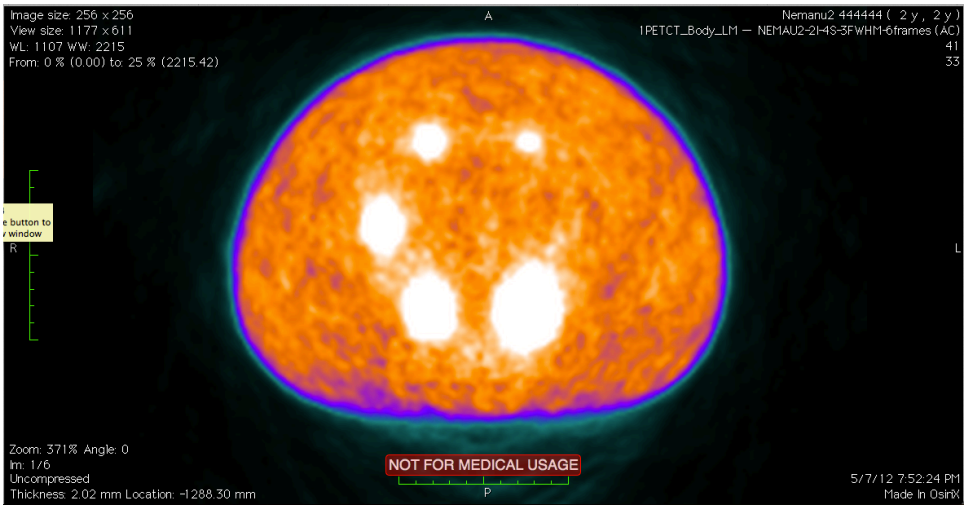


Figure 53: Example image with Medium noise amplitude, high noise frequency, somewhat distorted lesion geometry, medium edge distortion. (2 Iterations, 4 Subsets, 3mm filter)

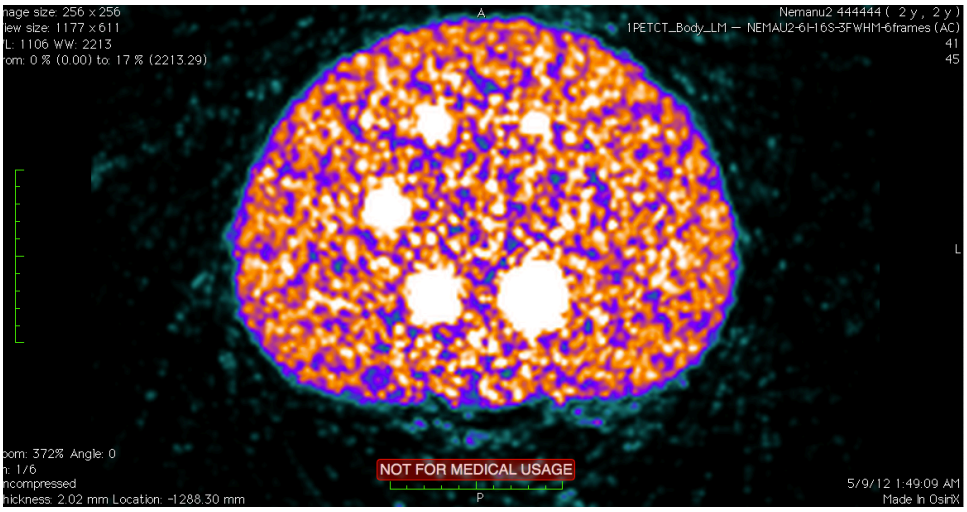


Figure 54: Example image with high noise amplitude, high noise frequency, circular lesion geometry and high edge distortion. (6 Iterations, 16 Subsets, 3mm filter)

During visual inspection of the image sets

- The size of two images under comparison was kept the same.
- The color look up table (CLUT) for two images was kept the same. All images were read in inverse black and white color scale as clinical scans are typically read with this.
- The window level for two images was set to an equal absolute value.

The main points considered while comparing two images in a Visual Inspection test are:

- The lesions in the phantom should have similar geometric shape.
- The distortion at the edges of the lesions should be alike.
- The lesions should have similar apparent size.

- The background of the images need not be an exact replica, but should have similar noise properties both in magnitude and spatial frequency.
- The apparent maximum magnitude of the lesions should be similar.

All the image sets are examined using the above mentioned parameters and allotted a rank as defined in Table [7]. Difference analysis was performed on these reconstructions for all five-texture parameters respectively. For example, when comparing the following reconstructions, NEMA 2I-4S-8mm (two iterations, four subsets and 8mm filter) and NEMA 6I-16S-5mm (six iterations, sixteen subsets and 5mm filter), a similarity rank was assigned to this image pair according to the scale mentioned in Table [7] and the mathematical difference between their individual texture parameters was calculated.

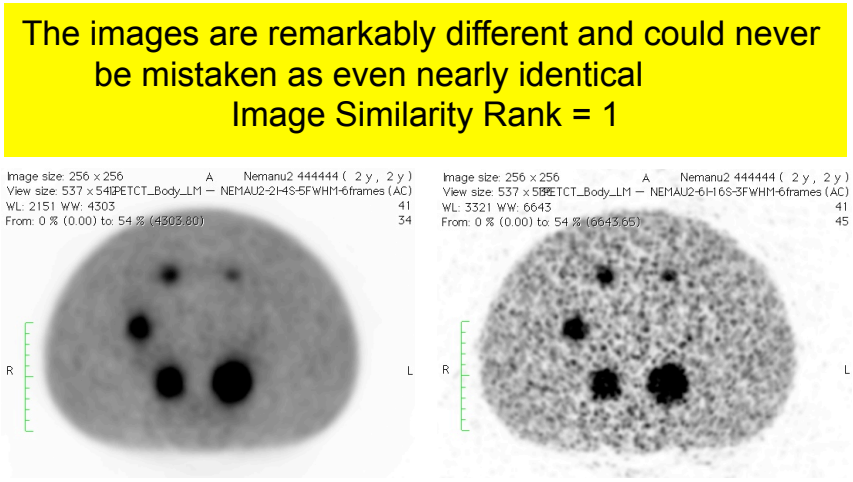


Figure 55: Dissimilar images will different reconstruction parameters (Rank allotted is 1 according to Table [7], as noise frequency, lesion shape, lesion distortion and background are different looking).

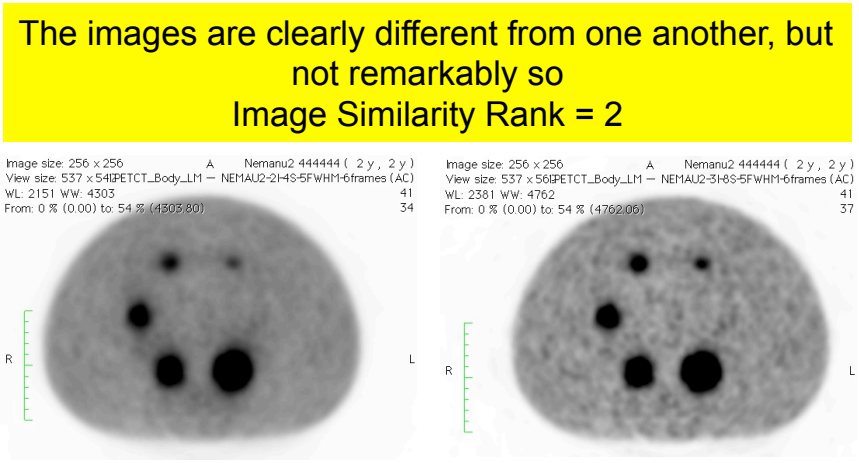


Figure 56: Dissimilar images will different reconstruction parameters (Rank allotted is 2 according to Table [7], as noise frequency, lesion shape, lesion distortion and background are relatively different looking).

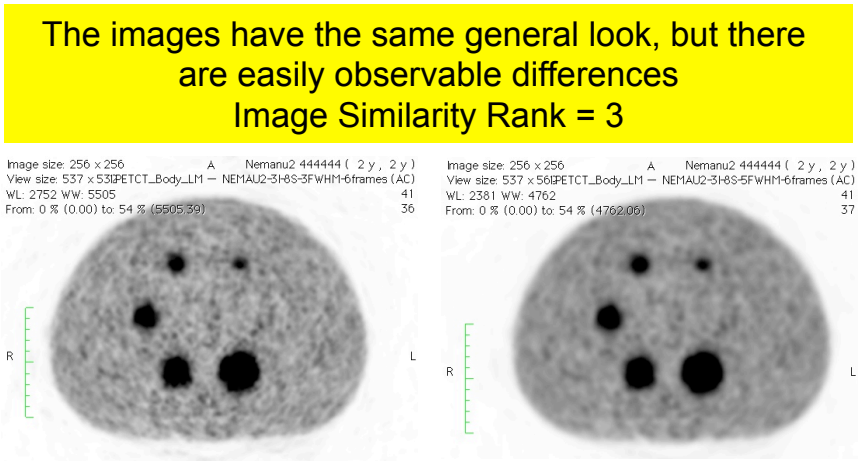


Figure 57: Images will different reconstruction parameters (Rank allotted is 3 according to Table [7], as lesion distortion and background are relatively different looking).

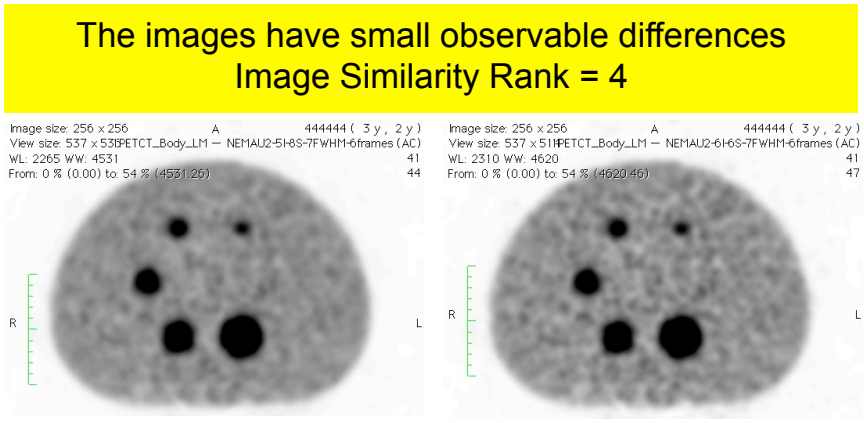


Figure 58: Images will different reconstruction parameters (Rank allotted is 4 according to Table [7], as background is somewhat different looking).

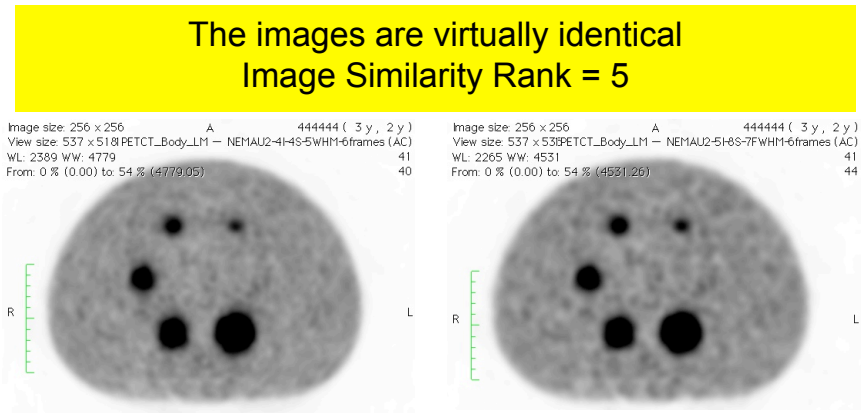


Figure 59: Similar images with different reconstruction parameters (Rank allotted is 5 according to Table [7], as noise frequency, lesion shape and background are close to same)

4.2.1.1 GLCM Parameters used for texture calculation

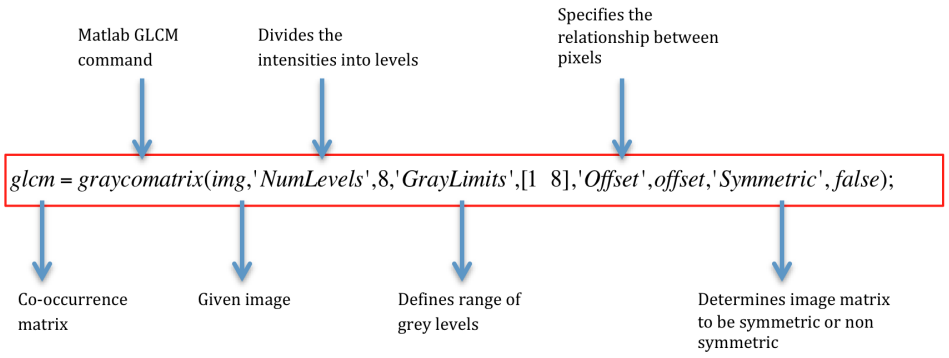


Figure 60: GLCM in MATLAB

As explained in Chapter 3.4 the GLCM is used to calculate the four texture metrics` Contrast, Energy, Homogeneity and Correlation. The figure below represents the MATLAB command to calculate GLCM

Offset as explained in Chapter 3.4 is the parameter that defines the relationship between the pixel of interest and its neighbor at varying distance and direction. It is not limited to a single direction or distance.

The following offset value was used for our analysis

Offset = [0 1, 0 2, 0 3, 0 4;

-1 1, -2 2, -3 3, -4 4;

-1 0, -2 0, -3 0, -4 0;

-1 -1, -2 -2, -3 -3, -4 -4]

This is a [16 X 2] matrix that spans for 1,2,3, and 4 pixel spaces in four directions.

A symmetric matrix is one in which the same values occur on the opposite sides of diagonal. It is a GLCM parameter that decided if the ordering of the values in pixel pairs would be considered or not. The symmetry parameter was set to 'true' as it is inline with the geometry of the PET scanners, thereby increasing the number of directions as specified by offset to 8.

The analysis was performed on the pixels that were 1, 2, 3, and 4-pixel distance apart in 8 directions, hence creating a [16 X 2] matrix. Calculations were performed with distances of 6-pixels and 8-pixels, however the resulting texture values did not provide any additional discriminatory information, and actually diluted discriminatory capabilities of the metrics.

The Number of Grey Levels or 'NumLevels' used for our analysis were 256 i.e. the images were scaled into 256 grey levels, hence creating a [256 X 256 X 16] matrix. 256 grey levels were chosen as they delivered best results. Experiments were performed using 128 and 64 grey levels respectively, but they did not perform any better. The ratio between the texture values obtained using different grey levels was found to be roughly constant, but with finer discriminating capability with the 256 gray levels.

The GLCM calculation with the pixel distances of 1, 2, 3, and 4 for NEMA NU-2 phantom are depicted in the figure below. A rectangular portion of the slice encompassing the lesions and a limited amount of background was selected as shown in Figure [49]. The values for offset and symmetry were used as explained above. A grey level value of 64 is used instead of 256 to make the graphs easier to interpret, however 256 levels were used in the actual calculations.

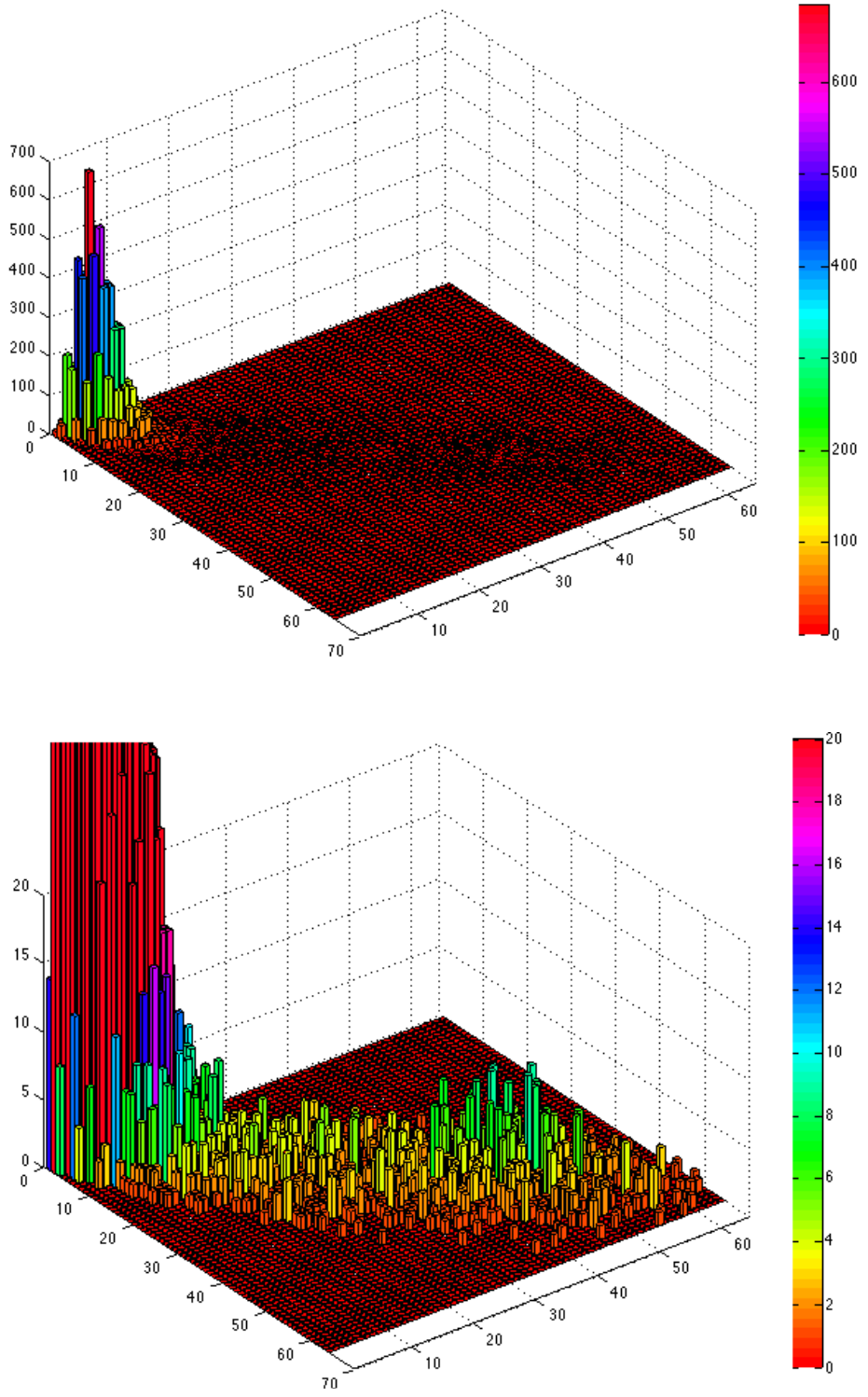


Figure 61: GLCM distribution for NEMA NU-2 phantom using 64 grey levels and pixel distance of 1

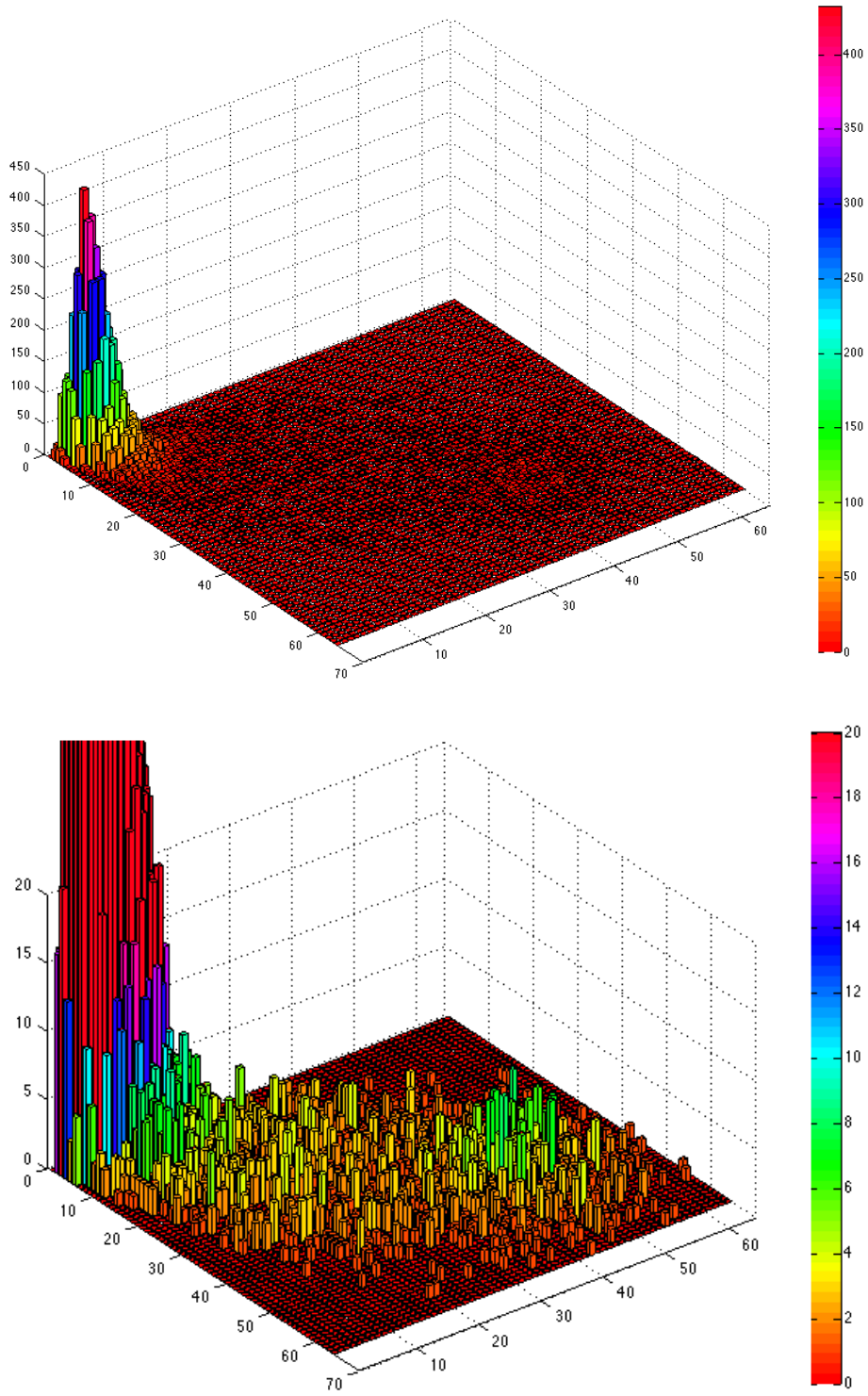


Figure 62: GLCM distribution for NEMA NU-2 phantom using 64 grey levels and pixel distance of 2

Figure [61] shows the GLCM for 64 grey levels and 1 pixel distance in 8 directions. As the distance of 1 pixel is used, we can see that the highest frequency of pairs is present near the origin where the difference between the pixel values is small i.e. (1,1), (1,2), (2,2) etc. The uniform background contributes towards the smaller values whereas the pixels inside lesions contribute to larger diagonal or near diagonal values like (40,40), (49,50) etc. As the values of most of the adjacent pixels are same, so most of the GLCM data points lie on the diagonal.

The figure [62] shows the GLCM for 64 grey levels and 2-pixel distance in 8 directions. As the distance of 2 pixels is used, we can see that the high frequencies are still present for the low difference values. But as compared to GLCM with 1 pixel distance, the chances of finding greater difference between the pixel values increases as we go further apart and hence the GLCM is more spread out from the diagonal.

The figure [63] below shows the GLCM for 64 grey levels and 3-pixel distance in 8 directions. As the distance of 3 pixels is used, we can see that the frequency for smaller differences in pixel values i.e. near the origin is still highest but the GLCM data points are relatively more spread out and off diagonal as compared to GLCMs with pixel distance of 1 or 2. Hence as we go farther pixels apart the chances of finding GLCM data points with greater pixel difference is higher i.e. we will find pixels with greater difference like (1,8), (1,15) etc. which will form the off diagonal components.

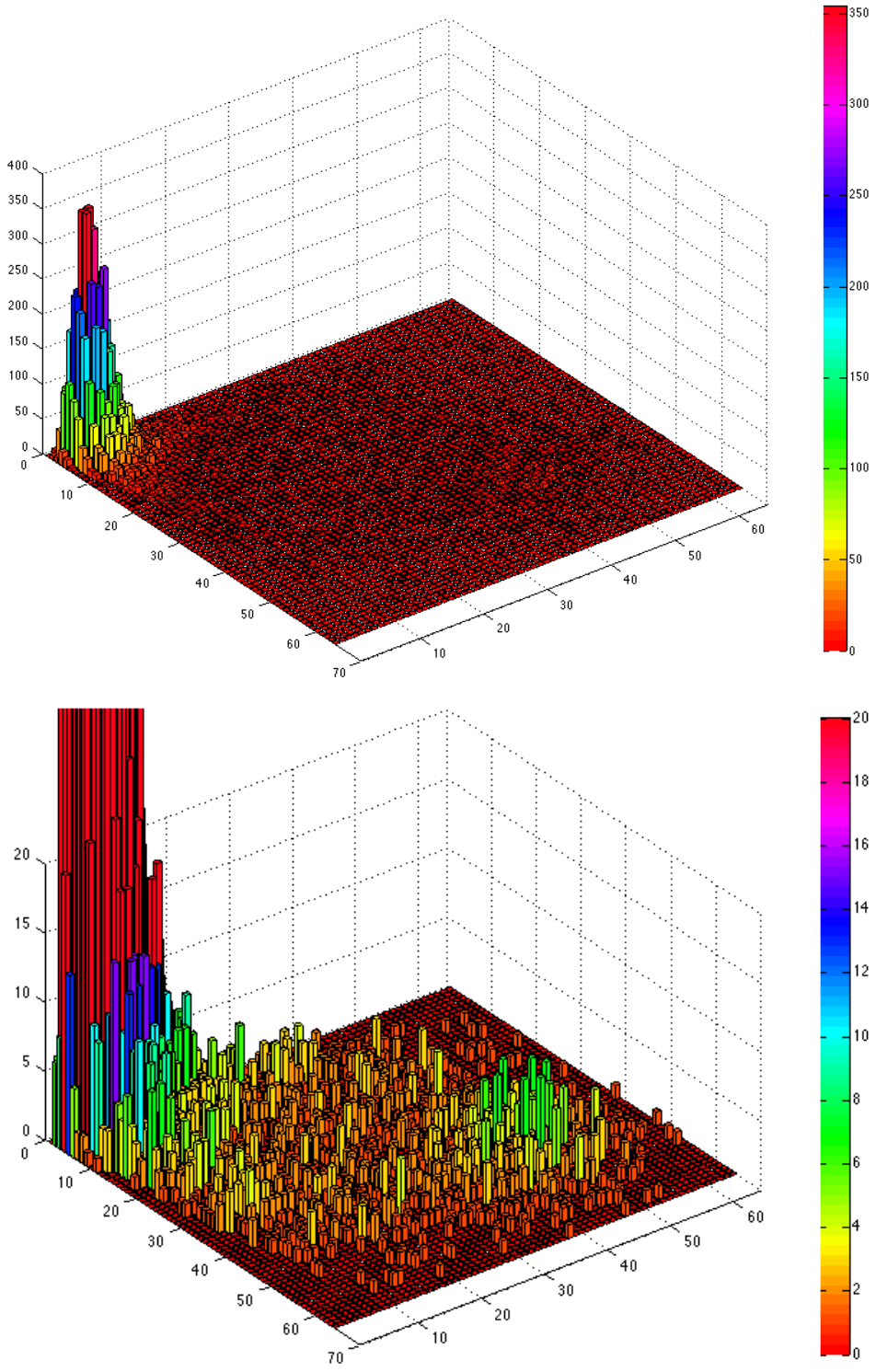


Figure 63: GLCM distribution for NEMA NU-2 phantom using 64 grey levels and pixel distance of 3

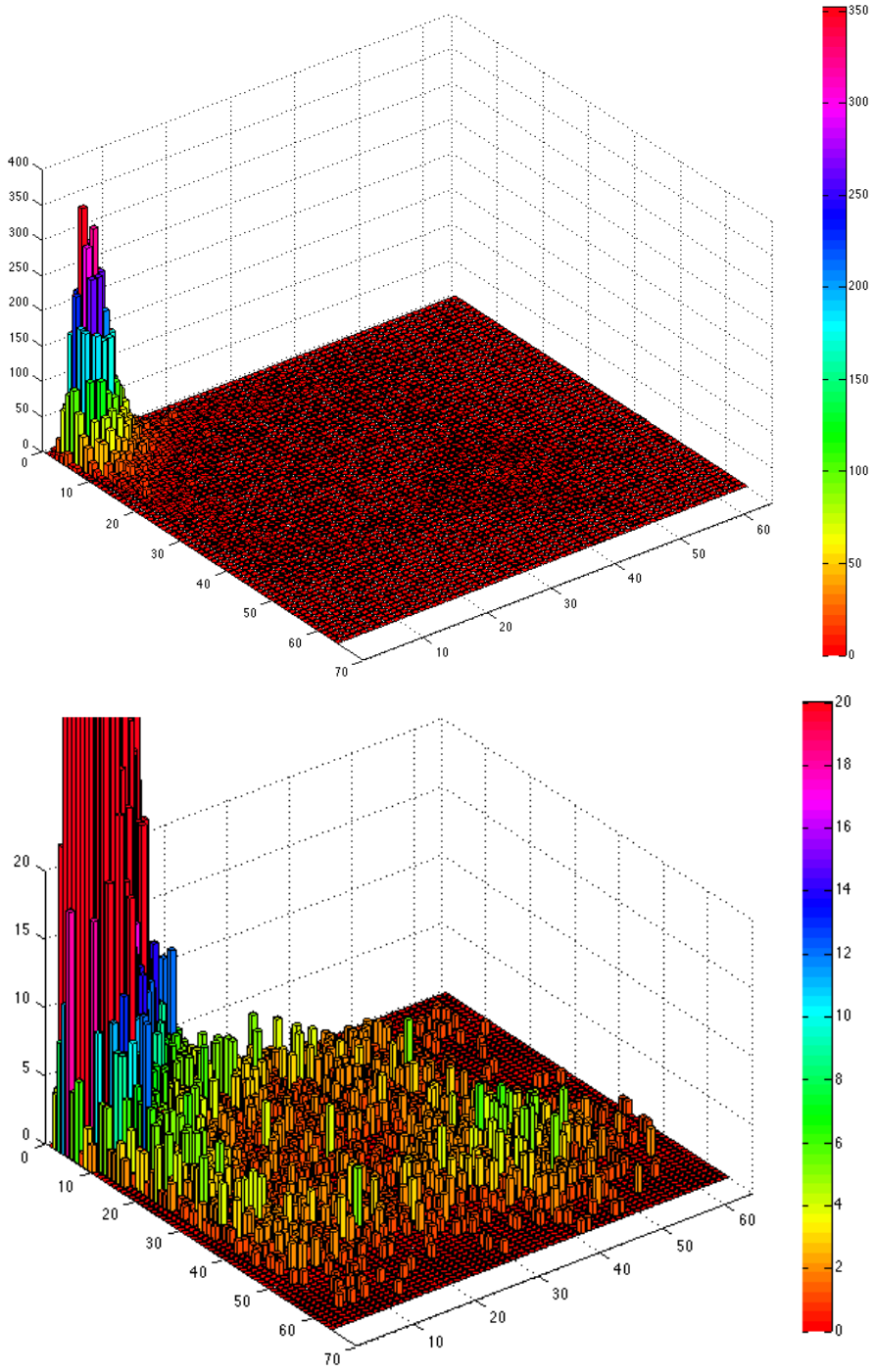


Figure 64: GLCM distribution for NEMA NU-2 phantom using 64 grey levels and pixel distance of 4

The Figure [64] shows the GLCM for 64 grey levels and 4-pixel distance in 8 directions. As the distance of 4 pixels is used, we can see that the frequency of occurrence of smaller pixel values has reduced. Also we can see that the distribution in the GLCM for pixel distance 3 and 4 are somewhat similar, which shows that after a certain pixel distance the GLCM does not provides significant additional information and the results do not have much difference between them.

Ideally, if two images are similar, the difference between their respective texture parameters should be minimum. Any texture parameter that does not demonstrate this property would not be a good predictor for image quality sameness. For any texture parameter the difference value between the similar image quality images determined from visual inspection are plotted against the difference value of the entire dataset in a histogram fashion e.g. when comparing group 1 from Table [8], we compared 15 reconstructions of NEMA NU-2 with Biograph 40 vs Biograph 40 resulting in 105 individual cases. Analyzing these 105 cases by visual inspection resulted in 30 reconstruction pairs with a rank of 4 or 5 as stated in Table [7] and hence similar image quality. Difference analysis was performed between the texture parameters of similar image quality pairs (30 image sets) and entire dataset (105 image sets) and the difference values plotted in the form of a histogram.

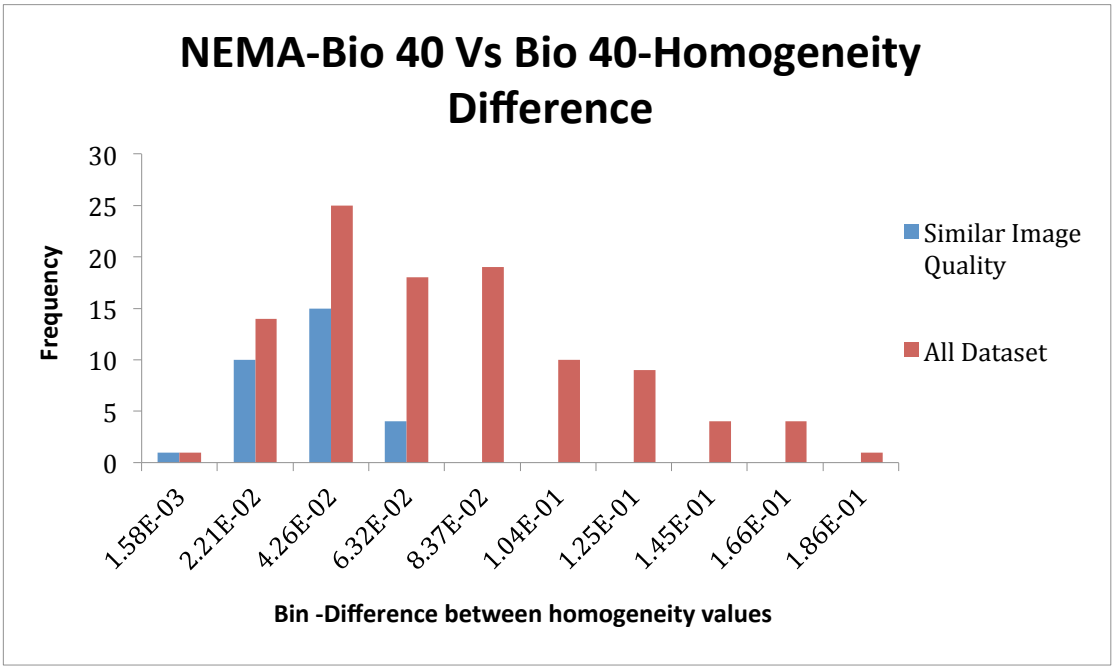


Figure 65: Homogeneity difference values for images with similar quality plotted against the difference value of the entire dataset from NEMA NU-2 with Biograph 40.

Figure [65] shows the histogram plot for difference values of the Homogeneity texture parameter with the NEMA NU-2 phantom when comparing Biograph 40 vs Biograph 40 reconstructions. The X-axis describes the bin difference of the histogram, which is the equal to the difference in homogeneity texture values between the reconstructions being compared (105 points). The Y-axis is the frequency of occurrence of image pairs in a particular difference bin.

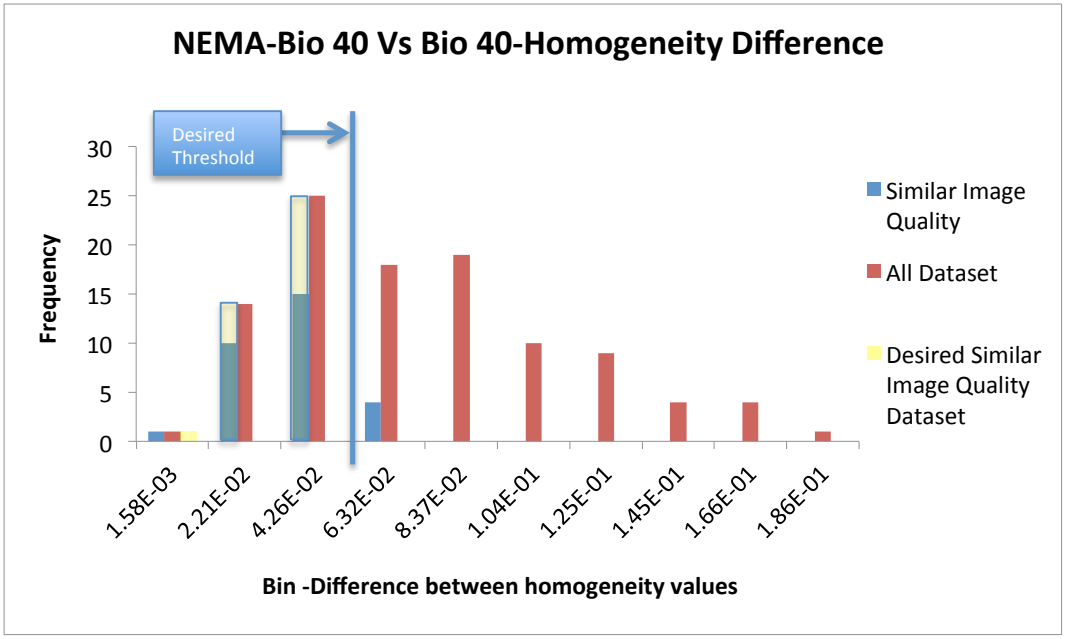


Figure 66: Homogeneity difference values for image with similar quality plotted against the difference value of the entire dataset from NEMA NU-2 with Biograph 40. Ideal case for similar image quality parameter is also described.

Red bars represent the frequency of occurrence of entire dataset (105 data points) and blue bars represent the frequency of occurrence of similar image quality pairs (30 data points as determined by visual inspection test). We can see that in the case of images with similar image quality the difference value between the homogeneity parameter is on the lower side of the histogram i.e. images with similar quality have smaller difference between their homogeneity texture parameter, hence suggesting that homogeneity is a good predictor of image quality sameness.

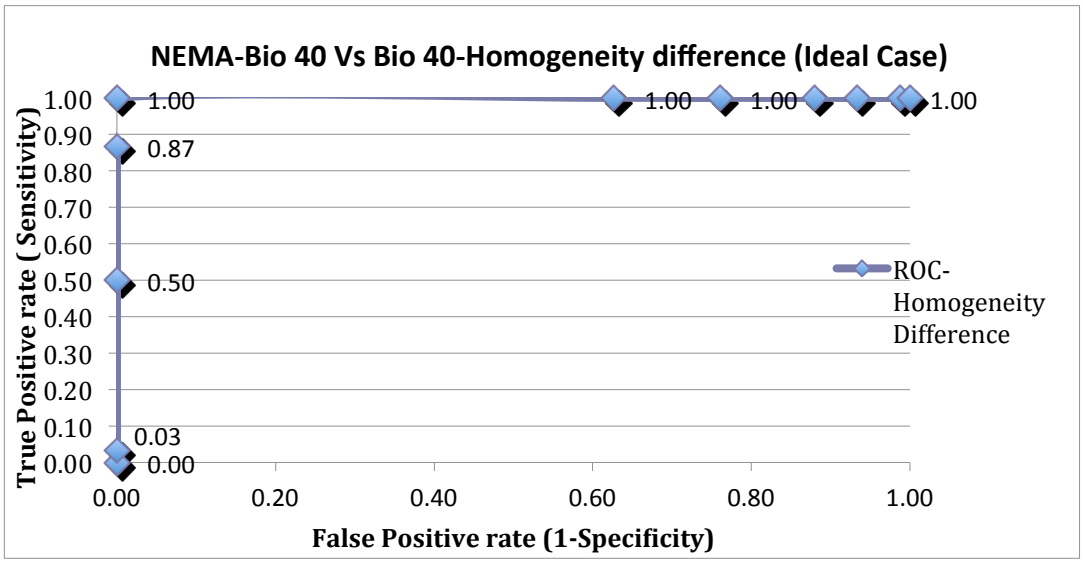


Figure 67: ROC curve illustrating the fictitious ideal similar image quality dataset case using NEMA phantom Bio 40 Vs Bio 40 with Homogeneity difference values

Figure [66] shows the same data as in Figure [65] with the addition of a fictitious "Ideal Similar Image Quality Dataset". This represents an ideal desired distribution we would like to achieve for the similar image quality pairs using any texture metric. In ideal case all the lower end values of the histogram or the smaller differences in the texture values will be occupied by the images that are visually similar. All the difference data below the desired threshold (4.26E-02) would be True Positive i.e. the number of False Positive below this threshold would be zero providing high sensitivity and high specificity. Beyond the threshold value the number of False negative will drop down to zero providing high sensitivity and low specificity.

The histogram plots obtained for other texture metrics for various comparison groups described in table [8] are shown in the figure below.

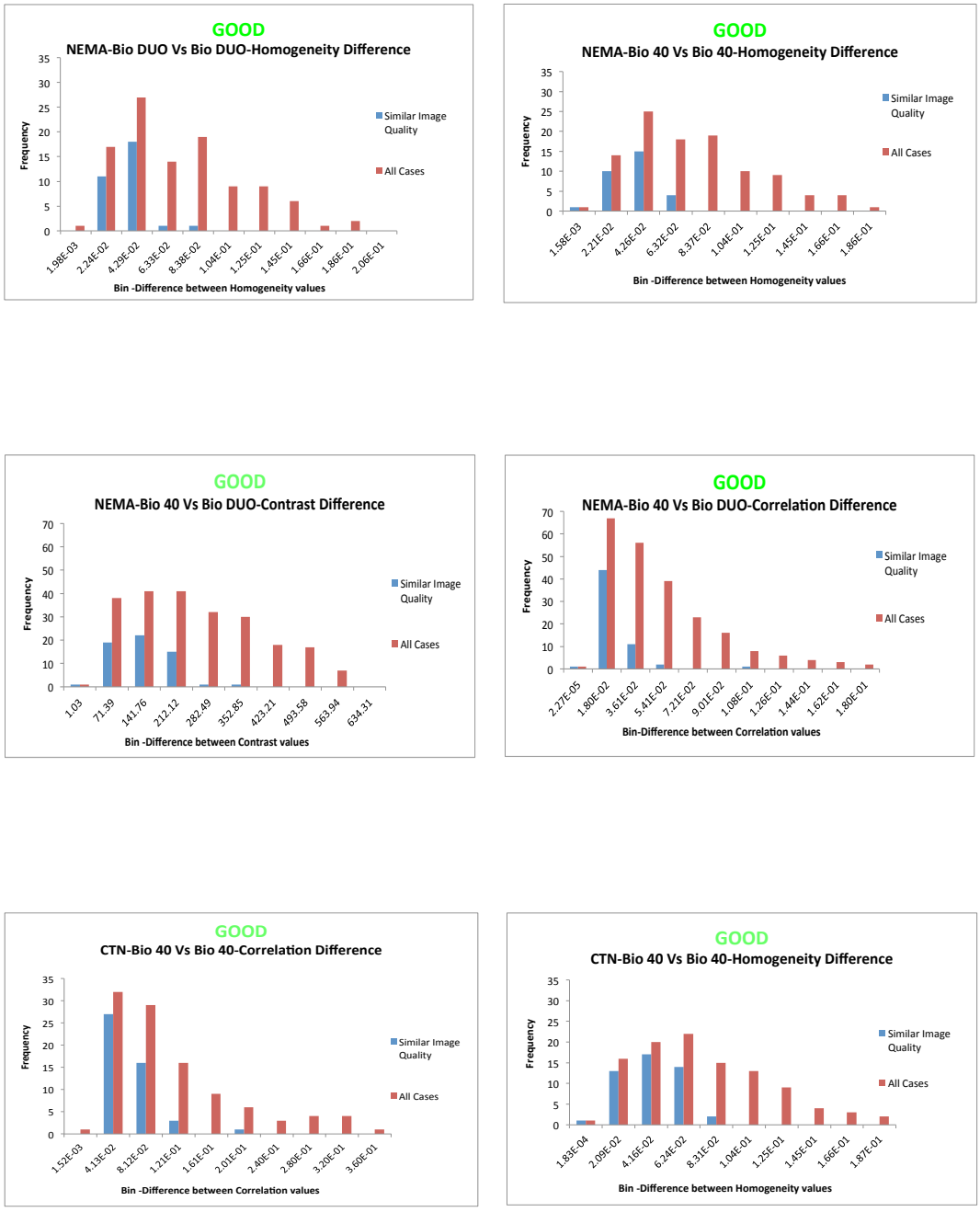


Figure 68: Histogram plots for images with similar image quality against the dataset for their respective comparison group. Comparison groups are depicted in table [8].

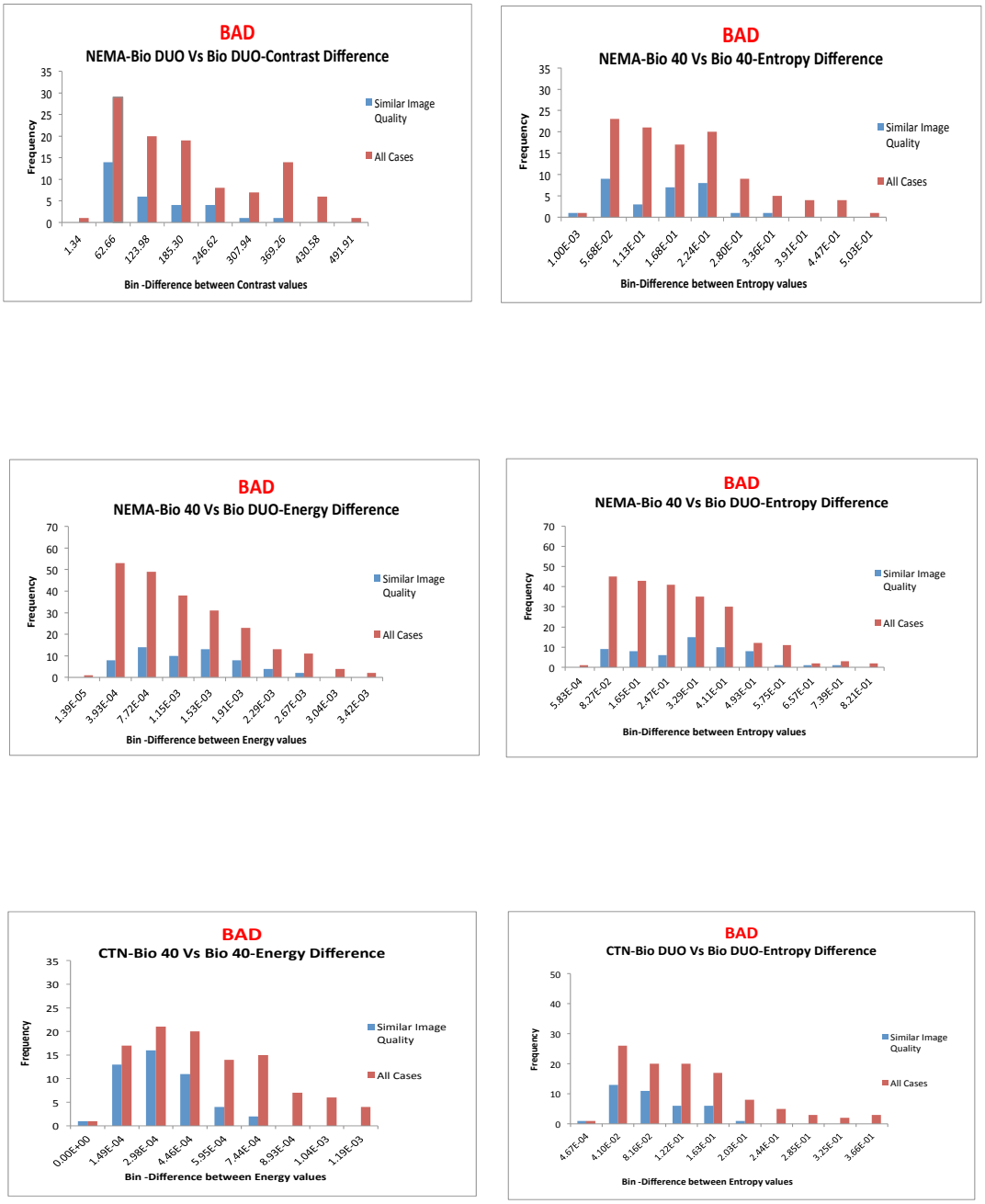


Figure 69: Histogram plots for images with similar image quality against the dataset for their respective comparison group. Comparison groups are depicted in table [8].

Images in Figure [68] depict the cases where the difference between the texture parameters for images with similar quality was minimum or on the lower end of histogram. These identified texture metrics (Contrast, Correlation, and Homogeneity) could provide useful information to predict image quality.

Images in Figure [69] illustrate cases where the difference between the texture parameters for images with similar quality was largely uniformly spread throughout the histogram range. These texture metrics do not provide any relevant information to predict image quality.

After studying the histograms for all the texture matrices we can see that homogeneity and correlation give the best results as these parameters have the least discrepancy in their values for the images that are visually similar. Contrast performs somewhat well, but Entropy and Energy under-performed and their parameter values for similar images are spread throughout the histogram range. For all histogram plots for NEMA NU-2 and CTN phantom see Appendix B.

4.2.2 Receiver Operating Characterstics (ROC) Curves

Following the analysis using histogram plots, ROC curves were calculated and plotted for the texture parameters. The ROC analysis will be used in determing which texture parameters are better predictors of similarity of images. ROC is a graph of sensitivity vs 1-specificity of a parameter and illustrates relative tradeoffs between true positives and false positives. Each point on the ROC curve represents a sensitivity/specificity pair that corresponds to a certain threshold value.

Sensitivity or true positive rate of a parameter describes the proportion of true positives that are identified as positive. Specificity, or true negative rate, of a parameter describes the proportion of true negatives that are identified as negative. 1-Specificity gives us the False positive rate.

TRUE	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
FALSE	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)
	Sensitivity=TP/(TP+FN)	Specificity=TN/(FP+TN)

Table 9: Table depicting Sensitivity and Specificity formulae

The area under the ROC curve is a measure of how well a parameter can distinguish between two groups i.e. SIMILAR/DISSIMILAR images in our study. Area Under Curve (AUC) is a portion of the area of the unit square and its value will always be between 0 and 1.

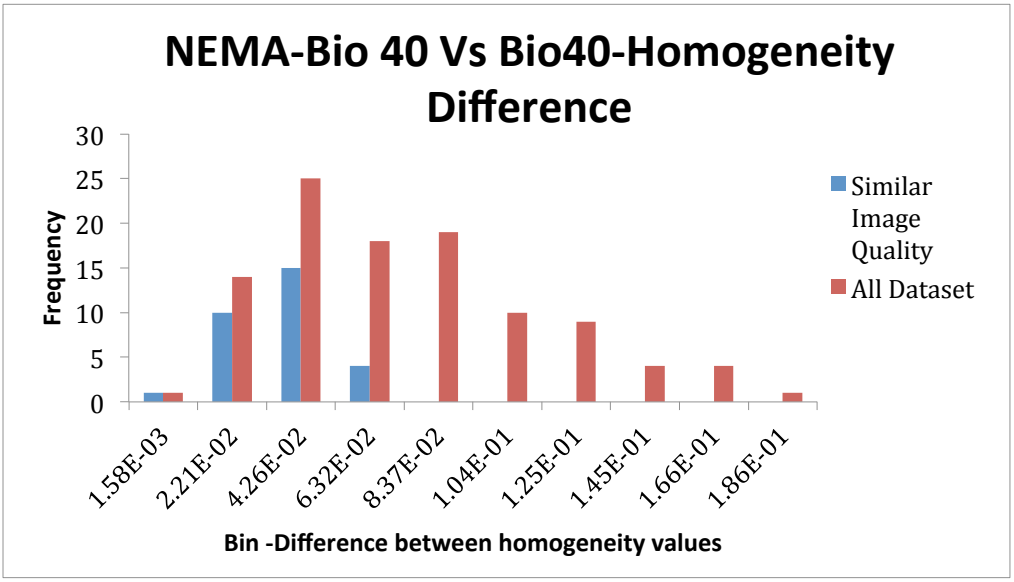


Figure 70: Homogeneity difference values for look alike image sets plotted against the difference value of the entire dataset for NEMA NU-2 with Biograph 40

The AUC of a parameter is equivalent to the probability that the parameter will rank a randomly chosen positive instance higher than a randomly chosen negative instance. Any parameter that appears in the lower right triangle of the ROC curve performs worse than random guessing. Figure [70] shows an ROC curve description using an exampe.

	Thresh hold	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	1-Specificity
n<=	1.00E-03	0	30	0	75	0.00	1.00	0.00
n<=	1.58E-03	1	29	0	75	0.03	1.00	0.00
n<=	2.21E-02	11	19	4	71	0.37	0.95	0.05
n<=	4.26E-02	26	4	14	61	0.87	0.81	0.19
n<=	6.32E-02	30	0	28	47	1.00	0.63	0.37
n<=	8.37E-02	30	0	47	28	1.00	0.37	0.63
n<=	1.04E-01	30	0	57	18	1.00	0.24	0.76
n<=	1.25E-01	30	0	66	9	1.00	0.12	0.88
n<=	1.45E-01	30	0	70	5	1.00	0.07	0.93
n<=	1.66E-01	30	0	74	1	1.00	0.01	0.99
n<=	1.86E-01	30	0	75	0	1.00	0.00	1.00

Table 10: Table illustrating calculations for plotting ROC curves. (This example shows data used to create the ROC curves for difference values using the homogeneity texture parameter using the NEMA NU-2 phantom with the Biograph 40)

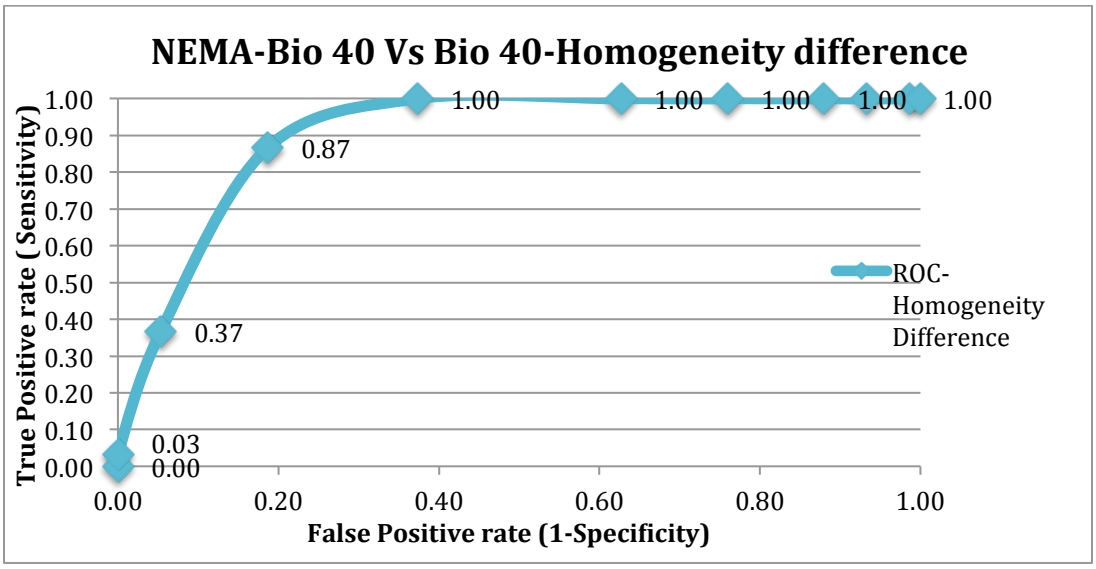


Figure 71: ROC curves for difference values in homogeneity texture parameter using the NEMA NU-2 phantom with the Biograph 40

TPs are the similar image pairs that are correctly identified as look-alikes (similar) for a given threshold value. For example while comparing group I (NEMA NU-2 Biograph 40 Vs NEMA NU-2 Biograph 40) from Table [8], the visually similar image pairs with differences below the threshold of 4.26E-02 are 26. So 26 is the number of TPs that are less than or equal to this threshold value.

FNs are defined as the number of truly similar images (as determined by visual analysis) identified as dissimilar via the texture metric analysis. For a given threshold value, FNs are the total number of similar image pairs as determined by the visual inspection test minus the number of TPs. For example while using comparison group I (NEMA NU-2 Biograph 40 Vs NEMA NU-2 Biograph 40) from Table [8], there are 26 total look-alike image sets below the threshold of 4.26E-02. But there were 30 data points

in total that were observed as similar by the visual inspection test. So ideally these 30 data points should have been considered as similar by the model for any given threshold value. Instead 26 points were considered as similar by the model for the 4.26E-02-threshold value. So these 30 visually alike images will yield 4 false positives.

FPs is the number of dissimilar images identified as similar. For a given threshold value, it is the total number of data points (positive + negative) minus the number of TPs. For example while using comparison group I (NEMA NU-2 Biograph 40 Vs NEMA NU-2 Biograph 40) from Table [8], the total data points (positive + negative) below 4.26E-02-threshold value were 40. So 40-TP will yield the value of 14 FP (40-26=14).

TNs are the number of dissimilar images identified as dissimilar. For a given threshold value, it is the number of negative data points that are correctly identified as negative. For example while using comparison group I (NEMA NU-2 Biograph 40 Vs NEMA NU-2 Biograph 40) from Table [2], the total data points in the data set were 105. We deduced TP=26, FP=14, FN=4. Therefore TN= 105-TP-FP-FN =61. Hence 61 data points or dissimilar images below the threshold value of 4.26E-02 were correctly identified as dissimilar.

$$\text{Sensitivity}= \frac{\text{Number of True Positives}}{\text{Number of True Positives}+\text{Number of False Negatives}}$$

$$\text{Sensitivity}= \frac{26}{(26+4)} = 87\%$$

$$\text{Specificity}= \frac{\text{Number of True Negatives}}{\text{Number of True Negatives}+\text{Number of False Positives}}$$

$$\text{Specificity} = \frac{61}{(61+14)} = 81\%$$

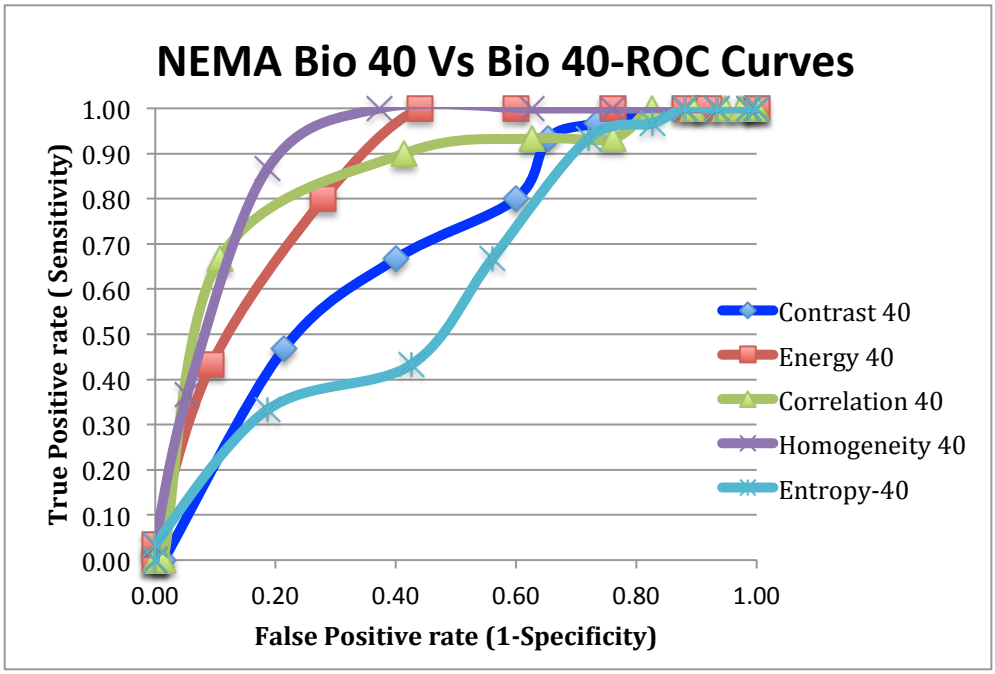
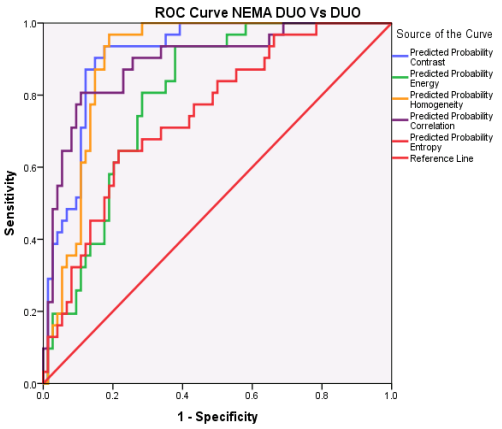
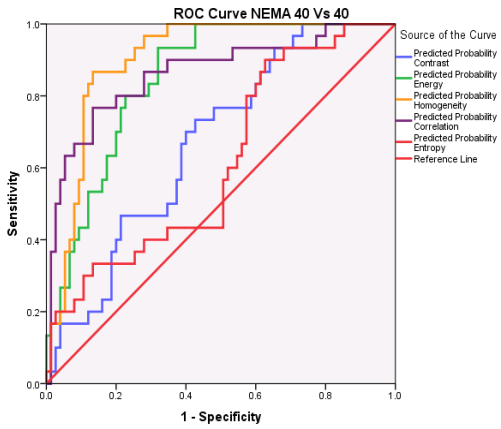


Figure 72: ROC curves for difference values in texture parameters using NEMA NU-2 with Biograph 40

The ROC curves for difference values of five texture parameters under various comparison sets are plotted in Figure [73].



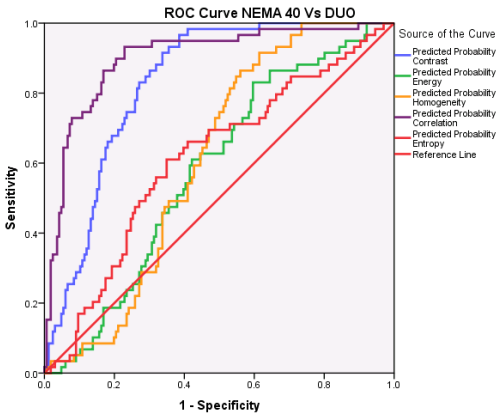


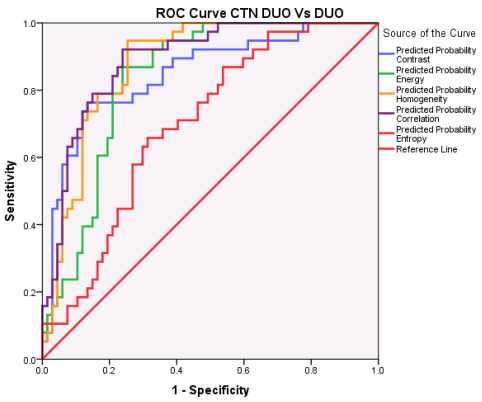
Area Under the Curve NEMA DUO Vs DUO

Test Result Variables	Area
Predicted probability Contrast Only	0.91
Predicted probability Energy Only	0.797
Predicted probability Homogeneity Only	0.897
Predicted probability Correlation Only	0.891
Predicted probability Entropy Only	0.747

Area Under the Curve NEMA 40 Vs 40

Test Result Variables	Area
Predicted probability Contrast Only	0.67
Predicted probability Energy Only	0.849
Predicted probability Homogeneity Only	0.901
Predicted probability Correlation Only	0.862
Predicted probability Entropy Only	0.616





Area Under the Curve NEMA 40 Vs DUO

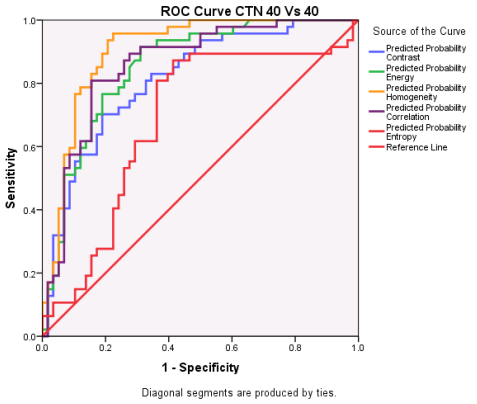
Test Result Variables	Area
Predicted probability Contrast Only	0.831
Predicted probability Energy Only	0.579
Predicted probability Homogeneity Only	0.608
Predicted probability Correlation Only	0.9
Predicted probability Entropy Only	0.613

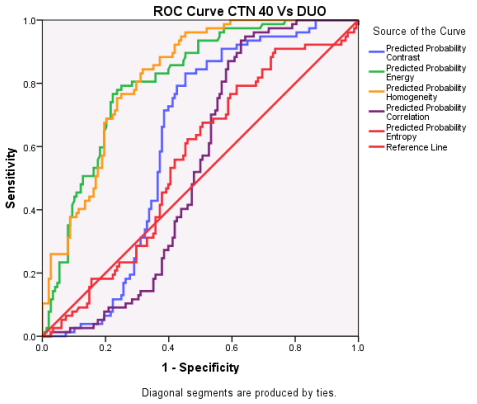
Area Under the Curve CTN DUO Vs DUO

Test Result Variables	Area
Predicted probability Contrast Only	0.848
Predicted probability Energy Only	0.835
Predicted probability Homogeneity Only	0.878
Predicted probability Correlation Only	0.889
Predicted probability Entropy Only	0.698

Figure 73: ROC curves and AUC for difference values in texture parameters for various image set comparisons using NEMA NU-2 and CTN phantoms

Figure 73 – Continued





Area Under the Curve CTN 40 Vs 40

Test Result Variables	Area
Predicted probability Contrast Only	0.813
Predicted probability Energy Only	0.852
Predicted probability Homogeneity Only	0.904
Predicted probability Correlation Only	0.858
Predicted probability Entropy Only	0.681

Area Under the Curve CTN 40 Vs DUO

Test Result Variables	Area
Predicted probability Contrast Only	0.617
Predicted probability Energy Only	0.808
Predicted probability Homogeneity Only	0.821
Predicted probability Correlation Only	0.535
Predicted probability Entropy Only	0.556

The Area Under the Curve (AUC) for the texture parameters provides their predicting probability in determining the similarity between image sets. For example in NEMA NU-2 phantom under Biograph 40 vs Biograph 40 the AUC for homogeneity parameter is 0.901. From the results obtained by plotting ROC curves for NEMA NU-2 and CTN phantom we can see that Homogeneity and Correlation are the most powerful predictors for similarity between images sets. Contrast is also a good predictor whereas Entropy and Energy do not seem to be relevant.

It is hypothesized that even better results might be achieved through the application of combination of texture parameters. The combined predictors that maximize the AUC would be an even better predictor of image similarity.

4.3 Optimization using logistic regression analysis

Logistic regression, also called a logit model, is used to model dependent categorical outcome variables from a set of predictor variables that can be continuous, discrete, dichotomous, or a mix. In the logit model the log of odds of success is modeled as a linear combination of the predictor variables. This regression analysis is mostly used in health science, social science and marketing fields.

Logistic regression is mostly used in problems in which the dependent variable is binary—that is, the number of available categories is two (Yes/No, Healthy/ill etc.) and problems with more than two categories are referred to as multinomial logistic regression.

Logistic regression defines the relationship between a categorical dependent variable and predictor independent variables, through conversion of the dependent variable into probability scores.

4.3.1 Null Hypothesis

The statistical null hypothesis is that the probability of a particular value of the nominal variable is not associated with the value of the predictor variable.

4.3.2 How Logistic Regression Analysis works

Logistic regression finds the equation that best fits the value of Y for each value of X, where Y is the probability of obtaining a particular value of a nominal variable. In our research, we want to determine the image pairs that are visually similar. The value of dependent nominal variables would be “Similar” or “Dissimilar”. The Y variable used in logistic regression would then be the probability of images being similar. This probability could take values from 0 to 1. The limited range of this probability would present

problems if used directly in a regression, so the odds ratio is used instead.

$$\text{Odds of Success} = \frac{Y}{1-Y}$$

For example, if the probability of an image pair being similar is 0.4, then the odds of success would be 0.4/(1-0.4)=2/3. Taking the natural log of the odds makes the variable more suitable for a regression. The natural logarithm of odds of success is called Logit. In Logistic regression Logit is the dependent variable.

$$\log(odds) = Logit(Y) = \ln\left(\frac{Y}{1-Y}\right)$$

So the result of a logistic regression with one independent predictor variable is an equation that looks like this:

$$\ln\left(\frac{Y}{1-Y}\right) = a + bX$$

$$\frac{Y}{1-Y} = e^{a+bX}$$

$$Y = \frac{e^{a+bX}}{1+e^{a+bX}}$$

For multivariate Logistic regression the equation will be modified to

$$Y = \frac{e^{a+b_1X_1+b_2X_2+.....+b_pXp}}{1+e^{a+b_1X_1+b_2X_2+.....+b_pXp}}$$

Here Y represents the probability of the dependent variable that depends on p-independent predictor variables.

$$LogitY = \ln\left[\frac{Y}{1-Y}\right] = \frac{\dfrac{e^{a+b_1X_1+b_2X_2+.....+b_pXp}}{1+e^{a+b_1X_1+b_2X_2+.....+b_pXp}}}{1-\dfrac{e^{a+b_1X_1+b_2X_2+.....+b_pXp}}{1+e^{a+b_1X_1+b_2X_2+.....+b_pXp}}}$$

$$= \ln[e^{a+b_1X_1+b_2X_2+.....+b_pXp}] = a + b_1X_1 + b_2X_2 + ... + b_pXp$$

The equation above shows that logistic regression is really just a standard linear regression model. This transform changes the range of **Y** from 0 to 1 to $-\infty$ to $+\infty$, as usual for linear regression. The slope (b) and intercept (a) of the best-fitting equation in a logistic regression are found via maximum-likelihood. Maximum likelihood is an iterative process that starts with random values of predictor variables to find the best linear combination of predictors in order to maximize the likelihood of obtaining the observed outcome frequencies.

In our calculations with Logistic Regression analysis we used the Similarity/Dissimilarity parameter as dependent categorical variables. The image sets were determined to be similar or dissimilar using the visual analysis explained earlier. For all the groups described in Table [8], the image sets were categorized as Similar or Dissimilar and their respective texture parameter differences were plotted against them.

The individual texture parameters were used as the independent predictor variables in order to find the probability of each data point in predicting the image sets to be similar or dissimilar. Using these probability values an ROC curve is generated and the Area Under the Curve is calculated. AUC provides us objective information of how well a texture parameter predicts the similarity between the image sets. ROC curves for all texture parameters were developed for the groups defined in Table [8], and are also shown in Figure [73]. As mentioned before Homogeneity, Contrast and Correlation were

found to be good predictors of image similarity whereas Energy and Entropy were irrelevant for the most cases.

We used the SPSS software package (by IBM) to perform Binary Logistic regression analysis. SPSS has a built-in Logistic regression function and provides the flexibility of using single or multiple independent predictor variables. An example of analysis using Logistic regression is explained below. In this case we are comparing image sets between group I on Table [8], i.e. NEMA NU-2 vs NEMA NU-2 using Biograph 40. The categorical variable SIMILAR/DISSIMILAR is used as the dependent variable and the difference between homogeneity texture parameter between the respective image sets is used as the independent continuous variable. The following is the output of the SPSS statistical package.

The case-processing summary in Table [11] shows that there were 105 total data points used in the analysis, i.e. 105 image sets were compared and there were no missing cases.

Case Processing Summary

Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	105	100.0
	Missing Cases	0	.0
	Total	105	100.0
Unselected Cases		0	.0
Total		105	100.0

Table 11: Case processing summary for homogeneity texture parameter being used as a predictor variable

Dependent Variable Encoding	
Original Value	Internal Value
DISSIMIL	0
SIMILAR	1

Table 12: Binary codes assigned to the categorical dependent variable during logistic regression analysis.

The dependent variable encoding table shows the binary coding that is assigned to the categorical data. SIMILAR is coded as 1 and DISSIMILAR is coded as 0.

Block 0: Beginning Block

Classification Table					
	Observed		Predicted		
			SIMILAR/ DISSIMILAR		Percentage Correct
			DISSIMIL	SIMILAR	
Step 0	SIMILAR/ DISSIMILAR	DISSIMIL	75	0	100.0
		SIMILAR	30	0	.0
	Overall Percentage				71.4

Table 13: Prediction probability calculation without using any predictor variable in logistic regression analysis.

NOTE: The cut off value is 0.500

Block 0 depicts the prediction probability when considering the null hypothesis that all the images are DISSIMILAR. In this case no predictor variable is used in the model. For this comparison we had 105 total image pairs out of which only 30 were SIMILAR. The accuracy of the model is calculated using the weights of individual data points in the two categories.

$$\text{Percentage Correct} = \frac{75*100+30*0}{105} = 71.4\%$$

This model will correctly identify the similarity among the image sets 71.4% of the time.

Classification Table					
	Observed		Predicted		
			SIMILAR/ DISSIMILAR		Percentage Correct
			DISSIMIL	SIMILAR	
Step 1	SIMILAR/ DISSIMILAR	DISSIMIL	67	8	89.3
		SIMILAR	8	22	73.3
	Overall Percentage				84.8

Table 14: Prediction probability calculation using homogeneity difference values as the independent predictor variable in logistic regression analysis.

NOTE: The cut off value is 0.500

The block in Table [14] shows the prediction probability when the homogeneity difference parameter is used as an independent predictor variable. The similarity or

dissimilarity of image pairs converted to binary form as shown in Table [12] acts as the dependent categorical variable. The accuracy of the model in identifying image sets as similar or dissimilar increases to 84.8%.

$$\text{Percentage Correct}=\frac{75*0.893+30*0.733}{105}=84.8\%$$

The program also returns the predicted probability of the homogeneity difference parameter to correctly identify the similar/dissimilar image sets for all 105 data points.

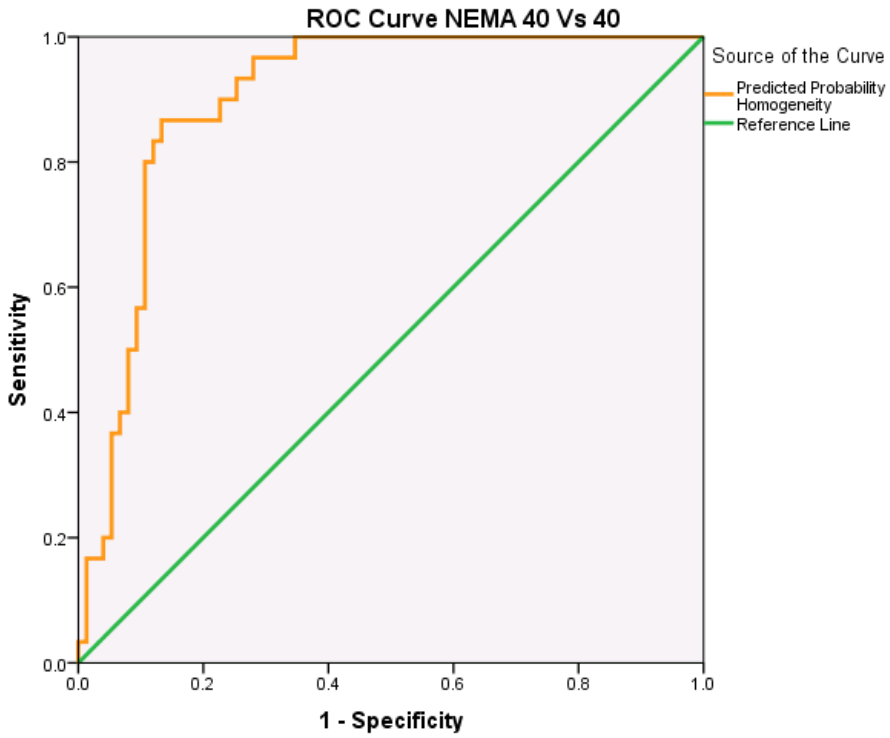


Figure 74: ROC curves depicting the prediction probability of homogeneity difference values using NEMA NU-2 phantom with Biograph 40

The ROC curves are plotted using this probability value as thresholds and sensitivity and specificity for the 105 data points are calculated. The area under the curve using homogeneity, as the predictor variable when comparing NEMA NU-2 with NEMA NU0-2 under Biograph 40 is 0.901

The homogeneity difference parameter has a large AUC and hence is a good predictor of similarity between image sets. Similar results are obtained for other texture parameters as depicted in Figure [73].

Dependent Variable Encoding

Dependent Variable	Independent Predictor Variable
SIMILAR/DISSIMILAR	Homogeneity + Correlation
SIMILAR/DISSIMILAR	Homogeneity + Contrast
SIMILAR/DISSIMILAR	Contrast + Correlation
SIMILAR/DISSIMILAR	Homogeneity + Contrast + Correlation
SIMILAR/DISSIMILAR	Homogeneity + Contrast + Correlation + Energy +Entropy

Table 15: Texture parameter combinations simultaneously used as independent predictor variables in Binary logistic regression analysis

To optimize the data and increase the prediction power of our model, we lastly used the texture parameters in combination. Various combinations were modeled together to increase the AUC. The best model would be one that gives a large AUC while using the minimum number of texture parameters. As Energy and Entropy were fairly irrelevant, they were not used in most combinations.

SPSS provides the provision to use multiple independent variables as predictor variables when performing Binary logistic regression analysis. As performed in the

previous analysis, SIMILAR/DISIMILAR categorical variable is used as the dependent variable. Various combinations of texture parameters differences are used as depicted in the table below.

The more complex case of analysis using Logistic regression is explained below. In this case we are comparing image sets between group I on Table [8], i.e. NEMA NU-2 vs NEMA NU-2 using Biograph 40. The categorical variable SIMILAR/DISSIMILAR is used, and the dependent variable and homogeneity difference and correlation difference are the independent predictor variables. The case processing summary table shows that there were 105 total data points used in the analysis.

Case Processing Summary

Case Processing Summary			
Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	105	100.0
	Missing Cases	0	.0
	Total	105	100.0
Unselected Cases		0	.0
Total		105	100.0

Table 16: Processing summary for homogeneity and correlation texture parameters being used as a predictor variable

The dependent variable encoding table shows the binary coding that is assigned to the categorical data. SIMILAR is coded as 1 and DISSIMILAR is coded as 0.

Dependent Variable Encoding

Dependent Variable Encoding	
Original Value	Internal Value
DISSIMIL	0
SIMILAR	1

Table 17: Binary codes assigned to the categorical dependent data during logistic regression analysis.

Block 0: Beginning Block

Classification Table					
	Observed		Predicted		
			SIMILAR/ DISSIMILAR		Percentage Correct
			DISSIMIL	SIMILAR	
Step 0	SIMILAR/ DISSIMILAR	DISSIMIL	75	0	100.0
		SIMILAR	30	0	0
	Overall Percentage				71.4

Table 18: Prediction probability calculation without using any predictor variable in logistic regression analysis.

Block 0 shows the prediction probability when considering the null hypothesis i.e. all the images are DISSIMILAR. In this case no predictor variable is used in the model.

For this comparison we had 105 total data points or image sets out of which only 30 were SIMILAR. The accuracy of model is calculated using the weights of individual data points in the two categories.

$$\text{Percentage Correct}=\frac{75*100+30*0}{105}=71.4\%$$

This simple model will correctly identify the Similarity or similarity among the image sets 71.4% of the times.

Classification Table					
	Observed		Predicted		
			SIMILAR/ DISSIMILAR		Percentage Correct
			DISSIMIL	SIMILAR	
Step 1	SIMILAR/ DISSIMILAR	DISSIMIL	68	7	90.7
		SIMILAR	7	23	76.7
	Overall Percentage				86.7

Table 19: Prediction probability calculation using both homogeneity and correlation difference values as the independent predictor variables in logistic regression analysis using 0.50 as a cutoff value

NOTE: The cut off value is 0.500

Table [19] depicts the prediction probability when considering the homogeneity difference parameter and correlation difference parameters as independent predictor variables. The similarity or dissimilarity of image pairs converted to binary form as

shown in Table [12] acts as the dependent categorical variable. The accuracy of the model in identifying image sets as similar or dissimilar increases to 86.7%.

$$\text{Percentage Correct} = \frac{75*0.907+30*0.767}{105} = 86.7\%$$

Homogeneity and Correlation - The plot in Figure [75] illustrates the combined predicted probability of the homogeneity difference parameter and correlation difference parameter taken together to correctly identify the similar/dissimilar image sets for all 105 data points.

The ROC curves are plotted using this probability value as thresholds and sensitivity and specificity for the 105 data points are calculated. Area under the curve using homogeneity and correlation, as the predictor variable when comparing NEMA NU-2 with NEMA NU-2 under Biograph 40 is 0.909

The combined homogeneity and correlation parameters have large AUC and hence are better predictor of similarity between image sets than homogeneity or correlation alone. Similar results are obtained for other combinations of texture parameters too as depicted in Figure [76].

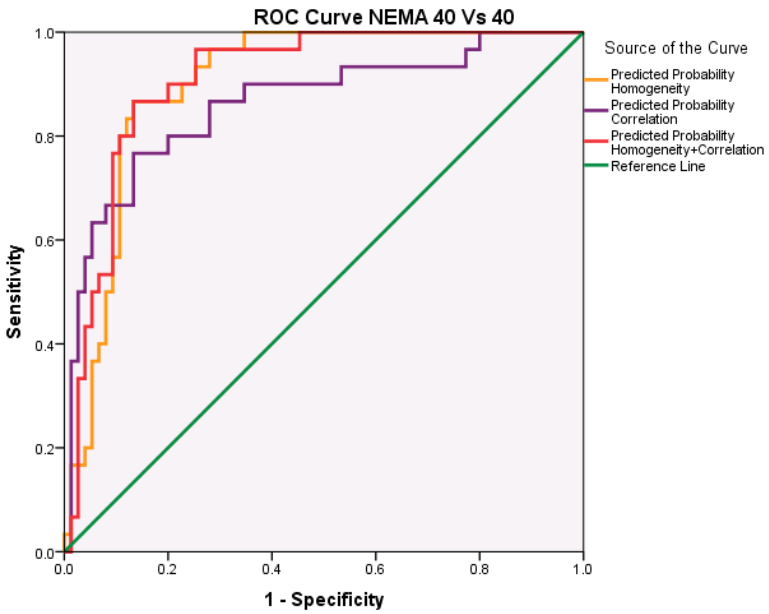


Figure 75: ROC curves for combined prediction probability of homogeneity and correlation difference using the NEMA NU-2 phantom

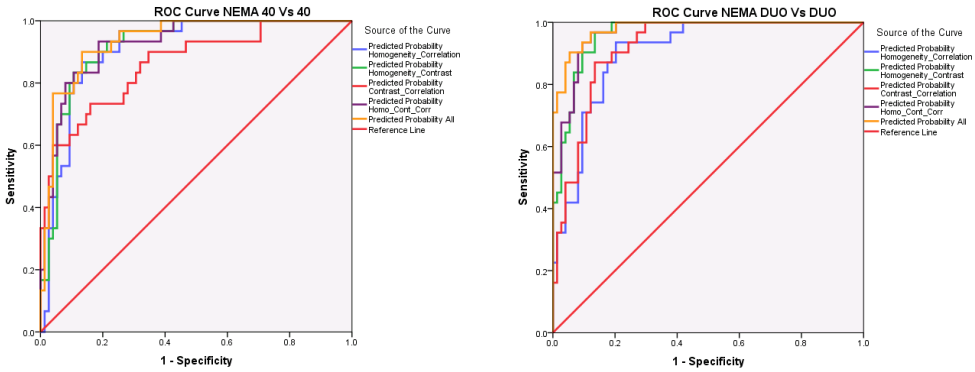


Figure 76: ROC curves and AUC for combination of difference values of texture parameters for various image set comparisons using NEMA NU-2 and CTN phantoms

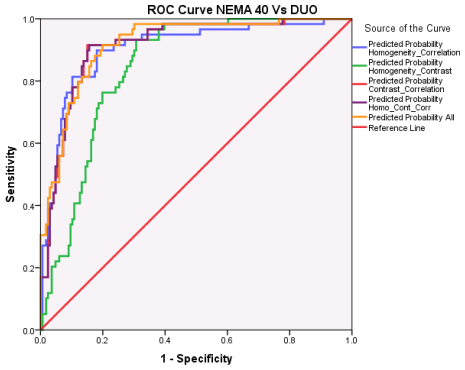
Figure 76 – Continued

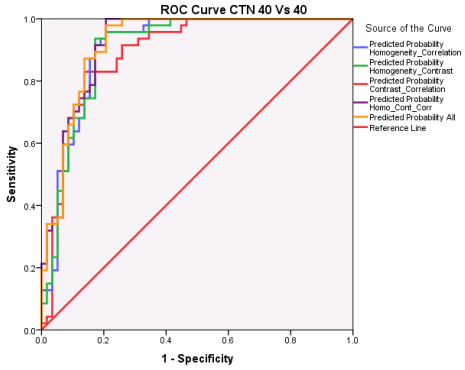
Area Under the Curve NEMA 40 Vs 40

Test Result Variables	Area
Predicted probability Homo_Corr	0.909
Predicted probability Homo_Cont	0.921
Predicted probability Cont_Corr	0.862
Predicted probability Homo_Corr_Cont	0.926
Predicted probability All or Combined Probability of all parameters	0.936

Area Under the Curve NEMA DUO Vs DUO

Test Result Variables	Area
Predicted probability Homo_Corr	0.905
Predicted probability Homo_Cont	0.961
Predicted probability Cont_Corr	0.918
Predicted probability Homo_Corr_Cont	0.966
Predicted probability All or Combined Probability of all parameters	0.98



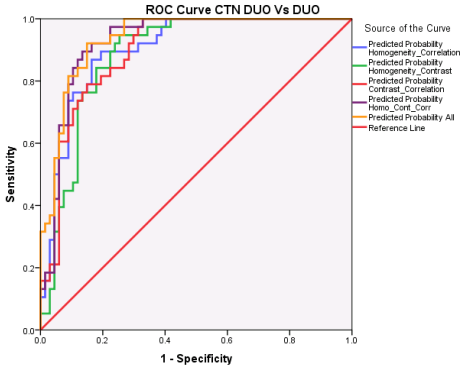


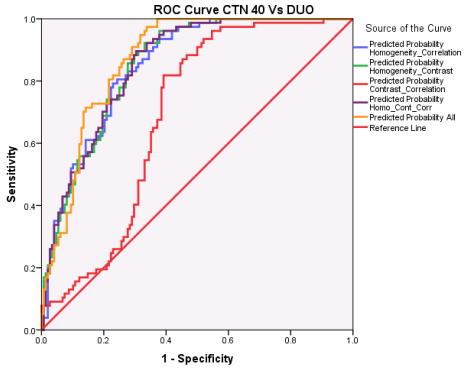
Area Under the Curve NEMA 40 Vs DUO

Test Result Variables	Area
Predicted probability Homo_Corr	0.904
Predicted probability Homo_Cont	0.842
Predicted probability Cont_Corr	0.912
Predicted probability Homo_Corr_Cont	0.913
Predicted probability All or Combined Probability of all parameters	0.918

Area Under the Curve VA 40 Vs 40

Test Result Variables	Area
Predicted probability Homogeneity_Correlation	0.906
Predicted probability Homogeneity_Contrast	0.902
Predicted probability Contrast_Correlation	0.892
Predicted probability Homo_Cont_Corr	0.92
Predicted probability All or Combined Probability of all parameters	0.921





Area Under the Curve VA DUO Vs DUO

Test Result Variables	Area
Predicted probability Homogeneity_Correlation	0.903
Predicted probability Homogeneity_Contrast	0.884
Predicted probability Contrast_Correlation	0.896
Predicted probability Homo_Cont_Corr	0.926
Predicted probability All or Combined Probability of all parameters	0.936

Area Under the Curve VA 40 Vs DUO

Test Result Variables	Area
Predicted probability Homogeneity_Correlation	0.851
Predicted probability Homogeneity_Contrast	0.853
Predicted probability Contrast_Correlation	0.689
Predicted probability Homo_Cont_Corr	0.853
Predicted probability All or Combined Probability of all parameters	0.871

4.4 Discussion

4.4.1 Quantitative Harmonization

As evident from the Figure [47], by showing the RC overlap area between the Biograph 40 and Biograph Duo, we can successfully harmonize two scanners of different make and model. This harmonization was achieved by identifying scanner specific reconstruction parameters that optimally aligned their respective RC curves. Quantitative harmonization could be very helpful in order to get reliable data when performing multi-center clinical trials from different sites using different scanners. As described in Chapter 3.5.4 various VOIs were used for a comparable RC curve generation.

4.4.1.1 VOI Strategies

Various VOIs used to quantitatively harmonize Biograph 40 and Biograph Duo were the SUVmax, 2X2X2 voxel cube, 3X3X3 voxel sphere and 5X5X5 voxel sphere. The harmonization zones for three of the VOI metrics for Biograph 40 and Biograph Duo are plotted to determine an area of RC overlap described by the shaded region in Figure [47]. We get a large harmonization region using SUVmax and 3-pixel diameter spherical VOI whereas the harmonization region when using the larger 5-pixel diameter spherical VOI is very limited. When using SUVmax as VOI, we are looking at the pixel with highest intensity value, which has high probability of being subjected to noise and as a result yields a wider or noisier harmonization region. When using a larger VOI, the data is smoothed out due to a larger volume resulting in narrower or less noisy harmonization region. Due to the difference in the resolution of Biograph 40 and Biograph Duo, arising from the difference in slice thickness, the area of RC curve overlap is narrow and limited to few reconstruction parameter combinations. In our study, the optimally harmonized reconstructions included the lowest resolution reconstruction of the higher performing scanner i.e. Biograph 40. If the Biograph Duo performance were a little worse, then the ability to harmonize reconstructions between these two scanners would have been absent.

It shows us that the quantitative harmonization between two scanners is limited by inherent scanner design properties. We can push the RC curve of a scanner higher by increasing the number of iterations and subsets, and reducing the FWHM of the smoothing filters, but within a limited range only. In our case, we had to degrade the performance of the higher resolution scanner (the Biograph 40) in order to match its RC curves with the lower performance scanner (the Biograph Duo).

One strategy to enhance the probability of quantitative harmonization in multi-center clinical trials, is to simply exclude the low performing scanners, lower RC curves force us to degrade the performance of other scanners in order to get matching performance. Hence, quantitative harmonization can help us in identification and elimination of the low performance scanners from the harmonization process.

To identify inherent variance within a scanner and its impact on quantitative harmonization, we took individual 6 minute scans using list mode feature in Biograph 40 and frame mode in Biograph Duo as described in Chapter 3.2. The 6 data sets were reconstructed using the same set of reconstruction parameters.

The Figure [77] above describes a RC curves for NEMA NU-2 phantom using Biograph 40. The reconstruction is performed using 8 iterations, 21 subsets and 2 mm Gaussian filter. Various RC curves are shown for respective VOIs used. The error bars with different VOIs represent the variance in the data when using 6 different frames under the same conditions. As described earlier, the variance is larger when using SUVmax and smaller when using larger VOIs (e.g. 5X5X5 voxel sphere).

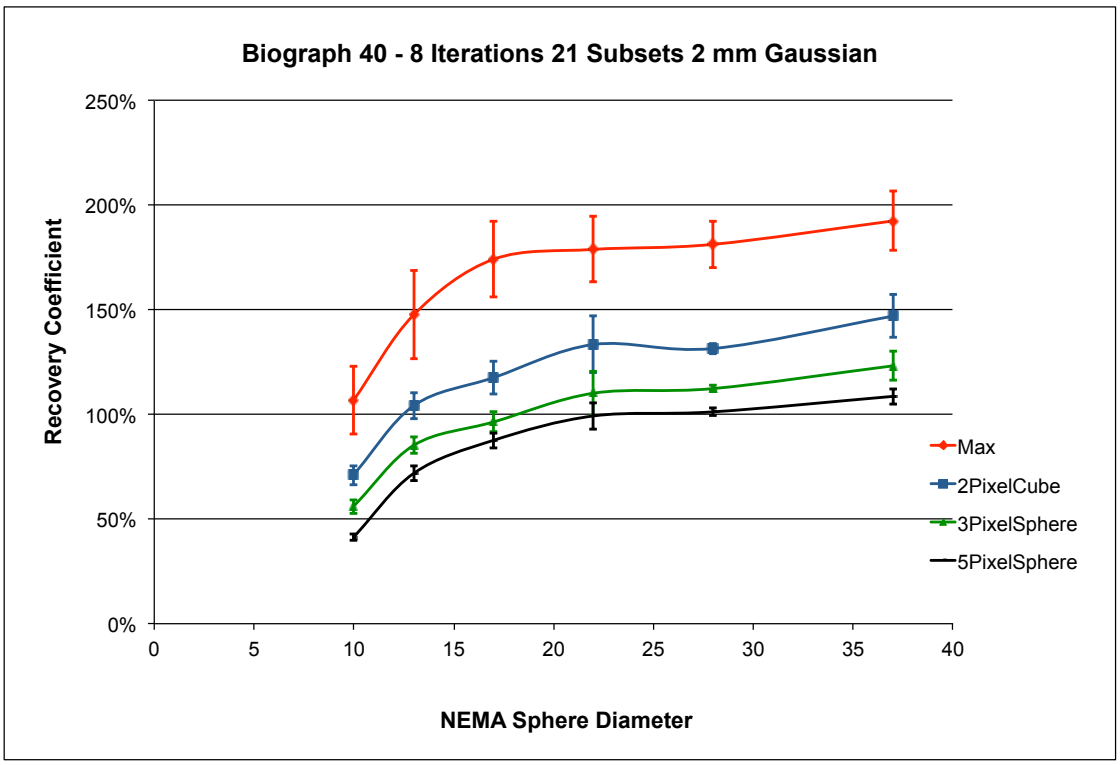


Figure 77: RC curves for NEMA NU-2 phantom using Biograph 40

Similar results are obtained for Biograph Duo with NEMA NU-2 phantom as shown in Figure [78]. The reconstruction is performed using 8 iterations, 32 subsets and 2 mm Gaussian filter. Various RC curves are shown for respective VOIs used. The error bars with different VOIs represent the variance in the data when using 6 different frames under the same conditions.

As expected, and similar to the Biograph 40 results, the variance in Biograph Duo is larger when using SUVmax and decreases with increasing VOI volume.

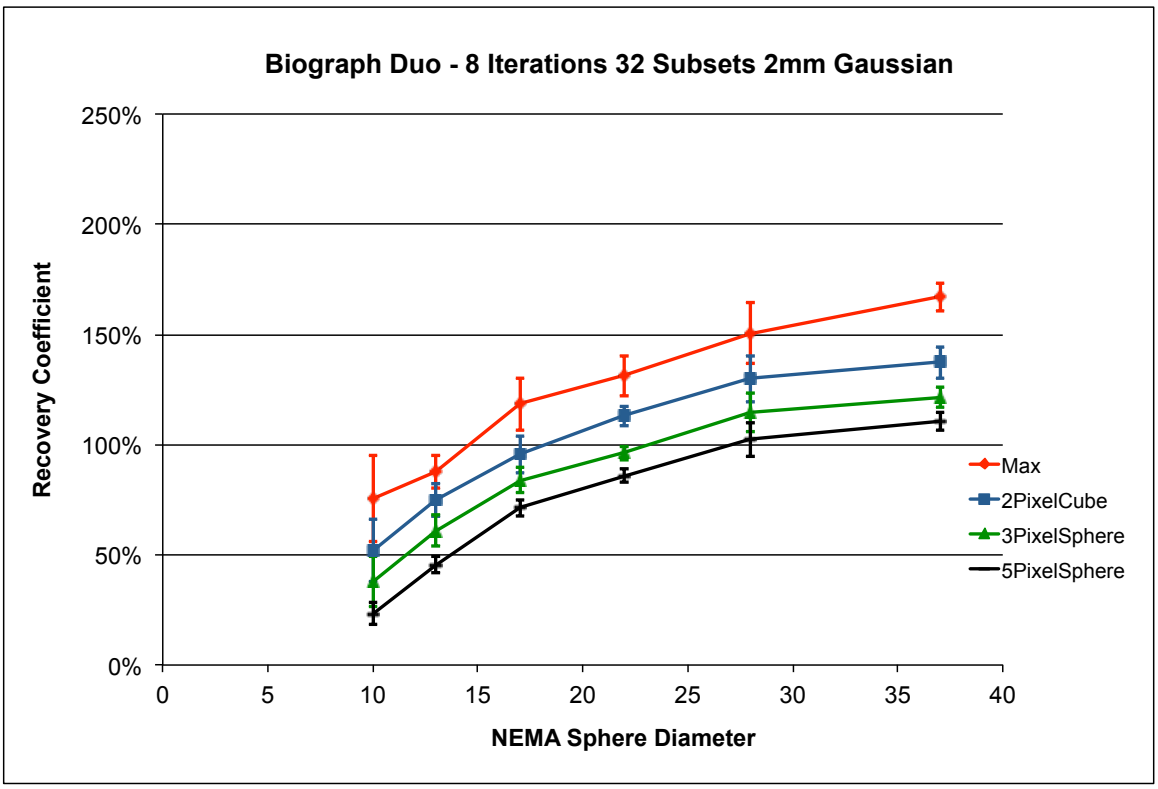


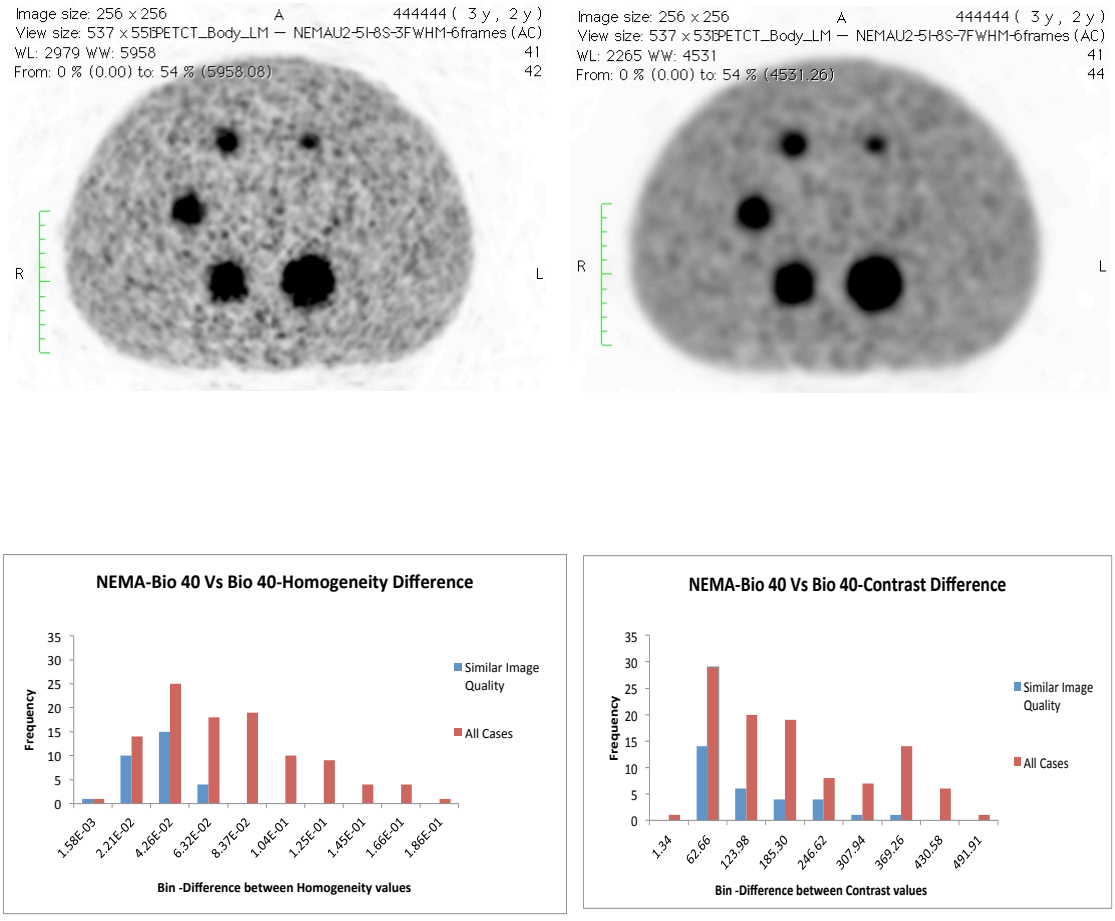
Figure 78: RC curves for NEMA NU-2 phantom using Biograph Duo

4.4.2 Qualitative Harmonization

For our purposes, qualitative harmonization is achieved when two images of the same object have similar qualitative features like lesion size, boundary shape, contrast, noise properties and lesion detectability. In this project we have tried to explore the use of texture metrics analysis as a tool to identify whether two images have similar image quality.

Five texture parameters were calculated for the different reconstructions and their differences were plotted as histograms. These difference histograms provided us with a

better understanding of the performance of each texture parameter. Ideally if two images are similar the difference between their respective texture parameters should be minimum. Any texture parameter that does not demonstrate this property would not be a good predictor for image quality sameness.



(a) (b)

Figure 79: Homogeneity and Contrast difference plotted, as histograms between images with image similarity rank=2. Comparison is performed between NEMA NU-2 Vs NEMA NU-2 with Biograph 40.

In our analysis, texture parameter difference values for scan pairs for images deemed to be of similar image quality are plotted with the difference value of the entire dataset in a histogram fashion (Figure [79). Based upon the histograms, the ROC curves were generated for each texture parameter. ROC is a graph plotting sensitivity versus (1-specificity) of a parameter and depicts relative tradeoffs between true positives and false positives. Each point on the ROC curve represents a sensitivity/specificity pair that corresponds to a certain threshold value.

The area under the ROC curve is a measure of how well a parameter can distinguish between two groups i.e. SIMILAR/DISSIMILAR images in our study. Results of histograms and ROC curves revealed that homogeneity, correlation and contrast were good predictors of image quality whereas energy and entropy were less relevant. The parameters homogeneity, correlation and contrast when used individually or together using logistic regression analysis did a good prediction of image quality sameness as described in the next Chapter 4.5.

The majority of time when any texture parameter failed to predict image quality or the majority of time when the method yielded a FP, it was almost always traced back to the images with an image similarity rank of 3. Images with image similarity rank of 4 or 5 never appeared as FP whereas images with similarity rank of 1 and 2 seldom appeared as FP.

For Example, as shown in the histogram (a) in Figure [79], the blue bars have an image similarity rank of 4 or 5 and the red bars below the threshold value of 6.32E-02 contributing to the FP have an image similarity rank of 3. None of the data point on these red bars below 6.32E-02 corresponds to the image similarity rank of 1 or 2.

A case where the texture analysis method fails and the FP is contributed by image quality rank of 1 and 2 along with 3 is also described in Figure [79]. For example in the histogram (b) shown in figure [79], regardless of the fact that contrast difference in the first bin is the smallest difference between any two images in the entire dataset, the

respective images have an image similarity rank of 2. Ideally images with small texture difference should look alike but in this case irrespective of small contrast difference, images have a low similarity rank. The homogeneity difference between the images is large, which suggests it to be a better predictor of image quality in this case. So in this case, using homogeneity as a predictor we get an accurate result (images are dissimilar) whereas using contrast we get an inaccurate result (images are similar).

Although using a single texture parameter often predicts visual similarity between images, per our visual similarity approach, it is clear from our data that this approach is not perfect. Accuracy increases when more than one parameter is used. In general, using multiple texture parameters increases substantially the area under the ROC curve.

4.4.2.1 Predicting Image Quality Similarity

Within the context of a hypothetical multi-center trial, where lesion detectability and image quality harmonization is important, our work demonstrates that performing a difference calculation using a combination of several texture metrics effectively determines whether two images are functionally of equivalent diagnostic quality. The accuracy of the method is shown in Table [20].

One example of a practical approach to using this image quality similarity technique using texture metrics is the following: In a multi-center clinical trial, ask each participating site to scan a standard phantom e.g. NEMA NU-2 and reconstruct the images with 5-6 clinically relevant reconstruction parameters. After receiving the reconstructed scans from different sites, we can run the reconstructions through the texture model and calculate respective texture metrics. Using these texture metrics, we will be able to identify those individual reconstructions from each site that are candidates of harmonized image quality. The same image data can be used to assess quantitative harmonization of the various reconstructed data sets. As multi-parametric texture assessment is not completely accurate verifying the image quality sameness by visual

inspection is still a necessary step, but not nearly so onerous as having to visually compare all image sets with one another. Using the texture analysis method we can reduce the number of reconstructions that need to be compared and validated for similar image quality by visual inspection. This method has the potential to save time and resources used for harmonization of results when using multiple scanners.

4.5 Conclusions

The optimal result of these experiments would identify the texture metric combination with the maximum AUC with minimum texture parameters. Homogeneity and Correlation combined with Contrast provides the maximum AUC when limiting calculations to only three parameters and hence represent a good estimate of the images being similar or dissimilar.

Sensitivity and Specificity were calculated as described in the previous sections. In order to calculate the accuracy associated with an optimal operating point on the ROC curve we had to find prevalence also known as prior probability. It is the probability of images having similar image quality in the absence of test data.

Prevalence =Probability of Image Sameness

= Probability of (Similar Images as per visual inspection) / Total Images

Accuracy = (TP+TN)/(TP+TN+FP+FN)

=(Sensitivity) (Prevalence) + (Specificity) (1-Prevalance)

The table below describes the sensitivity, specificity and accuracy values for the most optimal values on the ROC curves for various reconstruction comparisons depicted in table [8].

Reconstruction Groups	Sensitivity for Homogeneity, Correlation and Contrast combined	Specificity for Homogeneity, Correlation and Contrast combined	Accuracy for Homogeneity, Correlation and Contrast combined	Predicted Probability (AUC) for Homogeneity, Correlation and Contrast combined.
NEMA 40 Vs 40	93.3 %	78.7 %	82.69 %	0.926
NEMA DUO Vs DUO	93.5 %	90.5 %	91.2 %	0.966
NEMA 40 Vs DUO	91.5 %	84.3 %	86 %	0.913
VA 40 Vs 40	91.5 %	82.8 %	86.6 %	0.92
VA DUO Vs DUO	92.1 %	83.6 %	86.5 %	0.926
VA 40 Vs DUO	75.3 %	75.7 %	75.3 %	0.853

Table 20: ROC curves plotted using homogeneity, correlation and contrast difference values with various image sets

In most of the cases depicted in Table [20], for an optimal operating point on the ROC curve, the parameters can predict the similarity or dissimilarity between two images correctly with an accuracy of 85%. So we believe using this model, in phantom studies with images unseen we can get a good idea of the images with similar image quality. Also for a point on ROC curve, the Table [20] provides us with the expected values of sensitivity, specificity and accuracy

This model is a good predictor of image quality for phantom studies and in the future can also be extended to work with actual human images.

REFERENCES

[1] American Cancer Society: Facts and Figures 2012
http://www.cancer.org/research/cancerfactsfigures/cancerfactsfigures/cancer-facts-figures-2012

[2] H. Kaizer "A quantification of textures on aerial photographs", Boston Univ. Res. Lab., Boston, Masss. Tech. Note 121, 1955, AD69484.

[3] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein "Textural features for image classification" IEEE Transactions on systems, Man and Cybernetics, Vol. SMC-3, No. 6, November 1973, pp. 610-621

[4] Daeyoun Kim, J.JayLiu, ChonghunHan, "Determination of steel quality based on discriminating textural feature selection", Chemical Engineering Science, Chemical Engineering Science, 2011, Vol.66(23), pp.6264-6271

[5] Balaji Ganeshan, Sandra Abaleke, Rupert C.D. Young, Christopher R. Chatwin and Kenneth A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage", Cancer Imaging (2010) 10, pp.137-143

[6] Florent Tixier , Catherine Cheze Le Rest, Mathieu Hatt , Nidal Albarghach, Olivier Pradier, Jean-Philippe Metges, Laurent Corcos , and Dimitris Visvikis, "Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer". J Nucl Med 2011; 52:369- 378

[7] Huan Yu, Curtis Caldwell, Katherine Mah, Ian Poon, Judith Balogh, Robert Mackenzie, Nader Khaouam, Romeo Tirona, "Automated Radiation Targeting in head and neck cancer using region based texture analysis of PET and CT images", International Journal of Radiation Oncology, Biology, Physics, 2009, Vol.75 (2), pp.618-625

[8] The GLCM tutorial
Department of Geography, University of Calgary
http://www.fp.ucalgary.ca/mhallbey/tutorial.htm

[9] Eran Segal, Claude B Sirlin, Clara Ooi, Adam S Adler, Jeremy Gollub, Xin Chen, Bryan K Chan, George R Matcuk, Christopher T Barry, Howard Y Chang & Michael D Kuo, "Decoding global gene expression programs in liver cancer by noninvasive imaging", Nature Biotechnology, June, 2007, Vol.25 (6), p.675 (6)

[10] Master's thesis "Texture Analysis and Computer Aided Diagnosis for Prostate MRI", Jaochim Nilsson, University of Gothenburg, May 2011

[11] Arion F. Chatziioannou, Simon R. Cherry, Yiping Shao, Robert W. Silverman, Ken Meadors, Thomas H. Farquhar, Marjan Pedarsani and Michael E. Phelps, "Performance Evaluation of MicroPET: A High-Resolution Lutetium Oxyorthosilicate PET scanner for PET scanner for Animal Imaging", J Nucl Med 1999; 40:1164-1175

[12] G.B. Saha, Basics of PET Imaging: Physics, Chemistry, and Regulations, ISBN : 978-1-4419-0804-9

[13] Society of Nuclear Medicine and Molecular Imaging www.snm.org

[14] Mathworks MATLAB product documentation http://www.mathworks.com/help/toolbox/images/f11-27972.html#f11-29651

[15] Laws, K.I., Textured Image Segmentation. Ph.D Thesis, University of Southern California, 1980.

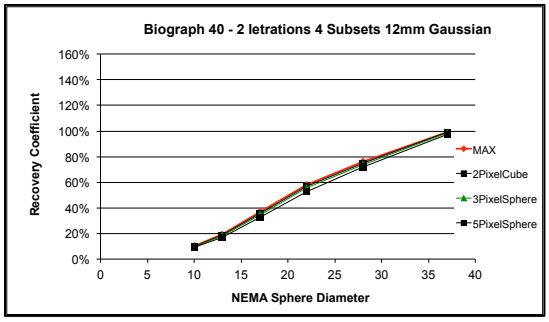
[16] Moses Amadasun, Robert King " Textural features corresponding to textural properties" IEEE Transactions on systems and Cybernetics, Vol. 19, No. 5, September/October 1989, pp.1264-

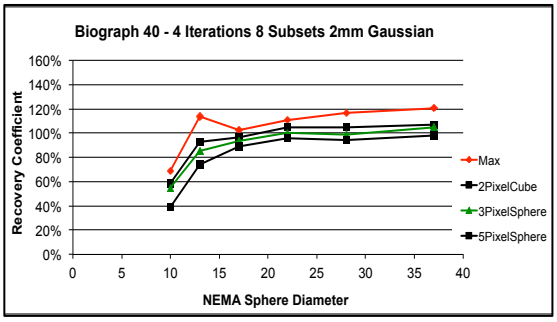
APPENDIX A – RC CURVES

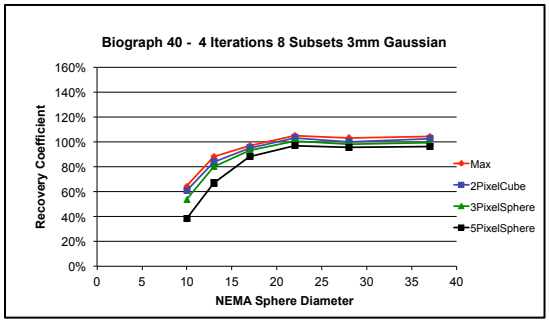
Plots describing RC curves with different VOIs, when using different reconstruction parameters with Biograph 40 and Biograph Duo are shown.

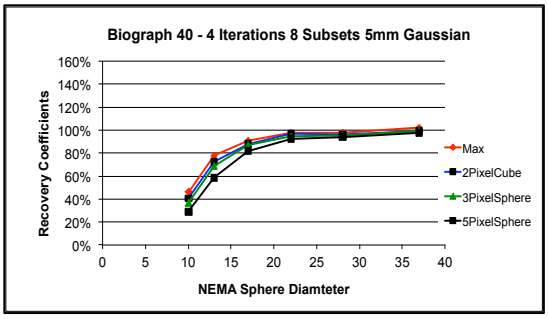
Variations in these plots provide us an idea of quantitative harmonization between two scanners.

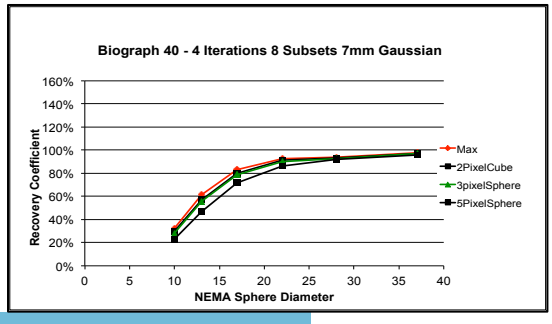
Biograph 40

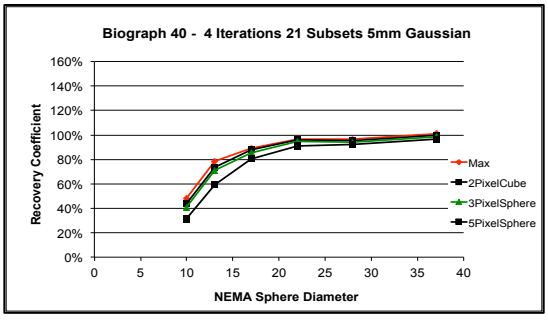


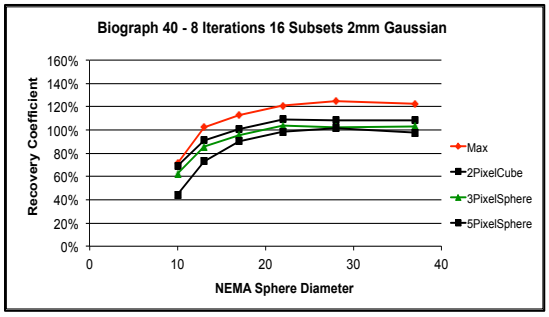


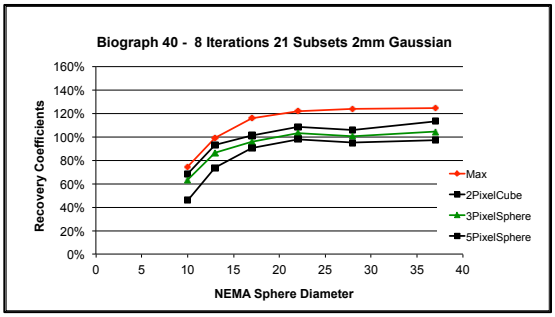


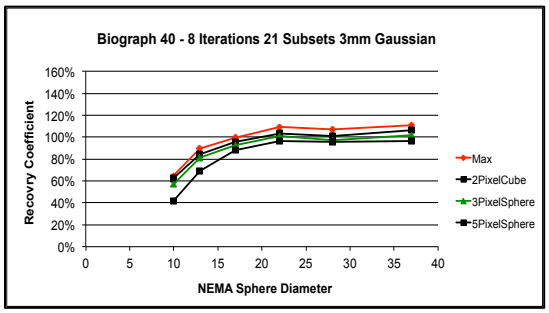




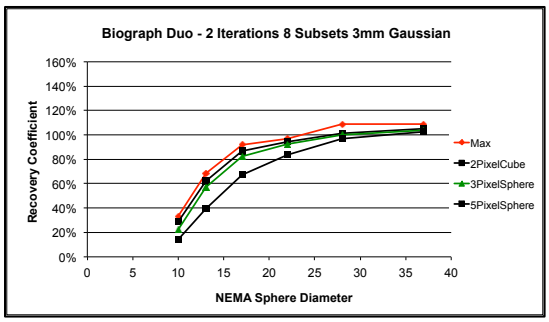


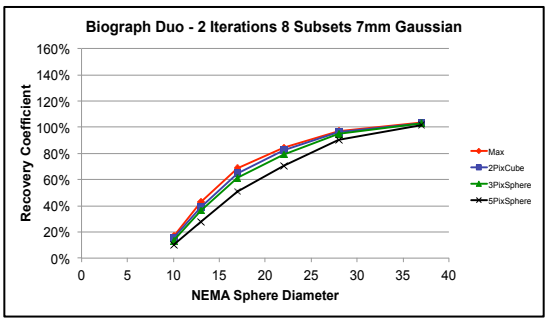


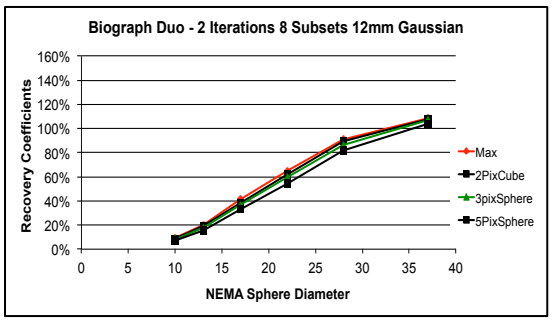


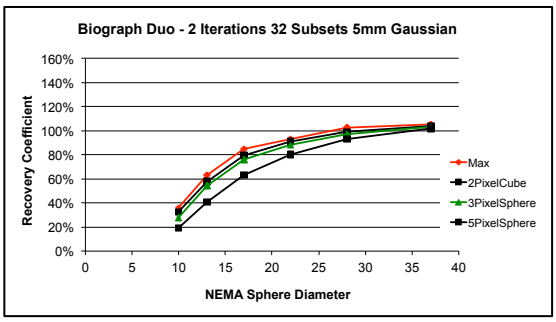


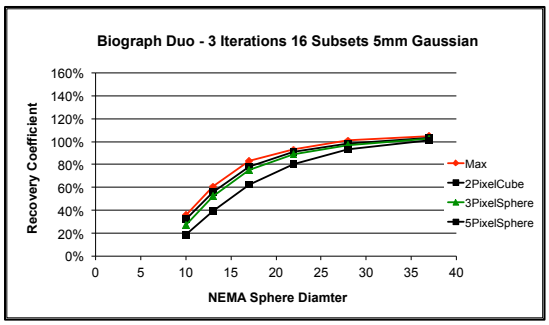
BIOGRAPH DUO

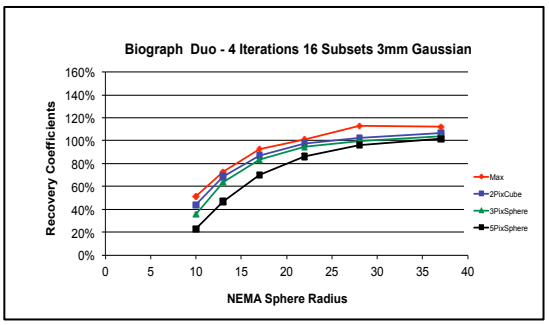


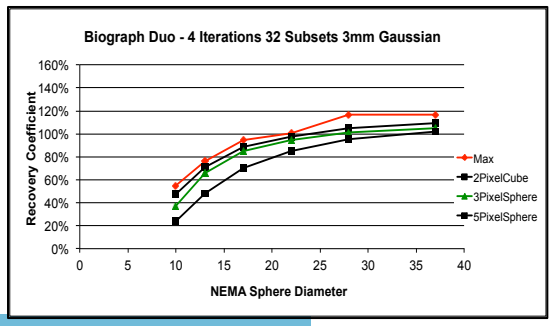


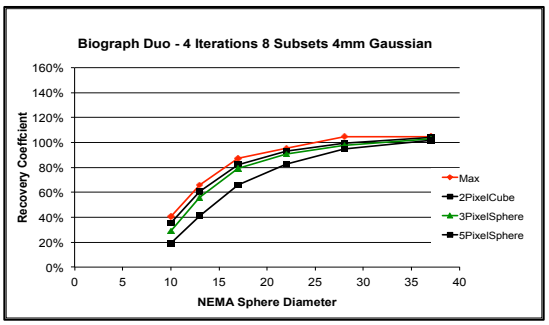


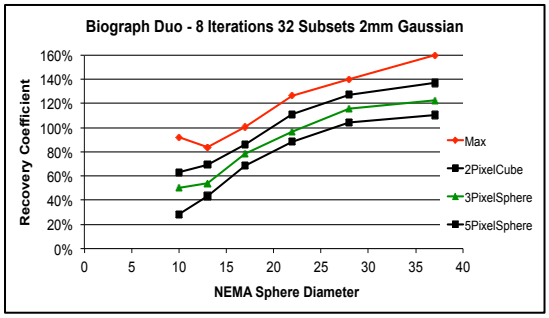












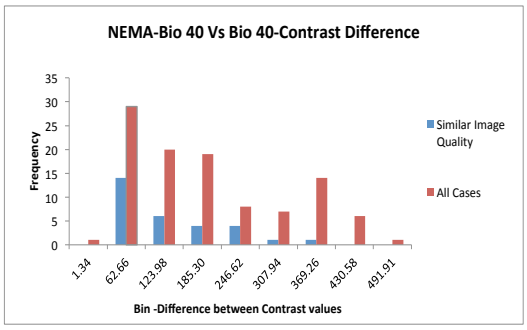
APPENDIX B - HISTOGRAMS

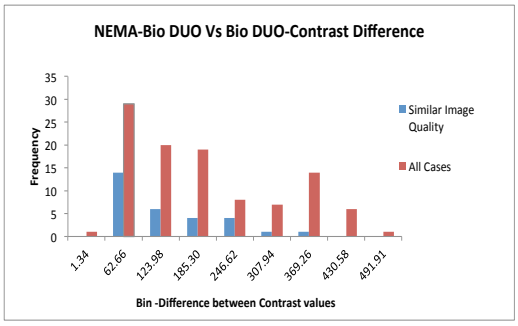
The results of plotting the difference in texture parameters of image pairs with similar image quality against their respective dataset is plotted in the histogram fashion as described in Chapter 4.2

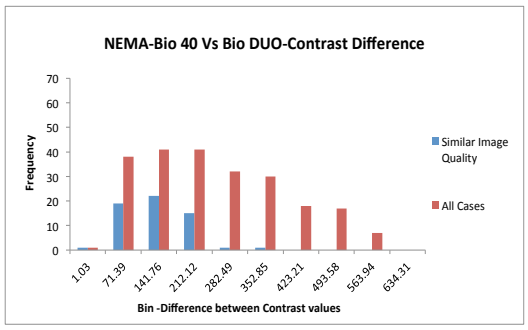
The plots are obtained for all the comparison groups depicted in table [8].

NEMA Phantom

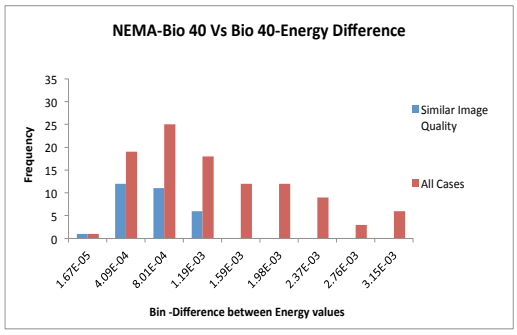
Contrast

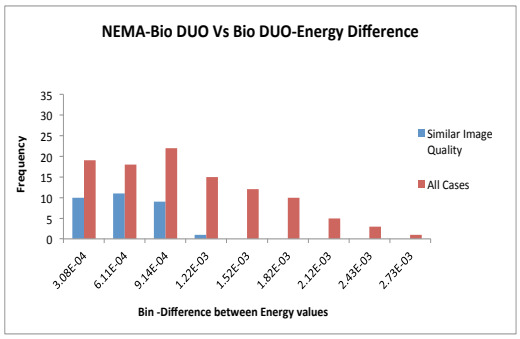


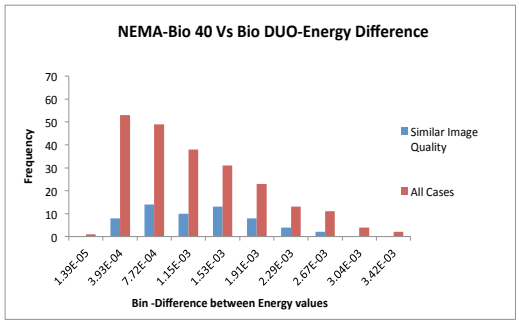




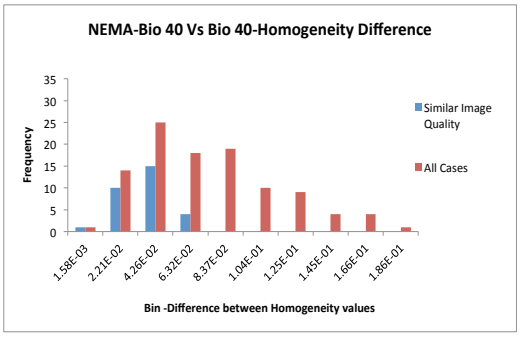
Energy

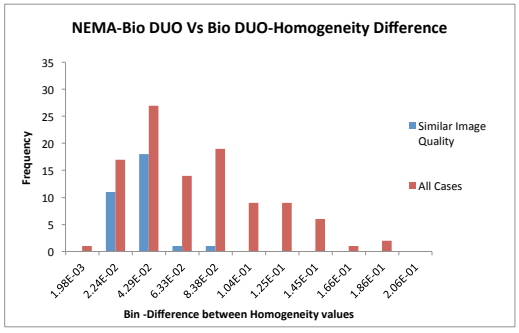


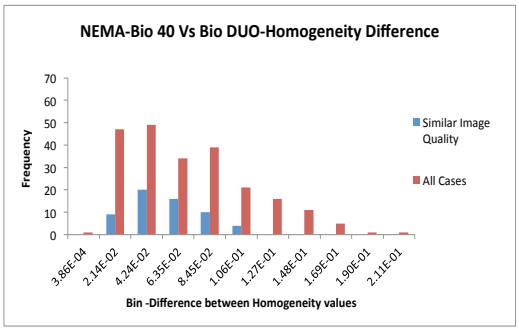




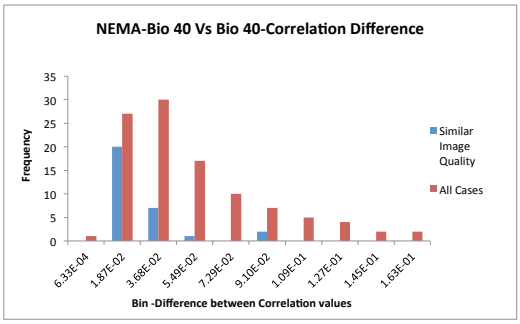
Homogeneity

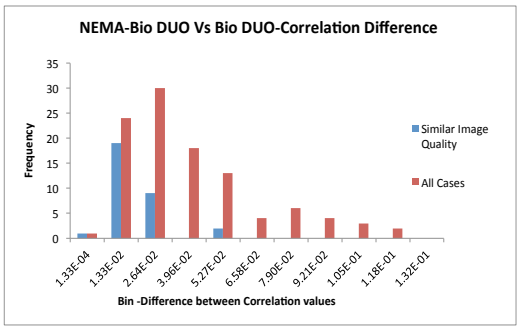


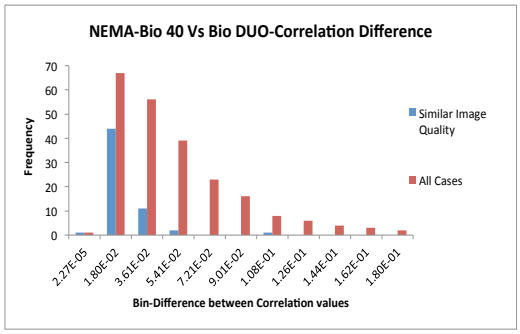




Correlation







Entropy