

Modeling decision-making in single- and multi-modal medical images

R. L. Canosa^{*a}, K. G. Baum^b

^aDept. of Computer Science, Rochester Institute of Technology, Rochester, NY USA 14623;

^bBiomedical and Materials Multimodal Imaging Laboratory, Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY USA 14623

ABSTRACT

This research introduces a mode-specific model of visual saliency that can be used to highlight likely lesion locations and potential errors (false positives and false negatives) in single-mode PET and MRI images and multi-modal fused PET/MRI images. Fused-modality digital images are a relatively recent technological improvement in medical imaging; therefore, a novel component of this research is to characterize the perceptual response to these fused images. Three different fusion techniques were compared to single-mode displays in terms of observer error rates using synthetic human brain images generated from an anthropomorphic phantom. An eye-tracking experiment was performed with naïve (non-radiologist) observers who viewed the single- and multi-modal images. The eye-tracking data allowed the errors to be classified into four categories: false positives, search errors (false negatives never fixated), recognition errors (false negatives fixated less than 350 milliseconds), and decision errors (false negatives fixated greater than 350 milliseconds). A saliency model consisting of a set of differentially weighted low-level feature maps is derived from the known error and ground truth locations extracted from a subset of the test images for each modality. The saliency model shows that lesion and error locations attract visual attention according to low-level image features such as color, luminance, and texture.

Keywords: Multi-modal fused images, eye-tracking, saliency maps, attention modeling

1. INTRODUCTION

Radiologists and radiological technicians are being called upon to extract information from a greater number of clinical images than ever before. In addition, these images are being generated from an ever increasing array of imaging modalities, including recently introduced multi-modal systems. With improvements in clinical imaging technology, the time it takes for a practitioner to interpret an image is generally longer than the time it takes to create the image¹. Also, an increasing variety of visualization choices can lead to a bewildering amount of visual information, some of it confusing, contradictory, or inaccessible to the practitioner. To make the best use of the abundance of information inherent in these images, health care professionals must be able to make diagnostic decisions quickly, efficiently, and most importantly, accurately. Therefore, fundamental research into the psychophysics of medical image interpretation and how the display of that information relates to interpretation is becoming increasingly crucial. Knowledge about which image features attract attention when distinguishing anomalous from normal tissue across an array of different imaging modalities may lead to a more effective use of medical imaging technology and, ultimately, to an improved health care system.

Eye movement monitoring is an effective and efficient means for extracting the pre-conscious behavioral and decision-making strategies of individuals engaged in complex, natural tasks. Specifically, when the task is to detect and localize pulmonary nodules in postero-anterior views of adult chest radiographs, experienced radiologists tend to have larger saccade amplitudes and restrict their search to a few, highly informative areas of the image². This suggests that experts use the strategy of “chunking” large sections of image locations under a single category, for example, “lesion likely” or “lesion not likely”. Radiology residents tend to have small saccade amplitudes and look at many different locations while searching for a potential lesion. After training, radiology residents look at fewer locations and spend less time scrutinizing the image before deciding. It is not surprising that practice enables more efficient search strategies and more correct detections; however, after training, residents tend to have increased false positive rates². Interestingly, false negatives (missed lesions) are usually fixated prior to decision, sometimes for as long as correctly detected lesions³.

*rlc@cs.rit.edu; phone: 1 585 475-5810

A long fixation duration (greater than 350 milliseconds, the length of time necessary for conscious recognition) classifies a diagnostic error as a *decision error*, as opposed to a *search error* (never fixated), or a *recognition error* (fixated, but not long enough for recognition)³. In one study, over 70% of missed malignancies in chest radiographs of lung nodules were found to have been fixated long enough for recognition⁴. Also, decision errors are the largest contributor of missed malignancies in pulmonary x-rays⁵ and mammograms⁶; inadequate search strategies play a minor role⁷. Therefore, knowledge of fixation locations as well as fixation durations can inform the development of new medical imaging display technology. For example, when used in a decision-support capacity, highlighting the location of potential errors with the implicit instruction to “look again” may improve diagnostic accuracy for experienced as well as resident radiologists.

The first part of this study is an examination of the frequency and types of errors (false positives, search, recognition, and decision) that people (non-radiologists) make when searching for anomalous features (simulated lesions) in simulated single and multi-modal fused positron emission tomography (PET) and magnetic resonance imaging (MRI) images. The goal of the first part is to compare the different types and frequencies of errors people make across the different modalities. The second part of this study introduces a saliency model for predicting the locations of the simulated lesions and the locations of potential errors in the images. The saliency model labels as highly salient specific locations in an image according to low-level image features that attract visual attention. The salient low-level features of the target and error locations are derived from eye position monitoring of participants during a search task.

2. METHODOLOGY

2.1 Participants

Nineteen observers (nine males and ten females) between the ages of 18 and 58 were recruited from the local college community, all naïve with respect to the purpose of the experiment, and none of which have had any experience with locating lesions in medical images. All participants were screened for normal color vision and all had normal or corrected-to-normal vision. Institutional Review Board approval was granted for the use of human participants in the eye-tracking experiment.

2.2 Eye Tracker Specification

An ASL Model 504 remote eye-tracker was used for this experiment, along with the ASL Eye-Trac 6 User Interface Software and Control Unit. The remote eye-tracker (Pan/Tilt Unit) has a vertical field of view of approximately 40° horizontally and 25° vertically and an accuracy of approximately $\pm 1/2^\circ$ visual angle.

The method used to detect eye movements for this study is based on the reflective properties of the eye. The retina of the eye is very reflective in the red and infra-red regions of the electromagnetic spectrum which enables a bright-pupil image of the eye to be detected on a sensor when the eye is illuminated with a co-axial light source of the proper wavelength. The reflection off the front surface of the cornea is known as the first Purkinje image (also known as the corneal reflection or “glint”) and can be used in conjunction with the bright-lit pupil to detect the magnitude and position of an eye movement. Either the pupil image or the first Purkinje image can be used alone to detect an eye movement, but these methods will be highly sensitive to movement of the head with respect to the detector. To circumvent the necessity of having the head remain completely immobile during the experiment, the pupil image and the corneal reflection image are used together. This enables the eye position to be calculated with respect to the head as the absolute difference (vector distance) between the centers of the pupil and corneal reflection images. The vector difference method works without securing the head because the distance between the two points remains constant whenever the head moves but the eye does not. As long as the head remains relatively stable with respect to the scene (for example by using a chin rest, as was used in this study), the gaze position with respect to the scene can be calculated and recorded.

A calibration procedure was performed for each participant before the experiment began and checked at the beginning and end of the experiment. To perform the calibration, a grid of nine points was projected onto the display used for the experiment. The screen resolution of the display was 1024 x 768 pixels and subtended a visual field of 39.2° horizontally and 29.5° vertically at a viewing distance of 52 cm. At this distance, approximately 26 pixels cover 1° of visual angle.

2.3 Procedure

The experiment consisted of monitoring and recording participants’ eye movements, fixation locations, and mouse clicks as they viewed a series of simulated brain images (66 test images and 11 training images). The images were generated

from single mode PET and MRI phantoms and multi-mode fused PET/MRI images (details of the fusion technique are given in Section 3). Three sets of fused images were used, each set using a different color look-up table (identified as 1a, 1b, and 2) for displaying the mixed modes. Each set of images (single-mode and fused) had between zero and five embedded simulated lesions. The fused images were sub-divided into three categories, depending upon whether the lesions were embedded in the PET image, the MRI image, or both. A summary of the image sets is given in Table 1. Figure 1 shows an example image from each set, where each example has five embedded lesions. In Figure 1, each lesion is surrounded by a white square (the squares were not visible to the participants during the experiment).

Table 1. Image sets used for eye-tracking experiment. Six images in each set.

Single Mode	Fused Mode 1a	Fused Mode 1b	Fused Mode 2
PETonly	fused1a_PET	fused1b_PET	fused2_PET
MRIonly	fused1a_MRI	fused1b_MRI	fused2_MRI
	fused1a_BOTH	fused1b_BOTH	fused2_BOTH

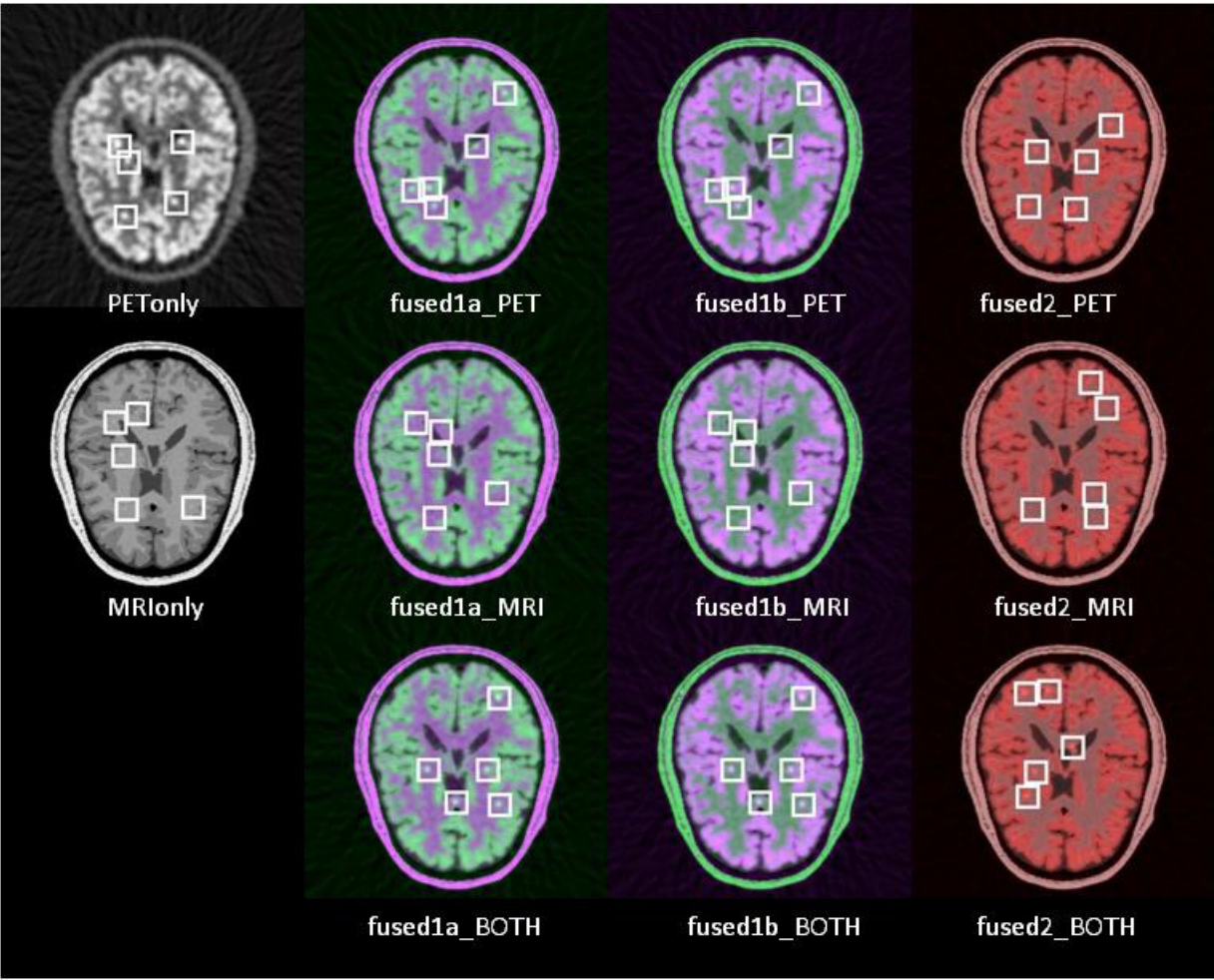


Figure 1. Example test images. Column 1: single-mode. Column 2: fusion color table 1a. Column 3: fusion color table 1b. Column 4: fusion color table 2. For all fused images, top image shows lesions embedded in PET image, middle image show lesions embedded in MRI image, and bottom image shows lesions embedded in both PET and MRI images. Lesion locations are surrounded by white squares here for illustrative purposes only; squares were not visible to participants during experiment.

The eye-tracking session lasted approximately 30 minutes per participant, including calibration time. Prior to the start of the experiment, each participant was given an instruction sheet with information about how the experiment would proceed. The instructions stated that, in general, a feature would appear as a circular spot in the image, and could be located anywhere within the anatomical portion of the image (i.e., a feature would never be located on the image border), and there may be between zero and five features in each image.

The images were divided into eleven sets, with six images in each set. Before each set, the participant was shown a training image to help him or her determine what a feature would look like. Each training image was preceded by a message of “Start Training”. When the training phase was completed, the testing phase began, preceded by a message of “Start Testing”. Each set of images had both a training phase and a testing phase. The experimental protocol is as follows:

“The purpose of the training phase is to show you what a feature looks like. The procedure for both training and testing is as follows: **click on the image wherever you believe a feature is located**. When you are finished clicking on the features, **hit the space bar to proceed to the next image**. It is possible that an image might not contain any features, in that case just hit the space bar without clicking on the image. The only difference between the training and testing phases is that for the training phase, after you hit the spacebar you will be shown where the features are actually located.”

Before each test image was displayed, a fixation target was shown on the screen. The participant was instructed to look at the center of the fixation target when it was shown. The next image was displayed automatically after the fixation target was shown, until all of the images had been displayed. At the completion of the experiment, the nine calibration points were displayed once again for a post-experiment calibration check. Data was collected that allowed for a determination of the number of true positives, false positives, false negatives of type A (search errors – never fixated), false negatives of type B (recognition errors - fixated less than 350 msec), and false negatives of type C (decision errors – fixated greater than 350 msec). Figure 2 shows example images from the eye-tracking data, indicating fixation locations, lesion locations, true positives, and each of the different types of errors.

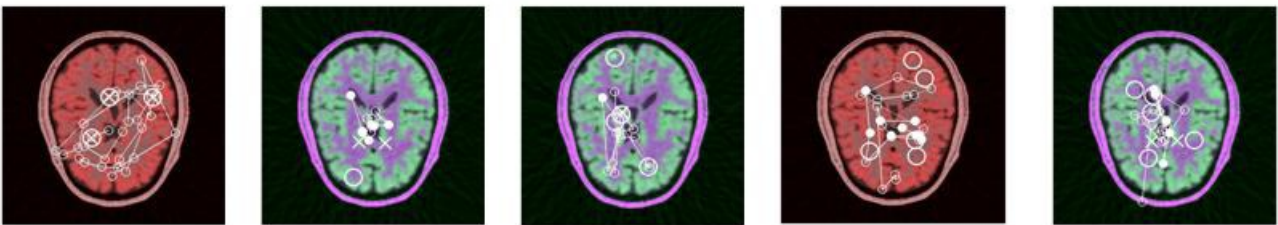


Figure 2. Examples of types of information collected from eye-tracking data. A large circle indicates location of a lesion, an X indicates a mouse click, a small, unfilled circle indicates a fixation less than 350 msec, and a small filled circle indicates a fixation longer than 350 msec. Left, participant was able to correctly locate all lesions. Second from left, two false positives and a false negative of type A. Third from left, one correct classification, two false negatives of type A and one of type B. Fourth from left, four false negatives of type A and one of type C, fifth from left, two false positives, three false negatives of type A, one of type B, and one of type C.

3. IMAGE FUSION AND SIMULATION

Artificial images created by medical imaging system modeling software were used for this study. These simulated images provided a data set with a known ground truth. The same background (an axial slice through the brain) for all images allowed naïve observers to participate in the study. Lesions with known size, shape, contrast, and location were inserted into the images.

The anthropomorphic brain phantom of subject four from BrainWeb was used (Figure 3)⁸. This high resolution (362x434x362, 0.5 mm³ pixels) phantom consists of 12 tissue classes: background, cerebrospinal fluid (CSF), grey matter, white matter, fat, muscle, muscle/skin, skull, blood vessels, connective tissue, dura matter, and bone marrow. Lesions were inserted into the white matter with a higher probability of being in a juxtacortical position, abutting the ventricles, approximating an individual with multiple sclerosis⁹.

In order to facilitate this, a probability mass function (pmf) used to determine lesion placement was derived from the white matter and CSF component images. The ventricles were first segmented from the CSF image after thresholding by selecting the 3D connected component with the proper volume (Figure 3)¹⁰. A forty-nine direction Euclidean distance map (EDM) specifying the distance from the ventricles was then generated¹¹. The pmf was formed by manipulating the EDM. First a power-law gray level transformation was applied to further weight locations abutting the ventricles¹⁰. The result was multiplied with the white matter component image, which segmented out the portion corresponding to the white matter and adjusted the weighting according to the proportion of white matter in each voxel. The intensities were then scaled to have a sum of unity providing the pmf (Figure 3). This pmf was used to randomly insert the desired number of lesions into the phantom before performing each simulation.

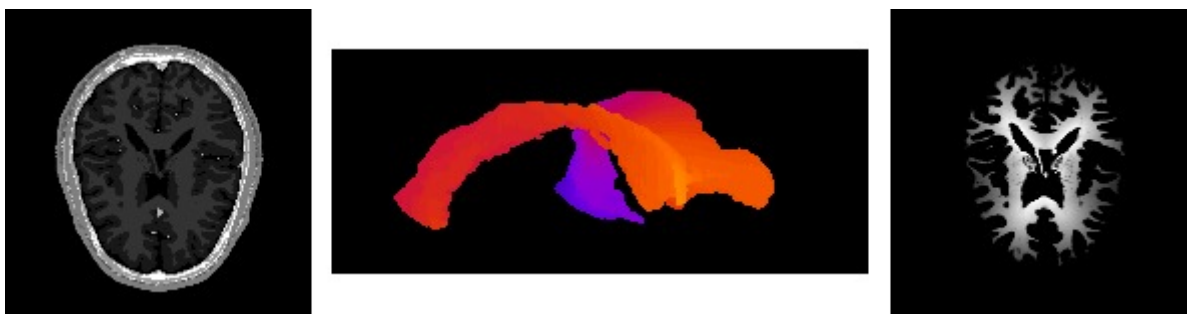


Figure 3. Left: axial slice through the brain phantom used when generating the PET and MRI images used in this study. Middle: surface rendering of the segmented ventricles. Right: probability mass function used for placement of the lesions.

The simulated MRI images were created using a modified version of the SIMRI software package, developed in CREATIS, Lyons, France¹². The phantom is treated as a set of magnetization vectors, whose evolution is governed by the Bloch equations. The rotation of the magnetization due to RF pulses, gradients, and relaxation is modeled. The final k-space signal is generated by summing the signal provided by each magnetization vector (isochromat summation), and the image is obtained by Fourier transforming the k-space data. If the fractional tissue components of each voxel are known, the magnetic resonance signal can be generated. It supports static field inhomogeneities due to improper shimming and tissue susceptibility, features efficient modeling of intravoxel inhomogeneities, and properly models the main artifacts such as susceptibility, wrap around, chemical shift, and partial volume effects. A spin-echo pulse sequence with TE = 20 ms, TR = 400 ms, and $B_0 = 1.5$ T was used and the lesion was assigned a $T_2 = 120$ ms which provided contrast to the white matter which had a $T_2 = 70$. The image was reconstructed to a 1 mm^3 resolution.

The simulated PET images were created using a modified version of the SimSET software package from the University of Washington¹³. SimSET uses Monte Carlo techniques to model the physical processes and instrumentation used in emission imaging. SimSET, which can be used to model both SPECT and PET, models the important physical phenomena including photoelectric absorption, Compton's scattering, coherent scattering, photon non-collinearity, and positron range. It supports a variety of collimator and detector designs, and already includes the attenuation properties for many common materials. The lesion was assigned an activity 8.5 times that of the white matter and the same attenuation properties. Data was collected in a 2D mode and reconstructed to have a resolution of 4.2 mm^3 , using a filtered back-projection technique with a Hamming window modulated Ram-Lak filter. Attenuation correction was performed using data from a simulated transmission scan.

The PET and MRI images were fused (merged into a single image) using three techniques. The advantage of a fused image lies in our inability to accurately visually judge spatial relationships between images when they are viewed side by side. Depending on background texture, mottle, shades, and colors, identical shapes and lines may appear to be different sizes¹⁴. This can be demonstrated by well-known simple optical illusions. The most obvious application is to combine a functional image that identifies a region of interest, but lacks structural information necessary for localization, with an anatomical image providing this information. It is hoped that the benefit gained by clearly displaying the spatial relationship between the two fused images outweighs the loss of information that occurs during the fusion process.

Fusion was performed through the use of two-dimensional color lookup tables. The color tables used for creating data sets 1a and 1b were generated by a genetic algorithm. This genetic algorithm searched for a color table which satisfied specific criteria. It was desired that this color table satisfy the order principle, the rows and columns principle, was

perceivably uniform, and had a high contrast. For more information on these properties and the genetic algorithm used to find color tables which satisfied them see^{15,16,17}. The color tables generated by the genetic algorithm and used in this study are shown in Figure 4. These fusion techniques have been shown to have benefits for localization tasks, and this study is evaluating their impact on detection^{16,18}.

The other fusion technique investigated uses a red/grey color table. The PET image determines the amount of red in the fused image, while the MRI determines the intensity. This technique was selected for study because it is used in a number of clinical applications. The color table for this technique is also shown in Figure 4.

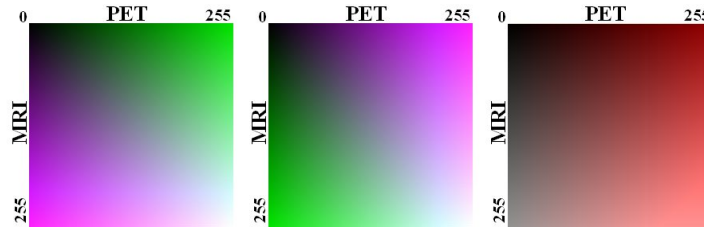


Figure 4. Left: two-dimensional color table used to generate data set 1a. Middle: two-dimensional color table used to generate data set 1b. This color table is a mirror of the one on the left. Both are used in this study since the genetic algorithm does not identify which source should be used for each axis. Right, the two-dimensional color table used to generate data set 2.

4. MODELING LESIONS AND ERRORS

4.1 Saliency map generation

This section describes the steps that were taken to construct the saliency map. The construction of the saliency map is an adaptation of a well-known model of visual saliency¹⁹. The saliency map consists of three essential feature maps – a color map, an intensity map, and an oriented edge (texture) map, as depicted in Figure 5. The feature maps along with the final merged saliency map are shown in Figure 5 as the shaded regions.

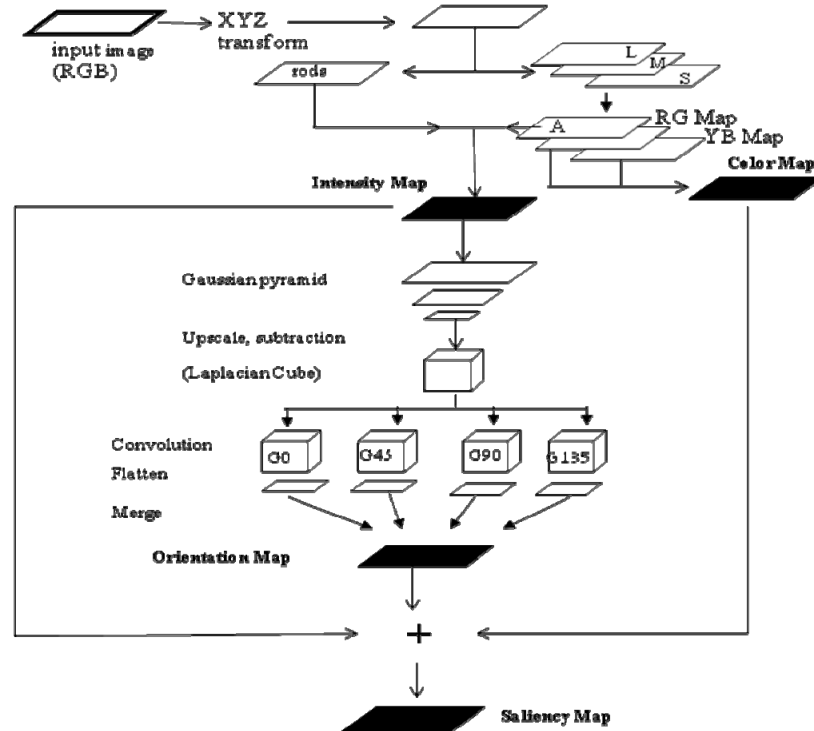


Figure 5. Schematic of the saliency map generation algorithm.

The pre-processing stage takes as input the original RGB formatted image and transforms that image from the RGB color space to the CIE tristimulus values, X, Y, and Z. The tristimulus values take into account the spectral properties of the display device and the color-matching functions of the CIE Standard Colorimetric Observer. Y corresponds to luminance and is equal to 100 when the object is a perfect white, and X and Z are used to calculate the chromaticity coordinates, which give each pixel's location in chromaticity space. In essence, X and Z are used to determine the perceived "colorfulness" of a stimulus, not including its luminance. The advantage of representing the input image in terms of XYZ tristimulus values rather than RGB digital counts is that the color and luminance of any input image is represented in a perceptually standardized way, independent of the display device. A second linear transformation represents the signal as rod (R) and cone (long L, medium M, and short S wavelength sensitive) responses. The rod and cone signals are calculated as the linear transform of the XYZ tristimulus values using transformation matrices²⁰. The rod and cone (L, M, S) responses are used to create the red/green and yellow/blue opponent color channels^{20,22}. These signals are also used in the CIE color appearance model of CIECAM97²³. In Figure 5, "A" refers to the achromatic color-opponent channel, "RG" refers to the red/green color-opponent channel, and "YB" refers to the yellow/blue color-opponent channel.

The color map takes as input the two chromatic signals of the cone responses, RG and YB. The resulting "colorfulness" at each pixel is defined as the vector distance from the origin (neutral point) to a point corresponding to the RG and YB signals. The intensity map is created by using a weighted sum of the output from the rod signal and the achromatic color-opponent channel. The oriented edge map is computed by first creating a seven level multi-resolution Gaussian pyramid²³ from the intensity signal. The second stage of processing involves subtracting adjacent levels of the pyramid, which simulates lateral inhibition and the center-surround organization of simple cells in the early stages of the primate visual system.

Since the human visual system has non-uniform sensitivity to spatial frequencies in an image, each level in the pyramid must be weighted by the human contrast sensitivity function²⁴. Contrast sensitivity is modeled by finding the frequency response of a set of difference-of-Gaussian convolution filters, and weighting each edge image in the Gaussian pyramid by the peak value of the response.

The next step is to refine the oriented edge map by representing the amount of edge information in the image at various spatial orientations. This is done by generating Gabor filters at four orientations: 0°, 45°, 90°, and 135°. In Figure 5, these are indicated by G0, G45, G90, and G135. A Gabor filter is derived from a Gaussian modulated sinusoid, and enables a local spatial frequency analysis to detect oriented edges at a particular scale. The biological justification for using this type of filter is that it simulates the structure of receptive fields in area V1 neurons. These neurons have been found to be tuned to particular orientations and have specific spatial frequencies²⁵. A Gabor function²⁶ is used to generate the filters, and each level of the Laplacian cube is convolved four times – once for each of the filters – creating four oriented Laplacian cubes. Each cube is flattened by summing the levels, which results in the oriented edge map. Once the color map, intensity map, and oriented edge map have been generated, they are summed to create a single low-level saliency map.

4.2 Scoring a Saliency Map

The saliency map described above assumes that each of the low-level feature maps, Color, Intensity, and Orientation, are weighted equally in the final summation step. In addition, the two color-opponent maps, RG and YB, are also assumed to have equal weighting when they are summed to create the Color feature map. The goal of this phase is to determine the optimal weights of the low-level feature maps such that the resulting saliency map will have a maximum response at the target locations. In other words, a saliency map should indicate high responses at the locations of the desired information (lesion, false positive, etc.) and low responses elsewhere.

In order to determine how heavily each feature map should be weighted for the optimal saliency map (given a target), a method must be found to compare the results using different sets of weights. To that end, a metric was developed to "score" a map for any given weight vector. The score of a map is defined simply as the ratio of the mean target saliency, S_t at some pre-defined locations to the mean saliency of the entire map, S_m .

$$\text{Score} = S_t / S_m \quad (1)$$

Mean target saliency S_t is found by first generating a saliency map (using a specific set of weights) for a particular input image. Next, the x,y-coordinates of a set of target locations are determined from eye-tracking data, ground-truth data, or from a record of observer responses (mouse clicks). Note that only a *subset* of the available targets (training data) should

be used to create the map, so that the remaining targets (testing data) can be used to test the resulting map for validity. For each target location in the training data, the x,y-coordinate is used as an index into the saliency map, and the saliency value at that location is extracted. A 7x7 pixel window (corresponding to $\frac{1}{4}^\circ$ visual_angle) is centered on the location, and all saliency values falling within the window are averaged together. This procedure is repeated for every target location in the map and the average of those values is used as the mean target saliency, S_t . The mean map saliency, S_m , is simply the average saliency over *all* locations in the map (target and non-target). The score of a map is then simply the ratio between the mean target saliency and the mean map saliency.

The map score can be used to determine how well a saliency map is able to locate a given type of response. If the score is close to one, then the map is not a good locator of the target – since S_t is nearly equal to S_m , any random location would do just as well at predicting the response. If, on the other hand, the score is greater than one, then the map is a good (better than random) model of the response because the target locations tend to be on regions of the image that the model has computed as being highly salient.

4.3 Feature Map Weight Generation

Determining the optimal weights to use for the low-level feature maps using an exhaustive search across the entire weight space is computationally prohibitive. Therefore, a genetic algorithm was developed to determine candidate weights, using the scoring metric described above as the fitness criteria.

Essentially, a genetic algorithm²⁷ “evolves” a solution by selecting the best candidate from a population of candidate solutions, over many generations. A good candidate from each generation is allowed to “mate” with another good candidate, reproducing offspring with variations of the attributes that contributed to the fitness of their parents. The goodness of a solution is determined by evaluating each candidate according to a pre-determined fitness criterion.

The algorithm is initialized with a first generation of ten random weight vectors (ten chromosomes). For every generation of ten chromosomes, the two best solutions (the two with the highest map score) are selected as parents and the remaining eight chromosomes are eliminated. The two parents mate (randomly exchange some of their weights) and produce eight new children with crossovers and mutations according to the established parameters. A total of 2,410 trials were run (300 x 8 plus original 10) for each image in the image database.

Figure 6 shows the final saliency map that was generated from the MRI image with five lesions. The figure shows a comparison between the un-weighted (or equally weighted) map, and the map with weights determined by the genetic algorithm. Notice that after thresholding the saliency maps at a level of 0.45, the lesions are not depicted as salient in the un-weighted saliency map, but are depicted as highly salient in the weighted saliency map. Note also the image on the far right of Figure 6. This image depicts the thresholded saliency map of the MRI image with three lesions using the weights that were determined from the MRI image with five lesions. This example shows that the weights can be determined from a training exemplar of the mode category and successfully used to detect targets in other images in the same category, without any *a-priori* information about the locations of targets in the test image. A follow-on study is in progress to determine the robustness of this observation across all imaging modalities studied here.

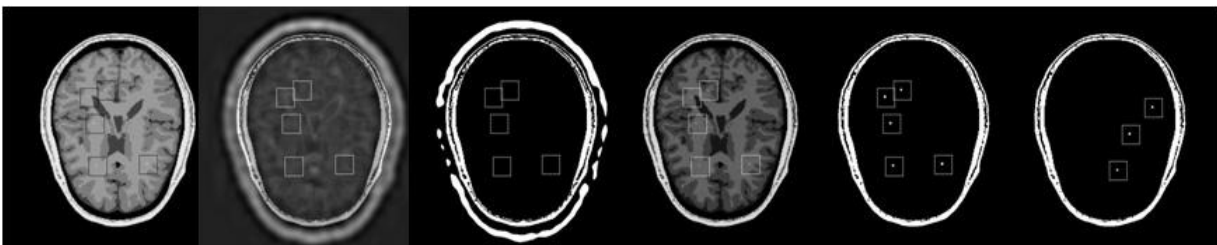


Figure 6. Un-weighted and weighted saliency maps. From left: original MRI image with five lesions, un-weighted saliency map of the same MRI image, un-weighted saliency map threshold of 0.45, weighted saliency map with weights determined by genetic algorithm, weighted saliency map threshold 0.45, weighted saliency map threshold 0.45 of MRI image with three lesions using the weights that were determined for the MRI image with five lesions. Weight vector = [0.003 0.000 0.001 0.996 0.000] for YB, RG, Color, Intensity, and Orientation. See Table 2 in Section 5 for weights used for all categories.

5. RESULTS AND DISCUSSION

5.1 Comparison of Types of Errors Across Different Modalities

Errors were classified as either False Positive (FP), False Negative A (FN/A – never fixated), False Negative B (FN/B – fixated less than 350 msec), and False Negative C (FN/C – fixated greater than 350 msec). Of the three types of false negatives, FN/A indicates a search error, FN/B indicates a recognition error, and FN/C indicates a decision error. The total number of false negatives possible per modality is 285 (15 lesions x 19 participants).

Figures 7, 8, and 9 show the total number of errors made by the participants in the eye-tracking study, by error category and according to the modality. Figure 7 shows a comparison of errors made in the single-mode PET images with the multi-mode fused images, where the lesion was embedded in the PET image. False positive errors dramatically declined in the fused images as compared to the single-mode images, whereas the false negative errors (of all types) either increased or remained the same, with the exception of fusion color table 2 (fused2_PET), which was low for all error categories. The significant decrease in false positives for the fused modalities may be due to the low resolution PET images being enhanced by the additional contextual information contributed from the MRI images.

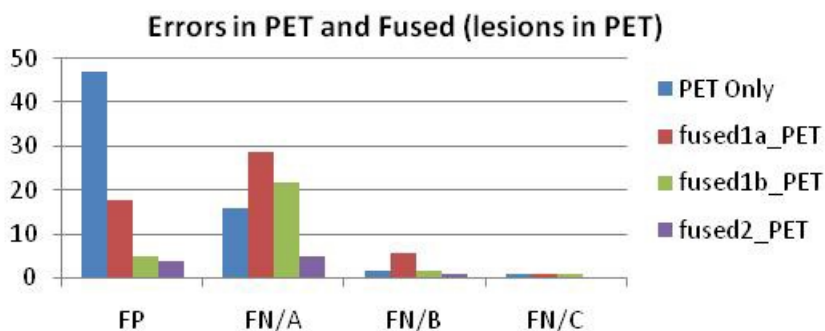


Figure 7. Total number of errors made by all 19 participants in the eye-tracking study, according to type of error and imaging modality. Errors in single-mode PET images are compared to errors in fused 1a, 1b, and 2 images with the lesions embedded in the PET image.

Figure 8 shows a comparison of the errors made in the single-mode MRI images and the fused images, where the lesion is embedded in the MRI image. For these images, the errors dramatically increased from the single-mode to the fused images, particularly for the fused 1a color table. The MRI lesions have a low contrast as compared to the higher dynamic range PET lesions, which may have caused them to be further obscured in the fused modalities, resulting in increased errors.

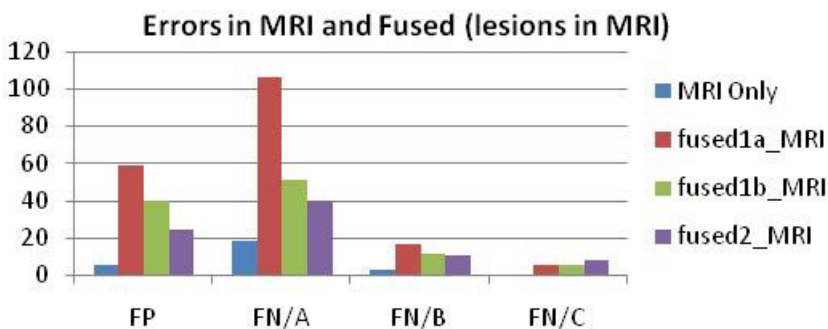


Figure 8. Total number of errors made by all 19 participants in the eye-tracking study, according to type of error and imaging modality. Errors in single-mode MRI images are compared to errors in fused 1a, 1b, and 2 images with the lesions embedded in the MRI image. Notice the range on the y-axis is greater in this figure than on the previous.

Figure 9 shows a comparison of the errors made in the single-mode PET and MRI images, and the fused images where the lesion is embedded in both the PET and MRI images. The results here are similar to the results shown in Figure 7, suggesting that embedding the lesion in the PET image enhances the ability of the viewer to extract meaningful information from the fused images, regardless of color table look-up. When identifying lesions in the multimodal images, viewers seem to rely more heavily on information provided

by the PET images, rather than the MRI images. It is also interesting to note that search errors (type A false negatives) are much more prevalent than recognition or decision errors (types B and C, respectively) for all modalities. Observers frequently do not fixate the lesion locations. This could be a result of the inexperience of the observers with viewing these types of images, and should be verified with a follow-up study using participants of varying levels of experience.

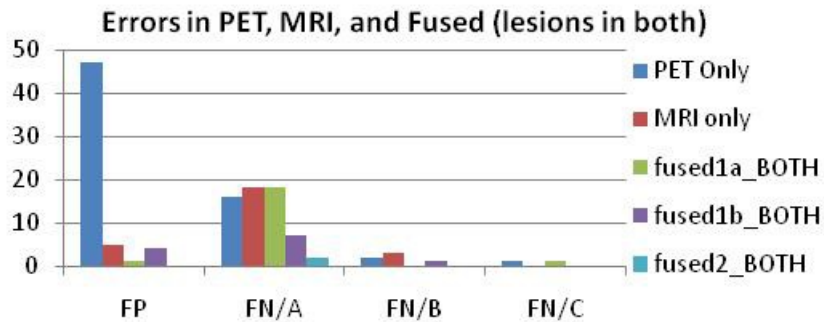


Figure 9. Total number of errors made by all 19 participants in the eye-tracking study, according to type of error and imaging modality. Errors in single-mode PET and MRI images are compared to errors in fused 1a, 1b, and 2 images with the lesions embedded in both the PET image and the MRI image.

Figure 10 shows a comparison of the average number of fixations and the average dwell time (total time spent) per image across all eleven modalities. In all cases, the MRI images (single-mode and fused) all had more fixations and longer dwell times than the PET images. This is an indication that the participants had a more difficult time locating lesions in the MRI images, regardless of fusion color table, which is in agreement with the number of errors made in the MRI images (Figure 8).

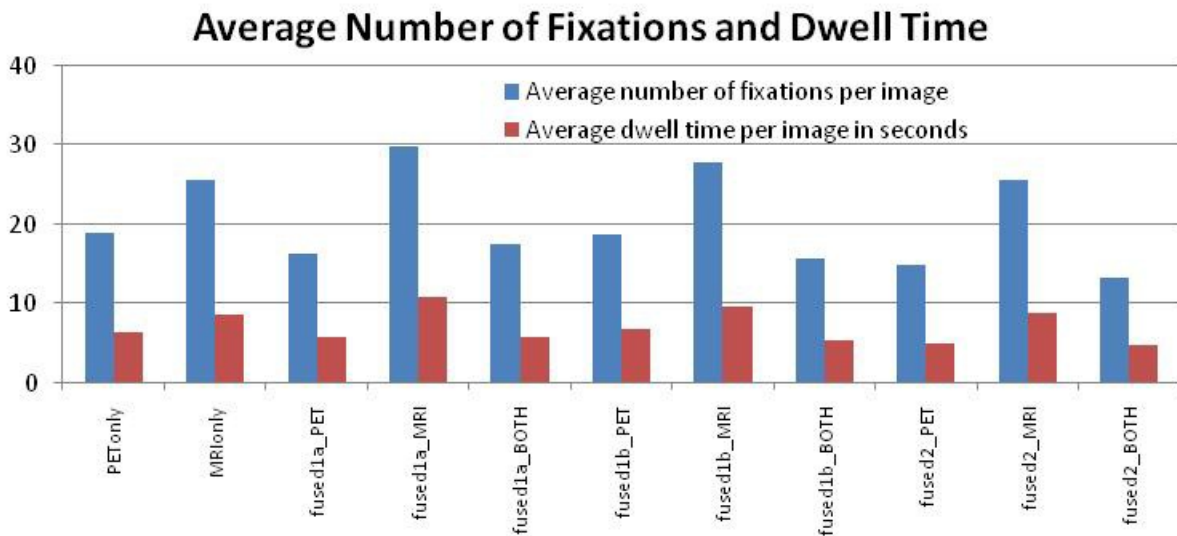


Figure 10. Average number of fixations per image and average dwell time per image (in seconds) across all 19 participants, for all eleven sets of images.

5.2 Saliency Map Weights for Lesions and Errors

The genetic algorithm was run as described in Section 4, with ground truth targets (GR) determined by lesion locations in the images with five embedded lesions. In addition, the genetic algorithm was run with FP targets, FN/A targets, FN/B targets, and FN/C targets. These target locations were found by analyzing the record of mouse clicks and fixations locations of the participants. Table 2 shows the feature map weights that were determined to be (near) optimal for locating targets of the specified type, for each of the eleven modalities. The feature map weight with the highest value for each target type/modality combination is highlighted.

Table 2. Feature map weights for ground truth and different types of errors, determined using genetic algorithm.

	YB	RG	Color	Inten	Orien		YB	RG	Color	Inten	Orien		YB	RG	Color	Inten	Orien		YB	RG	Color	Inten	Orien
PET Only Feature Map Weights						fused1a PET Feature Map Weights						fused1b PET Feature Map Weights						fused2 PET Feature Map Weights					
GR	0.337	0.228	0.222	0.180	0.033	GR	0.001	0.006	0.000	0.993	0.000	GR	0.321	0.001	0.227	0.445	0.006	GR	0.326	0.314	0.288	0.068	0.004
FP	0.954	0.011	0.035	0.000	0.000	FP	0.001	0.998	0.000	0.000	0.001	FP	0.000	0.997	0.002	0.000	0.000	FP	0.002	0.000	0.000	0.001	0.997
FN/A	0.979	0.005	0.015	0.000	0.000	FN/A	0.000	0.998	0.004	0.000	0.000	FN/A	0.983	0.005	0.012	0.000	0.000	FN/A	0.005	0.003	0.001	0.000	0.991
FN/B	0.940	0.014	0.046	0.000	0.000	FN/B	0.001	0.998	0.001	0.001	0.000	FN/B	0.002	0.026	0.971	0.000	0.001	FN/B	0.980	0.006	0.014	0.000	0.000
FN/C	0.917	0.012	0.071	0.000	0.000	FN/C	0.000	0.000	0.002	0.002	0.995	FN/C	0.002	0.993	0.005	0.000	0.000	FN/C	0.003	0.002	0.000	0.001	0.995
MRI Only Feature Map Weights						fused1a MRI Feature Map Weights						fused1b MRI Feature Map Weights						fused2 MRI Feature Map Weights					
GR	0.003	0.000	0.001	0.996	0.000	GR	0.350	0.048	0.303	0.271	0.028	GR	0.003	0.001	0.000	0.994	0.002	GR	0.000	0.002	0.003	0.995	0.000
FP	0.916	0.020	0.064	0.000	0.000	FP	0.002	0.991	0.007	0.001	0.000	FP	0.002	0.994	0.004	0.000	0.001	FP	0.000	0.000	0.002	0.001	0.997
FN/A	0.936	0.017	0.047	0.000	0.000	FN/A	0.352	0.070	0.282	0.280	0.017	FN/A	0.001	0.000	0.001	0.995	0.002	FN/A	0.001	0.001	0.002	0.995	0.001
FN/B	0.941	0.010	0.049	0.000	0.000	FN/B	0.371	0.052	0.264	0.299	0.014	FN/B	0.001	0.001	0.000	0.995	0.003	FN/B	0.003	0.000	0.001	0.002	0.995
FN/C	0.917	0.009	0.074	0.000	0.000	FN/C	0.299	0.035	0.369	0.295	0.002	FN/C	0.000	0.000	0.000	0.996	0.003	FN/C	0.000	0.001	0.000	0.001	0.998
						fused1a BOTH Feature Map Weights						fused1b BOTH Feature Map Weights						fused2 BOTH Feature Map Weights					
GR	0.000	0.001	0.000	0.996	0.002	GR	0.331	0.027	0.042	0.591	0.009	GR	0.454	0.186	0.235	0.112	0.013						
FP	0.001	0.995	0.004	0.000	0.000	FP	0.001	0.993	0.005	0.000	0.001	FP	0.003	0.001	0.001	0.003	0.993						
FN/A	0.000	0.993	0.001	0.006	0.000	FN/A	0.002	0.993	0.004	0.000	0.001	FN/A	0.003	0.000	0.003	0.000	0.994						
FN/B	0.000	0.993	0.006	0.000	0.000	FN/B	0.003	0.986	0.010	0.001	0.000	FN/B	0.000	0.000	0.001	0.001	0.997						
FN/C	0.000	0.997	0.000	0.000	0.003	FN/C	0.001	0.991	0.007	0.000	0.001	FN/C	0.003	0.000	0.001	0.001	0.995						

It is interesting to note that the feature weights for the ground truth (lesions) are frequently different from the feature weights of the different types of errors. For example, the highest weighted feature for lesions in the MRI-only set of images is intensity; however, for all of the error conditions in MRI-only, the highest weighted feature is the YB color opponent channel. This may account for the high error rates in the MRI images as depicted in Figure 8 – observers are just not looking at the “right” features in the image. This is an indication that errors may have specific attentional characteristics that differ from lesions, and this information can be exploited to warn observers of potential misses or false positives. This information might also be useful for a decision-support or computer-aided detection (CAD) system, to highlight locations with features corresponding to the different types of errors.

6. CONCLUSION

The goal of this study was to show that errors made by non-experts while locating lesions in simulated PET, MRI, and fused PET/MRI medical images can be characterized by modality as well as by low-level features in the image such as color, intensity, and texture. Fusion was found to be not always helpful for reducing errors at least for non-experts. When images are fused, information is lost which may hinder a detection task. When the lesion was present only in the PET image, fusion significantly reduced the number of false positives, most likely due to the participants’ ability to accurately localize the region of high metabolic activity within the white matter of the brain. When the lesion was present only in the MRI image, fusion with the PET image resulted in a higher false positive rate, most likely due to both the detection and the localization relying solely on the MRI features. In this case, the PET features seemed to obscure the MRI features. In general, viewers performed better with fusion technique 2 than with 1a or 1b. This may be because 1a and 1b used a more complex color table and observers had little training with the use of these color tables. A follow-on experiment with expert radiologists and radiology residents is necessary to verify the validity of characterizing the different modalities. What is clear, however, is that low-level features attract the attention of the human visual system during a search task, either consciously or pre-consciously, and those features are specific to certain types of targets. More research into the nature of decision-making at the level just below that of conscious awareness, as is enabled by eye-tracking experiments, will help to uncover the pre-conscious biases and strategies that contribute to image interpretation and misinterpretation.

REFERENCES

- [1] Robinson, P. A. J., “Radiology’s Achilles heel: Error and variation in the interpretation of the Roentgen image,” *British Journal of Radiology* 70, 1085-1098, (1997).

- [2] Manning, D., Ethell, S., Crawford, T., "An eye-tracking AFROC study of the influence of experience and training on chest X-ray interpretation," *Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment SPIE Vol. 5034*, (2003).
- [3] Krupinski, E. A. "The importance of perception research in medical imaging," *Radiation Medicine* 18 (6), 329-334, (2000).
- [4] Nodine, C.F., Kundel, H.L., "A visual dwell algorithm can aid search and recognition of missed lung nodules in chest radiographs." In Brogan, D. (Ed), *Visual Search* 1st edition, Taylor S. Francis, London, (1990).
- [5] Kundel, H.L., Nodine, C.F., and Carmody, D.P., "Visual scanning, pattern recognition, and decision making in pulmonary nodule detection," *Investigative Radiology* 13, 175-181, (1978).
- [6] Nodine, C.F., Mello-Thoms, C., Kundel, H.L., and Weinstein, "The time course of perception and decision making during mammographic interpretation," *American Journal of Roentgenology* 179, 917-923, (2002).
- [7] Manning, D., Barker-Mill, S.C., Donovan, T., and Crawford, T., "Time-dependent observer errors in pulmonary nodule detection," *The British Journal of Radiology*, 79(940), 342-346, (2006).
- [8] Aubert-Broche, B., Griffin, M., Pike, G. B., Evans, A. C., and Collins, D. L., "Twenty new digital brain phantoms for creation of validation image data bases," *IEEE Transaction on Medical Imaging* 25 (11), 1410-1416, (2006).
- [9] Calabresi, P. A., "Diagnosis and management of multiple sclerosis," *American Family Physician* 70 (10), 1935-1943, (2004).
- [10] Gonzalez, R. C., and Woods, R. E., "Digital Image Processing", Second Edition, Upper Saddle River, NJ: Prentice Hall (2002).
- [11] Russ, J. C., "The Image Processing Handbook", Fourth Edition, Boca Raton, FL: CRC Press (2002).
- [12] Benoit-Cattin, H., Collewet, G., Belaroussi, B., Saint-Jalmes, H., and Odet, C., "The SIMRI project: a versatile and interactive MRI simulator", *Journal of Magnetic Resonance* 173, 97-115, (2005).
- [13] Harrison, R. L., Vannoy, S. D., Haynor, D. R., Gillipsie, S. B., Kaplan, M. S., and Lewellen, T. K., "Preliminary experience with the photon history generator module of a public domain simulation system for emission tomography," *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference*, San Francisco, 1154-1158, (1993).
- [14] Rehm, K., Strother, S. C., Anderson, J. R., Schaper, K. A., and Rottenberg, D. A., "Display of merged multimodality brain images using interleaved pixels with independent color scales", *Journal of Nuclear Medicine* 35, 1815-21, (1994).
- [15] Baum, K. G., Helguera, M., and Krol, A., "Genetic algorithm automated generation of multivariate color tables for visualization of multimodal medical data sets," *Proceedings of IS&T/SID's Fourteenth Color Imaging Conference*, 138-143, (2006).
- [16] Baum, K. G., "Multimodal Breast Imaging: Registration, Visualization, and Image Synthesis", PhD Dissertation, Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science, (2008).
- [17] Trumbo, B. E., "Theory for coloring bivariate statistical maps," *The American Statistician* 35 (4), 220-226, (1981).
- [18] Baum, K. G., Schmidt, E., Rafferty, K., Helguera, M., Feiglin, D. H., Krol, A., "Investigation of PET/MRI image fusion schemes for enhanced breast cancer diagnosis," *IEEE Nuclear Science Symposium Conference Record* 5, 3774-3780, (2007).
- [19] Itti, L., Koch, C., and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11), 1254-1259, (1998).
- [20] Pattanaik, S.N., Ferwerda, J.A., Fairchild, M.D., and Greenberg, D.P., "A multi-scale model of adaptation and spatial vision for realistic image display," *Proceedings of the SIGGRAPH 98*, 287-298, (1998).
- [21] Hunt, R.W.G., "The reproduction of color", 5th edition. Kingston-upon-Thames, England: Fountain Press (1995).
- [22] Fairchild, M. D., "Color appearance models", Reading, MA: Addison-Wesley (1998).
- [23] Burt, P. J. and Adelson, E. H., "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications* 31 (4), 532-540, (1983).
- [24] Manno, J. L., and Sakrison, D. J., "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory* 20 (4), 525-535, (1974).
- [25] Hubel, D. H. and Wiesel, T. N., "Receptive fields and functional architecture of monkey striate cortex." *Journal of Physiology* 195, 215-243, (1968).
- [26] Gabor, D., "Theory of communication," *IEEE Proceedings* 93, 429 – 441, (1946).
- [27] Holland, J.H., "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence", Ann Arbor, MI: University of Michigan Press (1975).