Numerical Surrogates for Human Observers in Myocardial Motion Evaluation From SPECT Images

Thibault Marin, *Member, IEEE*, Mahdi M. Kalayeh, Felipe M. Parages, and Jovan G. Brankov*, *Senior Member, IEEE*

***Abstract*—In medical imaging, the gold standard for image-quality assessment is a task-based approach in which one evaluates human observer performance for a given diagnostic task (e.g., detection of a myocardial perfusion or motion defect). To facilitate practical task-based image-quality assessment, model observers are needed as approximate surrogates for human observers. In cardiac-gated SPECT imaging, diagnosis relies on evaluation of the myocardial motion as well as perfusion. Model observers for the perfusion-defect detection task have been studied previously, but little effort has been devoted toward development of a model observer for cardiac-motion defect detection. In this work, we describe two model observers for predicting human observer performance in detection of cardiac-motion defects. Both proposed methods rely on motion features extracted using previously reported deformable mesh model for myocardium motion estimation. The first method is based on a Hotelling linear discriminant that is similar in concept to that used commonly for perfusion-defect detection. In the second method, based on relevance vector machines (RVM) for regression, we compute average human observer performance by first directly predicting individual human observer scores, and then using multi reader receiver operating characteristic analysis. Our results suggest that the proposed RVM model observer can predict human observer performance accurately, while the new Hotelling motion-defect detector is somewhat less effective.**

***Index Terms*—Cardiac motion, cardiac-gated single photon emission computed tomography, image quality, machine learning, model observers, numerical observer.**

I. Introduction

IN MEDICAL imaging, an important step in the development of new imaging devices and algorithms is to optimize various design parameters, which should ideally be done using an appropriate image-quality figure of merit [1], [2]. Classical figures of merit such as mean-square error, or bias and variance, have been used extensively in the past. Unfortunately, they do not reflect the fact that the principal agent in medical decision-making is usually a human observer nor are they aware of the specific diagnostic task. Therefore, a consensus has emerged in the medical imaging field about the need to assess medical image quality by figures of merit that measure the performance of human observers for a given diagnostic task [1].

Cardiac-gated single photon emission computed tomography (SPECT), an important tool in the evaluation of coronary artery disease [3], is used to perform the following diagnostic tasks: detection and localization of perfusion defects; evaluation of severity, extent, and reversibility; evaluation of wall motion and thickening; and assessment of viability according to guidelines by the American Society of Nuclear Cardiology [4]. Information on cardiac motion is used for diagnosis, prediction of progression, and treatment of coronary artery disease. It is also used to extract useful diagnostic features such as ejection fraction, viability, etc., [5]–[8].

Proper evaluation of cardiac-gated SPECT should be based on human observer performance. However, studies involving human observers are difficult to implement, due to time and cost constraints. Therefore, numerical surrogates for human observers [1], [9], [10] are highly desirable tools in the early stages of design and optimization of devices and reconstruction methods.

At present, the most widely used model observer is the channelized Hotelling observer (CHO), a generalized likelihood ratio test, which has been successfully used as a surrogate for human observers in perfusion-defect detection tasks [9]–[21]. It has been shown that the CHO, with or without the inclusion of an internal-noise model, can predict human performance reasonably well in detection tasks for lesions [12], [13] and cardiac perfusion defects [14], [15]. Therefore, the CHO has been widely used to evaluate and optimize image reconstruction methods [11]–[21] using defect detectability as a figure of merit.

Inspired by the CHO approach, our group has developed a model observer methodology for perfusion defect detectability, which offers several advantages, most notably improved fitting and generalization accuracy. The approach was based on machine learning using nonlinear support vector machine (SVM) regression [22]. The term *generalization* describes the ability of a model observer to accurately predict the performance of a human observer over a wide range of image-reconstruction parameter settings on which data are not available for model observer tuning. The ability of a model observer to generalize to unseen data is of utmost importance since the purpose of the model observer is to evaluate new imaging devices or reconstruction methods for which human observer performance is as

Manuscript received June 25, 2013; revised August 14, 2013; accepted August 14, 2013. Date of publication August 22, 2013; date of current version December 27, 2013. This work was supported in part by the National Institutes of Health under Grant HL065425 and Grant HL091017. *Asterisk indicates corresponding author.*

T. Marin and F. M. Parages are with the Medical imaging Research Center, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: tmarin@iit.edu).

M. M. Kalayeh is with the Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816 USA (e-mail: mahdikalayeh@gmail.com).

*J. G. Brankov is with the Medical imaging Research Center; Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: brankov@iit.edu).

Digital Object Identifier 10.1109/TMI.2013.2279517

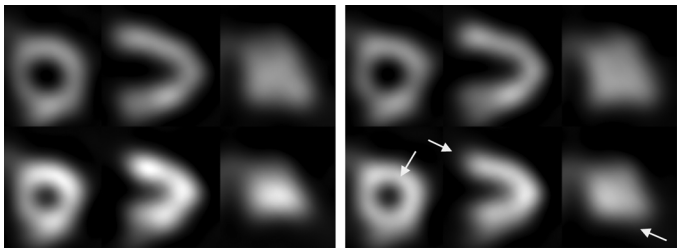

Fig. 1. Short-axis, vertical long-axis, and horizontal long-axis views of images reconstructed via FBP with spatial filtering and motion-compensated filtering (MC-FBP 3) (left) without motion defect and (right) with an anterior basal motion defect. Images are two frames from an image sequence: end diastole (ED; top row) and end systole (ES; bottom row). During human observer studies, the three views were used to display image sequences.

yet unknown; thus, images evaluated in practice by model observers are by definition unseen data and not merely different noise realizations of known types of data.

While model observers for the perfusion-defect detection task have been studied extensively, little research has been devoted to development of a model observer for motion-defect detection tasks. Model observers for perfusion defects are typically applied on static short-axis slices through the myocardium, and as such cannot capture motion information. Some research has been devoted to extension of the CHO to volumetric [23] and temporal [24] data. However, evaluation of motion visibility is arguably different from detection of a signal in a random intensity background. For motion analysis, initial motion observers were proposed in [25], [26], but these model observers are not specifically designed to correlate with a human observer. A more general approach to motion observers has been investigated by our group [27], but not directly in the context of cardiac SPECT. Therefore, in this paper, we developed and evaluated two new model observers for image-quality assessment of cardiac-gated perfusion SPECT performance in detection of motion defects.

The first method uses a Hotelling linear discriminant, a similar approach to that used in the perfusion-defect detection literature. The Hotelling discriminant does not have parameters to be tuned in order to match human observers performance. In the second approach, based on machine learning, the model observer aims specifically, by model tuning, to predict each human observer's performance in detection of motion defects (see Fig. 1 for an example of the cardiac motion defects considered in this work).

In our human observer studies for motion-defect detection, each human observer was asked to examine an image sequence and score his or her confidence as to the presence of a cardiac motion defect at a known location [e.g., on a scale of 1 (low confidence) to 6 (high confidence)]. The performance of the multiple human observers was then summarized using multiple-reader multiple-case (MRMC) analysis to obtain the area under the receiver operating characteristic curve (AUC) [1], [41].

In our second proposed approach, a model observer was trained to predict each human observer's confidence scores by using numerical features extracted from the image sequences. Once trained, the model observer yields predicted confidence scores that attempt to mimic those of each human observer individually, and these predicted scores can be used to determine the MRMC AUC in the same fashion as is normally done for human observers.

The success of the model observer in predicting human observers' confidence scores depends strongly on the choice of the numerical image features on which the model observer is based. In the perfusion-defect literature, bandpass channels generally serve as the features. In this paper, aimed at motion-defect detection, we propose a novel feature extraction method based on motion features and demonstrate that good prediction results are obtained.

The proposed motion features are extracted from a volumetric deformable mesh model (initially developed in [28] for purposes of image reconstruction), which is used to estimate a dense cardiac motion field. Selected features and human observer scores are then used to train a regression model based on a relevance vector machine (RVM), which is a modern Bayesian learning methodology [29]. Note that we have previously reported preliminary results using a Hotelling model observer (Hotelling-MO) [31], [32] and regression based on support vector machines [33].

In this paper, we first review the Hotelling-MO (which will be used as a basis for comparison). Next, we describe the RVM regression model observer (RVM-MO), the cardiac motion estimation scheme, and the proposed feature extraction method. Finally, we evaluate the performance of both proposed model observers and provide some discussion.

II. Methods

In this section, we describe the prediction model and the features that characterize the cardiac motion.

A. Prediction of Human Observer as a Regression

Consider a study in which a human observer is presented with an image sequence (see Fig. 1 for example image frames), for which he or she is asked to report a score r, corresponding to his or her level of confidence in the presence of a defect in myocardial motion at a specified location. The human observer's score, which is based on viewing the entire image sequence, is on a scale of 1 (low confidence) to 6 (high confidence). Now, in the proposed machine learning approach, we model the human observer's score as a parameterized regression function of a feature vector $\mathbf{x}$ extracted from the same sequence.

This can be summarized by the following regression model:

$$r = f(\mathbf{x}; \mathbf{w}) + \eta \tag{1}$$

in which $f(.)$ denotes the regression model and η is the modeling error. The parameterized regression model is trained, i.e., $\mathbf{w}$ are optimized, using a set of known human observer scores and features, $\{r_n, \mathbf{x}_n\}_{n=1}^{N_{tr}}$ where N_{tr} is the number of training scores for a given observer. Note that, in this work, the RVM-MO was trained separately for each human observer.

After the regression model was trained, the performance of the model observer was summarized in the same fashion as for the human observer, using the MRMC AUC analysis [1], [41].

B. Hotelling Model Observer (Hotelling-MO)

A model observer based on a Hotelling discriminant is often used for perfusion-defect detection analysis, where it is applied on features extracted via channels that mimic the human visual-system frequency response [9]. In this paper, we describe a different Hotelling-based model observer that uses features explicitly based on cardiac motion, as described later, *not the channelized features used in* [9]–[21].

Generically, a Hotelling model observer (Hotelling-MO) [9], [10], [31], is a generalized likelihood ratio test defined as

$$f_{\text{Hotelling}}(\mathbf{x}) = \left[\mathbf{K}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\right]^T x \tag{2}$$

where $\bar{\mathbf{x}}_i = E\left[\mathbf{x}|\mathcal{H}_i\right]$ is the expected value of the feature vector under hypothesis $\mathcal{H}_i$, where $\mathcal{H}_1$ signifies that a cardiac motion defect is present, and $\mathcal{H}_0$ signifies that no such defect is present. In (2), $\mathbf{K} = 1/2\left(\text{cov}[\mathbf{x}|\mathcal{H}_0] + \text{cov}[\mathbf{x}|\mathcal{H}_1]\right)$ denotes the average covariance matrix. Note that the Hotelling-MO only utilizes image features and the true classes labels, and depends neither on the human-observer performance, measured by AUC, nor on the scores. Therefore, it does not require a training step during which one would adjust model parameters. In practice, the expected values in (2) are replaced by sample estimates so that $\mathbf{K}$, $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ are calculated for each data set separately.

C. Relevance Vector Machine Model Observer (RVM-MO)

In this work, we used a relevance vector machine (RVM) to perform a nonlinear regression [29], [43]. An RVM is a so-called *kernel machine*, meaning that it is defined as a sum of kernel functions as follows:

$$\hat{r} = f_{\text{RVM}}(\mathbf{x};\mathbf{w}) = \sum_{n=1}^{N} w_n k(\mathbf{x},\mathbf{x}_n) + \eta \tag{3}$$

in which $\mathbf{x}_n, n = 1,\ldots,N$, are training examples, and $k(\mathbf{x},\mathbf{x}')$ is a kernel function. For notational simplicity, (3) can alternatively be rewritten as

$$\hat{r} = \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}) + \eta$$

in which

$$\boldsymbol{\varphi}(\mathbf{x}) = \left[1, k(\mathbf{x},\mathbf{x}_1),\ldots,k(\mathbf{x},\mathbf{x}_N)\right]^T.$$

In (3), one can think of the kernel function as a similarity metric that compares a data vector $\mathbf{x}$ to each of the training examples $\mathbf{x}_n$. These comparisons are then weighted by the coefficients w_n to form the prediction. Alternatively, one can think of (3) as a basis function expansion in which the basis functions are kernel functions involving the training example points $\mathbf{x}_n$.

The RVM is a very specific type of kernel machine, in which a probabilistic model is defined that promotes sparsity, meaning that most of the weights w_n will be zero. Thus, the RVM has an internal mechanism that helps avoid overfitting. The training examples $\mathbf{x}_n$ associated with nonzero weights w_n, known as *relevance vectors* (RVs), play the central role in this model, as they entirely define the summation in (3).

The RVM regression model assumes additive Gaussian modeling error and Gaussian distribution of the weights i.e.,

$$\hat{r} \sim \mathcal{N}(\hat{r}|\mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}),\sigma_\eta^2); \mathbf{w} \sim \prod_{n=1}^{N}\mathcal{N}(w_n|0,\alpha_n^{-1}).$$

Here, $\mathcal{N}(.|.)$ denotes a Gaussian probability density function and $\sim$ means "obeys."

The RVM methodology [29] is a probabilistic Bayesian kernel regression model that introduces a hierarchical prior model on previously defined scale parameters, σ_η^{-2} and α_n, that can be summarized as

$$\sigma_\eta^{-2} \sim \mathcal{G}(\sigma_\eta^{-2}|c,d); \boldsymbol{\alpha} \sim \prod\nolimits_{m=1}^{M}\mathcal{G}(\alpha_m|a,b). \tag{4}$$

Here, $\mathcal{G}(.|.)$ denotes the Gamma probability density function. To make the hyper-priors noninformative (i.e., flat), we fix their parameters to small values, i.e., $a = b = c = d = 10^{-4}$, which is typical in RVM implementations.

Here it is important to note that the true prior over the weights in $\mathbf{w}$, given as

$$p(w_n) = \int \mathcal{N}(w_n|0,\alpha_n^{-1})\mathcal{G}(\alpha_n|a,b)d\alpha_n \tag{5}$$

is a multidimensional Student t distribution. This distribution encourages sparsity of the weights in $\mathbf{w}$ [29] since it is sharply peaked around zero in each direction.

Now, this being a Bayesian approach, we desire to find a predictive distribution of $\hat{r}$ given the observed data $\{r_n,\mathbf{x}_n\}_{n=1}^N$ as

$$P(\hat{r}|\mathbf{r}) = \int p(\hat{r}|\mathbf{w},\sigma_\eta^2)p(\mathbf{w}|\mathbf{r},{\sigma_\eta}^2,\boldsymbol{\alpha}) \\ p(\sigma_\eta^2,\boldsymbol{\alpha}|\mathbf{r})d\mathbf{w}\; d\sigma_\eta^2\; d\boldsymbol{\alpha} \tag{6}$$

marginalized over all nuisance model parameters (parameters that are not of immediate interest but which define the regression model), where $\mathbf{r} = [r_1, r_2, \ldots, r_N]^T$ denotes the observed data and $\boldsymbol{\alpha} = [\alpha_1,\alpha_2,\ldots,\alpha_N]^T$ scale parameters. This is analytically intractable and we must seek an effective approximation as described next [29]. For clarity, the notation suppresses implicit conditioning upon the set of input vectors $\{\mathbf{x}_n\}$ in (6) and subsequent expressions.

Here, following RVM methodology, one assumes an approximation that the posterior of $p(\sigma_\eta^2,\boldsymbol{\alpha}|\mathbf{r})$ can be replaced by a delta function at its mode [29] or equivalently that: $\int p\left(\hat{r}|\mathbf{r},\sigma_\eta^2,\boldsymbol{\alpha}\right)p\left(\sigma_\eta^2,\boldsymbol{\alpha}|\mathbf{r}\right)d\sigma^2 d\boldsymbol{\alpha} \approx \int p\left(\hat{r}|\mathbf{r},\sigma_\eta^2,\boldsymbol{\alpha}\right)\delta\left(\sigma_{\text{MP}}^2,\boldsymbol{\alpha}_{\text{MP}}\right)d\sigma_\eta^2 d\boldsymbol{\alpha}$ is a good approximation. Using the fact that $p(\hat{r}|\mathbf{w},\sigma_\eta^2)$ and $p(\mathbf{w}|\mathbf{r},\sigma_\eta^2,\boldsymbol{\alpha})$ are multidimensional Gaussian distributions, and using (6) and

TABLE I
RVM ALGORITHM

0. For notational simplification, define a so-called design matrix $\mathbf{\Phi}$ as:

$$\mathbf{\Phi} = \left[\boldsymbol{\varphi}\left(\mathbf{x}_1\right),\ldots,\boldsymbol{\varphi}\left(\mathbf{x}_N\right)\right]^T. \tag{8}$$

1. Assign initial values to σ_η^2, $\boldsymbol{\alpha}$.

2. Compute $\mathbf{w}$ covariance matrix as:

$$\mathbf{\Sigma} = \left(\sigma_\eta^{-2}\mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{A}\right)^{-1}, \tag{9}$$

where $\mathbf{A} = diag(\boldsymbol{\alpha})$.

3. Compute mean of $\mathbf{w}$:

$$\boldsymbol{\mu} = \sigma_\eta^{-2}\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{r}. \tag{10}$$

4. Approximate hyperparameter values:

$$\gamma_n = 1 - \alpha_n\left[\mathbf{\Sigma}\right]_{nn}, \tag{11}$$

$$\alpha_n = \frac{\gamma_n + 2a}{\mu_n^2 + 2b}, \tag{12}$$

and:

$$\sigma_\eta^2 = \frac{\left\|\mathbf{r} - \mathbf{\Phi}\boldsymbol{\mu}\right\|^2 + 2d}{M - \sum_{n=1}^{N}\gamma_n + 2c}. \tag{13}$$

5. Repeat steps 2-4 until convergence.

the approximation described above, one can calculate the predictive distribution of $\hat{r}$ as

$$\hat{r} \sim p\left(\hat{r}|\mathbf{r}\right) = \mathcal{N}\left(\hat{r}|\boldsymbol{\mu}^T\boldsymbol{\varphi}\left(\mathbf{x}\right), \sigma_n^2 + \boldsymbol{\varphi}\left(\mathbf{x}\right)^T\mathbf{\Sigma}\boldsymbol{\varphi}\left(\mathbf{x}\right)\right) \tag{7}$$

where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are, respectively, the mean vector and covariance matrix of the weight vector $\mathbf{w}$. These parameters are estimated iteratively as described in Table I. Iterative steps in Table I are derived using expectation-maximization of the log-likelihood function and as such converge to a local maximum.

A complete description of the RVM methodology and derivation can be found in [29]. In this work we used the MATLAB Spider toolbox for implementation of the RVM training and testing [34]. Also note that in this work we used linear kernels, i.e., $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T\mathbf{x}'$ to limit model complexity, reducing the number of parameters that must be tuned for performance. For instance, radial basis functions (RBF) or polynomial kernels would introduce kernel parameters that must be tuned during training, typically using cross-validation procedures [22].

D. Feature Estimation

The proposed regression model uses a combination of features extracted from the image intensity and estimated cardiac motion from the image sequence on a 50 voxel region covering myocardium motion defect. The goal of this step is to build a dictionary of potentially relevant features from which the final features are selected in the feature selection step described below.

1) Feature Based on Image Sequence Intensity: The first feature we used is the *normalized myocardial regional maximum intensity*, which is similar to myocardial brightening features we used in [25] to capture wall thickening, and is routinely used in clinical myocardium evaluation. The normalization, which was done with respect to the maximum intensity in the whole data set, was needed because spatio-temporal filtering methods can produce images with different dynamic ranges.

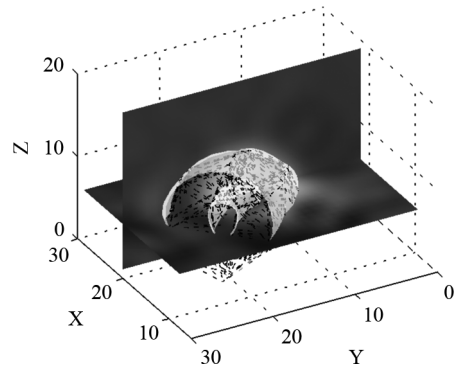

Fig. 2. Mesh model fitted to the left ventricle at end diastole. Cross sections correspond to sagittal and transverse views. Endocardial and epicardial surfaces are highlighted, with dotted lines delimiting tetrahedral mesh elements.

2) Features Based on Cardiac Motion Estimation: Since the model observer task is to predict a human observer's confidence level in the presence of a cardiac motion defect, we propose to use features extracted from an estimated cardiac motion field. For this motion estimation, we utilized a deformable mesh model as previously reported in [28] and [44]. Here, we briefly review the main points of this motion estimation procedure. First, we fitted a left ventricular shape model as described in [35]. This shape model directs the mesh structure creation so that it contains nodes located on the endocardial and epicardial surfaces, as shown in Fig. 2. In addition, the surface detection method also estimated the myocardium long axis that was used in a later step of feature extraction along axial, radial, and circumferential directions. Next, this volumetric mesh structure was deformed by iteratively minimizing the following image matching criterion:

$$J(\mathbf{D}^{k,l}) = (1-\lambda)E_{\mathcal{M}}(\mathbf{D}^{k,l}) + \lambda E_{\mathcal{S}}(\mathbf{D}^{k,l}) \tag{14}$$

where $\mathbf{D}^{k,l} = \{\mathbf{d}_v^{k,l}\}_{v=1}^{V}$ denotes the set of V nodal displacements between time-frames k and l, $E_{\mathcal{M}}$ is the intensity matching criterion, and $E_{\mathcal{S}}$ is a smoothing constraint and λ is a parameter that controls the tradeoff between the two criteria. The intensity-matching criterion is defined to account for myocardial brightening [28], [36]

$$E_{\mathcal{M}}(\mathbf{D}^{k,l}) = \sum_{b=1}^{B}\int_{\mathbf{r}\epsilon\mathcal{D}_b^k}\left[f_k(\mathbf{r}) - f_l(\mathbf{r} + \mathbf{d}^{k,l}(\mathbf{r}))\frac{|\mathcal{D}_b^l|}{|\mathcal{D}_b^k|}\right]^2 d\mathbf{r} \tag{15}$$

where $f_k(\mathbf{r})$ represents the intensity of the image at frame k and spatial location $\mathbf{r}$, $\mathbf{d}^{k,l}(\mathbf{r})$ is the displacement at location $\mathbf{r}$ between frames k and l, obtained by linear interpolation from a set $\mathbf{D}^{k,l}$ of nodal displacements and $|\mathcal{D}_b^l|$ denotes the volume of the bth tetrahedral mesh element $\mathcal{D}_b^l$ at frame k (where $b = 1, \ldots, B$

TABLE II
RESULTS OF FEATURE SELECTION PROCEDURE

	Feature from	Frame	Feature component	Quantity	*t*-statistic using FBP-SO data
1		2	tangential	std. dev.	4.723
2	Estimated cardiac	6	divergence	mean	4.433
3	motion vector-field	3	divergence	mean	4.401
4		6	tangential	maximum	4.260
5		5	intensity	maximum	6.034
6	Image intensity	9	intensity	maximum	4.746
7		7	intensity	maximum	4.707

is the total number of mesh elements and $k = 1, \ldots, K$ is the number to time gates). The smoothing constraint is defined by

$$E_{\mathcal{S}}(\mathbf{D}^{k,l}) = \sum_{v=1}^{V} ||\mathbf{d}_v^{k,l} - \bar{\mathbf{d}}_v^{k,l}||^2 \tag{16}$$

where $\mathbf{d}_v^{k,l}$ is the displacement of node $v (v = 1, \ldots, V$ where V is the total number of nodes) between frames k and l, and $\bar{\mathbf{d}}_v^{k,l}$ is the average displacement of the nodes connected to node v. Note that the weight λ was empirically set to $\lambda = 10^{-5}$ which yielded good mesh structures in a previous study by our group [28]. Finally, the volumetric motion field, defined by 3-D vectors, between end-diastolic and all other frames is linearly interpolated on a Cartesian voxel grid.

From the interpolated volumetric-vector-motion field we calculated the following quantities per voxel.

1) Norm $(\mathbf{D}_{\mathrm{N}})$.
2) Divergence $(\mathbf{D}_{\mathrm{D}})$.
3) Axial component $(\mathbf{D}_{\mathrm{A}})$.
4) Radial component $(\mathbf{D}_{\mathrm{R}})$.
5) Tangential component $(\mathbf{D}_{\mathrm{T}})$.

Next, for the 50 voxel region, covering the myocardial motion defect, for each frame and quantity listed above, we calculated regional *mean*, *median, histogram peak, maximum*, *minimum*, and *standard deviation* values.

The last feature added here was the *blood pool volume* estimated from the deformed mesh model endocardial surface.

The result of the feature extraction procedure yielded a total of $5 \times 6 + 2 = 32$ features, and each of these can be extracted for each of the 16 time frames in the sequence of myocardial images. Among these many features only seven were retained after the feature selection procedure described next.

E. Feature Selection

For each feature component we calculated the value of *independent two-sample t-test statistics* that measure feature separability of the two classes; in our case, between motion features extracted from image sequences with and without a motion defect, defined as

$$t_{\mathrm{score}} = \frac{|\bar{x}_1 - \bar{x}_0|}{S_{x_0,x_1}\sqrt{\frac{2}{N}}} \tag{17}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ correspond to the defect-absent and defect-present hypotheses, respectively, $\bar{x}_i = E[x|\mathcal{H}_i]$ is the expected value of a single feature under hypothesis $\mathcal{H}_i$, $S_{\mathbf{x}_0,\mathbf{x}_1}$ is the standard deviation pooled across the two hypotheses, and N is the number of samples in each group.

Only the 10 best features (those having the largest t-statistic values) were further considered. Next, from the best ten features we extracted a set of seven that had the smallest cross correlations among them. The feature selection was performed using *only* FBP-SO data, described in the following section, and selected features are reported in Table II with the corresponding t-statistics. Note that since the feature selection was based only on one reconstruction method, we eliminated bias when the model observer was tested on the other data set.

III. SIMULATED DATA AND OBSERVER STUDY

In this section, we describe the procedure to obtain simulated cardiac-gated SPECT data, followed by a description of the human observer studies.

A. Simulated Dataset

1) Phantom: In this work, we used the Mathematical Cardiac Torso (MCAT) phantom [37] wherein a cardiac motion defect was inserted as described previously in [38]. This was accomplished by replacing a segment of the anterior left ventricular (LV) myocardium by a segment contracting less vigorously or with abnormal myocardial motion. To obtain a continuous transition between the abnormal segment and the remainder of the LV myocardium, intermediate levels of motion were introduced. The phantom was simulated on a grid of $128 \times 128 \times 128$ voxels over 16 time frames (gates) with a voxel size of 0.317 cm. Note that, in order to isolate cardiac motion detection, the simulated MCAT phantom did not contain perfusion defects. Simulated phantoms with and without motion defect, as well as the difference image, are shown in Fig. 3.

2) Acquisition: The gated-SPECT acquisition was simulated using the SIMIND software package [39], including effects of nonuniform attenuation, camera response, and photon scatter. The projections were simulated at 64 $\times$64 bins with 64 camera angles over 360° and 16 gates. Acquisition noise was introduced at a level corresponding to 0.5M photon counts collected from the heart region. This simulation setup (count level, measurement dimensions and resolution) represents a typically observed study at University of Massachusetts Memorial Medical Center (UMMMC) using a low-energy high-resolution collimator.

This was repeated for 100 noise realizations with defect present and 100 with defect absent.

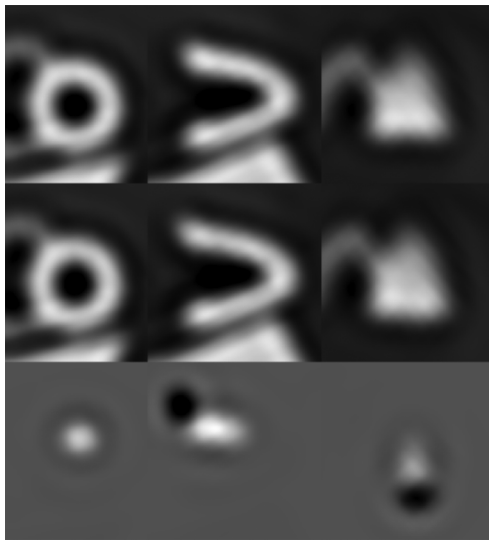

Fig. 3. Short-axis, vertical long-axis, and horizontal long-axis views (through the cardiac motion defect location) of the MCAT phantom at the end-systole for (top) data with no motion defect, (middle) data with a simulated motion defect in the anterior basal region, and (bottom) difference image.

3) Image Reconstruction: Image sequences were reconstructed using a conventional, frame-by-frame, filtered back-projection (FBP) method. Reconstructed images were further postprocessed by one of the following three filtering methods which were considered to illustrate the application of the proposed model observers.

1) *FBP-SO*. Each frame was separately filtered by a spatial Butterworth filter with order 5 and cutoff frequency 0.22 cycles per pixel [40], as routinely used at UMMMC in clinical patient evaluations.
2) *FBP-121*. In this method, also clinically used, each frame was first temporally filtered with the kernel ¼{1,2,1} followed by a spatial Butterworth filter identical to the one used for FBP-SO [30].
3) *MC-FBP*. In this approach, a motion-compensated spatio-temporal filtering method for cardiac SPECT image sequences, previously reported in [28], was applied. This method consists of two steps:
   - myocardial motion estimation;
   - motion-compensated spatio-temporal filtering of the FBP reconstructed image sequences.

In the first step of MC-FBP we used the motion-estimation procedure described in Section II-D for motion feature extraction. Specific details can be found in [28]. Once the deformation field $\boldsymbol{\delta}^{k,l}(\mathbf{r})$ was estimated, we applied the following motion-compensated spatio-temporal filter:

$$\tilde{f}_k(\mathbf{r}) = \sum_{l=1}^{K} h^{k,l} f_l(\mathbf{r} + \boldsymbol{\delta}^{k,l}(\mathbf{r})) \frac{|\mathcal{D}_b^l|}{|\mathcal{D}_b^k|}, \mathbf{r} \epsilon \mathcal{D}_b^k, b = 1, 2, \ldots, B \tag{18}$$

in which $\tilde{f}_k(\mathbf{r})$ is the kth frame (temporally filtered) $\boldsymbol{\delta}^{k,l}(\mathbf{r})$ represents the estimated displacement at location $\mathbf{r}$ between time frames k and l . Here, $h^{k,l}$ is a weighting window defined as

$$h^{k,l} = \mathrm{C}(1 - 2K^{-1}\min(|k-l|, K - |k-l|))^{\gamma} \tag{19}$$

where C is a constant, ensuring unit dc gain, and γ controls the strength of the temporal filter. Selection of the γ parameter is an important step in the evaluation of this reconstruction algorithm. The quantitative analysis reported in [28] showed image quality improvement compared to existing clinical methods, as measured using peak signal-to-noise ratio (PSNR) and region-of-interest bias-variance analysis. Using these figures of merit the "optimal" filter strength was found to be $\gamma = 0$. In this manuscript we evaluated $\gamma \in \{2, 3, 4\}$ denoted by MC-FBP 2, MC-FBP 3, and MC-FBP 4, respectively.

B. Human Observer Study

Five human observers (sometimes referred to as readers) performed an observer study using datasets described in the previous section. Human observers were familiar with cardiac SPECT images and familiar with the five types of image reconstruction studied in this paper but were not medical experts. For each postreconstruction method a number of image sequences were presented in a loop as a cine (see Fig. 1 for example of systolic and diastolic frames). The human observers were asked to score these sequences according to the level of confidence in the presence of an anterior (location known exactly) cardiac-motion defect (scores ranging from 1 to 6). For each reconstruction method, the study contained 180 images to score, preceded by training with 20 image sequence examples. A static random noise image was displayed between image sequences for 5 s. One study was performed per day to reduce the effects of fatigue, starting with the FBP-SO and FBP-121 datasets, followed by MC-FBP 2, 3, 4 studies. Images within each of the five datasets were presented in a random order for each observer.

The resulting performances for the five human observers were summarized using AUC with DBM-MRMC [41] analysis. The AUC values for each reconstruction method and corresponding standard deviations are reported in Fig. 4. Results suggest that motion-compensated filtering, at any of the temporal filtering level under consideration, produced images on which human observers can perform better than on FBP-SO or FBP-121 images. This observation is quantitatively supported by results shown in Table III where we report p-values of pair wise hypothesis test that reconstruction methods have equal AUC values. Here a smaller p-value indicates that it is less likely that the tested methods have equal AUC values. In addition, the presented p-values indicate that human performance was statistically indistinguishable between temporal smoothing strengths of $\gamma = 3$ and $\gamma = 4$. Table III also indicates that AUC values for $\gamma \in \{3, 4\}$ were only marginally better than for $\gamma = 2$.

In [28], using traditional nontask-based quantitative figures of merit such as peak-signal-to-noise ratio and bias-variance, the "optimal" filter strength of $\gamma = 0$ was found. By visual analysis, sequences for $\gamma = 0$ looked unnatural and somewhat erratic, which made it very difficult to assess motion. Therefore, in this paper we considered $\gamma \in \{2, 3, 4\}$ and the results presented here indicate a different optimal value around $\gamma \in \{3, 4\}$. This confirms that the traditional figures of merit, might point to a suboptimal filtering parameter value. Thus, image quality should be assessed by human observer or task-based model observer.

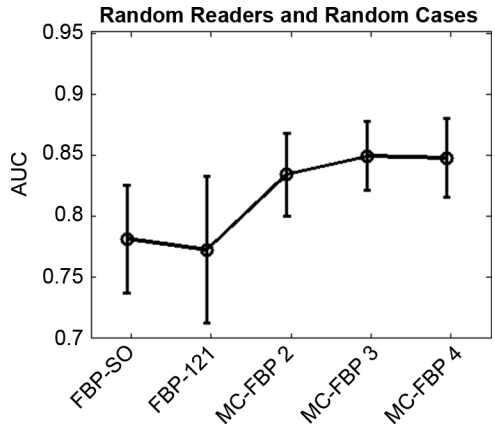

Fig. 4. Human observers' performance summarized by AUC for five different reconstruction methods and five observers. Error bars represent one standard deviation estimated using DBM MRMC software package [41] for random-reader random-case analysis.

TABLE III
HUMAN OBSERVER PERFORMANCE; P-VALUES OF PAIRWISE HYPOTHESIS TEST THAT RECONSTRUCTION METHODS HAVE EQUAL AUC VALUES

p-value					
	FBP-SO	FBP-121	MC-FBP 2	MC-FBP 3	MC-FBP 4
FBP-SO	-	0.74	0.06	0.02	0.02
FBP-121		-	0.03	0.01	0.01
MC-FBP 2			-	0.13	0.21
MC-FBP 3				-	0.73
MC-FBP 4					-

Note that the purpose of this work was not to evaluate the various postprocessing algorithms, but rather to use these algorithms as examples with which to evaluate the proposed model observers.

IV. EVALUATION OF THE MODEL OBSERVER

In this section, we present a comparison of the model observers in terms of generalization accuracy in predicting human-observer performance in motion defect detection. In the first experiment, two clinically used post-reconstruction filtering methods (i.e., FBP-SO and FBP-121) were considered. Next, in the second experiment we evaluated the model observer's utility for image quality assessment. Here, we used the proposed RVM model observer method, trained on images from two clinically used postreconstruction filtering methods, FBP-SO and FBP-121, to predict human observer performance using images derived by a postreconstruction motion-compensated temporal filtering method [28].

As we have pointed out previously [22], an important purpose of a model observer is to provide an estimate of a human-observer performance as a measure of image quality for reconstruction methods not yet evaluated by a human observer ROC study. Therefore, to be useful, a model observer must exhibit good *generalization* properties.

Bear in mind that the goal of our methods is to predict average human-observer performance, even though the proposed RVM-MO method achieves this by first predicting each individual human observer's performance as an intermediate step. Therefore, throughout we measured success by determining the extent to which the model observers could accurately predict average human-observer AUC.

TABLE IV
COMPARISON 1; AUC VALUES FOR FIVE OBSERVERS: RVM MODEL OBSERVER TRAINED ON FBP-121 DATASET AND TESTED ON FBP-SO AND TRAINED ON FBP-SO DATASET AND TESTED ON FBP-121. IN BOTH CASES, THE RVM-MO IS BETTER PREDICTING HUMAN OBSERVERS AUC VALUES THAN HOTELING-MO

	Tested on FBP-SO			Tested on FBP-121		
	Human observer	RVM-MO	Hotelling-MO	Human Observer	RVM-MO	Hotelling-MO
Observer #1	0.871	0.869	-	0.875	0.913	-
Observer #2	0.728	0.789	-	0.790	0.855	-
Observer #3	0.668	0.677	-	0.550	0.909	-
Observer #4	0.869	0.867	-	0.841	0.905	-
Observer #5	0.755	0.869	-	0.793	0.899	-
Average	**0.778**	**0.814**	**0.896**	**0.769**	**0.896**	**0.932**

1) Experiment 1: model observer generalization performance over FBP-SO and FBP-121 data sets.

We first evaluated the proposed RVM-MO's generalization properties by training it using the FBP-SO dataset and testing it on the FBP-121 dataset, and then reversing the roles of the two datasets, for each of the five human observers individually. The ability of the proposed RVM-MO to predict each human observer separately is valuable especially in cases when the observers perform very differently, as they did in our human observer study. The final accuracy of the model observers was measured in terms of accuracy in predicting the humans' averaged AUC. The proposed RVM-MO was also compared to the Hotelling-MO. Note that the Hotelling-MO does not depend directly on the human observers' scores and only aims to predict average human observer performance. In addition it does not require a training step during which one would adjust model parameters. In this model the expected values were replaced by sample estimates calculated for each dataset separately.

Table IV shows AUC values estimated using the DBM MRMC tool [41] for: five human observers, Hotelling-MO and RVM-MO for two possible combinations of training-testing data sets. Note that the RVM model observer had a reasonably close match in AUC values, and that the Hotelling-MO was not adjusting to each individual observer's performance. A more quantitative comparison of the AUC values is given in Table V, which shows p-values for testing the hypothesis that human-observer and model-observer AUC values are equal (p-values are also estimated using DBM-MRMC). One can observe that the Hotelling-MO had a low p-value, meaning that its AUC values were different from those of the human observers, while the RVM-MO had significantly higher p-values, indicating much better agreement with human observers.

2) Experiment 2: practical application

Here we consider the behavior of the proposed RVM-MO in a typical application. Specifically, after the RVM-MO was trained (tuned) using the FBP-SO and FBP-121 datasets, it was used to evaluate a recently proposed motion-compensated post-reconstruction strategy, MC-FBP. Here our goal was to find the optimal temporal filtering parameter (from a set of $\gamma \in \{2, 3, 4\}$) that maximized the model-observer performance in a cardiac-motion-defect detection task. Note that we also performed human observer studies so that we could evaluate how well the model observers predict human performance. The

TABLE V
COMPARISON 1; P-VALUES OF HYPOTHESIS TEST THAT HUMAN OBSERVER AND MODEL OBSERVERS AUC VALUES ARE EQUAL. IN BOTH CASES, THE RVM-MO HAS BETTER (LARGER) p-VALUES THAN THE HOTELLING-MO

p-value		
	FBP-SO	FBP-121
Human observers vs. Hotelling-MO	0.01	0.01
Human observers vs. RVM-MO	0.39	0.04

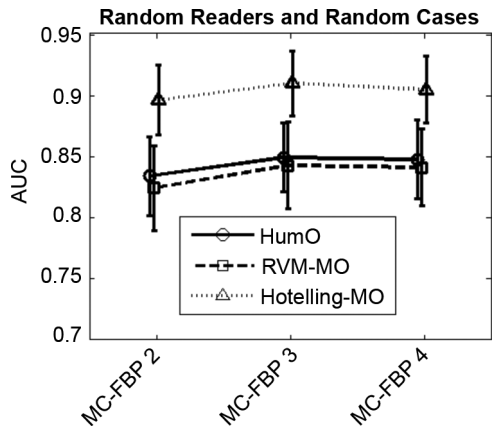

Fig. 5. *Comparison 2*; Area under the ROC curves (AUC) for MC-FBP datasets with different strength of motion-compensated filtering $\gamma \in \{2, 3, 4\}$ ("HumO" denotes "human observer"). Error-bars indicate the standard deviation estimated by the DBM MRMC tool [41].

results of this experiment are presented in Fig. 5, which shows the AUC of the human observers, together with predicted AUC of the model observers as a function of the filter strength, γ and an estimate of the AUC standard deviation, both calculated using the DBM MRMC software [41]. These results suggest that the RVM-MO captured human performance better than the Hotelling-MO. This observation can be evaluated more closely in Table VI where we show AUC values for five observers together with predicted AUC values.

It may at first appear paradoxical that the human observer sometimes achieves better detection performance than the Hotelling-MO (e.g., Observer #1 in MC-FBP 3 and MC-FBP 4; and Observer #4 in MC-FBP 2), which is based on the likelihood ratio. This may happen because the features used by the model observer do not capture all of the image information used by human observers, or that some of the Hotelling model assumptions are violated.

Next, we present a quantitative comparison in Table VII, which reports the p-values for a hypothesis test that human-observer and model-observer AUC values are equal. This table shows that the predictions made by the RVM-MO achieved very good agreement with human-observer performance, whereas those made by the Hotelling-MO did not.

In short, the results presented in Tables IV–VII suggest that the proposed RVM-MO performed well in capturing the performance of the human observers as measured by the AUC.

One may argue that a reasonably useful model might not reproduce absolute AUC values, as does RVM-MO, and it may be sufficient for the model to predict only the rankings. To the best of our knowledge there has not been a reported study exploring this issue, and no guidance has been proposed as to which metric is the most appropriate. However, if only the ranking is considered, both proposed methods (RVM-MO and Hotelling-MO) perform equally with rank correlation value of one.

V. DISCUSSION

This paper proposes two model observers for prediction of human performance in detection of cardiac motion defects. We argue that, since cardiac motion is an important diagnostic feature, task-based image quality assessment should include measures of cardiac motion visibility. For instance, temporal filtering methods for noise reduction in SPECT tend to reduce motion visibility and might have undesired effects on diagnostic performance, in particular for diagnosis based on motion assessment. However, measuring only perfusion-defect detectability at a given image frame, as has been done in the literature, might not reflect the degradation in motion visibility. Using an image-quality measure explicitly based on cardiac motion could assist in the development of motion-compensated reconstruction and processing methods for gated SPECT by offering a task-based evaluation of the performance. In particular, methods or parameters that excessively blur motion could be easily excluded from further consideration. Additionally, if it is possible to simulate both perfusion and motion defects with a given mathematical phantom (MCAT [37], NCAT [42]), then noise realizations could be generated once, and both image-quality measures (detectability of perfusion and motion defects) could be evaluated simultaneously. This would yield a metric closer to the actual diagnostic value of an image.

While the limited number of observers, reconstruction methods and images restricted the conclusions we could draw in this paper, our results strongly suggest that the proposed RVM model observer can correlate with human observer performance, even if trained on different types of reconstructed images. The results also suggest that the proposed method can adapt to different observers even if there is substantial variability in performance between the observers.

It is also worth emphasizing that both comparisons presented in Section IV use an RVM model observer trained and tested on fundamentally different images; RVM-MO was trained on FBP-SO and FBP-121 images, while testing was performed on motion-compensated filtered images, thus complicating the prediction problem. It would be interesting to see if the promising generalization properties presented in this paper extend to different types of images, such as images obtained via maximum-likelihood iterative reconstruction, and further between different types of iterative reconstruction with different types of prior information.

A broader set of data could provide a test bed to evaluate whether some important image details are not preserved in a certain set of reconstruction methods (e.g., because they have been blurred by temporal processing) while becoming visible and important for human diagnosis in another set (e.g., obtained using motion-compensated filtering). In this scenario, the model observer would likely benefit from training on a broad data set.

In addition, it will be important to explore various training set choices. For example, one could hypothesize that noniterative

TABLE VI
COMPARISON 2; AUC VALUE FOR FIVE OBSERVERS: RVM MODEL OBSERVER TRAINED ON FBP-SO AND FBP-121 DATASET AND TESTED ON FBP-MC 2, 3, 4

	MC-FBP 2			MC-FBP 3			MC-FBP 4		
	Human observer	RVM-MO	Hotelling-MO	Human observer	RVM-MO	Hotelling-MO	Human observer	RVM-MO	Hotelling-MO
Observer #1	0.858	0.842	-	0.912	0.874	-	0.908	0.867	-
Observer #2	0.819	0.821	-	0.822	0.863	-	0.853	0.842	-
Observer #3	0.739	0.754	-	0.789	0.747	-	0.743	0.773	-
Observer #4	0.926	0.851	-	0.895	0.865	-	0.880	0.861	-
Observer #5	0.828	0.854	-	0.829	0.865	-	0.855	0.862	-
Average	**0.834**	**0.824**	**0.896**	**0.849**	**0.842**	**0.910**	**0.847**	**0.841**	**0.905**

TABLE VII
COMPARISON 2; P-VALUES OF HYPOTHESIS TEST THAT HUMAN OBSERVER AND MODEL OBSERVERS AUC VALUES ARE EQUAL THE RVM-MO HAS SIGNIFICANTLY BETTER AGREEMENT WITH HUMAN OBSERVER THAN HOTELLING-MO

	p-value		
	MC-FBP 2	MC-FBP 3	MC-FBP 4
Human observer vs. Hotelling-MO	0.05	0.09	0.16
Human observer vs. RVM-MO	0.85	0.73	0.58

and iterative methods are different enough so that a complete training set should include images of both types. In a machine learning formulation, the performance of a classifier/regression is not guaranteed to be preserved when the testing samples lie far from the convex hull of the training samples. Further studies on a larger variety of reconstruction methods are needed to define a sufficient training set for cardiac-gated SPECT.

An advantage of using RVM over other regression models is that it reduces the complexity of the model (compared to the support vector machine, for example). Thus, even if a training set would include a large number of reconstruction methods, the computational cost associated with the training and testing procedures would remain reasonable. In particular, since there is no kernel parameter to optimize, there is no need for cross-validation during the training phase (as needed in [22]).

VI. CONCLUSION

In this paper, we present two numerical surrogates for human observers in task-based image quality assessment where the task is cardiac-motion-defect detection. Both proposed methods rely on motion features extracted using our previously reported deformable mesh model for myocardium motion estimation. In the first approach we used a Hotelling linear observer in a manner similar to that used traditionally for perfusion-defect detection. In the second proposed observer, the human observer modeling was approached as a Bayesian nonlinear regression problem using RVMs. In this regression we aimed to predict the human observer's level of confidence in the presence of a cardiac motion defect based on extracted image and motion features. Our results suggest that the RVM model observer provides good prediction of human-observer performance, and do so more accurately than a Hotelling linear observer using the same motion features, over a wide range of reconstruction methods. The proposed RVM model observer provides a new tool for image quality assessment of cardiac SPECT images. It focuses on visibility of cardiac-motion defects rather than that of perfusion defects.

REFERENCES

[1] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Hoboken, NJ: Wiley-Interscience, 2004, pp. 913–1001.

[2] ICRU Rep.54 J Int. Comm. Radiat. Units Meas., 1996.

[3] M. N. Wernick and J. N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT*. Boston, MA: Academic, 2004.

[4] C. L. Hansen, R. A. Goldstein, D. S. Berman, K. B. Churchwell, C. D. Cooke, J. R. Corbett, S. J. Cullom, S. T. Dahlberg, J. R. Galt, R. K. Garg, G. V. Heller, M. C. Hyun, L. L. Johnson, A. Mann, B. D. McCallister, R. Taillefer, R. P. Ward, and J. J. Mahmarian, Quality Assurance Comm. Am. Soc. Nucl. Cardiol., "Myocardial perfusion and function single photon emission computed tomography," *J. Nucl. Cardiol.*, vol. 133, pp. 97–120, 2006.

[5] A. I. Veress, J. A. Weiss, R. H. Huesman, B. W. Reutter, S. E. Taylor, A. Sitek, B. Feng, Y. Yang, and G. T. Gullberg, "Measuring regional changes in the diastolic deformation of the left ventricle of SHR rats using microPET technology and hyperelastic warping," *Ann. Biomed. Eng.*, vol. 36, pp. 1104–1117, 2008.

[6] G. S. Lin, H. H. Hines, G. Grant, K. Taylor, and C. Ryals, "Automated quantification of myocardial ischemia and wall motion defects by use of cardiac SPECT polar mapping and 4-dimensional surface rendering," *J. Nucl. Med. Technol.*, vol. 34, pp. 3–17, 2006.

[7] G. Germano, J. Erel, H. Lewin, P. B. Kavanagh, and D. S. Berman, "Automatic quantitation of regional myocardial wall motion and thickening from gated technetium-99m sestamibi myocardial perfusion single-photon emission computed tomography," *J. Am. Coll. Cardiol.*, vol. 30, pp. 1360–1367, 1997.

[8] B. Feng, A. I. Veress, A. Sitek, G. T. Gullberg, and D. G. Roy, "Estimation of mechanical properties from gated SPECT and cine MRI data using a finite-element mechanical model of the left ventricle," *IEEE Trans. Nucl. Sci.*, vol. 48, no. 3, pp. 725–733, Jun. 2001.

[9] J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE*, vol. 1768, pp. 161–168, 1992.

[10] K. J. Myers and H. H. Barrett, "Addition of a channel mecha to the ideal-observer model," *J. Opt. Soc. Am. A*, vol. 4, pp. 2447–2457, 1987.

[11] C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: Effects of regularization and object variability," *J. Opt. Soc. Am. A*, vol. 18, pp. 473–488, 2001.

[12] H. C. Gifford, M. A. King, D. J. d. Vries, and E. J. Soares, "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT MAP imaging," *J. Nucl. Med.*, vol. 41, pp. 541–521, 2000.

[13] J. Oldan, S. Kulkarni, Y. Xing, P. Khurd, and G. Gindi, "Channelized Hotelling and human observer study of optimal smoothing in SPECT MAP reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 3, pp. 733–741, Jun. 2004.

[14] S. D. Wollenweber, B. M. W. Tsui, D. S. Lalush, E. C. Frey, K. J. LaCroix, and G. T. Gullberg, "Comparison of Hotelling observer models and human observers in defect detection from myocardial SPECT imaging," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6, pp. 2098–2103, Dec. 1999.

[15] S. Sankaran, E. C. Frey, K. L. Gilland, and B. M. W. Tsui, "Optimum compensation method and filter cutoff frequency in myocardial SPECT: A human observer study," *J. Nucl. Med.*, vol. 43, no. 3, pp. 432–438, 2002.

[16] M. V. Narayanan, H. C. Gifford, M. A. King, P. H. Pretorius, T. H. Farncombe, P. P. Bruyant, and M. N. Wernick, "Optimization of iterative reconstructions of Tc99m cardiac SPECT study using numerical observers," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 5, pp. 2355–2360, Dec. 2002.

[17] E. C. Frey, K. L. Gilland, and B. M. W. Tsui, "Application of task-based measures of image quality to optimization and evaluation of three-dimensional reconstruction-based compensation methods in myocardial perfusion SPECT," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 1040–1050, Sep. 2002.

[18] C. Liu and B. M. W. Tsui, "Myocardial perfusion SPECT using a rotating multi-segment slant-hole collimator," *Med. Phys*, vol. 37, pp. 1610–1618, 2010.

[19] H. C. Gifford, R. G. Wells, and M. A. King, "A comparison of human observer LROC and numerical observer ROC for tumor detection in SPECT images," *IEEE Trans. Nucl. Sci*, vol. 46, no. 4, pp. 1032–1037, Aug. 1999.

[20] B. Liu, L. Zhou, and G. Gindi, "The efficiency of the human observer for lesion detection and localization in emission tomography," *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2651–2666, 2009.

[21] A. Lehovich, P. P. Bruyant, H. S. Gifford, P. B. Schneider, S. Squires, R. Licho, G. Gindi, and M. A. King, "Impact on reader performance for lesion-detection/localization tasks of anatomical priors in SPECT reconstruction," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1459–1467, Sep. 2009.

[22] J. G. Brankov, Y. Yang, L. Wei, I. E. Naqa, and M. N. Wernick, "Learning a channelized observer for image quality assessment," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 991–999, Jul. 2009.

[23] L. Platiša, B. Goossens, E. Vansteenkiste, S. Park, B. D. Gallas, A. Badano, and W. Philips, "Channelized Hotelling observers for the assessment of volumetric imaging data sets," *J. Opt. Soc. Am. A*, vol. 28, pp. 1145–1163, 2011.

[24] L. Caucci, H. H. Barrett, and J. J. Rodriguez, "Spatio-temporal Hotelling observer for signal detection from image sequences," *Opt. Exp.*, vol. 17, pp. 10946–10958, 2009.

[25] P. H. Pretorius, H. C. Gifford, M. V. Narayanan, S. T. Dahlberg, and M. A. King, "Comparison of detection accuracy of perfusion defects in SPECT for different reconstruction strategies using polar-map quantitation," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 5, pp. 1569–1574, Oct. 2003.

[26] J. Tang, T.-S. Lee, X. He, W. P. Segars, and B. M. W. Tsui, "Comparison of 3D OS-EM and 4D MAP-RBI-EM reconstruction algorithms for cardiac motion abnormality classification using a motion observer," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 5, pp. 2571–2577, Oct. 2010.

[27] F. Massanes and J. G. Brankov, "Motion perception in medical imaging," *Proc. SPIE Med. Imag.*, 2011.

[28] T. Marin and J. G. Brankov, "Deformable left-ventricle mesh model for motion-compensated filtering in cardiac gated SPECT," *Med. Phys.*, vol. 37, pp. 5471–5481, 2010.

[29] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[30] A.-D. Achtert, M. A. King, S. T. Dahlberg, P. H. Pretorius, K. J. LaCroix, and B. M. W. Tsui, "An investigation of the estimation of ejection fractions and cardiac volumes by a quantitative gated SPECT software package in simulated gated SPECT images," *J. Nucl. Cardiol.*, vol. 5, pp. 144–152, 1998.

[31] T. Marin, P. H. Pretorious, Y. Yang, M. N. Wernick, and J. G. Brankov, "Numerical observer for cardiac motion assessment using a linear discriminant," in *Proc. IEEE Nucl. Sci. Symp.*, Knoxville, TN, 2010, pp. 2352–2355.

[32] J. G. Brankov, T. Marin, P. H. Pretorius, Y. Yang, and M. N. Wernick, "Numerical observer for cardiac motion assessment," in *Proc. SPIE Comput. Imag. VIII*. San Jose, CA: , 2010, pp. 27–33.

[33] T. Marin, M. M. Kalayeh, P. H. Pretorius, M. N. Wernick, Y. Yang, and J. G. Brankov, "Numerical observer for cardiac motion assessment using machine learning," *Proc. SPIE 7966*, 2011.

[34] J. Weston, A. Elisseeff, G. Bakir, and F. Sinz, The Spider machine learning toolbox [Online]. Available: http://www.kyb.tuebingen.mpg.de/bs/people/spider/2005

[35] B. Feng, "Modeling of the left ventricle (LV) by using mechanical models and image data," Ph.D. dissertation, Univ. Utah, Salt Lake City, UT, 2002.

[36] J. R. Galt, E. V. Garcia, and W. L. Robbins, "Effects of myocardial wall thickness on SPECT quantification," *IEEE Trans. Med. Imag.*, vol. 9, no. 2, pp. 144–150, Jun. 1990.

[37] P. H. Pretorius, M. A. King, B. M. W. Tsui, K. J. LaCroix, and W. Xia, "A mathematical model of motion of the heart for use in generating source and attenuation maps for simulating emission imaging," *Med. Phys.*, vol. 26, pp. 2323–2332, 1999.

[38] P. H. Pretorius, M. A. King, and H. C. Gifford, "A five-dimensional mathematical model for regional and global changes in cardiac uptake and motion," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2634–2640, Oct. 2004.

[39] M. Ljungberg and S.-E. Strand, "A Monte Carlo program for the simulation of scintillation camera characteristics," *Comp. Methods Programs Biomed.*, vol. 29, pp. 257–272, 1989.

[40] P. H. Pretorius, M. A. King, H. C. Gifford, S. T. Dahlberg, F. Spencer, E. Simon, J. Rashkin, N. Botkin, W. Berndt, M.V. Narayanan, and J. A. Leppo, "Myocardial perfusion SPECT reconstruction: Receiver operating characteristic comparison of CAD detection accuracy of filtered backprojection reconstruction with all of the clinical imaging information available to readers and solely stress slices iteratively reconstructed with combined compensation," *J. Nucl. Cardiol.*, vol. 12, pp. 284–293, 2005.

[41] D. D. Dorfman, K. S. Berbaum, and C. E. Metz, Investigative Radiology "Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method," 1992, vol. 27, pp. 723–731.

[42] W. P. Segars, "Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom," Ph.D. dissertation, Dept. Biomed. Eng., Univ North Carolina, Chapel Hill, 2001.

[43] M. M. Kalayeh, T. Marin, P. H. Pretorius, M. N. Wernick, Y. Yang, and J. G. Brankov, "Channelized relevance vector machine as a numerical observer for cardiac perfusion defect detection task," *Proc. SPIE: Med. Imag.*, p. 79660E, 2011.

[44] F. M. Parages, M. N. Wernick, T. S. Denney, and J. G. Brankov, "Deformable mesh model of cardiac motion from tagged MRI data," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 213–216.