

# Machine Learning in Medical Imaging

Guest Editors: Kenji Suzuki, Pingkun Yan, Fei Wang, and Dinggang Shen





---

# **Machine Learning in Medical Imaging**

International Journal of Biomedical Imaging

---

## **Machine Learning in Medical Imaging**

Guest Editors: Kenji Suzuki, Pingkun Yan, Fei Wang,  
and Dinggang Shen



---

Copyright © 2012 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "International Journal of Biomedical Imaging." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Haim Azhari, Israel

K. Ty Bae, USA

Richard H. Bayford, UK

F. J. Beekman, The Netherlands

J. C. Chen, Taiwan

Anne Clough, USA

Carl Crawford, USA

Daniel Day, Australia

Eric Hoffman, USA

Jiang Hsieh, USA

M. Jiang, China

Marc Kachelrieß, Germany

Cornelia Laule, Canada

Seung W. Lee, Republic of Korea

A. K. Louis, Germany

Jayanta Mukherjee, India

Vasilis Ntziachristos, Germany

Scott Pohlman, USA

Erik L. Ritman, USA

Jay Rubinstein, USA

Peter Santago, USA

Lizhi Sun, USA

Kenji Suzuki, USA

Jie Tian, China

Michael W. Vannier, USA

Yue Wang, USA

Ge Wang, USA

Guo Wei Wei, USA

D. L. Wilson, USA

Sun K. Yoo, Republic of Korea

Habib Zaidi, Switzerland

Yantian Zhang, USA

Jun Zhao, China

Yibin Zheng, USA

Tiange Zhuang, China

Yu Zou, USA

# Contents

**Machine Learning in Medical Imaging**, Kenji Suzuki, Pingkun Yan, Fei Wang, and Dinggang Shen  
Volume 2012, Article ID 123727, 2 pages

**Pixel-Based Machine Learning in Medical Imaging**, Kenji Suzuki  
Volume 2012, Article ID 792079, 18 pages

**A Bayesian Hyperparameter Inference for Radon-Transformed Image Reconstruction**, Hayaru Shouno, Madomi Yamasaki, and Masato Okada  
Volume 2011, Article ID 870252, 10 pages

**Conditional Random Fields and Supervised Learning in Automated Skin Lesion Diagnosis**, Paul Wighton, Tim K. Lee, Greg Mori, Harvey Lui, David I. McLean, and M. Stella Atkins  
Volume 2011, Article ID 846312, 10 pages

**A Statistical Texture Model of the Liver Based on Generalized N-Dimensional Principal Component Analysis (GND-PCA) and 3D Shape Normalization**, Xu Qiao and Yen-Wei Chen  
Volume 2011, Article ID 601672, 8 pages

**Sparse Representation of Deformable 3D Organs with Spherical Harmonics and Structured Dictionary**, Dan Wang, Ahmed H. Tewfik, Yingchun Zhang, and Yunhe Shen  
Volume 2011, Article ID 658930, 17 pages

**Multiresolution Analysis Using Wavelet, Ridgelet, and Curvelet Transforms for Medical Image Segmentation**, Shadi AlZubi, Naveed Islam, and Maysam Abbod  
Volume 2011, Article ID 136034, 18 pages

**Biomedical Imaging Modality Classification Using Combined Visual Features and Textual Terms**, Xian-Hua Han and Yen-Wei Chen  
Volume 2011, Article ID 241396, 7 pages

**Automatic Segmentation of Dermoscopic Images by Iterative Classification**, Maciel Zortea, Stein Olav Skrøvseth, Thomas R. Schopf, Herbert M. Kirchesch, and Fred Godtlielsen  
Volume 2011, Article ID 972648, 19 pages

**Multiclass Sparse Bayesian Regression for fMRI-Based Prediction**, Vincent Michel, Evelyn Eger, Christine Keribin, and Bertrand Thirion  
Volume 2011, Article ID 350838, 13 pages

**Development of a New Tool for 3D Modeling for Regenerative Medicine**, Filippo Mattoli, Roberto Tiribuzi, Francesco D'Angelo, Ilaria di Girolamo, Mattia Quattrocchi, Simona Montesano, Lucia Crispoltoni, Vasileios Oikonomou, Maria Gabriella Cusella De Angelis, Peggy Marconi, Antonio Orlacchio, Maurilio Sampaolesi, Sabata Martino, and Aldo Orlacchio  
Volume 2011, Article ID 236854, 13 pages

## Editorial

# Machine Learning in Medical Imaging

**Kenji Suzuki,<sup>1</sup> Pingkun Yan,<sup>2</sup> Fei Wang,<sup>3</sup> and Dinggang Shen<sup>4</sup>**

<sup>1</sup> Department of Radiology, University of Chicago, Chicago, IL 60637, USA

<sup>2</sup> Xian Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Shaanxi, Xi'an 710119, China

<sup>3</sup> IBM Almaden Research Center, San Jose, CA 95120, USA

<sup>4</sup> Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Correspondence should be addressed to Kenji Suzuki, [suzuki@uchicago.edu](mailto:suzuki@uchicago.edu)

Received 22 December 2011; Accepted 22 December 2011

Copyright © 2012 Kenji Suzuki et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medical imaging is becoming indispensable for patients' healthcare. Machine learning plays an essential role in the medical imaging field, including computer-aided diagnosis, image segmentation, image registration, image fusion, image-guided therapy, image annotation, and image database retrieval. With advances in medical imaging, new imaging modalities and methodologies such as cone-beam/multislice CT, 3D ultrasound imaging, tomosynthesis, diffusion-weighted magnetic resonance imaging (MRI), positron-emission tomography (PET)/CT, electrical impedance tomography, and diffuse optical tomography, new machine-learning algorithms/applications are demanded in the medical imaging field. Because of large variations and complexity it is generally difficult to derive analytic solutions or simple equations to represent objects such as lesions and anatomy in medical images. Therefore, tasks in medical imaging require "learning from examples" for accurate representation of data and prior knowledge. Because of its essential needs, machine learning in medical imaging is one of the most promising, growing fields.

The main aim of this special issue is to help advance the scientific research within the broad field of machine learning in medical imaging. The special issue was planned in conjunction with the International Workshop on Machine Learning in Medical Imaging (MLMI 2010) [1], which was the first workshop on this topic, held at the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2010) in September, 2010, in Beijing, China. This special issue is one in a series of special issues of journals on this topic [2]; it focuses on major trends and challenges in this area, and it presents

work aimed at identifying new cutting-edge techniques and their use in medical imaging.

The quality level of the submissions for this special issue was very high. A total of 17 papers were submitted to this issue in response to the call for papers. Based on a rigorous review process, 10 papers (59%) were accepted for publication in the special issue. The special issue starts by a review of studies on a class of machine-learning techniques, called pixel/voxel-based machine learning, in medical imaging by K. Suzuki. A series of medical imaging applications of machine-learning techniques are presented. A large variety of applications are well represented here, including organ modeling by D. Wang et al. and X. Qiao and Y.-W. Chen, brain function estimation by V. Michel et al., image reconstruction by H. Shouno et al., lesion classification by P. Wighton et al., modality classification by X.-H. Han and Y.-W. Chen, lesion segmentation by M. Zortea et al., organ segmentation by S. Alzubi et al., and visualization of molecular signals by F. Mattoli et al. Also, the issue covers various biomedical imaging modalities, including MRI by D. Wang et al., CT by X. Qiao and Y.-W. Chen and H. Shouno et al., functional MRI by V. Michel et al., dermoscopy by P. Wighton et al. and M. Zortea et al., scintigraphy by X.-H. Han and Y.-W. Chen, ultrasound imaging by X.-H. Han and Y.-W. Chen, radiography by X.-H. Han and Y.-W. Chen, MR angiography by S. Alzubi et al., and microscopy by F. Mattoli et al. as well as a variety of organs, including the kidneys by D. Wang et al., liver by X. Qiao and Y.-W. Chen, brain by V. Michel et al. and F. Mattoli et al., chest by S. Alzubi et al., skin by P. Wighton et al. and M. Zortea et al., and heart by F. Mattoli et al. Various machine-learning techniques were

developed/used to solve the respective problems, including structured dictionary learning by D. Wang et al., generalized N-dimensional principal component analysis by X. Qiao et al., multiclass sparse Bayesian regression by V. Michel et al., Bayesian hyperparameter inference by H. Shouno et al., supervised learning of probabilistic models based on maximum a posteriori estimation and conditional random fields by P. Wighton et al., joint kernel equal contribution in support vector classification by X.-H. Han and Y.-W. Chen, and iterative hybrid classification by M. Zortea et al.

We are grateful to all authors for their excellent contributions to this special issue and to all reviewers for their reviews and constructive suggestions. We hope that this special issue will inspire further ideas for creative research, advance the field of machine learning in medical imaging, and facilitate the translation of the research from bench to bedside.

*Kenji Suzuki*  
*Pingkun Yan*  
*Fei Wang*  
*Dinggang Shen*

## References

- [1] F. Wang, P. Yan, K. Suzuki, and D. Shen, *Machine Learning in Medical Imaging*, vol. 6357 of *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2010.
- [2] K. Suzuki, "Special Issue on Machine Learning for Medical Imaging," *Algorithms*, vol. 2-3, 2009-2010.

## Review Article

# Pixel-Based Machine Learning in Medical Imaging

**Kenji Suzuki**

*Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL 60637, USA*

Correspondence should be addressed to Kenji Suzuki, [suzuki@uchicago.edu](mailto:suzuki@uchicago.edu)

Received 17 October 2011; Accepted 14 November 2011

Academic Editor: Dinggang Shen

Copyright © 2012 Kenji Suzuki. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine learning (ML) plays an important role in the medical imaging field, including medical image analysis and computer-aided diagnosis, because objects such as lesions and organs may not be represented accurately by a simple equation; thus, medical pattern recognition essentially require “learning from examples.” One of the most popular uses of ML is classification of objects such as lesions into certain classes (e.g., abnormal or normal, or lesions or nonlesions) based on input features (e.g., contrast and circularity) obtained from segmented object candidates. Recently, pixel/voxel-based ML (PML) emerged in medical image processing/analysis, which use pixel/voxel values in images directly instead of features calculated from segmented objects as input information; thus, feature calculation or segmentation is not required. Because the PML can avoid errors caused by inaccurate feature calculation and segmentation which often occur for subtle or complex objects, the performance of the PML can potentially be higher for such objects than that of common classifiers (i.e., feature-based MLs). In this paper, PMLs are surveyed to make clear (a) classes of PMLs, (b) similarities and differences within (among) different PMLs and those between PMLs and feature-based MLs, (c) advantages and limitations of PMLs, and (d) their applications in medical imaging.

## 1. Introduction

Machine learning (ML) plays an essential role in the medical imaging field, including medical image analysis and computer-aided diagnosis (CAD) [1, 2], because objects such as lesions and organs in medical images may be too complex to be represented accurately by a simple equation; modeling of such complex objects often requires a number of parameters which have to be determined by data. For example, a lung nodule is generally modeled as a solid sphere, but there are nodules of various shapes and nodules with internal inhomogeneities, such as spiculated nodules and ground-glass nodules [3]. A polyp in the colon is modeled as a bulbous object, but there are also polyps which exhibit a flat shape [4, 5]. Thus, diagnostic tasks in medical images essentially require “learning from examples (or data)” to determine a number of parameters in a complex model.

One of the most popular uses of ML in medical image analysis is the classification of objects such as lesions into certain classes (e.g., abnormal or normal, lesions or non-lesions, and malignant or benign) based on input features (e.g., contrast, area, and circularity) obtained from segmented

object candidates (This class of ML is referred to feature-based ML.). The task of ML here is to determine “optimal” boundaries for separating classes in the multidimensional feature space which is formed by the input features [6]. ML algorithms for classification include linear discriminant analysis [7], quadratic discriminant analysis [7], multilayer perceptron [8, 9], and support vector machines [10, 11]. Such ML algorithms were applied to lung nodule detection in chest radiography [12–15] and thoracic CT [16–19], classification of lung nodules into benign or malignant in chest radiography [20] and thoracic CT [21, 22], detection of microcalcifications in mammography [23–26], detection of masses in mammography [27], classification of masses into benign or malignant in mammography [28–30], polyp detection in CT colonography [31–33], determining subjective similarity measure of mammographic images [34–36], and detection of aneurysms in brain MRI [37].

Recently, as available computational power increased dramatically, pixel/voxel-based ML (PML) emerged in medical image processing/analysis which uses pixel/voxel values in images directly instead of features calculated from segmented regions as input information; thus, feature calculation or

TABLE 1: Classes of PMLs, their functions, and their applications.

PMLs	Functions	Applications
Neural filters (including neural edge enhancers)	Image processing	Edge-preserving noise reduction [38, 39]. Edge enhancement from noisy images [40]. Enhancement of subjective edges traced by a physician [41].
Convolution neural networks (including shift-invariant neural networks)	Classification	FP reduction in CAD for lung nodule detection in CXR [42–44]. FP reduction in CAD for detection of microcalcifications [45] and masses [46] in mammography. Face recognition [47]. Character recognition [48].
Massive-training artificial neural networks (MTANNs, including a mixture of expert MTANNs, a LAP-MTANN, an MTSVR)	Classification (image processing + scoring), pattern enhancement and suppression, object detection (pattern enhancement followed by thresholding or segmentation)	FP reduction in CAD for detection of lung nodules in CXR [57] and CT [17, 52, 63]. Distinction between benign and malignant lung nodules in CT [58]. FP reduction in CAD for polyp detection in CT colonography [53, 59–62]. Bone separation from soft tissue in CXR [54, 55]. Enhancement of lung nodules in CT [56].
Others	Image processing or classification	Segmenting posterior ribs in CXR [64]. Separation of ribs from soft tissue in CXR [65].

segmentation is not required. Because the PML can avoid errors caused by inaccurate feature calculation and segmentation which often occur for subtle or complex objects, the performance of the PML can potentially be higher for such objects than that of common classifiers (i.e., feature-based MLs). In this paper, PMLs are surveyed and reviewed to make clear (a) classes of PMLs, (b) the similarities and differences within different PMLs and those between PMLs and feature-based MLs, (c) the advantages and limitations of PMLs, and (d) their applications in medical imaging.

## 2. Pixel/Voxel-Based Machine Learning (PML)

**2.1. Overview.** PMLs have been developed for tasks in medical image processing/analysis and computer vision. Table 1 summarizes classes of PMLs, their functions, and their applications. There are three classes of PMLs: neural filters [38, 39] (including neural edge enhancers [40, 41]), convolution neural networks (NNs) [42–48] (including shift-invariant NNs [49–51]), and massive-training artificial neural networks (MTANNs) [52–56] (including multiple MTANNs [17, 38, 39, 52, 57, 58], a mixture of expert MTANNs [59, 60], a multiresolution MTANN [54], a Laplacian eigenfunction MTANN (LAP-MTANN) [61], and a massive-training support vector regression (MTSVR) [62]). The class of neural filters has been used for image-processing tasks such as edge-preserving noise reduction in radiographs and other digital pictures [38, 39], edge enhancement from noisy images [40], and enhancement of subjective edges traced by a physician in left ventriculograms [41]. The class of convolution NNs has been applied to classification tasks such as false-positive (FP) reduction in CAD schemes for detection of lung nodules in chest radiographs (also known as chest X-rays; CXRs) [42–44], FP reduction in CAD schemes for detection of microcalcifications [45] and masses [46] in mammography, face recognition [47], and character

recognition [48]. The class of MTANNs has been used for classification, such as FP reduction in CAD schemes for detection of lung nodules in CXR [57] and CT [17, 52, 63], distinction between benign and malignant lung nodules in CT [58], and FP reduction in a CAD scheme for polyp detection in CT colonography [53, 59–62]. The MTANNs have also been applied to pattern enhancement and suppression such as separation of bone from soft tissue in CXR [54, 55] and enhancement of lung nodules in CT [56]. There are other PML approaches in the literature. An iterative, pixel-based, supervised, statistical classification method called iterated contextual pixel classification has been proposed for segmenting posterior ribs in CXR [64]. A pixel-based, supervised regression filtering technique called filter learning has been proposed for separation of ribs from soft tissue in CXR [65].

**2.2. Neural Filters.** In the field of signal/image processing, supervised nonlinear filters based on a multilayer ANN, called neural filters, have been studied [38, 39]. The neural filter employs a linear-output ANN model as a convolution kernel of a filter. The inputs to the neural filter are an object pixel value and spatially/spatiotemporally adjacent pixel values in a subregion (or local window). The output of the neural filter is a single pixel value. The neural filter is trained with input images and corresponding “teaching” (desired or ideal) images. The training is performed by a linear-output backpropagation algorithm [40] which is a back-propagation algorithm modified for the linear-output ANN architecture. The input, output, and teacher (desired output) for neural filters are summarized in Table 2. Neural filters can acquire the functions of various linear and nonlinear filtering through training. Neural filters have been applied to reduction of the quantum noise in X-ray fluoroscopic and radiographic images [38, 39]. It was reported that the performance of the neural filter was superior to that of well-known nonlinear filters such as an adaptive weighted

TABLE 2: Classification of ML algorithms by their input, output, and teacher (desired output).

ML algorithms	Input	Output	Teacher
Neural filters	Pixel values in a subregion (local window) in a given image	Single pixel value (image is formed by collecting pixels)	Desired pixel value
MTANNs	Pixel values in a subregion (local window) in a given image	Single pixel value (image is formed by collecting pixels; likelihood score for the given image is obtained by use of the scoring method)	Likelihood of being a specific pattern at each pixel
Convolution NNs	Pixel values in a given image	Class to which the given image belongs	Nominal class label for the given image
Shift-invariant NNs	Pixel values in a given image	Class to which each pixel belongs	Nominal class label for each pixel
Multilayer perceptron for character recognition	Pixel values in a given binary image (character)	Class to which the given image belongs	Nominal class label for the given image
Classifiers (e.g., linear discriminant analysis, NNs, support vector machines)	Features extracted from a segmented object in a given image	Class to which the segmented object belongs	Nominal class label for the segmented object

averaging filter [66]. A study [38] showed that adding features from the subregion to the input information improved the performance of the neural filter. Neural filters have been extended to accommodate the task of enhancement of edges, and a supervised edge enhancer (detector), called a neural edge enhancer, was developed [40]. The neural edge enhancer can acquire the function of a desired edge enhancer through training. It was reported that the performance of the neural edge enhancer in the detection of edges from noisy images was far superior to that of well-known edge detectors such as the Canny edge detector [67], the Marr-Hildreth edge detector [68], and the Huckel edge detector [69]. In its application to the contour extraction of the left ventricular cavity in digital angiography, it has been reported that the neural edge enhancer can accurately replicate the subjective edges traced by a cardiologist [41].

### 2.3. Massive-Training Artificial Neural Network (MTANN).

An MTANN was developed by extension of neural filters to accommodate various pattern-recognition tasks [52]. A two-dimensional (2D) MTANN was first developed for distinguishing a specific opacity (pattern) from other opacities (patterns) in 2D images [52]. The 2D MTANN was applied to reduction of FPs in computerized detection of lung nodules on 2D CT slices in a slice-by-slice way [17, 52, 63] and in CXR [57], the separation of ribs from soft tissue in CXR [54, 55, 70], and the distinction between benign and malignant lung nodules on 2D CT slices [58]. For processing of three-dimensional (3D) volume data, a 3D MTANN was developed by extending the structure of the 2D MTANN, and it was applied to 3D CT colonography data [53, 59–62].

The generalized architecture of an MTANN which unifies 2D and 3D MTANNs is shown in Figure 1. The input, output, and teacher for MTANNs are shown in Table 2. An MTANN consists of an ML model such as a linear-output ANN regression model and a support vector regression model, which is capable of operating on pixel/voxel data directly [40]. The linear-output ANN regression model employs a linear function instead of a sigmoid function as

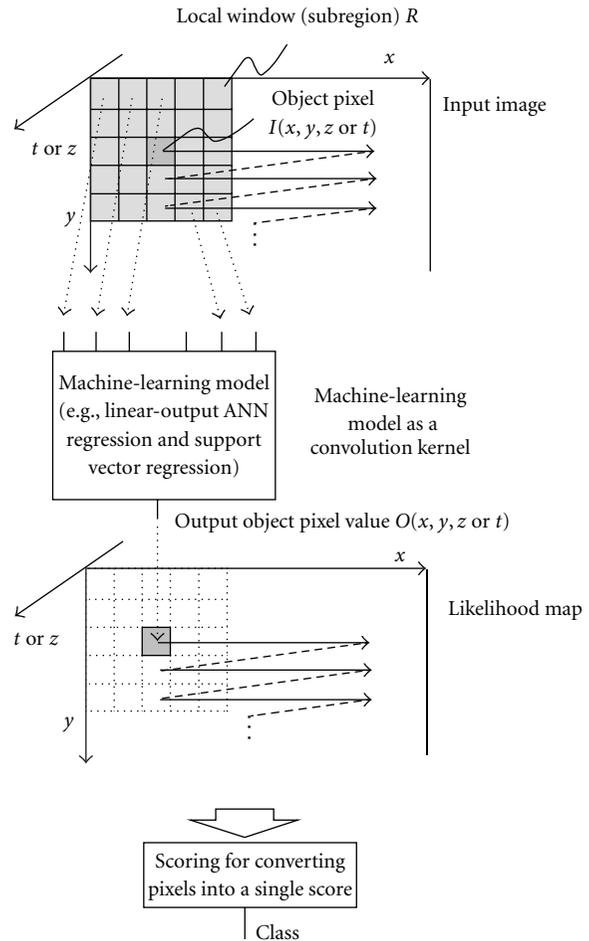


FIGURE 1: Generalized architecture of an MTANN (a class of PML) consisting of an ML model (e.g., linear-output ANN regression and support vector regression) with subregion input and single-pixel output. All pixel values in a subregion extracted from an input image are entered as input to the ML model. The ML model outputs a single pixel value for each subregion, the location of which corresponds to the center pixel in the subregion. Output pixel value is mapped back to the corresponding pixel in the output image.

the activation function of the unit in the output layer because the characteristics of an ANN were improved significantly with a linear function when applied to the continuous mapping of values in image processing [40]. Note that the activation functions of the units in the hidden layer are a sigmoid function for nonlinear processing, and those of the unit in the input layer are an identity function, as usual. The pixel/voxel values of the input images/volumes may be normalized from 0 to 1. The input to the MTANN consists of pixel/voxel values in a subregion/subvolume,  $R$ , extracted from an input image/volume. The output of the MTANN is a continuous scalar value, which is associated with the center voxel in the subregion, and is represented by

$$\begin{aligned} O(x, y, z \text{ or } t) \\ = \text{ML}\{I(x - i, y - j, z - k \text{ or } t - k) \mid (i, j, k) \in R\}, \end{aligned} \quad (1)$$

where  $x$ ,  $y$ , and  $z$  or  $t$  are the coordinate indices,  $\text{ML}(\cdot)$  is the output of the ML model, and  $I(x, y, z \text{ or } t)$  is a pixel/voxel value of the input image/volume. A three-layer structure may be selected as the structure of the ANN, because it has been proved that any continuous mapping can be approximated by a three-layer ANN [71, 72]. The structure of input units and the number of hidden units in the ANN may be designed by use of sensitivity-based unit-pruning methods [73, 74]. Other ML models such as support vector regression [10, 11] can be used as a core part of the MTANN. ML regression models rather than ML classification models would be suited for the MTANN framework, because the output of the MTANN is continuous scalar values (as opposed to nominal categories or classes). The entire output image/volume is obtained by scanning with the input subvolume of the MTANN on the entire input image/volume. The input subregion/subvolume and the scanning with the MTANN can be analogous to the kernel of a convolution filter and the convolutional operation of the filter, respectively.

The training of an MTANN is shown in Figure 2. The MTANN is trained with input images/volumes and the corresponding “teaching” images/volumes for enhancement of a specific pattern and suppression of other patterns in images/volumes. The “teaching” images/volumes are ideal or desired images for the corresponding input images/volumes. For enhancement of lesions and suppression of nonlesions, the teaching volume contains a map for the “likelihood of being lesions,” represented by

$$T(x, y, z \text{ or } t) = \begin{cases} \text{a certain distribution,} & \text{for a lesion,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

To enrich the training samples, a training region,  $R_T$ , extracted from the input images is divided pixel by pixel into a large number of overlapping subregions. Single pixels are extracted from the corresponding teaching images as teaching values. The MTANN is massively trained by use of each of a large number of input subregions together with each of the corresponding teaching single pixels, hence the

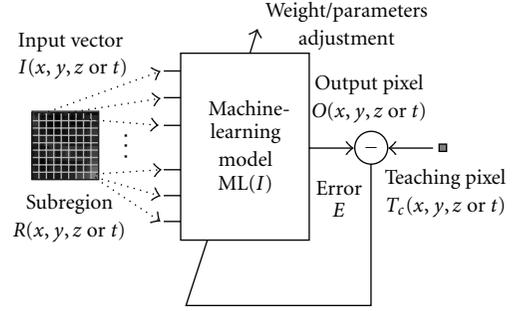


FIGURE 2: Training of an MTANN (a class of PML). An input vector is entered as input to the ML model. An error is calculated by subtracting of a teaching pixel from an output pixel. The parameters such as weights between layers in an ANN model are adjusted so that the error becomes small.

term “massive-training ANN.” The error to be minimized by training of the MTANN is represented by

$$E = \frac{1}{P} \sum_c \sum_{(x, y, z \text{ or } t) \in R_T} \{T_c(x, y, z \text{ or } t) - O_c(x, y, z \text{ or } t)\}^2, \quad (3)$$

where  $c$  is a training case number,  $O_c$  is the output of the MTANN for the  $c$ th case,  $T_c$  is the teaching value for the MTANN for the  $c$ th case, and  $P$  is the number of total training voxels in the training region for the MTANN,  $R_T$ . The expert 3D MTANN is trained by a linear-output backpropagation (BP) algorithm [40] which was derived for the linear-output ANN model by use of the generalized delta rule [8]. After training, the MTANN is expected to output the highest value when a lesion is located at the center of the subregion of the MTANN, a lower value as the distance from the subregion center increases, and zero when the input subregion contains a nonlesion.

A scoring method is used for combining output pixels from the trained MTANNs. A score for a given region of interest (ROI) from the MTANN is defined as

$$S = \sum_{(x, y, z \text{ or } t) \in R_E} f_w(x, y, z \text{ or } t) \times O(x, y, z \text{ or } t), \quad (4)$$

where  $f_w$  is a weighting function for combining pixel-based output responses from the trained MTANN into a single score, which may often be the same distribution function used in the teaching images, and with its center corresponding to the center of the region for evaluation,  $R_E$ ; and  $O$  is the output image of the trained MTANN, where its center corresponds to the center of  $R_E$ . This score represents the weighted sum of the estimates for the likelihood that the ROI (e.g., a lesion candidate) contains a lesion near the center; that is, a higher score would indicate a lesion and a lower score would indicate a non-lesion. Thresholding is then performed on the scores for distinction between lesions and non-lesions.

2.4. Convolution Neural Network (NN). A convolution NN has first been proposed for handwritten ZIP-code recognition [75]. The architecture of a convolution NN is illustrated

in Figure 3. The input, output, and teacher for convolution NNs are summarized in Table 2. The convolution NN can be considered as a simplified version of the Neocognitron model [76–78] which was proposed to simulate the human visual system in 1980 [78]. The input and output of the convolution NN are images and nominal class labels, respectively. The convolution NN consists of one input layer, several hidden layers, and one output layer. The layers are connected with local shift-invariant interconnections (or convolution with a local kernel). Unlike the neocognitron, the convolution NN has no lateral interconnections or feedback loops, and the error BP algorithm [8] is used for training of the convolution NN. Units (neurons) in any hidden layer are organized in groups. Each unit in a subsequent layer is connected with the units of a small region in each group in the preceding layer. The groups between adjacent layers are interconnected by weights that are organized in kernels. For obtaining the shift-invariant responses, connection weights between any two groups in two layers are constrained to be shift-invariant; in other words, forward signal propagation is similar to a shift-invariant convolution operation. The signals from the units in a certain layer are convolved with the weight kernel, and the resulting value of the convolution is collected into the corresponding unit in the subsequent layer. This value is further processed by the unit through an activation function and produces an output signal. The activation function between two layers is a sigmoid function. For deriving the training algorithm for the convolution NN, the generalized delta rule [8] is applied to the architecture of the convolution NN. For distinguishing an ROI containing a lesion from an ROI containing a non-lesion, a class label (e.g., 1 for a lesion, 0 for a non-lesion) is assigned to an output unit.

Variants of the convolution NN have been proposed. The dual-kernel approach, which employs central kernels and peripheral kernels in each layer [43], was proposed for distinction between lung nodules and nonnodules in chest radiographs [42, 43] and distinction between microcalcifications and other anatomic structures in mammograms [43]. This dual-kernel-based convolution NN has several output units (instead of one or two output units in the standard convolution NN) for two-class classification. The fuzzy association was employed for transformation of output values from the output units to two classes (i.e., nodules or nonnodules; microcalcifications or other anatomic structures). A convolution NN which has subsampling layers has been developed for face recognition [47]. Some convolution NNs have one output unit [48, 79], some have two output units [80], and some have more than two output units [42, 43, 45, 47] for two-class classification.

Shift-invariant NNs [50, 51] are mostly the same as convolution NNs except for the output layer, which outputs images instead of classes. The shift-invariant NNs were used for localization (detection) of lesions in images, for example, detection of microcalcifications in mammograms [50, 51] and detection of the boundaries of the human corneal endothelium in photomicrographs [81].

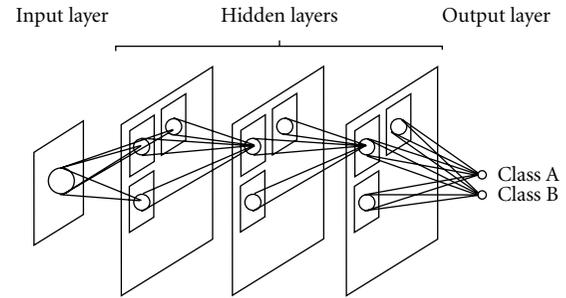


FIGURE 3: Architecture of a convolution NN (a class of PML). The convolution NN can be considered as a simplified version of the Neocognitron model, which was proposed to simulate the human visual system. The layers in the convolution NN are connected with local shift-invariant inter-connections (or convolution with a local kernel). The input and output of the convolution NN are images and nominal class labels (e.g., Class A and Class B), respectively.

**2.5. Multilayer Perceptron for Character Recognition.** A multilayer perceptron has been proposed for character recognition from an optical card reader [82, 83]. The architecture of the multilayer perceptron for character recognition is shown in Figure 4. The input, output, and teacher for the multilayer perceptron for character recognition are summarized in Table 2. The input and output of the multilayer perceptron are pixel values in a given binary image that contains a single character (e.g., a, b, or c) and a class to which the given image belongs, respectively. The number of input units equals the number of pixels in the given binary image (e.g.,  $16 \times 16$  pixels). The number of output units equals the number of classes (i.e., 26 for small-letter alphabetic characters). Each output unit is assigned to one of the classes. The class to which the given image belongs is determined as the class of the output unit with the maximum value. In the teaching data, a class label of 1 is assigned the corresponding output unit when a training sample belongs to that class; 0 is assigned to the other output units. Characters in given images are geometrically normalized before they are entered to the multilayer perceptron, because the architecture is not designed for being scale-invariant. Because character recognition with this type of the multilayer perceptron architecture is not shift-, rotation-, or scale-invariant, a large number of training samples is generally required. To enrich training samples, shifting, rotating, and scaling of training characters are often performed.

This type of multilayer perceptron has been applied to the classification of microcalcifications in mammography [23]. In this application, input images are not binary but gray-scale images. Pixel values in ROIs in mammograms or those in the Fourier-transformed ROIs were entered as input to the multilayer perceptron. In that study, the performance of the multilayer perceptrons based on ROIs in the spatial domain and the Fourier domain was found to be comparable.

**2.6. Non-PML Feature-Based Classifiers.** One of most popular uses of ML algorithms would probably be classification. In this use, an ML algorithm is called a classifier. A standard

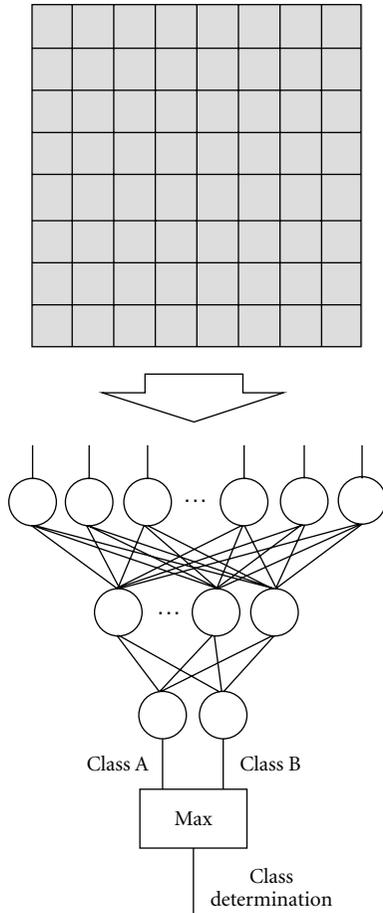


FIGURE 4: Architecture of a multilayer perceptron for character recognition. The binary pixel values in an image are entered as input to the multilayer perceptron. The class to which the given image belongs is determined as the class of the output unit with the maximum value.

classification approach based on a multilayer perceptron is illustrated in Figure 5. The input, output, and teacher for a classifier with features are summarized in Table 2. First, target objects are segmented by use of a segmentation method. Next, features are extracted from the segmented objects. Then, extracted features are entered as input to an ML model such as linear discriminant analysis [7], quadratic discriminant analysis [7], a multilayer perceptron [8, 9], and a support-vector machine [10, 11]. The ML model is trained with sets of input features and correct class labels. A class label of 1 is assigned to the corresponding output unit when a training sample belongs to that class, and 0 is assigned to the other output units. After training, the class of the unit with the maximum value is determined to be the corresponding class to which an unknown sample belongs. For details of feature-based classifiers, refer to one of many textbooks in pattern recognition such as [6–8, 10, 84].

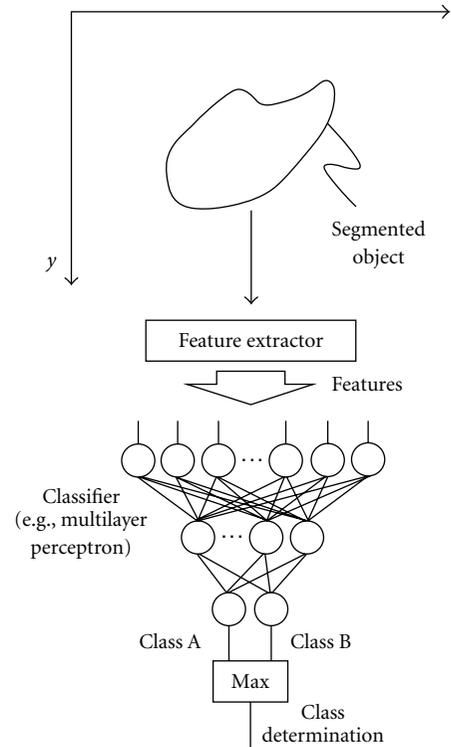


FIGURE 5: Standard classifier approach to classification of an object (i.e., feature-based ML). Features (e.g., contrast, effective diameter, and circularity) are extracted from a segmented object in an image. Those features are entered as input to a classifier such as a multilayer perceptron. Class determination is made by taking the class of the output unit with the maximum value.

### 3. Similarities and Differences

**3.1. Within Different PML Algorithms.** MTANNs [52] were developed by extension of neural filters to accommodate various pattern-recognition tasks. In other words, neural filters are a subclass or a special case of MTANNs. The applications and functions of neural filters are limited to noise reduction [38, 39] and edge enhancement [40, 41], whereas those of MTANNs were extended to include classification [52–54, 57–62], pattern enhancement and suppression [54], and object detection [56]. The input information to MTANNs, which is the pixel values in a subregion, is the same as that to neural filters. However, the output of (thus, teacher for) neural filters is the desired pixel values in a given image, whereas that of MTANNs is a map for the likelihood of being a specific pattern in a given image, as summarized in Table 2.

Both convolution NNs and the perceptron used for character recognition are in the class of PML. Input information to the convolution NNs and the perceptron is the pixel values in a given image, whereas the output of (thus, teacher for) both algorithms is a nominal class label for the given image. Thus, the input and output information are the same for both algorithms. However, the input images for the perceptron for character recognition are limited to be binary, although the perceptron itself is capable of processing

gray-scale images. The major difference between convolution NNs and the perceptron used for character recognition is their internal architectures. Units in layers of the perceptron are fully connected, whereas the connections in the convolution NN are spatially (locally) limited. Because of this architecture, forward signal propagation in the convolution NN is realized by a convolution operation. This convolution operation offers a shift-invariant property which is desirable for image classification. The applications and functions of the perceptron are limited to character recognition such as zip code recognition and optical character recognition, whereas those of convolution NNs are general classification of images into known classes such as classification of lesion candidates into lesions or nonlesions [42–46], classification of faces [47], and classification of characters [48].

Convolution NNs, shift-invariant NNs, and MTANNs perform convolution operations. In convolution NNs and shift-invariant NNs, convolution operations are performed within the network, as shown in Figure 3, whereas the convolutional operation is performed outside the network in the MTANN, as shown in Figure 1.

*3.2. Between PML Algorithms and Ordinary Classifiers.* The major difference between PMLs and ordinary classifiers (i.e., feature-based classifiers) is the input information. Ordinary classifiers use features extracted from a segmented object in a given image, whereas PMLs use pixel values in a given image as the input information. Although the input information to PMLs can be features (see addition of features to the input information to neural filters in [38], i.e.), these features are obtained pixel by pixel (rather than by object). In other words, features for PMLs are features at each pixel in a given image, whereas features for ordinary classifiers are features from a segmented object. In that sense, feature-based classifiers may be referred to as object-based classifiers. Because PMLs use pixel/voxel values in images directly instead of features calculated from segmented objects as the input information, feature calculation or segmentation is not required. Although the development of segmentation techniques has been studied for a long time, segmentation of objects is still challenging, especially for complicated objects, subtle objects, and objects in a complex background. Thus, segmentation errors may occur for complicated objects. Because, with PMLs, errors caused by inaccurate feature calculation and segmentation can be avoided, the performance of PMLs can be higher than that of ordinary classifiers for some cases, such as complicated objects.

The output information from ordinary classifiers, convolution NNs, and the perceptron used for character recognition is nominal class labels, whereas that from neural filters, MTANNs, and shift-invariant NNs is images. With the scoring method in MTANNs, output images of the MTANNs are converted to likelihood scores for distinguishing among classes, which allow MTANNs to do classification. In addition to classification, MTANNs can perform pattern enhancement and suppression as well as object detection, whereas the other PMLs cannot.

## 4. Applications of PML Algorithms in Medical Images

*4.1. Edge-Preserving Noise Reduction by Use of Neural Filters.* Quantum noise is dominant in low-radiation-dose X-ray images used in diagnosis. For training a neural filter to reduce quantum noise in diagnostic X-ray images while preserving image details such as edges, noisy input images and corresponding “teaching” images are necessary. When a high radiation dose is used, X-ray images with little noise can be acquired and used as the “teaching” images. A noisy input image can be synthesized by addition of simulated quantum noise (which is modeled as signal-dependent noise) to a noiseless original high-radiation-dose image  $f_o(x, y)$ , represented by

$$f_N(x, y) = f_o(x, y) + n[\sigma\{f_o(x, y)\}], \quad (5)$$

where  $n[\sigma\{f_o(x, y)\}]$  is noise with standard deviation  $\sigma\{f_o(x, y)\} = k_N\sqrt{f_o(x, y)}$  and  $k_N$  is a parameter determining the amount of noise. A synthesized noisy X-ray image obtained with this method and a noiseless original high-radiation-dose X-ray image are illustrated in Figure 6(a). They are angiograms of coronary arteries. They were used as the input image and as the teaching image for training of a neural filter. For sufficient reduction of noise, the input region of the neural filter consisted of  $11 \times 11$  pixels. For efficient training of the entire image, 5,000 training pixels were sampled randomly from the input and teaching images. The training of the neural filter was performed for 100,000 iterations. The output image of the trained neural filter for a nontraining case is shown in Figure 6(b). The noise in the input image is reduced while image details such as the edges of arteries are maintained. When an averaging filter was used for noise reduction, the edges of arteries were blurry, as shown in Figure 6(b).

*4.2. Edge Enhancement from Noisy Images by Use of Neural Edge Enhancer.* Although conventional edge enhancers can very well enhance edges in images with little noise, they do not work well on noisy images. To address this issue, a neural edge enhancer has been developed for enhancing edges from very noisy images [40]. The neural edge enhancer is based on a neural filter and can be trained with input images and corresponding “teaching” edge images. Figure 7(a) shows a way of creating noisy input images and corresponding “teaching” edge images from a noiseless image for training of a neural edge enhancer. Simulated quantum noise was added to original noiseless images to create noisy input images. A Sobel edge enhancer [85] was applied to the original noiseless images to create “teaching” edge images. The key here is that the Sobel edge enhancer works very well for noiseless images. The neural edge enhancer was trained with the noisy input images together with the corresponding teaching edge images. For comparison, the trained neural edge enhancer and the Sobel edge enhancer were applied to nontraining noisy images. The resulting nontraining edge-enhanced images are shown in Figure 7(b). Edges are enhanced clearly in the output image of the neural edge

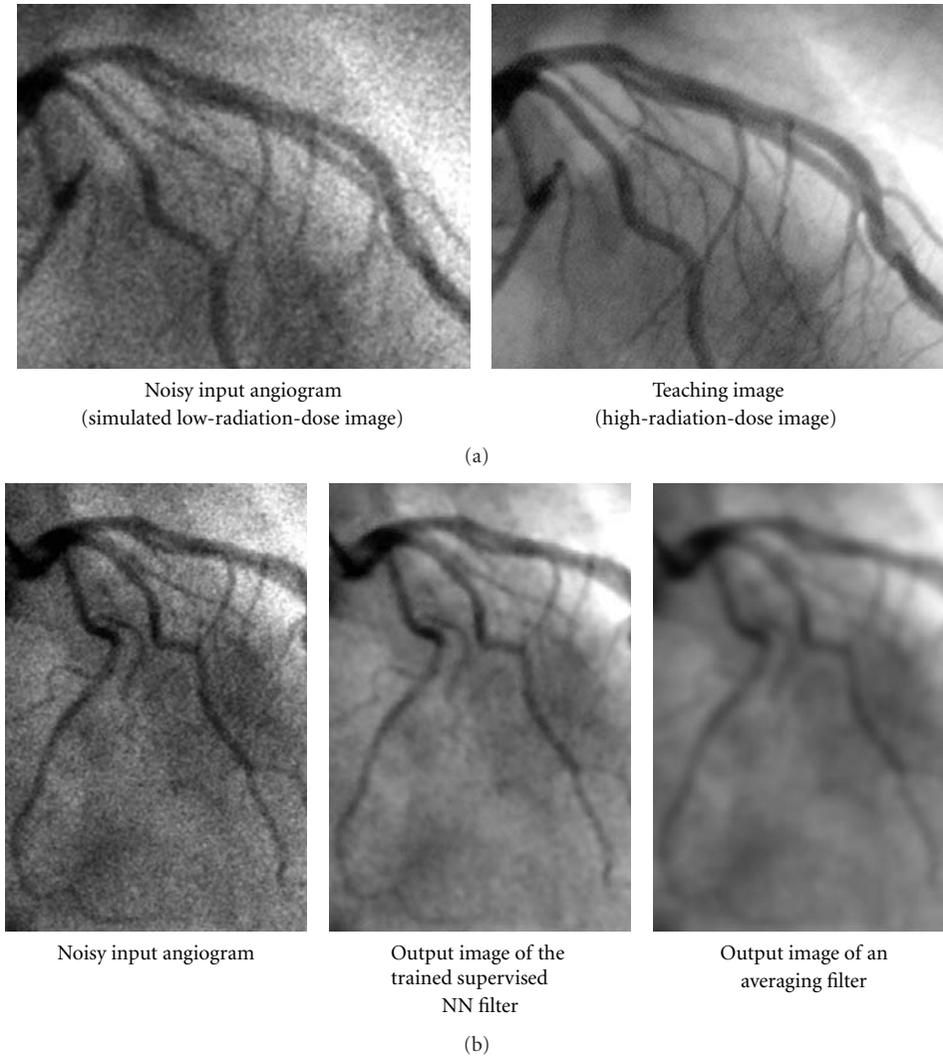


FIGURE 6: Reduction of quantum noise in angiograms by using a supervised NN filter called a “neural filter.” (a) Images used for training of the neural filter. (b) Result of an application of the trained neural filter to a nontraining image and a comparison result with an averaging filter.

enhancer while noise is suppressed, whereas the Sobel edge enhancer enhances not only edges but also noise.

*4.3. Bone Separation from Soft Tissue in Chest Radiographs (CXRs) by Use of MTANNs.* CXR is the most frequently used diagnostic imaging examination for chest diseases such as lung cancer, tuberculosis, and pneumonia. More than 9 million people worldwide die annually from chest diseases [86]. Lung cancer causes 945,000 deaths and is the leading cause of cancer deaths in the world [86] and in countries such as the United States, the United Kingdom, and Japan [87]. Lung nodules (i.e., potential lung cancers) in CXR, however, can be overlooked by radiologists in from 12 to 90% of cases that have nodules visible in retrospect [88, 89]. Studies showed that 82 to 95% of the missed lung cancers were partly obscured by overlying bones such as ribs and/or a clavicle [88, 89]. To address this issue, dual-energy imaging has been

investigated [90, 91]. Dual-energy imaging uses the energy dependence of the X-ray attenuation by different materials; it can produce two tissue-selective images, that is, a “bone” image and a “soft-tissue” image [92–94]. Major drawbacks of dual-energy imaging, however, are that (a) the radiation dose can be double, (b) specialized equipment for obtaining dual-energy X-ray exposures is required, and (c) the subtraction of two-energy images causes an increased noise level in the images.

For resolving the above drawbacks with dual-energy images, MTANNs have been developed as an image-processing technique for separation of ribs from soft tissue [54, 70]. The basic idea is to train the MTANN with soft-tissue and bone images acquired with a dual-energy radiography system [92, 95, 96]. For separation of ribs from soft tissue, the MTANN was trained with input CXRs and the corresponding “teaching” dual-energy bone images, as illustrated in Figure 8(a). Figure 8(b) shows a nontraining

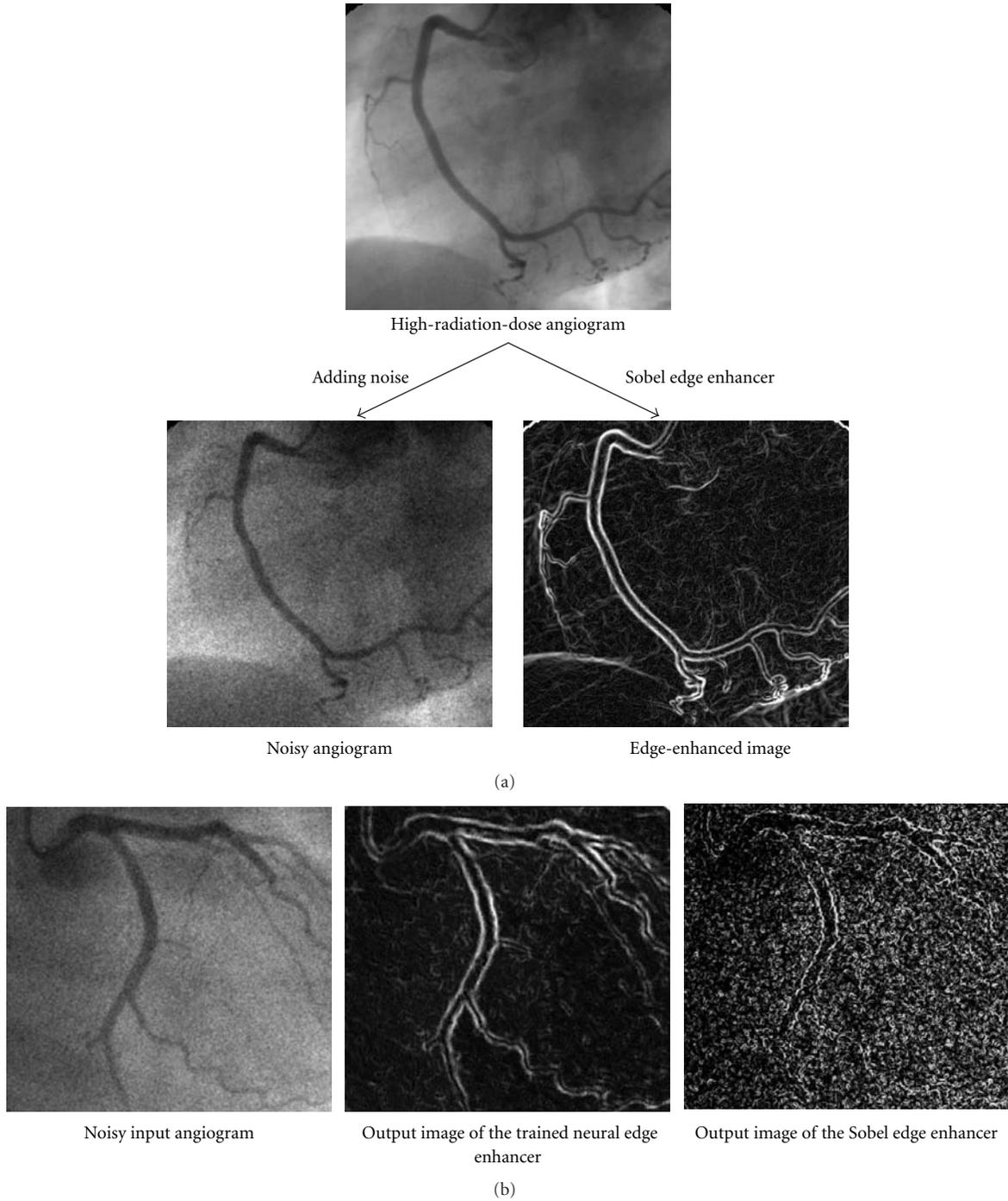


FIGURE 7: Enhancement of edges from noisy images by use of a supervised edge enhancer called a “neural edge enhancer.” (a) A way to create noisy input images and corresponding “teaching” edge images from noiseless images for training a neural edge enhancer. (b) Result of an application of the trained neural edge enhancer to a nontraining image and a comparison result with a Sobel edge enhancer.

original CXR and a soft-tissue image obtained by use of the trained MTANN. The contrast of ribs is suppressed substantially in the MTANN soft-tissue image, whereas the contrast of soft tissue such as lung vessels is maintained. There is another PML approach called filter learning to do the same task [64].

4.4. *Enhancement and Detection of Lesions by Use of MTANNs.* Computer-aided diagnosis (CAD) has been an active area of study in medical image analysis [1, 2, 97, 98]. Some CAD schemes employ a filter for enhancement of lesions as a preprocessing step for improving sensitivity and specificity, but some do not employ such a filter. The filter enhances

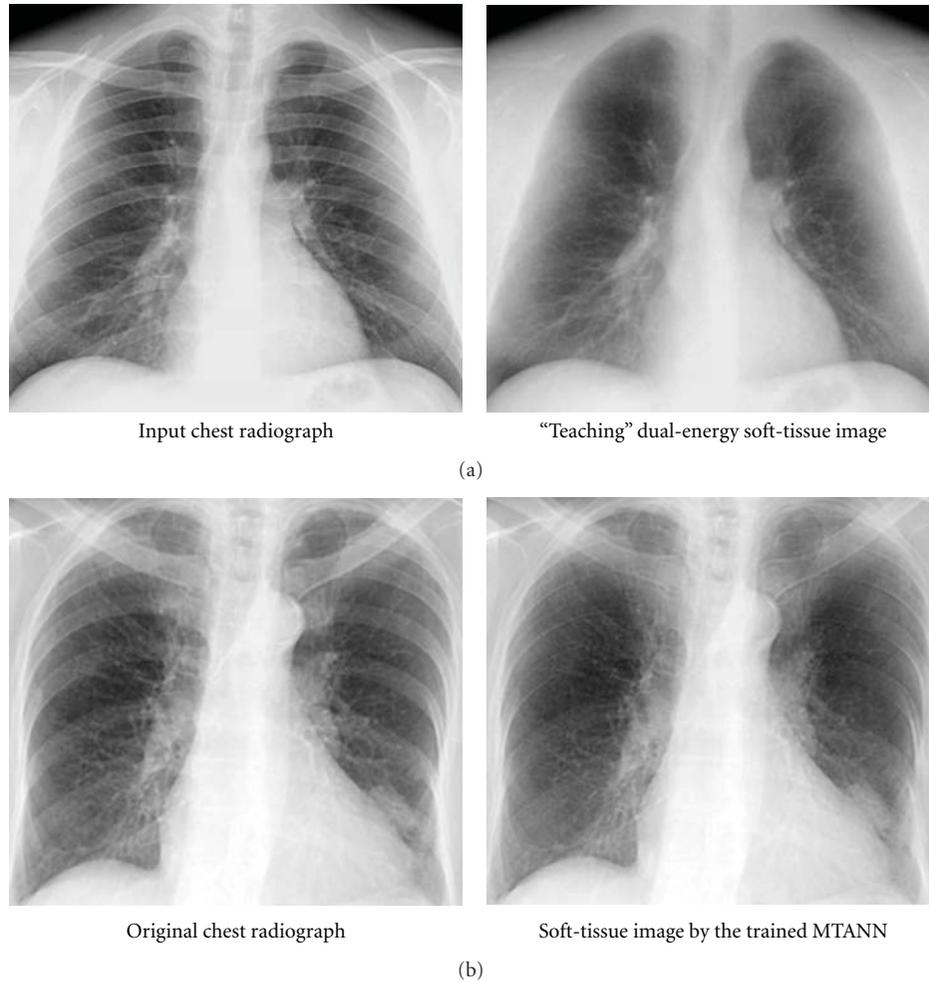


FIGURE 8: Separation of bones from soft tissue in CXRs by use of an MTANN. (a) Images used for training the MTANN. (b) Result of an application of the trained MTANN to a nontraining CXR.

objects similar to a model employed in the filter; for example, a blob-enhancement filter based on the Hessian matrix enhances sphere-like objects [99]. Actual lesions, however, often differ from a simple model; for example, a lung nodule is generally modeled as a solid sphere, but there are nodules of various shapes and inhomogeneous nodules such as nodules with spiculation and ground-glass nodules. Thus, conventional filters often fail to enhance such actual lesions.

To address this issue, a "lesion-enhancement" filter based on MTANNs has been developed for enhancement of actual lesions in a CAD scheme for detection of lung nodules in CT [56]. For enhancement of lesions and suppression of nonlesions in CT images, the teaching image contains a map for the "likelihood of being lesions." For enhancement of a nodule in an input CT image, a 2D Gaussian distribution was placed at the location of the nodule in the teaching image, as a model of the likelihood of being a lesion. For testing of the performance, the trained MTANN was applied to nontraining lung CT images. As shown in Figure 9, the nodule is enhanced in the output image of the trained MTANN filter, while normal structures such as lung vessels

are suppressed. Note that small remaining regions due to vessels can easily be separated from nodules by use of their area information which can be obtained by use of connected-component labeling [100–102].

#### 4.5. Classification between Lesions and Nonlesions by Use of Different PML Algorithms

4.5.1. MTANNs. Shift-invariant NNs are mostly the same as convolution NNs except for the output layer, which outputs images instead of classes. The shift-invariant NNs can be used for localization (detection) of objects in images in addition to classification [50, 51]. A major challenge in CAD development is to reduce the number of FPs [27, 103–107], because there are various normal structures similar to lesions in medical images. To address this issue, an FP-reduction technique based on an MTANN has been developed for a CAD scheme for lung nodule detection in CT [52]. For enhancement of nodules (i.e., true positives) and suppression of nonnodules (i.e., FPs) on CT images, the teaching

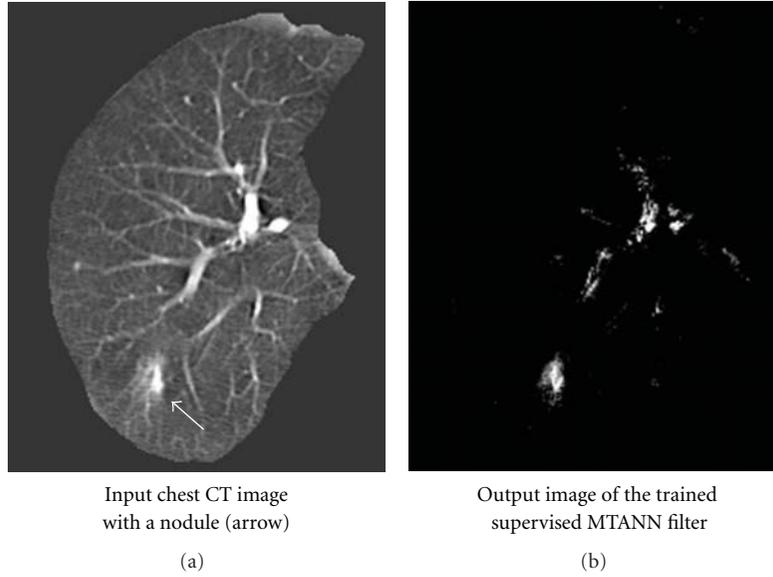


FIGURE 9: Enhancement of a lesion by use of the trained lesion-enhancement MTANN filter for a nontraining case. (a) Original chest CT image of the segmented lung with a nodule (indicated by an arrow). (b) Output image of the trained lesion-enhancement MTANN filter.

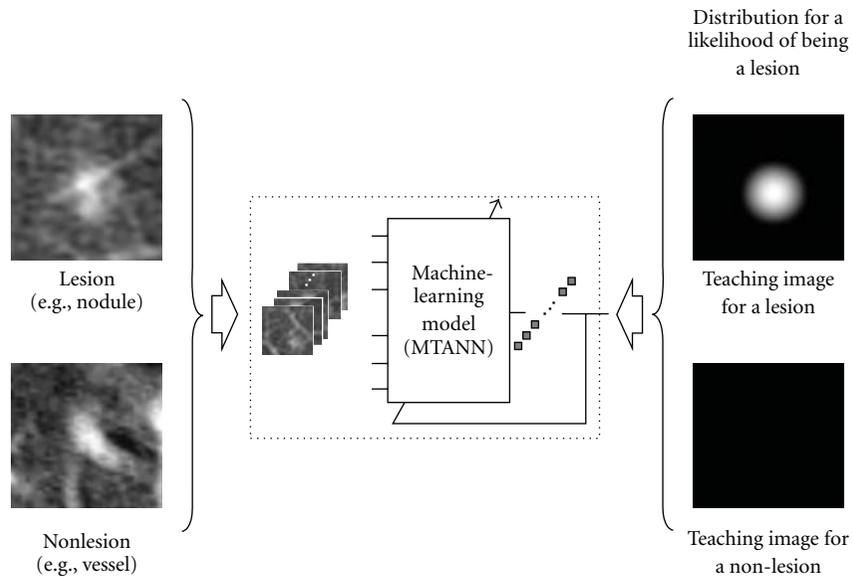


FIGURE 10: Training of an MTANN for distinction between lesions and non-lesions in a CAD scheme for detection of lesions in medical images. The teaching image for a lesion contains a Gaussian distribution; that for a non-lesion contains zero (completely dark). After the training, the MTANN expects to enhance lesions and suppress non-lesions.

image contains a distribution of values that represent the “likelihood of being a nodule.” For example, the teaching volume contains a 3D Gaussian distribution with standard deviation  $\sigma_T$  for a lesion and zero (i.e., completely dark) for non-lesions, as illustrated in Figure 10. This distribution represents the “likelihood of being a lesion”:

$$T(x, y, z \text{ or } t) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left\{-\frac{(x^2 + y^2 + z^2 \text{ or } t^2)}{2\sigma_T^2}\right\}, & \text{for a lesion,} \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

A 3D Gaussian distribution is used to approximate an average shape of lesions. The MTANN involves training with a large number of subvolume-voxel pairs, which is called a massive-subvolumes training scheme.

A scoring method is used for combining of output voxels from the trained MTANNs, as illustrated in Figure 11. A score for a given ROI from the MTANN is defined as

$$S = \sum_{(x,y,z \text{ or } t) \in R_E} f_W(x, y, z \text{ or } t) \times O(x, y, z \text{ or } t), \quad (7)$$

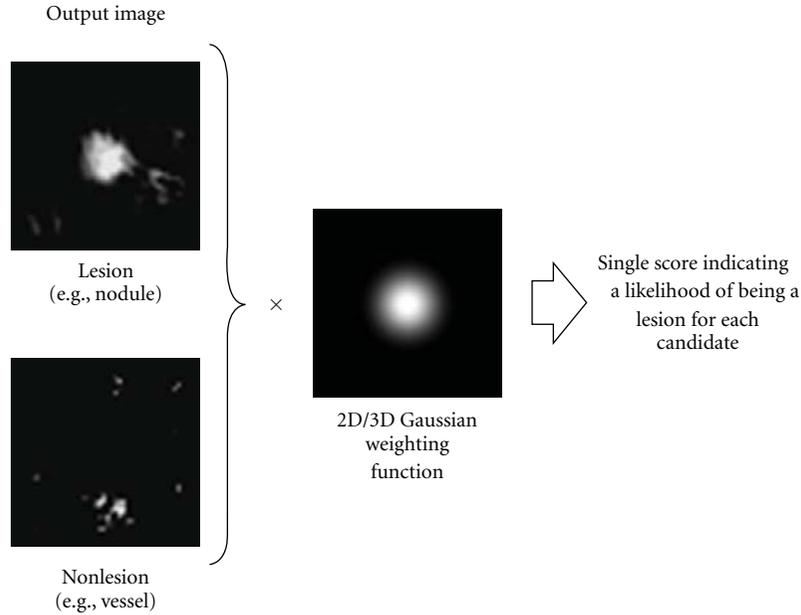


FIGURE 11: Scoring method for combining pixel-based output responses from the trained MTANN into a single score for each ROI.

where

$$f_W(x, y, z \text{ or } t) = f_G(x, y, z \text{ or } t; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x^2+y^2+z^2 \text{ or } t^2)/2\sigma^2} \quad (8)$$

is a 3D Gaussian weighting function with standard deviation  $\sigma$  and with its center corresponding to the center of the volume for evaluation,  $R_E$ , and  $O$  is the output image of the trained MTANN, where its center corresponds to the center of  $R_E$ . The use of the 3D Gaussian weighting function allows us to combine the responses (outputs) of a trained MTANN as a 3D distribution. A 3D Gaussian function is used for scoring, because the output of a trained MTANN is expected to be similar to the 3D Gaussian distribution used in the teaching images. This score represents the weighted sum of the estimates for the likelihood that the ROI (lesion candidate) contains a lesion near the center; that is, a higher score would indicate a lesion and a lower score would indicate a nonlesion. Thresholding is then performed on the scores for distinction between lesions and non-lesions.

An MTANN was trained with typical nodules and typical types of FPs (nonnodules) and corresponding teaching images. The trained MTANN was applied to 57 true positives (nodules) and 1,726 FPs (nonnodules) produced by a CAD scheme [52]. Figure 12 shows various types of nodules and nonnodules and the corresponding output images of the trained MTANN. Nodules such as a solid nodule, a part-solid (mixed-ground-glass) nodule, and a non-solid (ground-glass) nodule are enhanced, whereas nonnodules such as different-sized lung vessels and soft-tissue opacity are suppressed around the centers of ROIs. For combining output pixels into a single score for each nodule candidate, a scoring method was applied to the output images for distinction between a nodules and a nonnodule. Thresholding of

scores was done for classification of nodule candidates into nodules or nonnodules. Free-response receiver operating characteristic (FROC) analysis [108] was carried out for evaluation of the performance of the trained MTANN. The FROC curve for the MTANN indicates 80.3% overall sensitivity (100% classification performance) and a reduction in the FP rate from 0.98 to 0.18 per section, as shown in Figure 13.

**4.5.2. Convolution NNs and Shift-Invariant NNs.** Convolution NNs have been used for FP reduction in CAD schemes for lung nodule detection in CXRs [42–44]. A convolution NN was trained with 28 chest radiographs for distinguishing lung nodules from nonnodules (i.e., FPs produced by an initial CAD scheme). The trained convolution NN reduced 79% of FP detections (which is equivalent to 2–3 FPs per patient), while 80% of true-positive detections were preserved. Convolution NNs have been applied to FP reduction in CAD schemes for detection of microcalcifications [45] and masses [46] in mammography. A convolution NN was trained with 34 mammograms for distinguishing microcalcifications from FPs. The trained convolution NN reduced 90% of FP detections, which resulted in 0.5 FP detections per image, while a true-positive detection rate of 87% was preserved [45].

Shift-invariant NNs have been used for FP reduction in CAD for detection of microcalcifications [50, 51]. A shift-invariant NN was trained to detect microcalcifications in ROIs. Microcalcifications were detected by thresholding of the output images of the trained shift-invariant NN. When the number of detected microcalcifications was greater than a predetermined number, the ROI was considered as a microcalcification ROI. With the trained shift-invariant NN, 55% of FPs was removed without any loss of true positives.

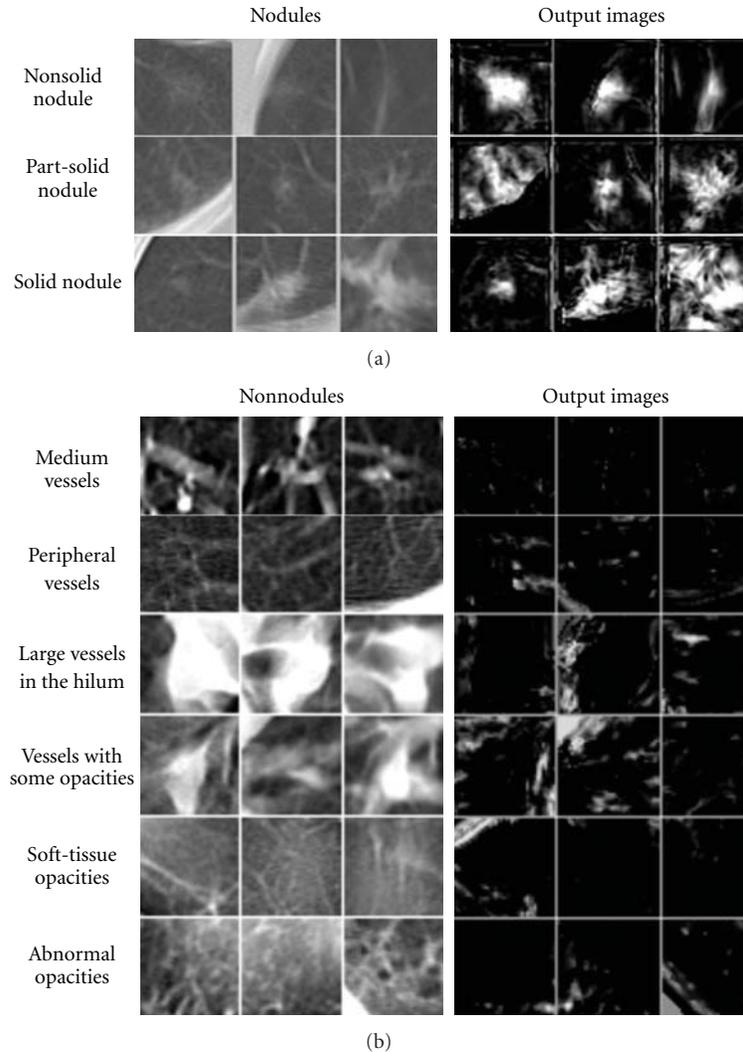


FIGURE 12: Illustrations of various types of nontraining nodules and nonnodules and corresponding output images of the trained MTANN. Nodules are represented by bright pixels, whereas nonnodules are almost dark around the centers of ROIs.

## 5. Advantages and Limitations of PML Algorithms

As described earlier, the major difference between PMLs and ordinary classifiers is the direct use of pixel values with PML. In other words, unlike ordinary classifiers, feature calculation from segmented objects is not necessary. Because the PML can avoid errors caused by inaccurate feature calculation and segmentation, the performance of the PML can potentially be higher than that of ordinary feature-based classifiers for some cases. PMLs learn pixel data directly, and thus all information on pixels should not be lost before the pixel data are entered into the PML, whereas ordinary feature-based classifiers learn the features extracted from segmented lesions and thus important information can be lost with this indirect extraction; also, inaccurate segmentation often occurs for complicated patterns. In addition, because feature calculation is not required for PML, development and

implementation of segmentation and feature calculation, and selection of features are unnecessary.

Ordinary classifiers such as linear discriminant analysis, ANNs, and support vector machines cannot be used for image processing, detection (localization) of objects, or enhancement of objects or patterns, whereas MTANNs can do those tasks. For example, MTANNs can separate bones from soft tissue in CXRs [54], and MTANN can enhance and detect lung nodules on CT images [56].

The characteristics of PMLs which use pixel data directly should differ from those of ordinary feature-based classifiers. Therefore, combining an ordinary feature-based classifier with a PML would yield a higher performance than that of a classifier alone or a PML alone. Indeed, in previous studies, both classifier and PML were used successfully for classification of lesion candidates into lesions and non-lesions [17, 45, 46, 49–53, 58–63].

A limitation of PMLs is the relatively long time for training because of the high dimensionality of input data. Because

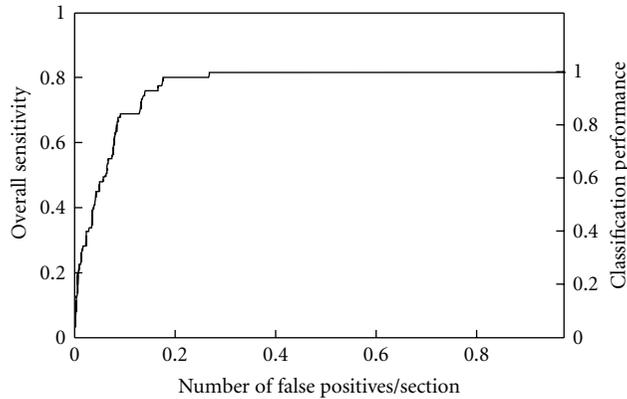


FIGURE 13: FROC curve indicating the performance of the MTANN in distinction between 57 true positives (nodules) and 1.726 FPs (nonnodules).

PMLs use pixel data in images directly, the number of input dimensions is generally large. For example, a 3D MTANN for 3D CT data requires 171 dimensions for its input [53, 60]. The ordinary feature-based classifiers are more efficient than PMLs. In an application of PMLs and feature-based classifiers to CAD schemes, a feature-based classifier should be applied first, because the number of lesion candidates that need to be classified is larger at an earlier stage. After the number of lesion candidates is reduced by use of the feature-based classifier, a PML should be applied for further reduction of FPs. Indeed, previous studies employed this strategy [17, 52, 53, 58–61].

To address the issue of training time for PML, dimensionality reduction methods for PML have been proposed [61]. With the use of the Laplacian-eigenfunction-based dimensionality reduction of the input vectors to a 3D MTANN, the training time was reduced by a factor of 8.5.

## 6. Conclusion

In this paper, PMLs were surveyed and compared with each other as well as with other non-PML algorithms (i.e., ordinary feature-based classifiers) to make the similarities, differences, advantages, and limitations clear. The major difference between PMLs and non-PML algorithms (e.g., classifiers) is a need for segmentation and feature calculation with non-PML algorithms. The major advantage of PMLs over non-PML algorithms is that no information is lost due to inaccurate segmentation and feature calculation, which would result in a higher performance for some cases such as complicated patterns. With the combination of PMLs with non-PML algorithms, the performance of a system can be improved substantially. In addition to a classification task, MTANNs can be used for enhancement (and suppression) and detection (i.e., localization) of objects (or patterns) in images.

## Acknowledgments

The author is grateful to all members in the Suzuki laboratory, that is, postdoctoral scholars, computer scientists, visiting scholars/professors, medical students, graduate/undergraduate students, research technicians, research volunteers, and support staff, in the Department of Radiology at the University of Chicago, for their invaluable assistance and contributions in the studies, to colleagues and collaborators for their valuable suggestions, and to Ms. E. F. Lanzl for improving the paper. This work was partly supported by Grant number R01CA120549 from the National Cancer Institute/National Institutes of Health and by the NIH S10 RR021039 and P30 CA14599. PML technologies developed at the University of Chicago have been licensed to companies including Riverain Medical, Deus Technology, and Median Technologies. It is the policy of the University of Chicago that investigators disclose publicly actual or potential significant financial interests that may appear to be affected by research activities.

## References

- [1] M. L. Giger and K. Suzuki, "Computer-aided diagnosis (CAD)," in *Biomedical Information Technology*, D. D. Feng, Ed., pp. 359–374, Academic Press, New York, NY, USA, 2007.
- [2] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *British Journal of Radiology*, vol. 78, supplement 1, pp. S3–S19, 2005.
- [3] F. Li, S. Sone, H. Abe, H. MacMahon, S. G. Armato III, and K. Doi, "Lung cancers missed at low-dose helical CT screening in a general population: comparison of clinical, histopathologic, and imaging findings," *Radiology*, vol. 225, no. 3, pp. 673–683, 2002.
- [4] A. Lostumbo, C. Wanamaker, J. Tsai, K. Suzuki, and A. H. Dachman, "Comparison of 2D and 3D views for evaluation of flat lesions in CT colonography," *Academic Radiology*, vol. 17, no. 1, pp. 39–47, 2010.
- [5] R. M. Soetikno, T. Kaltenbach, R. V. Rouse et al., "Prevalence of nonpolypoid (flat and depressed) colorectal neoplasms in asymptomatic and symptomatic adults," *Journal of the American Medical Association*, vol. 299, no. 9, pp. 1027–1035, 2008.
- [6] R. O. Duda, P. E. Hart, and D. G. Stork, in *Pattern Recognition*, Wiley Interscience, Hoboken, NJ, USA, 2nd edition, 2001.
- [7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, Calif, USA, 2nd edition, 1990.
- [8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Parallel Distributed Processing*, vol. 1, pp. 318–362, 1986.
- [10] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, Berlin, Germany, 1995.
- [11] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.
- [12] J. Shiraiishi, Q. Li, K. Suzuki, R. Engelmann, and K. Doi, "Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: localized search method based

- on anatomical classification," *Medical Physics*, vol. 33, no. 7, pp. 2642–2653, 2006.
- [13] G. Coppini, S. Diciotti, M. Falchini, N. Villari, and G. Valli, "Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 4, pp. 344–357, 2003.
  - [14] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs," *Medical Image Analysis*, vol. 12, no. 3, pp. 240–258, 2008.
  - [15] S. Chen, K. Suzuki, and H. MacMahon, "Development and evaluation of a computer-aided diagnostic scheme for lung nodule detection in chest radiographs by means of two-stage nodule enhancement with support vector classification," *Medical Physics*, vol. 38, no. 4, pp. 1844–1858, 2011.
  - [16] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: preliminary results," *Medical Physics*, vol. 28, no. 8, pp. 1552–1561, 2001.
  - [17] H. Arimura, S. Katsuragawa, K. Suzuki et al., "Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening," *Academic Radiology*, vol. 11, no. 6, pp. 617–629, 2004.
  - [18] X. Ye, X. Lin, J. Dehmshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 7, Article ID 5073252, pp. 1810–1820, 2009.
  - [19] T. W. Way, B. Sahiner, H. P. Chan et al., "Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features," *Medical Physics*, vol. 36, no. 7, pp. 3086–3098, 2009.
  - [20] M. Aoyama, Q. Li, S. Katsuragawa, H. MacMahon, and K. Doi, "Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images," *Medical Physics*, vol. 29, no. 5, pp. 701–708, 2002.
  - [21] M. Aoyama, Q. Li, S. Katsuragawa, F. Li, S. Sone, and K. Doi, "Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images," *Medical Physics*, vol. 30, no. 3, pp. 387–394, 2003.
  - [22] S. K. Shah, M. F. McNitt-Gray, S. R. Rogers et al., "Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancement features," *Academic Radiology*, vol. 12, no. 10, pp. 1310–1319, 2005.
  - [23] Y. Wu, K. Doi, M. L. Giger, and R. M. Nishikawa, "Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks," *Medical Physics*, vol. 19, no. 3, pp. 555–560, 1992.
  - [24] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.
  - [25] S. N. Yu, K. Y. Li, and Y. K. Huang, "Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model," *Computerized Medical Imaging and Graphics*, vol. 30, no. 3, pp. 163–173, 2006.
  - [26] J. Ge, B. Sahiner, L. M. Hadjiiski et al., "Computer aided detection of clusters of microcalcifications on full field digital mammograms," *Medical Physics*, vol. 33, no. 8, pp. 2975–2988, 2006.
  - [27] Y. T. Wu, J. Wei, L. M. Hadjiiski et al., "Bilateral analysis based false positive reduction for computer-aided mass detection," *Medical Physics*, vol. 34, no. 8, pp. 3334–3344, 2007.
  - [28] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," *Academic Radiology*, vol. 5, no. 3, pp. 155–168, 1998.
  - [29] P. Delogu, M. E. Fantacci, P. Kasae, and A. Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," *Computers in Biology and Medicine*, vol. 37, no. 10, pp. 1479–1491, 2007.
  - [30] J. Shi, B. Sahiner, H. P. Chan et al., "Characterization of mammographic masses based on level set segmentation with new image features and patient information," *Medical Physics*, vol. 35, no. 1, pp. 280–290, 2008.
  - [31] H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1261–1274, 2001.
  - [32] A. K. Jerebko, R. M. Summers, J. D. Malley, M. Franaszek, and C. D. Johnson, "Computer-assisted detection of colonic polyps with CT colonography using neural networks and binary classification trees," *Medical Physics*, vol. 30, no. 1, pp. 52–60, 2003.
  - [33] S. Wang, J. Yao, and R. M. Summers, "Improved classifier for computer-aided polyp detection in CT colonography by nonlinear dimensionality reduction," *Medical Physics*, vol. 35, no. 4, pp. 1377–1386, 2008.
  - [34] C. Muramatsu, Q. Li, R. A. Schmidt et al., "Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings," *Medical Physics*, vol. 34, no. 7, pp. 2890–2895, 2007.
  - [35] C. Muramatsu, Q. Li, R. Schmidt et al., "Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: observer study results," *Medical Physics*, vol. 33, no. 9, pp. 3460–3468, 2006.
  - [36] C. Muramatsu, Q. Li, K. Suzuki et al., "Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results," *Medical Physics*, vol. 32, no. 7, pp. 2295–2304, 2005.
  - [37] H. Arimura, Q. Li, Y. Korogi et al., "Computerized detection of intracranial aneurysms for three-dimensional MR angiography: feature extraction of small protrusions based on a shape-based difference image technique," *Medical Physics*, vol. 33, no. 2, pp. 394–401, 2006.
  - [38] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Neural filter with selection of input features and its application to image quality improvement of medical image sequences," *IEICE Transactions on Information and Systems*, vol. E85-D, no. 10, pp. 1710–1718, 2002.
  - [39] K. Suzuki, I. Horiba, and N. Sugie, "Efficient approximation of neural filters for removing quantum noise from images," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1787–1799, 2002.
  - [40] K. Suzuki, I. Horiba, and N. Sugie, "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1582–1596, 2003.
  - [41] K. Suzuki, I. Horiba, N. Sugie, and M. Nanki, "Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector," *IEEE Transactions on Medical Imaging*, vol. 23, no. 3, pp. 330–339, 2004.
  - [42] S. B. Lo, S. A. Lou, J. S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE*

- Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, 1995.
- [43] S. C. B. Lo, H. P. Chan, J. S. Lin, H. Li, M. T. Freedman, and S. K. Mun, “Artificial convolution neural network for medical image pattern recognition,” *Neural Networks*, vol. 8, no. 7-8, pp. 1201–1214, 1995.
- [44] J. S. Lin, B. Shih-Chung, A. Hasegawa, M. T. Freedman, and S. K. Mun, “Reduction of false positives in lung nodule detection using a two-level neural classification,” *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 206–217, 1996.
- [45] S. C. B. Lo, H. Li, Y. Wang, L. Kinnard, and M. T. Freedman, “A multiple circular path convolution neural network system for detection of mammographic masses,” *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 150–158, 2002.
- [46] B. Sahiner, H. P. Chan, N. Petrick et al., “Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images,” *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 598–610, 1996.
- [47] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, “Face recognition: a convolutional neural-network approach,” *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.
- [48] C. Neubauer, “Evaluation of convolutional neural networks for visual recognition,” *IEEE Transactions on Neural Networks*, vol. 9, no. 4, pp. 685–696, 1998.
- [49] D. Wei, R. M. Nishikawa, and K. Doi, “Application of texture analysis and shift-invariant artificial neural network to microcalcification cluster detection,” *Radiology*, vol. 201, pp. 696–696, 1996.
- [50] W. Zhang, K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt, “An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms,” *Medical Physics*, vol. 23, no. 4, pp. 595–601, 1996.
- [51] W. Zhang, K. Doi, M. L. Giger, Y. Wu, R. M. Nishikawa, and R. A. Schmidt, “Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network,” *Medical Physics*, vol. 21, no. 4, pp. 517–524, 1994.
- [52] K. Suzuki, S. G. Armato III, F. Li, S. Sone, and K. Doi, “Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography,” *Medical Physics*, vol. 30, no. 7, pp. 1602–1617, 2003.
- [53] K. Suzuki, H. Yoshida, J. Nappi, and A. H. Dachman, “Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: suppression of rectal tubes,” *Medical Physics*, vol. 33, no. 10, pp. 3814–3824, 2006.
- [54] K. Suzuki, H. Abe, H. MacMahon, and K. Doi, “Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN),” *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 406–416, 2006.
- [55] S. Oda, K. Awai, K. Suzuki et al., “Performance of radiologists in detection of small pulmonary nodules on chest radiographs: effect of rib suppression with a massive-training artificial neural network,” *American Journal of Roentgenology*, vol. 193, no. 5, pp. W397–W402, 2009.
- [56] K. Suzuki, “A supervised ‘lesion-enhancement’ filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD),” *Physics in Medicine and Biology*, vol. 54, no. 18, pp. S31–S45, 2009.
- [57] K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi, “False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network,” *Academic Radiology*, vol. 12, no. 2, pp. 191–201, 2005.
- [58] K. Suzuki, F. Li, S. Sone, and K. Doi, “Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network,” *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1138–1150, 2005.
- [59] K. Suzuki, D. C. Rockey, and A. H. Dachman, “CT colonography: advanced computer-aided detection scheme utilizing MTANNs for detection of ‘missed’ polyps in a multicenter clinical trial,” *Medical Physics*, vol. 37, no. 1, pp. 12–21, 2010.
- [60] K. Suzuki, H. Yoshida, J. Nappi, S. G. Armato III, and A. H. Dachman, “Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography,” *Medical Physics*, vol. 35, no. 2, pp. 694–703, 2008.
- [61] K. Suzuki, J. Zhang, and J. Xu, “Massive-training artificial neural network coupled with laplacian-eigenfunction-based dimensionality reduction for computer-aided detection of polyps in CT colonography,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, Article ID 5491180, pp. 1907–1917, 2010.
- [62] J. W. Xu and K. Suzuki, “Massive-training support vector regression and Gaussian process for false-positive reduction in computer-aided detection of polyps in CT colonography,” *Medical Physics*, vol. 38, no. 4, pp. 1888–1902, 2011.
- [63] F. Li, H. Arimura, K. Suzuki et al., “Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization,” *Radiology*, vol. 237, no. 2, pp. 684–690, 2005.
- [64] M. Loog and B. van Ginneken, “Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 5, pp. 602–611, 2006.
- [65] M. Loog, B. van Ginneken, and A. M. R. Schilham, “Filter learning: application to suppression of bony structures from chest radiographs,” *Medical Image Analysis*, vol. 10, no. 6, pp. 826–840, 2006.
- [66] M. K. Ozkan, I. M. I. Sezan, and A. M. Tekalp, “Adaptive motion-compensated filtering of noisy image sequences,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 4, pp. 277–290, 1993.
- [67] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [68] D. Marr and E. Hildreth, “Theory of edge detection,” *Proceedings of the Royal Society of London—Biological Sciences B*, vol. 207, no. 1167, pp. 187–217, 1980.
- [69] M. H. Hueckel, “An operator which locates edges in digitized pictures,” *Journal of Alternative and Complementary Medicine*, vol. 18, no. 1, pp. 113–125, 1971.
- [70] K. Suzuki, H. Abe, F. Li, and K. Doi, “Suppression of the contrast of ribs in chest radiographs by means of massive training artificial neural network,” in *Proceedings of the SPIE Medical Imaging (SPIE MI '04)*, pp. 1109–1119, San Diego, Calif, USA, February 2004.
- [71] A. R. Barron, “Universal approximation bounds for superpositions of a sigmoidal function,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930–945, 1993.

- [72] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [73] K. Suzuki, "Determining the receptive field of a neural filter," *Journal of Neural Engineering*, vol. 1, no. 4, pp. 228–237, 2004.
- [74] K. Suzuki, I. Horiba, and N. Sugie, "A simple neural network pruning algorithm with application to filter synthesis," *Neural Processing Letters*, vol. 13, no. 1, pp. 43–53, 2001.
- [75] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, and W. Hubbard, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [76] S. Deutsch, "A simplified version of Kunihiko Fukushima's neocognitron," *Biological Cybernetics*, vol. 42, no. 1, pp. 17–21, 1981.
- [77] K. Fukushima, "Neocognitron capable of incremental learning," *Neural Networks*, vol. 17, no. 1, pp. 37–46, 2004.
- [78] K. Fukushima, "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.
- [79] M. N. Gurcan, B. Sahiner, H. P. Chan, L. Hadjiiski, and N. Petrick, "Selection of an optimal neural network architecture for computer-aided detection of microcalcifications—comparison of automated optimization techniques," *Medical Physics*, vol. 28, no. 9, pp. 1937–1948, 2001.
- [80] H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network," *Medical Physics*, vol. 22, no. 10, pp. 1555–1567, 1995.
- [81] A. Hasegawa, K. Itoh, and Y. Ichioka, "Generalization of shift invariant neural networks: image processing of corneal endothelium," *Neural Networks*, vol. 9, no. 2, pp. 345–356, 1996.
- [82] C. M. Bishop, "An example—character recognition," in *Neural Networks for Pattern Recognition*, C. M. Bishop, Ed., pp. 1–4, Oxford University Press, New York, NY, USA, 1995.
- [83] D. F. Michaels, "Internal organization of classifier networks trained by backpropagation," in *Neural Networks in Vision and Pattern Recognition*, J. Skrzypek and W. Karplus, Eds., World Scientific, Singapore, 1992.
- [84] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1995.
- [85] I. Pitas, "Edge detection," in *Digital Image Processing Algorithms and Applications*, pp. 242–249, Wiley-Interscience, New York, NY, USA, 2000.
- [86] C. J. Murray and A. D. Lopez, "Mortality by cause for eight regions of the world: global burden of disease study," *The Lancet*, vol. 349, no. 9061, pp. 1269–1276, 1997.
- [87] G. E. Goodman, "Lung cancer. 1: prevention of lung cancer," *Thorax*, vol. 57, no. 11, pp. 994–999, 2002.
- [88] J. H. Austin, B. M. Romney, and L. S. Goldsmith, "Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesion evident in retrospect," *Radiology*, vol. 182, no. 1, pp. 115–122, 1992.
- [89] P. K. Shah, J. H. Austin, C. S. White et al., "Missed non-small cell lung cancer: radiographic findings of potentially resectable lesions evident only in retrospect," *Radiology*, vol. 226, no. 1, pp. 235–241, 2003.
- [90] R. Glocker and W. Frohnmayer, "Über die röntgenspektroskopische bestimmung des gewichtsanteiles eines elementes in gemengen und verbindungen," *Annalen der Physik*, vol. 76, pp. 369–395, 1925.
- [91] B. Jacobson and R. S. Mackay, "Radiological contrast enhancing methods," *Advances in Biological and Medical Physics*, vol. 6, pp. 201–261, 1958.
- [92] T. Ishigaki, S. Sakuma, and Y. Horikawa, "One-shot dual-energy subtraction imaging," *Radiology*, vol. 161, no. 1, pp. 271–273, 1986.
- [93] T. Ishigaki, S. Sakuma, and M. Ikeda, "One-shot dual-energy subtraction chest imaging with computed radiography: clinical evaluation of film images," *Radiology*, vol. 168, no. 1, pp. 67–72, 1988.
- [94] B. K. Stewart and H. K. Huang, "Single-exposure dual-energy computed radiography," *Medical Physics*, vol. 17, no. 5, pp. 866–875, 1990.
- [95] D. L. Ergun, C. A. Mistretta, D. E. Brown et al., "Single-exposure dual-energy computed radiography: improved detection and processing," *Radiology*, vol. 174, no. 1, pp. 243–249, 1990.
- [96] G. J. Whitman, L. T. Niklason, M. Pandit et al., "Dual-energy digital subtraction chest radiography: technical considerations," *Current Problems in Diagnostic Radiology*, vol. 31, no. 2, pp. 48–62, 2002.
- [97] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [98] M. L. Giger, "Update on the potential role of CAD in radiologic interpretations: are we making progress?" *Academic Radiology*, vol. 12, no. 6, pp. 669–670, 2005.
- [99] A. F. Frangi, W. J. Niessen, R. M. Hoogeveen, T. van Walsum, and M. A. Viergever, "Model-based quantitation of 3-D magnetic resonance angiographic images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 946–956, 1999.
- [100] L. He, Y. Chao, K. Suzuki, and K. Wu, "Fast connected-component labeling," *Pattern Recognition*, vol. 42, no. 9, pp. 1977–1987, 2009.
- [101] L. He, Y. Chao, and K. Suzuki, "A run-based two-scan labeling algorithm," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 749–756, 2008.
- [102] K. Suzuki, I. Horiba, and N. Sugie, "Linear-time connected-component labeling based on sequential local operations," *Computer Vision and Image Understanding*, vol. 89, no. 1, pp. 1–23, 2003.
- [103] L. Boroczky, L. Zhao, and K. P. Lee, "Feature subset selection for improving the performance of false positive reduction in lung nodule CAD," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 504–511, 2006.
- [104] A. S. Roy, S. G. Armato III, A. Wilson, and K. Drukker, "Automated detection of lung nodules in CT scans: false-positive reduction with the radial-gradient index," *Medical Physics*, vol. 33, no. 4, pp. 1133–1140, 2006.
- [105] Z. Hongbin, L. Zhengrong, J. P. Perry et al., "Increasing computer-aided detection specificity by projection features for CT colonography," *Medical Physics*, vol. 37, no. 4, pp. 1468–1481, 2010.
- [106] G. Lordanescu and R. M. Summers, "Reduction of false positives on the rectal tube in computer-aided detection for CT colonography," *Medical Physics*, vol. 31, no. 10, pp. 2855–2862, 2004.
- [107] J. Yao, J. Li, and R. M. Summers, "Employing topographical height map in colonic polyp measurement and false positive reduction," *Pattern Recognition*, vol. 42, no. 6, pp. 1029–1040, 2009.

- [108] P. C. Bunch, J. F. Hamilton, G. K. Sanderson, and A. H. Simmons, "A free-response approach to the measurement and characterization of radiographic-observer performance," *Journal of Applied Photographic Engineering*, vol. 4, no. 4, pp. 166–171, 1978.

## Research Article

# A Bayesian Hyperparameter Inference for Radon-Transformed Image Reconstruction

Hayaru Shouno,<sup>1</sup> Madomi Yamasaki,<sup>1</sup> and Masato Okada<sup>2,3</sup>

<sup>1</sup>Department of Informatics, Graduate School of Informatics and Engineering, University of Electro-Communications, Chofugaoka 1-5-1, Chofu, Tokyo 182-8585, Japan

<sup>2</sup>Division of Transdisciplinary Science, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan

<sup>3</sup>Okanoya Emotional Information Project, RIKEN Brain Science Institute (BSI), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Correspondence should be addressed to Hayaru Shouno, shouno@ice.uec.ac.jp

Received 24 February 2011; Revised 3 August 2011; Accepted 18 August 2011

Academic Editor: Dinggang Shen

Copyright © 2011 Hayaru Shouno et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We develop a hyperparameter inference method for image reconstruction from Radon transform which often appears in the computed tomography, in the manner of Bayesian inference. Hyperparameters are often introduced in Bayesian inference to control the strength ratio between prior information and the fidelity to the observation. Since the quality of the reconstructed image is controlled by the estimation accuracy of these hyperparameters, we apply Bayesian inference into the filtered back-projection (FBP) reconstruction method with hyperparameters inference and demonstrate that the estimated hyperparameters can adapt to the noise level in the observation automatically. In the computer simulation, at first, we show that our algorithm works well in the model framework environment, that is, observation noise is an additive white Gaussian noise case. Then, we also show that our algorithm works well in the more realistic environment, that is, observation noise is Poissonian noise case. After that, we demonstrate an application for the real chest CT image reconstruction under the Gaussian and Poissonian observation noises.

## 1. Introduction

In the field of medical imaging and noninvasive measurement, computed tomography (CT) plays an important role in diagnosis. The tomography image is reconstructed from a series of projection data, which are transmitted signals throughout an object, such as X-rays, in multiple directions. A lot of algorithms have been proposed to reconstruct tomography images [1–4]. Radon transform is usually used in mathematical formulations to describe the generating process of the observation data, and inverse of the Radon transform is considered as one of the frameworks for the image reconstruction from observation data; unfortunately, this reconstruction formulation does not care about noisy observations.

In order to improve image quality occurred by noisy observation, several image restoration methods based on the

Bayesian inference are discussed in the field of image processing [5, 6]. The purpose of image restoration lends itself naturally to the Bayesian formulation, which infers a posterior probability for the original image using the prior probability of an assumed model for the original image and the corruption process. One well-known strategy for Bayesian image restoration is to adopt the image that maximizes the posterior probability; this is called the maximum a posteriori (posterior) probability (MAP) inference. In MAP inference, the quality of a restoration image is controlled by the strengths ratio between fidelity of the observation process and the prior strength of the model. Hyperparameters are often introduced to describe these strengths of the ratio; however, these hyperparameters inference is a hard problem in the MAP framework. In order to estimate hyperparameters in the MAP framework, the cross-validation method is considered as effective; however, we consider that there exists several

problems. The first point is computational cost. In general, the cross-validation requires high computational cost. And the second point is to determine the cost function for the hyperparameters. In the field of image restoration, several types of methods are compared [7]; however, it is difficult to choose a cost function that is suitable for our problem.

In contrast, from the viewpoint of the Bayesian inference, the hyperparameter inference problem can be expressed naturally. For example, in the field of the image restoration, Molina et al. demonstrated several hyperparameter inference methods in the Bayesian manner in the manner of a hierarchical Bayes inference [8]. Pryce and Bruce and MacKay et al. proposed marginal likelihood maximization to infer those hyperparameters, which is called “evidence” framework or type 2 marginal likelihood maximization [6, 9–13].

In typical conventional methods, which use MAP inference for the computed tomography, a cost function that consists of data-fitting terms and several smoothness constraints has been introduced, and a minimization of the cost function is carried out in order to obtain the reconstructed image from the noisy observation data. Unfortunately, there have been few discussions related to the inference of a proper ratio between the data fitting and the constraints within the MAP framework. On the contrary, from the Bayesian inference point of view, it is natural to discuss the hyperparameter inference for image restoration using an evidence framework [14–16].

In our previous work, we proposed a CT image reconstruction in the manner of Bayes inference with a hyperparameter inference method from the noisy Radon-transformed observation by the evidence framework [12, 13]. In the previous work [12], however, we only showed that the Bayesian inference framework works well in the specific environment, that is, we assumed the additive white Gaussian noise for the 2-dimensional object observation. Gaussian noise is one of the tractable models for a mathematical formulation; however, in the X-ray CT or positron emission tomography (PET) image observation, we should assume Poissonian noise for the observation. Thus, in this study, we show that our reconstruction algorithm also works well under the Poissonian noise as well as under the Gauss noise case. Considering the Poissonian noise case for the observation which is different from our assuming model, we show a kind of robustness of our reconstruction model.

Moreover, we apply our reconstruction model into the real CT image data. Shepp and Logan phantom, which is usually used for evaluation of CT/PET image reconstruction, is a simple model of the axial cross-section human body. The internal organ of human body is not so much simple, so we use a real CT image data for reconstruction.

## 2. Formulation

In order to explain our Bayesian inference method, we show the conventional CT reconstruction method using filtered backprojection (FBP) under the formulation of the Radon transform. After that, we introduce Bayesian inference into the reconstruction process.

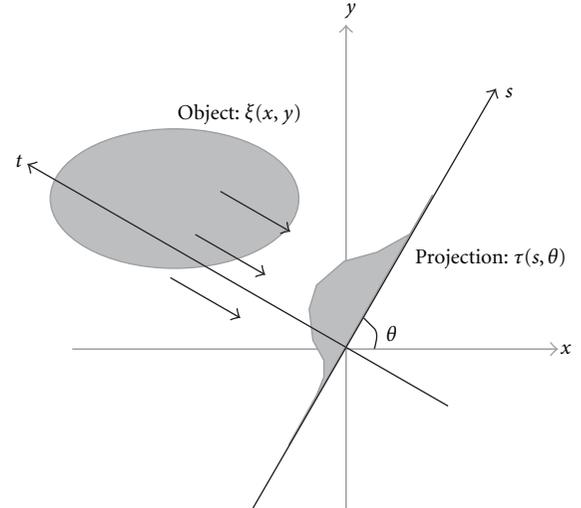


FIGURE 1: Schematic diagram of the Radon transform. Detectors are aligned on the  $s$  axis, which has an angle described as  $\theta$ .

**2.1. Radon Transform.** Briefly, the Radon transform assumes that the observed signals are transmitted through the target object. Figure 1 shows the schematic diagram of the Radon transform. We describe the target object density as the function of the  $(x, y)$  coordinate and assume that the detectors are aligned along the  $s$  axis that is rotated in  $\theta$ . We can thus denote the relationship between the  $(x, y)$  and  $(s, t)$  coordinates as a rotation

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (1)$$

We describe the density of the target as  $\xi(x, y, z)$ , that is,  $\xi(x, y, z)$  represents the absorption coefficients of the X-ray in the case of X-ray CT observation. The detectors are aligned on the  $s$  axis, so we describe the observation  $\tau(s, \theta, z)$  as the following formulation, called Radon transform:

$$\tau(s, \theta, z) = \int dt \xi(x, y, z) = \int dt \xi(x(s, t), y(s, t), z). \quad (2)$$

**2.2. FBP Reconstruction.** Before introducing the Bayes inference, we formulate the conventional filtered backprojection (FBP) method. This reconstruction method is mainly formulated on the frequency domain, so we introduce the 2-dimensional Fourier transform of the reconstruction image  $\sigma(x, y)$  and its inverse transform pair as

$$\tilde{\sigma}(\tilde{x}, \tilde{y}) = \iiint dx dy \sigma(x, y) e^{-2\pi j(x\tilde{x} + y\tilde{y})}, \quad (3)$$

$$\sigma(x, y) = \iint d\tilde{x} d\tilde{y} \tilde{\sigma}(\tilde{x}, \tilde{y}) e^{2\pi j(x\tilde{x} + y\tilde{y})}, \quad (4)$$

where the  $(\tilde{x}, \tilde{y})$  represents the frequency space coordinate.

Meanwhile, we can apply a 1-dimensional Fourier transform for the  $s$  of the observed data  $\tau(s, \theta)$  as  $\tilde{\tau}(\tilde{s}, \theta)$ . The

$\tilde{\tau}(\tilde{s}, \theta)$  satisfies the following relationship, which is called a projection theorem:

$$\tilde{\tau}(\tilde{s}, \theta, \tilde{z}) = \tilde{\xi}(\tilde{s} \cos \theta, \tilde{s} \sin \theta, \tilde{z}). \quad (5)$$

The FBP method is derived as a coordinate transformation from Cartesian coordinate  $(\tilde{x}, \tilde{y})$  into the polar coordinate  $(\tilde{s}, \theta)$  in the inverse Fourier transform (4)

$$\sigma(x, y, z) = \int_0^\pi d\theta \int_{-\infty}^\infty d\tilde{s} |\tilde{s}| \tilde{\sigma}(\tilde{s} \cos \theta, \tilde{s} \sin \theta) e^{2\pi j s \tilde{s}} \quad (6)$$

$$= \int_0^\pi d\theta g(s, \theta), \quad (7)$$

where

$$g(s, \theta) = \int d\tilde{s} |\tilde{s}| \tilde{\tau}(\tilde{s}, \theta) e^{2\pi j s \tilde{s}}, \quad (8)$$

since we can assume that the reconstruction image  $\sigma(x, y)$  should be identical to the original image  $\xi(x, y)$  without the observation noise, and we can apply the projection theorem in (5).

Thus, the reconstructed image  $\sigma(x, y)$  can be obtained by substituting the coordinate relationship  $s = x \cos \theta + y \sin \theta$ , that is derived from the rotation coordinate in (1) into (7). We call this reconstruction method the FBP method [1, 2].

**2.3. Stochastic Model.** In this section, we introduce a stochastic observation process into the FBP method. Of course, it is natural to consider Poissonian noise for observation in a realistic model; however, introducing Poissonian process makes it hard to solve the reconstruction in analytic form. We consider that a solvable model is important for understanding the reconstruction process. So in our theoretical framework, we introduced additive white Gaussian noise for observation on the signal  $\xi(x, y)$ . When we consider the Gaussian noise  $n_p(x, y)$  on the image  $\xi(x, y)$ , the observation through the Radon transform  $\tau(s, \theta)$  can be described as

$$\tau(s, \theta) = \int dt (\sigma(x, y) + n_p(x, y)) = \int dt \sigma(x, y) + N_p(s, \theta), \quad (9)$$

where  $N_p(s, \theta) = \int dt n_p(x, y)$ , and we also treat it as Gaussian noise. In the manner of the conventional image restoration method proposed by Tanaka and Inoue, we also introduce the energy function  $H_n(\tau | \sigma)$  as follows [14, 16]:

$$H_n(\tau | \sigma) = 4\pi^2 \int_0^\pi d\theta \int ds \left( \tau(s, \theta) - \int dt \sigma(x, y) \right)^2. \quad (10)$$

The important point of (10) is that the energy function  $H_n(\tau | \sigma)$  is defined as a kind of quadrature form of the difference between observation  $\tau(s, \theta)$  and the Radon transform of the reconstruction image  $\int dt \sigma(x, y)$ . We can thus denote the observation process as

$$p(\tau | \sigma) = \frac{1}{Z_n(\gamma)} \exp(-\gamma H_n(\tau | \sigma)), \quad (11)$$

$$Z_n(\gamma) = \sum_\tau \exp(-\gamma H_n(\tau | \sigma)), \quad (12)$$

where  $Z_n(\gamma)$  is to normalize a factor called the partition function. The hyperparameter  $\gamma$  represents a precision parameter that is proportionate to the inverse of the variance of the Gaussian noise  $N_p(s, \theta)$ , that is, the large  $\gamma$  indicates a good S/N ratio in the observation. Moreover, introducing both a polar coordinate for the frequency domain and Planchrel's theorem, we can drive the following expression:

$$p(\tau | \sigma) = \frac{1}{Z_n(\gamma)} \exp\left(-4\pi^2 \gamma \int d\theta \int d\tilde{s} |\tilde{\tau}_{\tilde{s}, \theta} - \tilde{\sigma}_{\tilde{s}, \theta}|^2\right), \quad (13)$$

where  $\tilde{\tau}_{\tilde{s}, \theta} = \tilde{\tau}(\tilde{s}, \theta)$  and  $\tilde{\sigma}_{\tilde{s}, \theta} = \tilde{\sigma}(\tilde{s} \cos \theta, \tilde{s} \sin \theta)$ . In the following formulation, we adopt these expressions for the polar coordinate in the frequency domain description for the sake of convenience.

To reconstruct an image from noisy data, using Bayes inference, we also denote the prior distribution. At first, we introduce the following energy function  $H_{\text{MRF}}(\sigma)$  for smoothness of the image:

$$H_{\text{MRF}}(\sigma) = \iint dx dy \|\nabla \sigma(x, y)\|^2, \quad (14)$$

where  $\nabla$  means gradient operator  $\nabla = (\partial/\partial x, \partial/\partial y)$ . This energy plays a role in the Markov random field (MRF) like a constraint since the gradient operation in the discretized space can be regarded as the difference between the neighboring pixels. So, this constraint controls neighboring pixel values to become similar to the target pixel. Then, we also introduce the following energy constraint to avoid taking large absolute pixel values:

$$H_{L_2}(\sigma) = \iint dx dy \|\sigma(x, y)\|^2, \quad (15)$$

which are sometimes called "L<sub>2</sub> constraint." Hence, we treat the prior as Gibbs-Boltzmann distribution of the linear combination of energies  $H_{\text{MRF}}(\sigma)$  and  $H_{L_2}(\sigma)$

$$p(\sigma) = \frac{1}{Z_{\text{pri}}(\beta, h)} \exp(-\beta H_{\text{MRF}}(\sigma) - 4\pi^2 h H_{L_2}(\sigma)), \quad (16)$$

$$Z_{\text{pri}}(\beta, h) = \sum_\sigma \exp(-\beta H_{\text{MRF}}(\sigma) - 4\pi^2 h H_{L_2}(\sigma)). \quad (17)$$

The hyperparameters  $\beta$  and  $h$  control the strength of each constraint. The prior probability can thus be described as follows when we adopt the polar coordinate in the frequency domain:

$$p(\sigma) = \frac{1}{Z_{\text{pri}}(\beta, h)} \exp\left(-4\pi^2 \int d\theta \int d\tilde{s} (\beta \tilde{s}^2 + h) |\tilde{s}| |\tilde{\sigma}_{\tilde{s}, \theta}|^2\right). \quad (18)$$

From (13) and (18), we can derive the posterior probability with Bayes theorem  $p(\tau | \sigma) = p(\tau | \sigma) p(\sigma) / \sum_\sigma p(\tau | \sigma) p(\sigma)$ . Then, we can describe the posterior as

$$p(\tau | \sigma) \propto \exp\left(-4\pi^2 \int_0^\pi d\theta \int d\tilde{s} F_{\tilde{s}} \left| \tilde{\sigma}_{\tilde{s}, \theta} - \frac{\gamma}{F_{\tilde{s}}} \tilde{\tau}_{\tilde{s}, \theta} \right|^2\right), \quad (19)$$

where  $F_{\tilde{s}} = (\beta \tilde{s}^2 + h) |\tilde{s}| + \gamma$ .

In order to calculate the denominator value called partition function, we discretize the integral description in the partition function over polar coordinate in frequency domain. When we denote the sampling width for radial direction and polar angle as  $\Delta_{\tilde{s}}$  and  $\Delta_{\theta}$ , respectively, the discretized sampling point  $(\tilde{s}_k, \theta_l)$  can be described as  $\tilde{s}_k = \tilde{k}\Delta_{\tilde{s}}$  and  $\theta_l = l\Delta_{\theta}$ , respectively, where  $\tilde{k}$  and  $l$  represent the indexes of the radial direction and the polar angle. The angle  $\theta_l$  corresponds to the detectors array angle in the observation. Therefore, we assume that the observation is carried out  $N_{\theta}$  times in the angle  $[0, \pi]$ , that is,  $\Delta_{\theta} = \pi/N_{\theta}$ . The coordinate value  $\tilde{s}_k$  represents the position in the radial direction, which means the spatial frequency described in the Fourier transform. From the Nyquist frequency, we can denote  $\Delta_{\tilde{s}} = 1/N_s\Delta_s$ , where  $\Delta_s$  is an interspace of the detectors in the array. We assume the length of detectors array as  $L$ , and  $N_s$  detectors are assigned with the same interspace in the array, so  $\Delta_s = L/N_s$ .

When we discretize the integral  $\int d\tilde{s}$  in the posterior as  $\sum_{\tilde{k}}^{N_s-1} \Delta_{\tilde{s}}$ , we can derive the marginalized posterior probability as a Gaussian distribution

$$p(\tilde{\sigma}_{k,l} | \tau) = \mathcal{N}\left(\tilde{\sigma}_{k,l} | \frac{\gamma}{F_k} \tilde{\tau}_{k,l}, \frac{N_s}{8\pi^2 \Delta_{\theta} \Delta_s F_k}\right), \quad (20)$$

where the descriptions  $\tilde{\sigma}_{k,l}$ ,  $\tilde{\tau}_{k,l}$ , and  $F_k$  represent  $\tilde{\sigma}_{k,l} = \tilde{\sigma}(\tilde{s}_k \cos \theta_l, \tilde{s}_k \sin \theta_l)$ ,  $\tilde{\tau}_{k,l} = \tilde{\tau}(\tilde{s}_k, \theta_l)$ , and  $F_k = F_{\tilde{s}_k} = (\beta \tilde{s}_k^2 + h) |\tilde{s}_k| + \gamma$ , respectively.

**2.4. Image Reconstruction.** We adopt the marginalized posterior mean  $\langle \sigma(x, y) \rangle$  for the image reconstruction solution. The posterior mean can be denoted as

$$\langle \sigma(x, y) \rangle = \int_0^{\pi} d\theta \int_{-\infty}^{\infty} d\tilde{s} |\tilde{s}| \langle \tilde{\sigma}_{\tilde{s}, \theta} \rangle e^{2\pi j \tilde{s}(x \cos \theta + y \sin \theta)}. \quad (21)$$

Thus,  $\{\langle \tilde{\sigma}_{\tilde{s}, \theta} \rangle\}$ , which represents an average set of Fourier expressions, is required to obtain the mean pixel value over the posterior  $\langle \sigma(x, y) \rangle$ . We can evaluate  $\langle \tilde{\sigma}_{\tilde{s}, \theta} \rangle$  by discretizing the coordinate as in the previous section, thereby obtaining

$$\langle \tilde{\sigma}_{k,l} \rangle = \frac{\gamma}{F_k} \tilde{\tau}_{k,l}. \quad (22)$$

This solution, called the posterior mean (PM) solution, provides identical result as the MAP does, that is, energy function  $H_n(\sigma)$  minimization with the constraint of the smoothness of  $H_{\text{MRF}}(\sigma)$  and  $H_{L2}(\sigma)$ ,

$$\begin{aligned} \sigma_{\text{MAP}} &= \operatorname{argmax}_{\sigma} \ln p(\tau | \sigma) p(\sigma) \\ &= \operatorname{argmin}_{\sigma} (4\pi^2 \gamma H_n(\tau | \sigma) \\ &\quad + \beta H_{\text{MRF}}(\sigma) + 4\pi^2 h H_{L2}(\sigma)). \end{aligned} \quad (23)$$

Of course, PM solution is not identical to MAP solution in general; however, in this case, the PM solution and the MAP solution are identical, because the posterior distribution is denoted as a Gaussian distribution.

**2.5. Hyperparameter Inference.** To reconstruct an appropriate tomography image with our Bayesian inference, we need to assign proper values to the hyperparameters  $\beta$ ,  $h$ , and  $\gamma$ . These hyperparameters  $\beta$  and  $h$  control the strength of constraints, while  $\gamma$  controls the fidelity of the observation. We infer these hyperparameters by using maximization of marginal log likelihood, which is sometimes called evidence framework [9–11]. The marginal log-likelihood is denoted as the linear combination of log partition functions,

$$\ln p(\tau | \beta, h, \gamma) = \ln Z_{\text{post}}(\beta, h, \gamma) - \ln Z_n(\gamma) - \ln Z_{\text{pri}}(\beta, h), \quad (24)$$

where  $Z_n(\gamma)$  is also denoted as (12),  $Z_{\text{pri}}(\beta, h)$  is denoted as (17), and, for the posterior, we introduce  $Z_{\text{post}}(\beta, h, \gamma)$ ;

$$\begin{aligned} Z_{\text{post}}(\beta, h, \gamma) &= \sum_{\sigma} \exp(-4\pi^2 \gamma H_n(\tau | \sigma) \\ &\quad - \beta H_{\text{MRF}}(\sigma) - 4\pi^2 h H_{L2}(\sigma)). \end{aligned} \quad (25)$$

We use discretization to evaluate each partition function and obtain

$$\begin{aligned} \ln Z_{\text{pri}}(\beta, h) &= -\frac{N_{\theta}}{2} \sum_{\tilde{k}=0}^{N_s-1} \ln(\beta \tilde{s}_k^2 + h), \\ \ln Z_n(\gamma) &= -\frac{N_{\theta} N_s}{2} \ln \gamma, \\ \ln Z_{\text{post}}(\beta, h, \gamma) &= -\frac{4\pi^2 \Delta_{\theta} \Delta_s}{N_s} \sum_{\tilde{k}=0}^{N_s-1} \gamma \left(1 - \frac{\gamma}{F_k}\right) \\ &\quad \times \left| \tilde{\tau}_{k,l} \right|^2 - \frac{N_{\theta}}{2} \sum_{\tilde{k}=0}^{N_s-1} \ln F_k. \end{aligned} \quad (26)$$

To maximize the marginal log likelihood (24), we adopt a naive gradient method corresponding to the hyperparameters  $\beta$ ,  $h$ , and  $\gamma$ , that is, we update hyperparameters using the following rule:

$$\begin{pmatrix} \ln \beta^{t+1} \\ \ln h^{t+1} \\ \ln \gamma^{t+1} \end{pmatrix} = \begin{pmatrix} \ln \beta^t \\ \ln h^t \\ \ln \gamma^t \end{pmatrix} + \eta \begin{pmatrix} \frac{\partial \ln p(\tau | \beta^t, h^t, \gamma^t)}{\partial \ln \beta} \beta^t \\ \frac{\partial \ln p(\tau | \beta^t, h^t, \gamma^t)}{\partial \ln h} h^t \\ \frac{\partial \ln p(\tau | \beta^t, h^t, \gamma^t)}{\partial \ln \gamma} \gamma^t \end{pmatrix}, \quad (27)$$

where  $\eta$  is a sufficiently small value. Those update rules (27) are denoted for  $\ln \beta$ ,  $\ln h$ , and  $\ln \gamma$ , since  $\beta$ ,  $h$ , and  $\gamma$  should be nonnegative values.

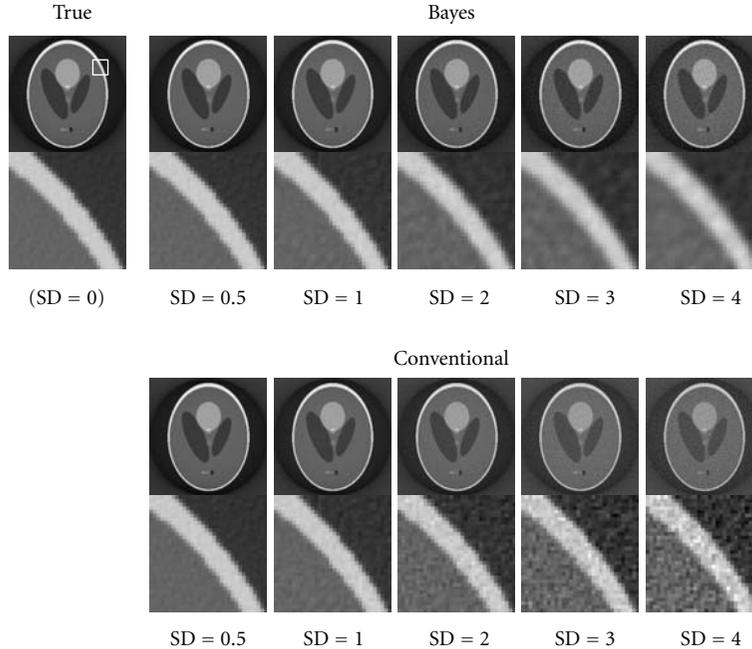


FIGURE 2: Comparison of the reconstructed tomography images derived using the Bayesian method and conventional FBP. The top row shows the Bayesian FBP methods, and the bottom one shows the conventional one. Each column corresponds to the strength of the observation Gaussian noise standard deviations. We show the magnification of a part of the reconstructed images around the edge of the phantom, whose location is indicated by white rectangle in the true image.

### 3. Evaluation by a Computer Simulation

**3.1. Phantom Image Reconstruction.** In the computer simulation, we created the Shepp and Logan phantom image in  $N_x \times N_y$  (pixels) and mapped the image into an origin-centered square with an edge length set to  $L$ , that is, the area is set to  $[-L/2, -L/2] \times [L/2, L/2]$ . In the square, the area, which takes distance from the origin larger than  $L/2$ , is sometimes unobservable by the detectors from several angles, and we therefore ignore this area during our evaluation. For each angle  $\theta_l$ , we assume the  $s$  axis as Figure 1, and the origin in the  $(x, y)$  coordinate projects to the point  $s = 0$  in any angle. We set the sampling parameters as  $N_x = N_y = N_\theta = N_s = 256$ , and the length of the detectors array as  $L = 1$ .

For hyperparameter inference, we adopt a gradient method that requires initial state of these parameters. In the following simulations, the initial state of  $\beta^{(0)}$ ,  $h^{(0)}$ , and  $\gamma^{(0)}$  is set to be 1. And the number of iterations is limited to the 10000 times.

**3.1.1. Gaussian Noise Case.** In order to evaluate the performance of the hyperparameter inference, we carry out the simulation in the additive white Gaussian noise environment at first. We assumed that the Gaussian noise  $n_p(x, y)$  was added during the observation process (see (9)) and controlled the noise standard deviation (SD) in the range of 0 to 6. A small SD means the low noise level in the

observation process, and the larger SD becomes, the higher additive Gaussian noise level becomes. On the other hand, the large SD observation makes a lot of information loss for reconstruction. The MRF like prior (16) plays a roll of compensation for the information loss. In the simulation, Gaussian noise value sometimes makes fluctuation to the result, so we evaluated the average performance over 10 trials.

The computational cost is mainly consumed by hyperparameters inference. In this study, we adopted gradient method for the hyperparameter inference, so the computational cost depends on the initial state of these hyperparameters and learning coefficients  $\eta$ . In typical cases, about 1000~2000 iterations are required to converge for the  $\eta = 10^{-6}$ . It takes 1~2 minutes for Intel Xeon E5530 2.40 GHz with 24 GiB memory.

Figure 2 shows typical results of the reconstruction images. The most left image shows the “true” which means a reconstruction image without any observation noise (SD = 0.0). The top part shows the result using our Bayesian inference with inferred hyperparameters, and the bottom one shows the result using the conventional FBP method [1]. Each column corresponds to the SD of the additive Gaussian noise  $n_p(x, y)$ . In Figure 2, we show magnification of each reconstructed image around the edge whose location is located as the white rectangle in the “true” image. The degradation of the image in the conventional FBP result when the noise SD is large is clearly visible, whereas the contrast of the image has been maintained in the Bayesian reconstruction result.

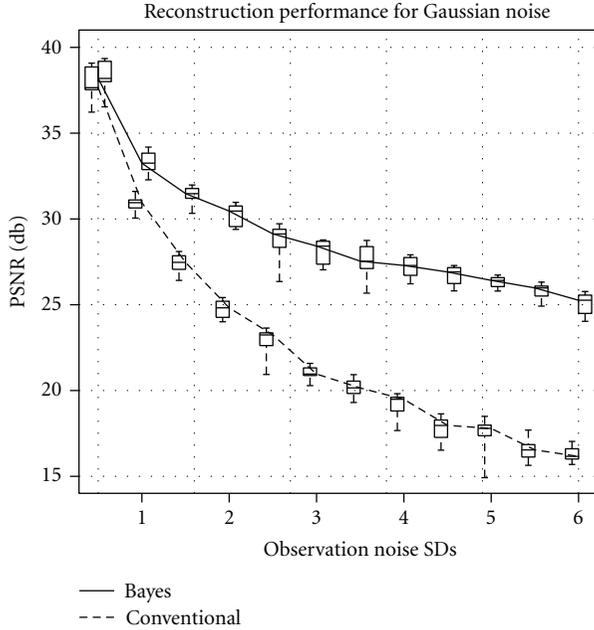


FIGURE 3: Qualities of reconstruction images measured by PSNR. The horizontal axis shows the SD of the Gaussian noise. The vertical axis shows the PSNR. The solid line shows the median of the 10 trials of our Bayesian inference results, and box plot shows quartile deviation. The dashed line shows the results of the conventional FBP method.

We used the peak signal-to-noise ratio (PSNR) to evaluate the quality of the reconstructed image. The result of this evaluation is shown in Figure 3. The horizontal axis indicates the SD of the  $n_p(x, y)$ , and the vertical shows the PSNR between the reconstructed image for both a noised and noiseless reconstruction images. The solid line shows the median of the Bayesian inference reconstruction results for 10 trials, and each box plot shows the quartiles deviations. The dashed one shows those of the conventional FBP results. The Bayesian inference maintained high reconstruction quality compared to the conventional FBP method. Even when the SD of the noise was 4.0, the PSNR value remained 27.5 (dB). On the other hand, the PSNR of the conventional FBP method was degraded and became 27.7 [dB] when the SD is only 1.5. This demonstrated that the Bayesian inference is more robust to the observation noise rather than the conventional FBP method.

Figure 4 shows the reconstruction performance against the hyperparameter  $\beta$ . The horizontal axis shows the value of the hyperparameter  $\beta$ , and the vertical one shows the PSNR. We fixed other hyperparameters,  $h$  and  $\gamma$ , to the estimated value. Each image in the figure shows the reconstruction result with corresponding hyperparameter setting. The hyperparameter  $\beta$  controls the smoothness of the image in the prior equation (16), so too much large  $\beta$  makes excessive blurring. Our hyperparameter inference algorithm, shown in the filled rectangle in the figure, looks to provide optimal value.

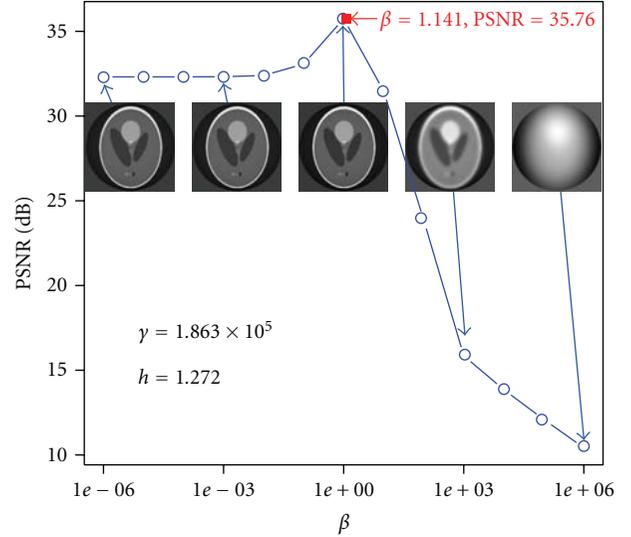


FIGURE 4: Reconstruction performance against hyperparameter  $\beta$ . The horizontal axis shows the  $\beta$ , and the vertical one shows the PSNR. Other hyperparameters,  $h$  and  $\gamma$ , are fixed to the estimated value. The filled rectangle shows the result of our hyperparameter inference for  $\beta$ .

3.1.2. *Poissonian Noise Case.* Gauss noise observation is the assumed model in our formulation equation (9); however, the CT/PET observation process is usually described as the Poissonian process. Thus, we should evaluate the reconstruction quality for the Poissonian noise case for the more realistic environment. Of course, our model is designed for the Gaussian noise case, so the performance of reconstruction for the Poissonian process observation might become worse; however, quantitative evaluation is important in the meaning of the approximation.

In the computer simulation, we used R PET package for Poissonian noise sampling [17]. The Poissonian noise value is generated by acceptance-rejection method [18, 19]. Hence, the number of the samplings determines the noise strength property corresponding to the SDs in the Gaussian case, that is, less number of the samplings make low signal-to-noise ratio. The computational cost is also consumed by hyperparameter inference, and it takes about 1000 times iterations for the  $\eta = 10^{-6}$ , that is, it requires  $\sim 1$  minute for the convergence in our computational environment.

Figure 5 shows the reconstructed image using our Bayesian method and conventional FBP method. The top part shows the result of our Bayesian reconstruction images, and the bottom one shows the conventional FBP result. The most left image shows also the “true” image that means a reconstructed image without any Poissonian noise. In other columns, we show the image with Poissonian noise whose strength is controlled by sampling levels, that is, the S/N ratio becomes worse when sampling level becomes low [20]. In the figure, the noise strength becomes large for the right direction.

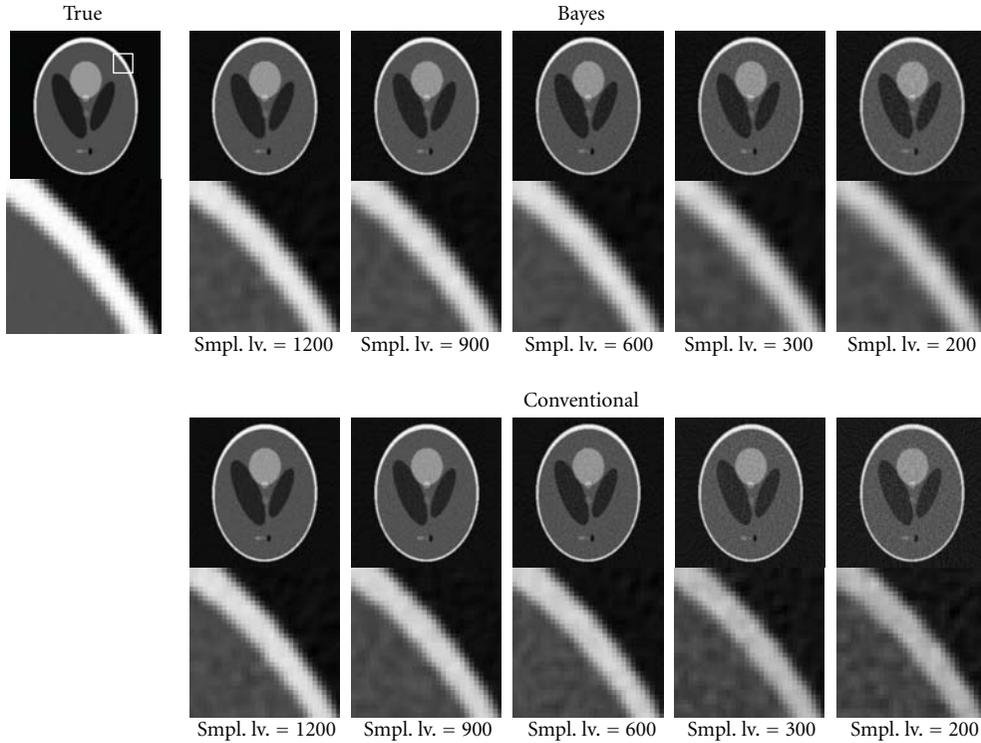


FIGURE 5: Comparison of the reconstructed tomography images derived using the Bayesian method and conventional FBP under the Poissonian noise. The top row shows the results of our method, and the bottom one shows the results of the conventional FBP method. Each column corresponds to the strength of the observation noise which can be denoted as the number of sampling in the acceptance rejection method. We show the magnification of a part of the reconstructed images around the edge of the phantom, whose location is indicated by a white rectangle in the true image.

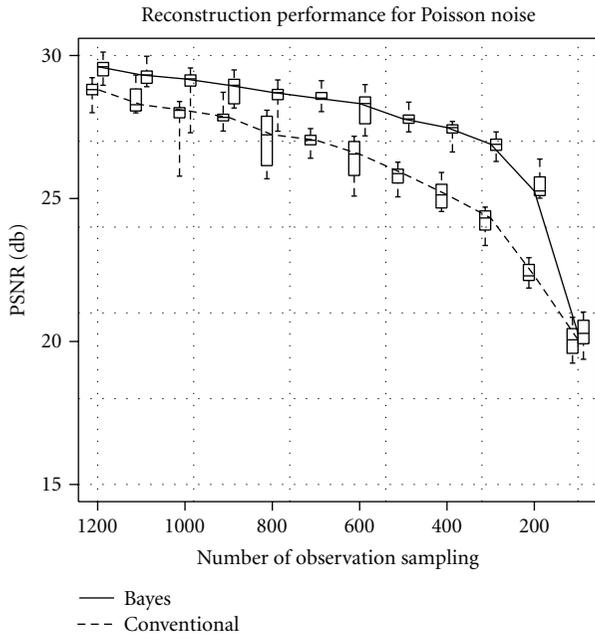


FIGURE 6: Qualities of reconstruction images measured by PSNR. The horizontal axis shows the SD of the sampling level, which means the inverse of the Poissonian noise. The vertical axis shows the PSNR. The solid line shows the median of the 10 trials of our Bayesian inference results, and box plot shows quartile deviation. The dashed line shows the results of the conventional FBP method.

Figure 6 shows the quantity evaluation result in the meaning of the PSNR against the sampling level of the observation. The horizontal axis shows the sampling level, and the vertical one shows the PSNR. The solid line shows the median of 10 trials for our Bayesian reconstruction method, and the box plots are quartiles for each sampling levels. The dashed one shows the result of the conventional FBP method. Roughly speaking, the Bayes reconstruction shows better result in the meaning of the PSNR.

3.2. *Real CT Image Reconstruction.* In order to evaluate the performance of our method for the CT/PET image, we applied our method to a real CT image reconstruction.

We prepare several real CT images provided by Tokushima University Hospital. The acquisition parameters of those HRCT images are as follows: Toshiba “Aquilion 16” is used for imaging device, and each slice image consists of  $512 \times 512$  pixels, and pixel size corresponds to  $0.546 \sim 0.826$  mm; slice thickness is 1 mm. Thus, we set the sampling parameters as  $N_x = N_y = N_\theta = N_s = 512$ .

In order to obtain noise-corrupted data  $\tau$ , we simulate Gaussian and Poissonian noised observation for these CT images in the same manner with phantom images.

Figure 7 shows a reconstruction result for the real chest CT image with Gaussian noise. The top row shows our Bayesian method, and the bottom one shows the conventional FBP results. Each column corresponds to the additive

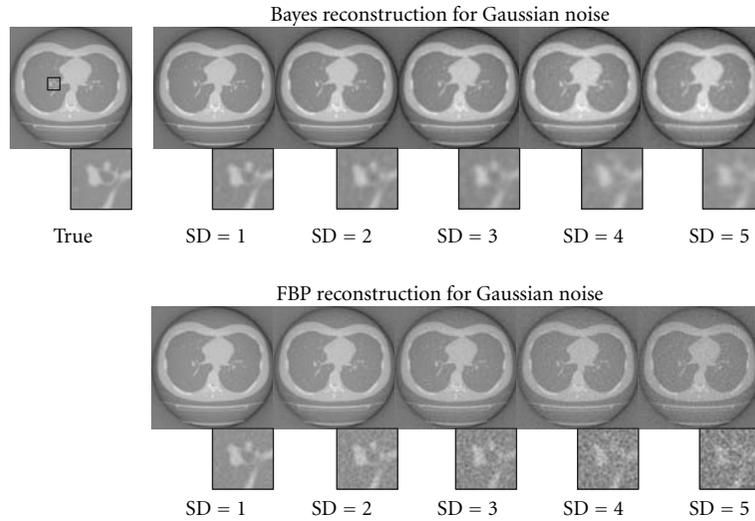


FIGURE 7: Comparison of the reconstructed images using real CT data with Gaussian noise between Bayes method and conventional FBP method. The top row shows the results of our method, and the bottom one shows the conventional FBP results. Each column corresponds to the strength of the observation noise that can be denoted as standard deviation (SD) of adding noise. We also show the magnification of a part around the bronchus, whose location is indicated by black rectangle in the true image.

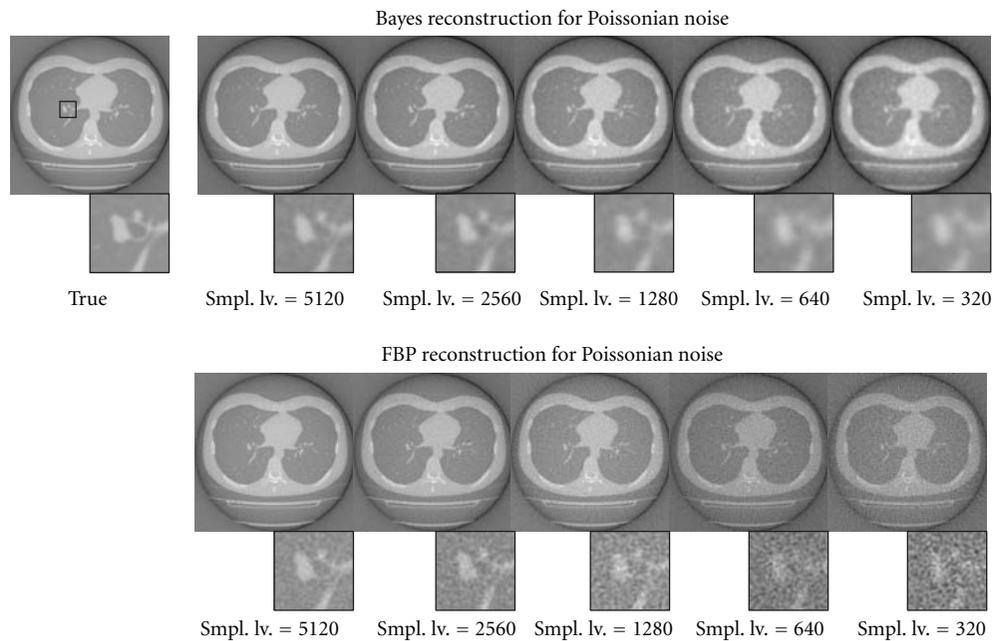


FIGURE 8: Comparison of the reconstructed images using real CT data with Poissonian noise between Bayes method and conventional FBP method. The top row shows the results of our method, and the bottom one shows the conventional FBP results. Each column corresponds to the strength of the observation noise that can be denoted as the number of sampling in the acceptance-rejection method. We also show the magnification of around bronchus indicated by black rectangle in the true image.

Gaussian noise strength for pseudo-observation. In each image, we show a magnification part around bronchus, whose location is described as a black rectangle in the true image. In the Bayesian reconstruction, our MRF prior makes a blurring effect for edge components on the image. The hyperparameter inference mechanism would try to compensate for the information loss, which is caused by the

observation noise, by use of the MRF prior. As a result, the large SD makes strong blurring effect to the image. In the magnification image of the Bayesian inference, the bronchus parts are hard to identify around  $SD > 4.0$ , however, vessels along the bronchus are able to identify for these SDs. In contrast, in the conventional FBP results, both of those parts are just difficult to identify for these SDs. Figure 8

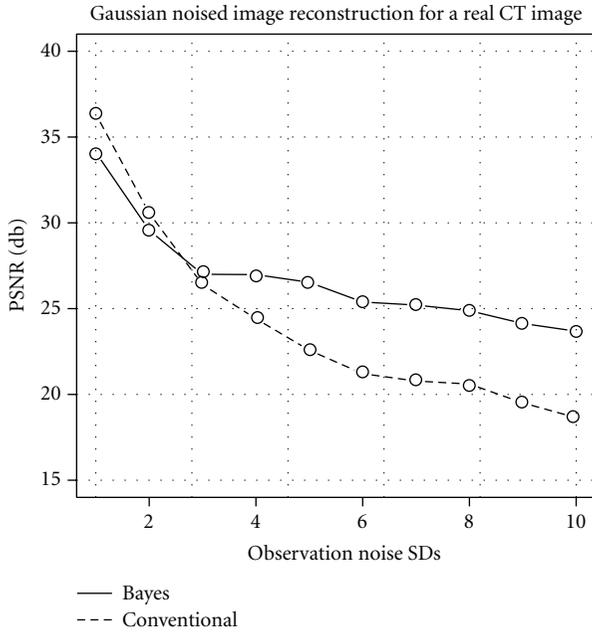


FIGURE 9: Qualities of reconstruction images measured by PSNR for real CT image reconstruction under the Gaussian noise. The horizontal axis shows the observation SD of Gaussian noise, and the vertical axis shows the PSNR. The solid line shows the result of our Bayesian method, and the dashed one shows that of the conventional FBP method.

also shows a Poissonian noise case for the chest CT image. We can see the similar tendency to the Gaussian case. In the magnification images, we can identify the bronchus over 1280 sampling levels. In contrast, low sampling level makes large blurring effect by the MRF prior. As a result, bronchus part is hard to identify at fewer than 640 sampling levels. However, the reconstruction result looks better than those of the conventional FBP method.

Moreover, we evaluate the quantitative reconstruction performance by PSNR for the real CT image. Figure 9 shows the result for the Gaussian noise case, and Figure 10 shows the one for the Poissonian case. Each horizontal axis means the noise strength control variable, and the vertical shows the reconstruction performance by PSNR. In both of these results, the Bayes reconstruction method shows better performances in the strong noise area. In contrast, in the weak noise area, the Bayes reconstruction result is just worse than that of the conventional method. We can see that the real CT image is more complex than the Shepp and Logan phantom image like Figure 5, and simple MRF like prior (16) prefers smooth image. Thus, in the weak noise area, complex shape in the real image makes overestimate for the prior strength  $\beta$ , which controls blurring effect by the prior. As a result, our Bayesian reconstruction method prefers too much smooth image in the weak noise area; however, the PSNR value stays around 30 (db) for the  $SD = 2$  in the Gaussian case and around 28 (db) for the 2560 samplings in the Poissonian case.

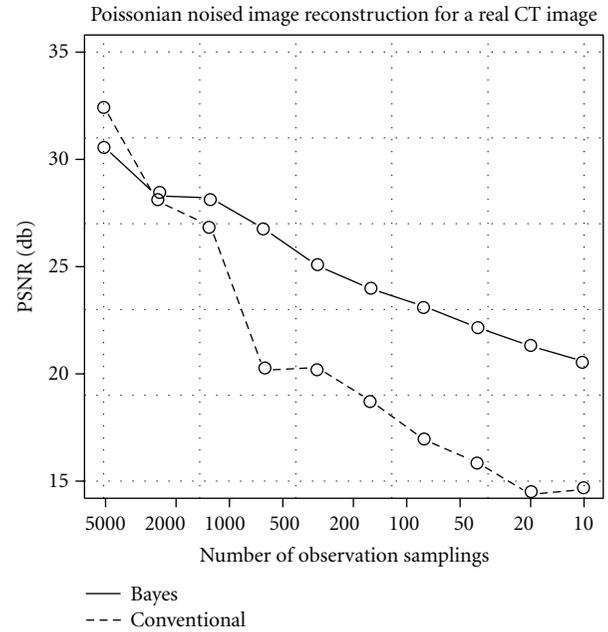


FIGURE 10: Qualities of reconstruction images measured by PSNR for real CT image reconstruction under the Poissonian noise. The horizontal axis shows the sampling level, and the vertical axis shows the PSNR. The solid line shows the result performance of Bayesian method, and the dashed line shows the results of the conventional FBP method.

## 4. Conclusion

We proposed a hyperparameter inference based on the Bayesian inference in order to reconstruct tomography image formulated by Radon transform. As a stochastic model, we introduced a simple MRF-like distribution  $p(\sigma)$  for the prior and formulated the observation process  $p(\tau | \sigma)$  by assuming the Gaussian noise channel.

We discretized the image signals in the frequency domain expressed by the polar coordinate in order to evaluate the posterior distribution analytically, resulting in the ability to conduct posterior mean for the reconstructed image. Using the marginal-likelihood maximization method, we show that the hyperparameters introduced as  $\beta$ ,  $h$ , and  $\gamma$ , which allows us to maintain a balance between observation fidelity and prior constraint, could be determined automatically. And using those hyperparameters, we could obtain a higher-quality reconstructed image than when using the conventional FBP method.

In order to evaluate the performance of our method, we simulated two observation noise cases, that is, Gaussian and Poissonian noises. We controlled noise strength by SD for Gaussian noise and sampling levels for Poissonian noise. In the phantom simulation for the Gaussian noise, we confirmed that our hyperparameter inference worked well against the PSNR, and the performance for the reconstruction was better than that of the conventional FBP. The computational cost for the hyperparameter inference depend on the initial state of them; however, about 1200~2000 times iterations made convergence to them for typical cases. In

the Poissonian cases, the tendency of the reconstruction performance is similar to the Gaussian case. Our Bayesian method made better performance than the conventional FBP in any noise strength area. However, in the strong Poissonian noise case, that is, the noise could not approximate well by Gaussian noise, we confirmed that the performance of the reconstruction was not good enough for diagnosing. Moreover, we evaluated the performance by a real chest CT image. The real image has a little complex shape against the phantom image. Thus, in the low-noise strength area for both noise cases, the prior components worked too much for the smoothness effect. As a result, the PSNR was just worse than the conventional FBP in such area. However, detail structure of the organ was easy to identify in the obtained image of our model.

In this study, we demonstrate applying our algorithm to the only 2-dimensional image reconstruction. We consider the algorithm easy to extend for 3-dimensional case. Thus, we would reformulate our algorithm for applying to the 3-dimensional image reconstruction and confirm the performance in the future work.

## Acknowledgments

This work is supported by Grant-in-Aids for Scientific Research (C) 21500214 and Innovative Areas 21103008, MEXT, Japan. The authors thank Professor Shoji Kido, Yamaguchi University. He provided a lot of advice about this study and made effort to prepare the CT images. They also thank Professor Junji Ueno, Tokushima University. He provided them with several high-resolution CT image for this study.

## References

- [1] G. N. Ramachandran and A. V. Lakshminarayanan, "Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, no. 9, pp. 2236–2240, 1971.
- [2] L. A. Shepp and B. F. Logan, "Reconstructing interior head tissue from x-ray transmissions," *IEEE Transactions on Nuclear Science*, vol. 21, pp. 228–236, 1974.
- [3] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.
- [4] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Transactions on Medical Imaging*, vol. 9, no. 1, pp. 84–93, 1990.
- [5] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [6] J. M. Pryce and A. D. Bruce, "Statistical mechanics of image restoration," *Journal of Physics A*, vol. 28, no. 3, article 009, pp. 511–532, 1995.
- [7] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.
- [8] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 231–246, 1999.
- [9] D. J. C. Mackay and Cavendish Laboratory, "Hyperparameters: optimize, or integrate out," in *In Maximum Entropy and Bayesian Methods*, pp. 43–60, Kluwer, Santa Barbara, Calif, USA, 1996.
- [10] D. J. C. Mackay, *Information Theory, Inference and Learning Algorithm*, Cambridge University Press, New York, NY, USA, 2003.
- [11] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [12] H. Shouno and M. Okada, "Bayesian image restoration for medical images using Radon transform," *Journal of the Physical Society of Japan*, vol. 79, no. 7, Article ID 074004, 2010.
- [13] H. Shouno and M. Okada, "A hyper-parameter inference for radon transformed image reconstruction using bayesian inference," in *Machine Learning in Medical Imaging*, F. Wang, P. Yan, K. Suzuki, and D. Shen, Eds., Lectures Notes in Computer Science, pp. 26–33, Springer, New York, NY, USA, 2010.
- [14] J. I. Inoue and K. Tanaka, "Dynamics of the maximum marginal likelihood hyperparameter estimation in image restoration: Gradient descent versus expectation and maximization algorithm," *Physical Review E*, vol. 65, no. 1, Article ID 016125, pp. 016125/1–016125/11, 2002.
- [15] K. Tanaka, "Statistical-mechanical approach to image processing," *Journal of Physics A*, vol. 35, no. 37, pp. R81–R150, 2002.
- [16] K. Tanaka, H. Shouno, M. Okada, and D. M. Titterton, "Accuracy of the Bethe approximation for hyperparameter estimation in probabilistic image processing," *Journal of Physics A*, vol. 37, no. 36, pp. 8675–8695, 2004.
- [17] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Mendeley, Vienna, Austria, 2009.
- [18] J. Gentle, *Elements of Computational Statistics*, Springer, New York, NY, USA, 2002.
- [19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition, The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 3rd edition, 2007.
- [20] J. Schulz, *Diploma thesis: analyse von PET daten unter ein-satz adaptiver Glaettungsverfahren*, Ph.D. thesis, Humboldt-Universitatet zu Berlin, Berlin, Germany, 2006.

## Research Article

# Conditional Random Fields and Supervised Learning in Automated Skin Lesion Diagnosis

**Paul Wighton,<sup>1,2,3</sup> Tim K. Lee,<sup>1,2,3</sup> Greg Mori,<sup>1</sup> Harvey Lui,<sup>2,3</sup> David I. McLean,<sup>2</sup> and M. Stella Atkins<sup>1,2</sup>**

<sup>1</sup>Department of Computing Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

<sup>2</sup>Department of Dermatology and Skin Science, Photomedicine Institute, University of British Columbia and Vancouver Coastal Health Research Institute, Vancouver, BC, Canada V5Z 4E8

<sup>3</sup>Cancer Control Research Program and Integrative Oncology Department, BC Cancer Research Centre, Vancouver, BC, Canada V5Z 4E6

Correspondence should be addressed to Paul Wighton, paul.wighton@gmail.com

Received 2 March 2011; Revised 21 May 2011; Accepted 23 July 2011

Academic Editor: Fei Wang

Copyright © 2011 Paul Wighton et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many subproblems in automated skin lesion diagnosis (ASLD) can be unified under a single generalization of assigning a label, from an predefined set, to each pixel in an image. We first formalize this generalization and then present two probabilistic models capable of solving it. The first model is based on independent pixel labeling using maximum a-posteriori (MAP) estimation. The second model is based on conditional random fields (CRFs), where dependencies between pixels are defined using a graph structure. Furthermore, we demonstrate how supervised learning and an appropriate training set can be used to automatically determine all model parameters. We evaluate both models' ability to segment a challenging dataset consisting of 116 images and compare our results to 5 previously published methods.

## 1. Introduction

Incidence rates of melanoma are increasing rapidly in the western world, growing faster than any other cancer [1]. Since there is no effective therapy for patients with advanced melanoma [2], educational campaigns attempt to encourage high-risk individuals to undergo routine screening so that melanomas can be identified early while they are still easily treatable [3]. While worthwhile, these educational campaigns generate a large amount of referrals to dermatologists, whose services are already undersupplied [4].

Automated skin lesion diagnosis (ASLD) is expected to alleviate some of this burden. By acting as a screening tool, ASLD can reject obviously benign lesions, while referring more suspicious ones to an expert for further scrutiny. Most ASLD methods adopt the standard computer-aided diagnosis (CAD) pipeline illustrated in Figure 1. First an image is acquired with a digital dermoscope. Next, undesirable artifacts (such as hair or oil bubbles) are identified and, if necessary, replaced with an estimate of the underlying skin

color. After this, the lesion is segmented, and discriminative features are then extracted. Finally, supervised learning is used to classify previously unseen images.

Our previous work demonstrated how the use of supervised learning, under the proper generalization (of assigning labels to pixels), was able to solve several tasks in this pipeline including detecting occluding hair, segmenting the lesions, and detecting the dermoscopic structure *pigment network* [5]. Our method was relatively simple; it labeled pixels in an image independently using modest features, linear discriminant analysis (LDA) for supervised dimensionality reduction, and maximum a-posteriori (MAP) estimation. Nevertheless, in spite of its simplicity, our model was able to perform comparably to other previously published, nongeneral methods for lesion segmentation [6–10] and hair detection [11].

In this paper, we seek to expand on this generalization by replacing the per-pixel (PP) estimation model with a conditional random field (CRF) model. The largest criticism levied at the PP approach is that pixels are labeled independently,

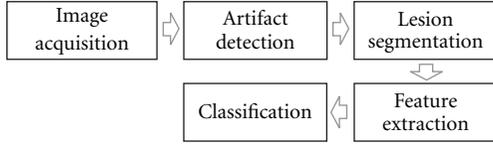


FIGURE 1: Typical computer aided diagnosis (CAD) pipeline usually adopted for automated skin lesion diagnosis (ASLD). Our goal is to (1) generalize the artifact detection, segmentation as well as a portion of the feature extraction stage into a single mathematical framework and (2) propose and evaluate probabilistic models which employ supervised learning to quickly and automatically “learn” to perform these tasks.

regardless of the label assigned to their neighbors. This assumption of independence is clearly not valid, as there is a high degree of correlation between neighboring pixels in any image (any image other than pure noise, i.e.). CRF-based models attempt to relax this assumption of independence by creating a graphical model which defines the dependencies between pixels.

In order to apply a CRF model, a parameter vector specifying the relative contribution of the input features is required. Often, these parameters are determined in an ad hoc fashion via trial and error. Since our goal is a *general* method, easily applicable to a variety of problems, it is crucial that these parameters be determined automatically based on observations. We, therefore, apply the maximum likelihood estimator for the parameter vector and describe a gradient-based method for its computation. We also address many practical considerations encountered during the implementation.

The paper is organized as follows: in Section 2, we briefly review previous work. In Section 3, we formulate the generalization in Section 3.1, review our previous PP model [5] in Section 3.2, and present our CRF model in Section 3.3. In Section 4, we present results. Finally, we conclude in Section 5.

## 2. Previous Work

Our original PP model was based on the work by Debeir et al. [12] who also attempts to generalize many tasks in ASLD. Our model was found to perform comparably to many published lesion segmentation algorithms including K-means++ (KPP) [6], J-image segmentation (JSEG) [7], dermatologist-like tumor area extraction algorithm (DTEA) [8], statistical region merging (SRM) [9], and threshold fusion (FSN) [10]. It also performed comparably to DullRazor [11] for detecting occluding hair and was able to identify the dermoscopic structure *pigment network*. Our PP model is briefly reviewed in Section 3.2; however, we refer readers to our previous study for further details, as well as a more comprehensive review of previous work in ASLD, including the methods against which we compare [5].

We defer the review of the CRF model until Section 3.3, where we examine it in detail.

## 3. Method

This section is divided into 3 parts. We begin in Section 3.1 by formalizing the generalization that is capable of performing a variety of tasks in ASLD. In Section 3.2, we briefly review our previous PP model [5]. Finally, in Section 3.3, we outline our CRF model.

**3.1. The Generalization.** We are given a set of observations  $\{x^m, y^m\}$ , consisting of images ( $x$ ) and corresponding ground truths labeling ( $y$ ). Using the notation of Szummer et al. [13], the superscript  $x^m$  or  $y^m$  indexes a specific image/labeling in the set and the subscript  $x_i$  or  $y_i$  indexes a specific pixel. Let  $N_I$  represent the number of images, and  $N_p^m$  represents the number of pixels in image  $x^m$ . An imageset can contain any number of channels (or features), which we denote by  $N_C$ . Valid values for each entry in the label field ( $y_i$ ) are defined by the label set  $L = \{l_1, \dots, l_{N_L}\}$ , where  $N_L$  is the number of possible labels.

Given our training set  $\{x^m, y^m\}$ , we use supervised learning to predict the label fields for previously unseen images.

Formally, we are given

$$\begin{aligned} &\{x^m, y^m\}; \quad m = 1, \dots, N_I; \text{ i.i.d.} \\ &L = \{l_i\}; \quad i = 1, \dots, N_L; l_i \in \mathbb{N}, \\ &x^m \in \mathbb{R}^{N_p^m \times N_C}, \\ &y^m \in L^{N_p^m}. \end{aligned} \quad (1)$$

And our task is to use the information in  $\{x^m, y^m\}$  to infer the function  $f : x \rightarrow y$  that produces the best possible label field.

**3.2. The PP Model.** In this section, we briefly review our per-pixel (PP) estimation model [5]. An overview of the training and testing phases of the model is illustrated in Figures 2 and 3, respectively. Under this model, we assign the most probable label to each pixel independently

$$y_i^* = \arg \max_{l_j} [P(y_i = l_j | x_i)]; \quad i = 1, \dots, N_p, \quad (2)$$

Applying Bayes’ rule and simplifying, we arrive at the standard maximum likelihood formulation

$$y_i^* = \arg \max_{l_j} [P(x_i | y_i = l_j)P(y_i)]; \quad i = 1, \dots, N_p. \quad (3)$$

We model the posterior  $P(x | y = l)$  as a set of  $N_L$  multivariate normal distributions  $P(x | y = l_j) = N(\mu_{l_j}, \Sigma_{l_j})$ , whose parameters  $(\mu_{l_j}, \Sigma_{l_j})$  are estimated using the training set  $\{x^m, y^m\}$ . We model  $P(y)$  as a discrete distribution. Let  $N_{Y_i}$  represent the number of elements in  $y^m$  that assume the value  $l_i$ , then

$$P(y_i) = \frac{N_{Y_i}}{\sum_{j=1}^{N_L} N_{Y_j}}. \quad (4)$$

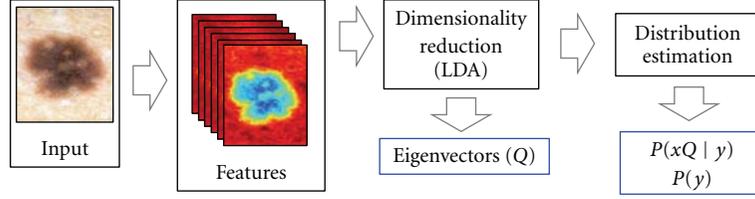


FIGURE 2: The training phase of our per-pixel (PP) model. Features are first computed, then the dimensionality of the featurespace is reduced using LDA. Posterior probabilities in this subspace are then estimated.

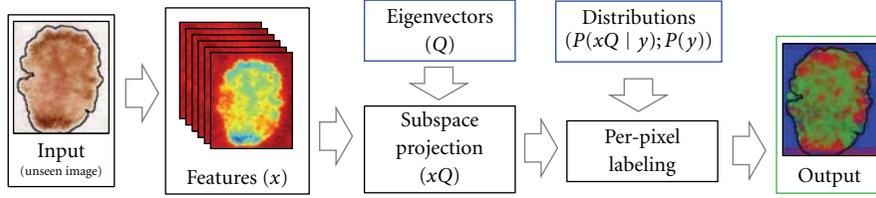


FIGURE 3: The testing phase of our per-pixel (PP) model. Features are computed as in the training phase. The projection  $Q$  is used to transform the features into the subspace determined in the training phase. Maximum a-posteriori estimation, using the posteriors estimated in the training phase, is then used to generate the label.

We also normalize the probabilities across the label set, which are later used as features in the CRF model. The normalized likelihood that a pixel  $i$  is associated with the label  $l_j$  is

$$\mathcal{L}_{i,j} = \frac{P(x_i | y_i = l_j)}{\sum_{k=1}^{N_L} P(x_i | y_i = l_k)}. \quad (5)$$

In order to examine the model's performance across the entire sensitivity/specificity range, we consider many thresholds  $T$  on  $\mathcal{L}_{i,j}$  over the range  $[0, 1]$  and label pixels accordingly.

As the number of channels ( $N_C$ ) in the images grows, we perform supervised dimensionality reduction on the observations  $x$  to focus the predictive power of our dataset onto a smaller subset of parameters. Linear discriminant analysis (LDA) is used to determine the subspace of  $x$  which best separates the labels [14].

LDA performs an eigenvalue decomposition of a scatter matrix representing the ratio of between-class covariance to within-class covariance. It returns a matrix of eigenvectors  $Q \in \mathbb{R}^{N_C \times N_L - 1}$  which projects observations ( $x$ ) from  $N_C$  dimensions to  $N_L - 1$

$$\begin{aligned} Q &= \text{eig}(S_w^{-1} S_b), \\ S_w &= \sum_{i=1}^{N_L} \Sigma_{l_i}, \\ S_b &= \sum_{i=1}^{N_L} (\mu_{l_i} - \mu)(\mu_{l_i} - \mu)^T, \end{aligned} \quad (6)$$

where  $\mu$  is the overall mean of  $x$  across all images and classes. Once the projection  $Q$  is determined, the posteriors

are estimated, likelihoods are computed, and inference is performed in this subspace ( $xQ$ )

$$\begin{aligned} P(xQ | y = l_j) &= N(\mu_{l_j}^Q, \Sigma_{l_j}^Q); \quad j = 1, \dots, N_L, \\ y_i^* &= \arg \max_{l_j} [P(x_i Q | y_i = l_j) P(y_i)]; \quad i = 1, \dots, N_P, \end{aligned} \quad (7)$$

$$\mathcal{L}_{i,j} = \frac{P(x_i Q | y_i = l_j)}{\sum_{k=1}^{N_L} P(x_i Q | y_i = l_k)}, \quad (8)$$

where the superscript  $Q$  ( $\mu_{l_j}^Q, \Sigma_{l_j}^Q$ ) is used to differentiate the label means/covariances in this subspace from the original space in which the observations were performed ( $\mu_{l_j}, \Sigma_{l_j}$ ).

**3.3. The CRF Model.** In this section, we seek to improve upon the PP model developed in previous work [5] and described in Section 3.2. We present an overview of conditional random fields (CRFs) in Section 3.3.1. In Section 3.3.2, we describe how the CRF parameters can be determined empirically using maximum likelihood estimation (MLE) [15]. In Section 3.3.3, we discuss practical considerations for finding these parameters, including how to estimate the partition function [16] and how to regularize the likelihood expression [15]. In Section 3.3.4, we solve the MLE formulation via gradient-based methods. An overview of the training and testing phases of our CRF model is illustrated in Figures 4 and 5, respectively.

**3.3.1. Overview.** The CRF model is an undirected graphical model that is naturally suited to represent and exploit the dependencies between observations, such as neighboring pixels in an image [15]. The probability that a label field  $y$

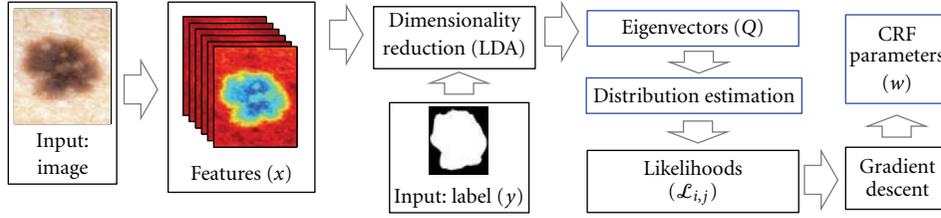


FIGURE 4: The training phase of our CRF model. We follow the same procedure as in our PP model up until the posteriors are estimated. We then calculate pixel likelihoods and use these as node features in our CRF model. We infer CRF parameters using gradient descent.

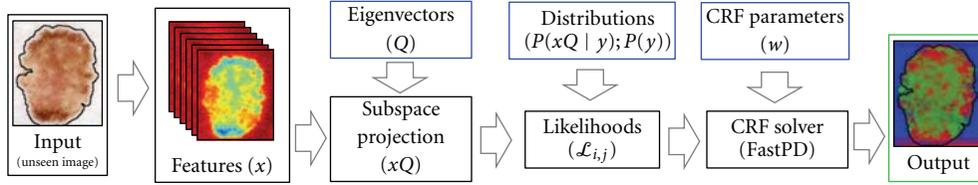


FIGURE 5: The testing phase of our CRF model. After the likelihoods are computed, we use the CRF parameters from the training phase, and the software FastPD to generate label fields.

is associated with the image  $x$  under model parameters  $w$  is given by

$$P(y | x; w) = \frac{1}{Z(x, w)} \exp(-E(y, x; w)), \quad (9)$$

where the function  $Z(x, w)$ , known as the partition function, is used to normalize the probabilities for given values of  $x$  and  $w$

$$Z(x, w) = \sum_y \exp(-E(y, x; w)). \quad (10)$$

The energy function  $E$  represents the linear combination of features employed by the model and is parameterized by the weight vector  $w$

$$E(y, x; w) = \sum_{k=1}^{N_w} w_k \Phi_k(y, x). \quad (11)$$

Given an undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  represents the nodes (i.e., pixels) of an observation,  $\mathcal{E}$  represents the *dependencies* between nodes (throughout this document,  $\mathcal{E}$  is the 4-connected set of neighboring pixels), and the energy function  $E$  is the weighted sum of features  $\Phi_i(y, x)$ . Features can either operate over the nodes of the graph ( $\Phi^V$ ), or over its edges ( $\Phi^E$ )

$$\begin{aligned} \Phi^V(y, x) &= \sum_{i \in \mathcal{V}} \phi(y_i, x_i), \\ \Phi^E(y, x) &= \sum_{(i,j) \in \mathcal{E}} \phi(y_i, y_j, x_i, x_j), \end{aligned} \quad (12)$$

In order for the model to be tractable, edge features  $\Phi_i^E(y, x)$ , and their corresponding weights must adhere to

certain constraints. Let  $\mathbb{E}$  represent the set of edge features. The following constraints must be satisfied [17]

$$\begin{aligned} w_i &> 0 \quad \forall i \in \mathbb{E}, \\ \phi^E(y_i, y_j, x_i, x_j) &= 0 \quad \forall (y_i, y_j) \quad \text{s.t. } y_i = y_j. \end{aligned} \quad (13)$$

Strictly speaking, the second constraint can be replaced with the more general constraint that edge feature functions be *submodular* [18]. However, throughout this document, we will impose this stricter constraint which can be interpreted as “an edge cost is only incurred across nodes with differing labels.”

A CRF solver is one that, given observations  $x$  and parameters  $w$ , can find the most likely labeling  $y^*$

$$y^* \leftarrow \arg \max_y P(y | x; w). \quad (14)$$

We use the software FastPD [19, 20], which can exactly solve (14), under the constraints imposed above.

**3.3.2. Determining MLE Parameters.** Since the emphasis of our work is on a general model capable of performing a variety of tasks, it is crucial that model parameters ( $w$ ) be determined automatically from training data via empirical means. In this section, we derive the partial derivatives of the likelihood function which can be used by gradient-based methods to compute  $w$ .

Since the observations  $\{x^m, y^m\}$  are assumed to be independent. The likelihood of the data, given the set of parameters, is equal to the product of the probabilities in the observed set, under those parameters

$$\ell(w) = \prod_{m=1}^{N_I} P(y^m | x^m; w). \quad (15)$$

The maximum likelihood estimator is then

$$w^* = \arg \max_w \prod_{m=1}^{N_I} P(y^m | x^m; w). \quad (16)$$

If we can find the partial derivatives  $\partial \ell / \partial w_i$ , we can optimize  $w$  using gradient-based methods. We begin by expressing the likelihood function  $\ell(w)$  in terms of  $w$

$$\begin{aligned} w^* &= \arg \max_w \prod_{m=1}^{N_I} P(y^m | x^m; w) \\ &= \arg \max_w \sum_{m=1}^{N_I} \ln(P(y^m | x^m; w)) \\ &= \arg \max_w \sum_{m=1}^{N_I} (-E(y^m, x^m, w) - \ln(Z(x, w))) \\ &= \arg \min_w \sum_{m=1}^{N_I} \left( \sum_{k=1}^{N_W} w_k \Phi_k(y^m, x^m) \right. \\ &\quad \left. + \ln \left[ \sum_y \exp \left( - \sum_{k=1}^{N_W} w_k \Phi_k(y, x^m) \right) \right] \right). \end{aligned} \quad (17)$$

Solving for the partial derivatives, we get the following expression for the gradients of the likelihood function:

$$\begin{aligned} \frac{\partial \ell}{\partial w_i} &= \sum_{m=1}^{N_I} \left( \Phi_i(y^m, x^m) \right. \\ &\quad \left. + \frac{\sum_y -\Phi_i(y, x^m) \exp \left( - \sum_{k=1}^{N_W} w_k \Phi_k(y, x^m) \right)}{\sum_y \exp \left( - \sum_{k=1}^{N_W} w_k \Phi_k(y, x^m) \right)} \right). \end{aligned} \quad (18)$$

However, we now come to an impasse. The second term of (18) would have us iterating over all possible label fields  $y$ . For a binary classification task over a modestly sized image of  $256 \times 128$ , this would require a summation over  $2^{256 \times 128} \approx 2 \times 10^{9000}$  labelings. Clearly this is intractable, and we must resort to estimating this second term.

**3.3.3. Practical Considerations.** In order to derive CRF parameters with grid-structured models for even modestly sized images, a method to estimate the partition function is required. Inspired by [21], we employ one of the simplest estimation methods and approximate the partition function using saddle-point approximation (SPA) [16]

$$\begin{aligned} \sum_y \Phi(y, x) &\approx \Phi(y^*, x), \\ y^* &\leftarrow \arg \max_y P(y | x; w). \end{aligned} \quad (19)$$

We also introduce an additional practical consideration. Since gradient-based methods will be used to determine  $w$ ,

we regularize the likelihood function ( $\ell(w)$ ) by the squared L2 norm of the parameters [15] to penalize large weight vectors (since scalar multiples of a weight vector produce identical results). This makes the resulting likelihood function *strictly convex*. The regularized likelihood is then

$$\begin{aligned} \ell(w) &= \sum_{m=1}^{N_I} \left( \sum_{k=1}^{N_W} w_k \Phi_k(y^m, x^m) - \ln \sum_y \exp \sum_{k=1}^{N_W} w_k \Phi_k(y, x^m) \right) \\ &\quad - \frac{\|w\|^2}{2\sigma^2} \end{aligned} \quad (20)$$

And the gradients become

$$\begin{aligned} \frac{\partial \ell}{\partial w_i} &= \sum_{m=1}^{N_I} \left( \Phi_i(y^m, x^m) \right. \\ &\quad \left. + \frac{\sum_y -\Phi_i(y, x^m) \exp \left( - \sum_{k=1}^{N_W} w_k \Phi_k(y, x^m) \right)}{\sum_y \exp \left( - \sum_{k=1}^{N_W} w_k \Phi_k(y, x^m) \right)} \right) - \frac{w_i}{\sigma^2}. \end{aligned} \quad (21)$$

Which under SPA becomes

$$\frac{\partial \ell}{\partial w_i} \approx \sum_{m=1}^{N_I} (\Phi_i(y^m, x^m) - \Phi_i(y^*, x^m)) - \frac{w_i}{\sigma^2}. \quad (22)$$

**3.3.4. Implementation.** We are now ready to implement a gradient-based method to estimate the CRF parameter vector  $w$ . Given an initial weight vector  $w^0$ , the gradients of the likelihood function are estimated as per (22). These gradients are used to update the weight vector, which in turn is used to estimate a new set of gradients. This process is repeated until convergence.

We have observed (as does [21]) that gradient methods using saddle point approximation lead to oscillating behavior. Therefore, we keep a record of the best empirical set of parameters found, rather than the parameters of the final iteration. We also enforce the constraint from (13) that weights for edge-based features must remain positive.

In addition to the training set ( $\{x, y\}$ ), the algorithm also requires an initial weight vector ( $w^0$ ), a regularization factor ( $\sigma^2$ ), a step size ( $\gamma$ ), and termination conditions (convergence criteria:  $\epsilon$ ; maximum number of iterations:  $N_{\text{itr}}$ ). The algorithm has been found to be robust to these additional parameters. Pseudocode of our implementation is presented in Algorithm 1.

## 4. Results

Previous work has demonstrated our model's ability to generalize to many applications [5]. Here, we focus on a single application (lesion segmentation) and present results for our two models. We also compare our results to 5 previously published methods (KPP [6], JSEG [7], DTEA [8], SRM [9], and FSN [10]).

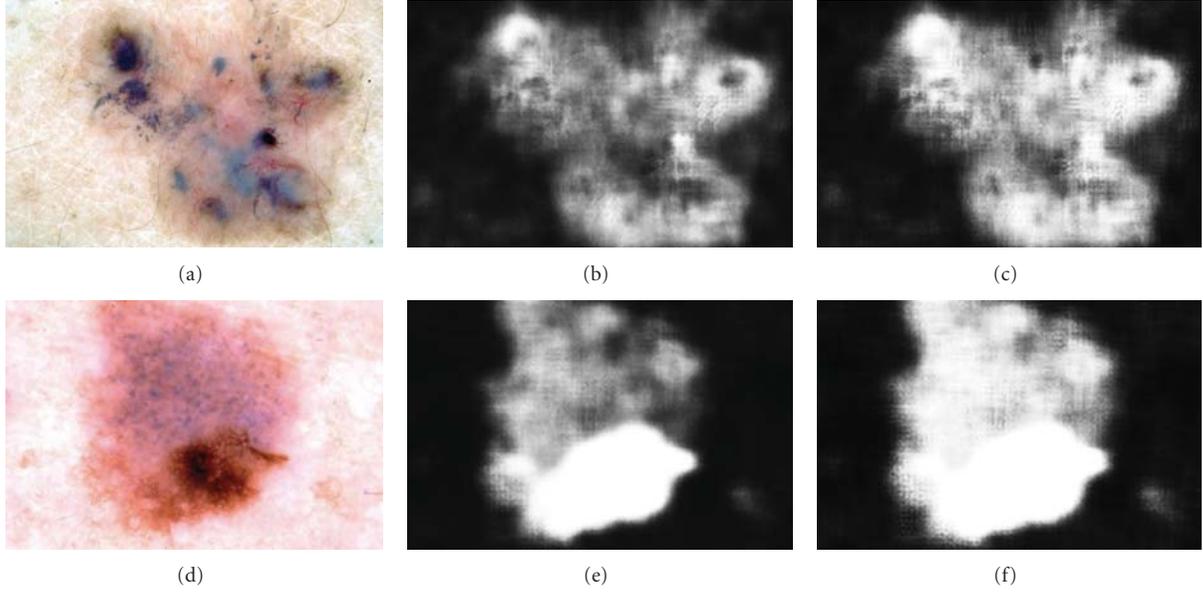


FIGURE 6: The effect of  $L^*$  normalization on the segmentation likelihoods. left column: original dermoscopic image; middle: segmentation likelihoods ( $\mathcal{L}_{i,\text{lesion}}$ ) before  $L^*$  normalization; right: after  $L^*$  normalization.

```

Require:  $x, y, w^0, \sigma^2 > 0, \gamma < 0, \epsilon > 0, N_{itr} > 0$ 
max ← 0
for  $i \leftarrow 1$  to  $N_{itr}$  do
   $g^i \leftarrow 0$ 
   $a \leftarrow 0$ 
  for  $m \leftarrow 1$  to  $N_I$  do
     $y^* \leftarrow \arg \max_y P(y \mid x^m; w^{i-1})$ 
     $a \leftarrow a + \text{accuracy}(y^*, y^m)/N_I$ 
     $g^i \leftarrow g^i + \Phi(y^m, x^m) - \Phi(y^*, x^m) - w^{i-1}/(N_I \sigma^2)$ 
  end for
  if  $a > \text{max}$  then
    max ←  $a$ 
     $w^* \leftarrow w^{i-1}$ 
  end if
   $w^i \leftarrow w^{i-1} + \gamma g^i$ 
  for all  $j \in \mathbb{E}$  do
    if  $w_j^i < 0$  then
       $w_j^i \leftarrow 0$ 
    end if
  end for
  if  $\|w^i - w^{i-1}\| < \epsilon$  then
    break
  end if
end for
return  $w^*$ 

```

ALGORITHM 1: Calculating the CRF parameter vector  $w$  using gradient descent and saddle-point approximation.

The dataset consists of 116 images from dermoscopy atlases [22, 23], which were acquired by a several dermatologists in separate practices using differing equipment. The images have not been properly color calibrated. Since the goal was to create a difficult dataset, 100 of the 116 lesions

were selected to be particularly challenging to segmentation algorithms [7]. We intentionally chose a simplistic featureset to emphasize the power of the models under consideration.

The features employed were 5 Gaussian, and 5 Laplacian of Gaussian filters applied a various scales ( $\sigma = [1.25, 2.5, 5, 10, 20]$ ) in each channel of the image in CIE  $L^*a^*b^*$  space. The responses of these filters represent the observations  $x$  (where  $N_C = 30$ ). Each image was expertly segmented by a dermatologist. These ground truth labelings are denoted as  $y$ .

For all experiments, 10-fold cross-validation was employed. The dataset was randomly divided into 10 groups, and label fields for each group of images were determined using model parameters which were estimated from the observations in the 9 other groups. In both the PP and CRF models, all steps after the computation of features (refer to Figures 2 and 4) were included within the cross-validation loop including determining the projection  $Q$ , estimating the prior/posteriors, determining CRF parameter vector  $w$ , and so forth.

*4.1. The PP Model.* We begin by summarizing previous results on how our PP model faired on this dataset. A more detailed analysis, including the relative contribution of various aspects of the model (including features, dimensionality reduction, and classification method), can be found in our previous work [5].

Since that time, we have discovered that we can partially compensate for the lack of color calibration by subtracting the mean of the  $L^*$  channel before computing features. While not as desirable as full color and lighting calibration [24], this procedure at least compensates for various camera exposure levels, as can be seen in the resulting PP likelihood maps in Figure 6 (as calculated by (8)). Figure 7 illustrates a ROC

TABLE 1: Comparison of our PP model’s ability to segment lesions to our CRF model and 5 previously published methods.

Method	$n$	Sens	Performance			
			Method	Spec	PP (nearest pt.)	
CRF	116	0.845		0.924	0.843	0.921
KPP [6]	116	0.765		0.770	0.941	0.763
JSEG [7]	91	0.627		0.987	0.677	0.980
DTEA [8]	116	0.597		0.986	0.638	0.985
SRM [9]	112	0.790		0.946	0.773	0.957
FSN [10]	116	0.808		0.934	0.814	0.939

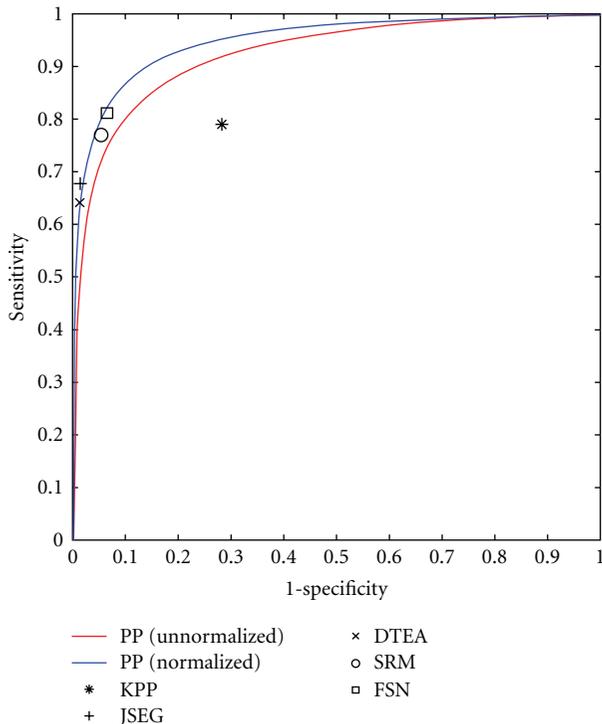


FIGURE 7: ROC curve comparing our PP model before normalization (red line) and after normalization (blue line) to 5 previously published methods.

curve comparing the performance of our PP model (before and after normalization) to the segmentation algorithms KPP [6], JSEG [7], DTEA [8], SRM [9], and FSN [10]. Our method performs comparably to JSEG, DTEA, SRM, and FSN and outperforms KPP although only KPP, DTEA, and FSN algorithms were able to generate results for all 116 images. Table 1 summarizes the results.

**4.2. The CRF Model.** As described in Section 3.3, the CRF model operates over an undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  and consists of node features  $(\Phi^V(y, x))$  and edge features  $(\Phi^E(y, x))$ . The graph structure employed was the 4-connected set of neighboring pixels. Our featureset contains 2 features: one over the nodes and one over the edges. The

node features are the likelihoods as computed by (8) of the PP model as in Section 4.1, and the edge features are set to the CIE  $L^*$  intensity difference between neighboring pixels, if the labels of said pixels differ

$$\Phi_1^V(y, x) = \sum_{i \in \mathcal{V}} \frac{P(x_i Q | y_i)}{\sum_{j=1}^{N_L} P(x_i Q | y_i = l_j)},$$

$$\Phi_2^E(y, x) = \sum_{(i,j) \in \mathcal{E}} |L^*(x_i) - L^*(x_j)| \mathbf{1}_{y_i \neq y_j},$$
(23)

where we use  $\mathbf{1}_{y_i \neq y_j}$  to denote the indicator function (i.e.,  $\mathbf{1}_{y_i \neq y_j}$  evaluates to 1 if  $y_i \neq y_j$ ; 0 otherwise)

While the method described in Section 3.3 is general enough to handle an arbitrary number of node and edge features, there are 2 reasons why we chose only one of each. To begin, we seek to make the comparison between the PP model and the CRF model as meaningful as possible. Using the likelihoods from the PP model as the node feature is an elegant way to evaluate the improvements realized by the CRF model. Note that with this choice of features, the CRF model with weight vector  $w = [1, 0]$  gives identical results to the PP model. Additionally, the saddle-point method for approximating the partition function seems to degrade as the number of features increases. We note, however, that even in studies where the partition function can be computed exactly (because the CRF graph contains no loops), the loss incurred by such *piecewise training* methods is negligible [25].

Figure 8 compares some segmentations produced by the PP and CRF model. By relaxing the assumption of independence in the PP model, the CRF model is able to smooth over small areas of discontinuity, filling in “gaps” in segmentations, and removing noise. In Figures 8(a) and 8(c), the “holes” in the resulting PP segmentations do not manifest in the CRF segmentations (Figures 8(b) and 8(d)) due to the model’s holistic search for the best label *field*, rather than best individual label. Additionally, while the PP model is already fairly robust to occluding hair (Figure 8(e)), the CRF model is even more robust, able to smooth over misclassifications due to artifacts.

We also tested the stability of the CRF model with respect to regularization and the hyperparameter  $\sigma^2$ . Varying  $\sigma^2$  (to assume values in the range  $[10^{-6}, \text{Inf}]$ ) had little effect on

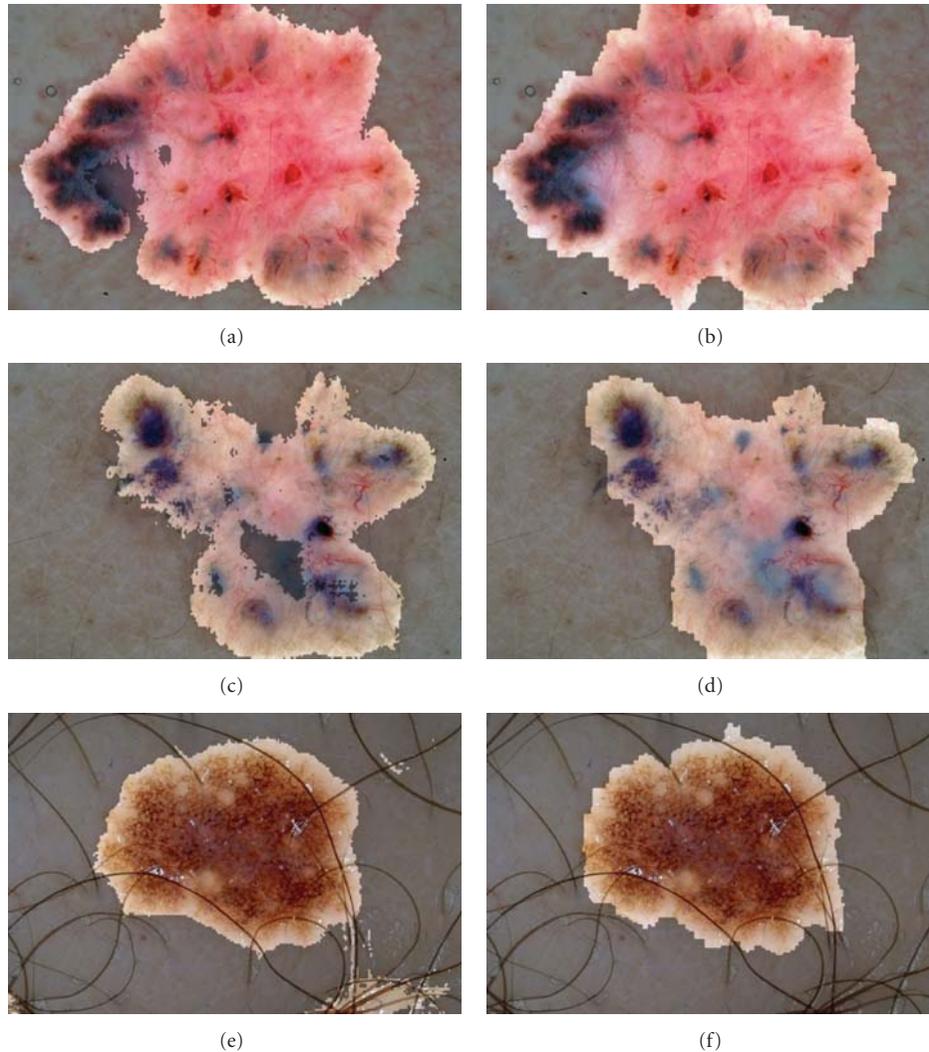


FIGURE 8: Comparing segmentations from our PP model (left) and CRF model (right). Since the CRF model relaxes the assumption of pixel independence in the PP model, it is able to smooth over local discontinuities. The result is better segmentations which fill in “holes” and remove “noise.”

performance of the model on this particular dataset. In spite of the seemingly ineffectual nature of this parameter, we do not remove it from the model since the emphasis of this work is on *general* models for ASLD. The effect of  $\sigma^2$  in general (over many tasks in ASLD) has yet to be determined.

While subjectively, the CRF model offers substantial improvements; objectively, the CRF model is a marginal improvement over the PP model. Figure 9 shows an ROC curve comparing the CRF’s performance to that of the PP model and previously published methods, and Table 1 summarizes the results.

## 5. Conclusions

In this paper, we have generalized several common problems in ASLD into a single formulation. We also presented 2 probabilistic models capable of solving the formulation, and described how supervised learning can be used to

determine all model parameters. Since the parameters for the resulting models can all be determined automatically from training data, it is hoped that these models can be applied quickly and effectively to a variety of relevant tasks in ASLD.

While both methods perform comparably to previously published methods, the qualitative improvements realized by CRF model aren’t reflected in the quantitative score. Unlike the PP model, the CRF model does not assign labels to pixels independently. Rather, the CRF model selects the best label *field* to assign to an image. This allows the CRF model to fill in “holes” and smooth out noise that would otherwise appear.

The discrepancy between the objective and subjective performance of the CRF model implies that our evaluation metric (pixel-wise sensitivity and specificity) may be less than ideal. Therefore, future work will explore the use of alternate evaluation metrics [26, 27].

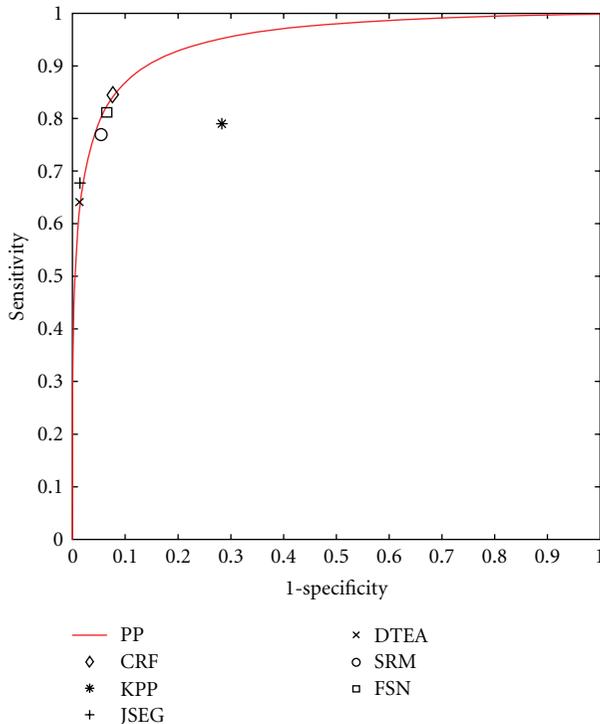


FIGURE 9: ROC curve comparing our CRF model (diamond) to our PP model (line) and 5 previously published methods.

Even though the models presented are competitive, there are many potential directions in which they can be improved upon even further. In our grid-structured CRF model, we must resort to approximating the partition function due to the computational complexity of calculating it exactly. Imposing a tree-based structure over the image [25] would enable the exact computation of the partition function via dynamic programming and should lead to more reliable CRF parameters. Replacing our gradient-based method for determining CRF parameters with a max-margin formulation [13] is another possible way to increase the reliability of the resulting parameters. We can also induce non-linearities into the model by replacing the linear dimensionality reduction step (LDA) with its nonlinear counterparts (i.e., KLDA [28]). Finally, the use of semi-supervised learning techniques may be used to decrease the cost of acquiring expertly annotated datasets [29].

## Acknowledgments

This work was supported in part by the Canadian Institutes of Health Research (CIHR) Skin Research Training Centre, the Canadian Dermatology Foundation, and Collaborative Health Research Projects Program, a grant jointly funded by the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research. The Authors would like to thank Dr. M. Emre Celebi for making available results of the alternative segmentation algorithms.

## References

- [1] C. Erickson and M. Driscoll, "Melanoma epidemic: facts and controversies," *Clinics in Dermatology*, vol. 28, no. 3, pp. 281–286, 2010.
- [2] M. Lens and M. Dawes, "Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 150, no. 2, pp. 179–185, 2004.
- [3] R. MacKie and D. Hole, "Audit of public education campaign to encourage earlier detection of malignant melanoma," *British Medical Journal*, vol. 304, no. 6833, pp. 1012–1015, 1992.
- [4] J. Resneck and A. B. Kimball, "The dermatology workforce shortage," *Journal of the American Academy of Dermatology*, vol. 50, no. 1, pp. 50–54, 2004.
- [5] P. Wighton, T. Lee, H. Lui, D. McLean, and M. Atkins, "Generalizing common tasks in automated skin lesion diagnosis," *IEEE Transactions on Information Technology in BioMedicine*, vol. 15, no. 4, pp. 622–629, 2011.
- [6] H. Zhou, M. Chen, L. Zou et al., "Spatially constrained segmentation of dermoscopy images," in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '08)*, pp. 800–803, Paris, France, 2008.
- [7] M. E. Celebi, Y. Aslandogan, W. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Research and Technology*, vol. 13, no. 4, pp. 454–462, 2007.
- [8] H. Iyatomi, H. Oka, M. E. Celebi et al., "An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 566–579, 2008.
- [9] M. E. Celebi, H. Kingravi, H. Iyatomi et al., "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, no. 3, pp. 347–353, 2008.
- [10] M. E. Celebi, S. Hwang, H. Iyatomi, and G. Schaefer, "Robust border detection in dermoscopy images using threshold fusion," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2541–2544, Hong Kong, 2010.
- [11] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor®: a software approach to hair removal from images," *Computers in Biology and Medicine*, vol. 27, no. 6, pp. 533–543, 1997.
- [12] O. Debeir, C. Decaestecker, J. Pasteels, I. Salmon, R. Kiss, and P. van Ham, "Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions," *Cytometry*, vol. 37, no. 4, pp. 255–266, 1999.
- [13] M. Szummer, P. Kohli, and D. Hoiem, "Learning CRFs using graph cuts," in *Proceedings of the 10th European Conference on Computer Vision (ECCV '08)*, pp. 582–595, Marseille, France, 2008.
- [14] A. Martínez and A. Kak, "Pca versus lda," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2002.
- [15] C. Sutton and A. McCallum, "An introduction to conditional random fields for relational learning," in *Introduction to Statistical Relational Learning*, p. 93, 2007.
- [16] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRFs: surface reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 5, pp. 401–412, 2002.

- [17] H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1333–1336, 2003.
- [18] V. Kolmogorov, "Primal-dual algorithm for convex Markov random fields," Microsoft Research MSR-TR-2005-117, 2005.
- [19] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1436–1453, 2007.
- [20] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic MRFs: setting the state of the art with primal-dual strategies," *Computer Vision and Image Understanding*, vol. 112, no. 1, pp. 14–29, 2008.
- [21] S. Kumar, J. August, and M. Hebert, "Exploiting inference for approximate parameter learning in discriminative fields: an empirical study," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 153–168, Springer, New York, NY, USA, 2005.
- [22] G. Argenziano and H. Soyer, *Interactive Atlas of Dermoscopy (Book and CD-ROM)*, Edra medical publishing and new media, Milan, Italy, 2000.
- [23] H. Soyer and G. Argenziano, *Dermoscopy of Pigmented Skin Lesions. An Atlas based on the Consensus Net Meeting on Dermoscopy*, Edra medical publishing and new media, Milan, Italy, 2000.
- [24] P. Wighton, T. Lee, H. Lui, D. McLean, and M. Atkins, "Chromatic aberration correction: an enhancement to the calibration of low-cost digital dermoscopes," *Skin Research and Technology*, vol. 17, no. 3, pp. 339–347, 2011.
- [25] S. Nowozin, P. Gehler, and C. Lampert, "On parameter learning in CRF-based approaches to object class image segmentation," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, pp. 98–111, Crete, Greece, 2010.
- [26] M. Celebi, G. Schaefer, H. Iyatomi, W. Stoecker, J. Malter, and J. Grichnik, "An improved objective evaluation measure for border detection in dermoscopy images," *Skin Research and Technology*, vol. 15, no. 4, pp. 444–450, 2009.
- [27] T. Lee, D. McLean, and M. Atkins, "Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions," *Medical Image Analysis*, vol. 7, no. 1, pp. 47–64, 2003.
- [28] D. Cai, X. He, and J. Han, "Efficient kernel discriminant analysis via spectral regression," in *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM '07)*, pp. 427–432, Omaha, Neb, USA, 2007.
- [29] C. Lee, S. Wang, F. Jiao, D. Schuurmans, and R. Greiner, "Learning to model spatial dependency: semi-supervised discriminative random fields," *Advances in Neural Information Processing Systems*, vol. 19, pp. 793–800, 2007.

## Research Article

# A Statistical Texture Model of the Liver Based on Generalized N-Dimensional Principal Component Analysis (GND-PCA) and 3D Shape Normalization

**Xu Qiao and Yen-Wei Chen**

*College of Information Science and Engineering, Ritsumeikan University, 525-0054 Kusatsu, Japan*

Correspondence should be addressed to Xu Qiao, qiaoxugala@gmail.com

Received 27 February 2011; Revised 2 July 2011; Accepted 22 July 2011

Academic Editor: Kenji Suzuki

Copyright © 2011 X. Qiao and Y.-W. Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a method based on generalized N-dimensional principal component analysis (GND-PCA) and a 3D shape normalization technique for statistical texture modeling of the liver. The 3D shape normalization technique is used for normalizing liver shapes in order to remove the liver shape variability and capture pure texture variations. The GND-PCA is used to overcome overfitting problems when the training samples are too much fewer than the dimension of the data. The preliminary results of leave-one-out experiments show that the statistical texture model of the liver built by our method can represent an untrained liver volume well, even though the mode is trained by fewer samples. We also demonstrate its potential application to classification of normal and abnormal (with tumors) livers.

## 1. Introduction

In the *recent* years, digital atlases of human anatomy have become popular and important topics in medical image analysis research [1, 2]. For interpretation of images of structures and variations in the organs of the human body, it is important to have a model of the way organ volumes can be represented.

The digital atlas can be categorized as a statistical shape atlas (statistical shape model) and a statistical appearance (volume) atlas (statistical appearance (volume) model). The statistical shape model focuses on the shape information, such as feature points and volume surface [3]. It is a useful tool for study of variations in anatomic shape and has been widely used in medical image analysis, for example, medical image segmentation [4–6] and shape registration [7]. The statistical appearance model is focused on both shape and texture (voxel intensity) information. Inspired from the works of active shape models (ASMs) [3], the authors of [5, 8] proposed 3D ASMs for construction of 3D statistical models for segmentation of the left ventricle of the heart. In [9], the authors extended the work on active appearance models (AAMs) [10], and propose the

use of 3D AAMs for the segmentation of cardiac MR and ultrasound images. Also, work [11] was done to build the 3D statistical deformation models (SDMs) for 3D MR brain images. Radiologists are mainly depending on the intensity variations (texture information) in livers on medical images to identify modules or tumors and make a diagnostic decision. However, there has been little research on applications of digital atlas to computer-assisted diagnosis (CAD). We have shown the potential application of statistical shape models to the classification of normal and cirrhotic livers [12]. Because many diseases will change the texture (voxel value) of the organ significantly, we need to capture not only shape variations, but also texture (voxel value) variations. Compared to statistical shape modeling, statistical texture modeling usually faces overfitting problems, and the statistical texture modeling for medical volumes is a challenging task because the dimensions of the medical volume are very high, while the training samples are much fewer than the dimensions of the data.

In our previous work, we have proposed a tensor-based subspace learning method named generalized N-dimensional principal component analysis (GND-PCA) for statistical

appearance modeling of medical volumes [13]. The high-dimensional volume is treated as a 3rd-order tensor, and the optimal subspace on each mode is calculated simultaneously by minimizing of the square error between the original tensor (volume) and the reconstructed tensor (volume), based on the subspace with an iteration algorithm. As an improvement on our previous work [13], we propose a framework for capturing texture variations of the liver by using GND-PCA together with a 3D shape normalization technique (a nonrigid registration technique). The GND-PCA is used to overcome the overfitting problem, and the 3D shape normalization technique is used for normalizing liver shapes to remove the liver shape variability and capture pure texture variations. The leave-one-out experiments show that the statistical texture model of the liver built by our method can represent an untrained liver volume well, even though the model is trained by fewer samples. The preliminary results also show that the features extracted by the statistical texture model have the capability of discrimination for different types of volume data, such as normal and abnormal (with tumors).

The rest of the paper is organized as follows. In Section 2, we introduce our methodology. In Section 3, we present the experimental evaluation of our approach after introducing the datasets we used. Section 4 concludes the paper by summarizing the main points of our contribution.

## 2. Methodology

Our proposed method for statistical texture modeling consists of two steps: (1) employing a nonrigid transformation for 3D shape normalization and (2) applying the GND-PCA method for feature extraction. The basic scheme is presented in Figure 1.

**2.1. 3D Shape Normalization.** In order to remove shape variations, we apply a nonrigid transformation based on mathematical forms for normalizing all of the datasets to the same shape. This is because the mathematical nonrigid transformations are simpler, and they can make the registration faster. Additionally, we do not need to assume the physical parameters, which are difficult to guess in practice. Hence, we adopted the mathematical nonrigid transformation in our research.

Here, we applied rigid transformation for global transformation and B-spline transformation for local transformation. The combination of global and local transformations can be represented by

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}_{\text{Global}}(\mathbf{x}) + \mathbf{T}_{\text{Local}}(\mathbf{x}), \quad (1)$$

where  $\mathbf{x} = [x, y, z]^T$  is the coordinate of a 3D point.

A rigid transformation is expressed by

$$\mathbf{T}_{\text{Global}}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}, \quad (2)$$

where  $\mathbf{R}$  is the rotation matrix which can be calculated from the rotation angles  $\boldsymbol{\theta} = [\theta_x, \theta_y, \theta_z]^T$  around each axis.  $\mathbf{t}$  is the translation vector  $\mathbf{t} = [t_x, t_y, t_z]^T$  along each axis. There are 6 parameters that should be estimated.

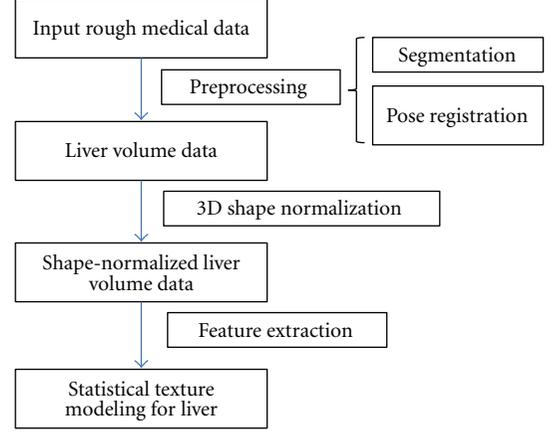


FIGURE 1: Basic scheme of statistical texture modeling.

The local motion is described by cubic B-spline-based free-form deformation (FFD) modeling [14, 15]. FFD is based on locally controlled functions such as the B spline and has been applied successfully to image registration. The basic idea of FFD is to deform an object by manipulating an underlying mesh of control points. B spline transformation is defined on a regular mesh of control points with uniform spacing. Let  $\boldsymbol{\rho} = [\rho_x, \rho_y, \rho_z]^T$  be the spacing of the control points along each axis. The coordinate of a control point can be expressed by

$$\boldsymbol{\varphi}_{ij} = [\varphi_{ijk,x}, \varphi_{ijk,y}, \varphi_{ijk,z}]^T = [i\rho_x, j\rho_y, k\rho_z]^T, \quad (3)$$

where  $i, j, k$  are the sequence number of the control points. Given the coefficients (translations) of the control points denoted as  $\boldsymbol{\lambda}_{ij} = [\lambda_{ijk,x}, \lambda_{ijk,y}, \lambda_{ijk,z}]^T$ , the B-spline transformation of a point  $\mathbf{x}$  can be expressed by

$$\begin{aligned} \mathbf{T}_{\text{Local}}(\mathbf{x}) = & \sum_{ijk} \boldsymbol{\lambda}_{ijk} \beta^{(3)}\left(\frac{x - \varphi_{ijk,x}}{\rho_x}\right) \beta^{(3)}\left(\frac{y - \varphi_{ijk,y}}{\rho_y}\right) \\ & \times \beta^{(3)}\left(\frac{z - \varphi_{ijk,z}}{\rho_z}\right), \end{aligned} \quad (4)$$

where  $\beta^{(3)}(a)$  is the third order cubic B-spline kernel. The coefficients of the control points,  $\boldsymbol{\lambda}_{ijk}$ , are the parameters of the B-spline transformation.

The parameters of global and local transformation are optimized separately [16]. We applied software in matlab named nonrigid B-spline grid image registration toolbox [17], which is based on FFD.

**2.2. GND-PCA Method.** Modeling for medical images is an important task in medical image analysis. The principal component analysis (PCA) method [18] is an efficient method for building statistical appearance models. In the PCA-based face representation and recognition methods, the 2D face image matrices must be previously transformed into 1D image vectors column by column [19]. Such an unfolding process causes two problems; one is the huge calculation cost and another is the poor performance to be generalized.

To overcome these problems, a new technique called 2-dimensional principal component analysis (2D-PCA) [20] has been proposed, which directly computes eigenvectors of the covariance matrix of the image without matrix-to-vector conversion. It was reported that the recognition accuracy with 2D-PCA on several face databases was higher than that with conventional 1D-PCA. However, the main disadvantage of 2D-PCA is that it needs many more coefficients than that with 1D-PCA for image representation. A method called generalized 2-dimensional principal component analysis (G2D-PCA) [21] has been proposed for finding the optimal basis for both row- and column-mode subspaces.

Recently, a method called N-dimensional PCA (ND-PCA) was proposed for high-dimensional data analysis [22]. In this method, the high-dimensional data are treated as a higher-order tensor which is directly trained to obtain the bases on one mode subspace by higher-order singular value decomposition (HOSVD) [23, 24]. This method was applied to 3D scanning data. Because ND-PCA only compresses the data on one mode subspace, it also suffered from the problem that the data cannot be represented efficiently, similar to the problem of 2D-PCA.

Inspired by the framework of generalized 2-dimensional principal component analysis [21] and N-dimensional principal component analysis [22], in our previous work, we proposed a method called generalized N-dimensional principal component analysis (GND-PCA). The high-dimensional data are treated as a series of higher-order tensors, and the optimal subspace on each mode is simultaneously calculated by minimizing the square error between the original tensor and the reconstructed tensor based on the subspace with an iteration algorithm.

*Algorithm 1.* GND-PCA is formalized as follows. Given a series of the  $N$ -order tensors with zero means,  $\mathcal{A}_i \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_N}$ ,  $i = 1, 2, \dots, M$ ,  $M$  is the number of samples. We hope to get another series of low-rank  $(J_1, J_2, \dots, J_N)$  tensors  $\mathcal{A}_i^*$  which accurately approximate the original tensors, where  $J_n \leq I_n$ . The new series is decomposed by the matrices  $\mathbf{U}^{(n)} \in \mathbf{R}^{I_n \times J_n}$  with orthogonal columns according to Tucker's model [24], which is shown by

$$\mathcal{A}_i^* = \mathcal{B}_i \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times \dots \times_n \mathbf{U}^{(n)} \times \dots \times_N \mathbf{U}^{(N)}, \quad (5)$$

where  $\mathcal{B}_i \in \mathbf{R}^{J_1 \times J_2 \times \dots \times J_N}$  are core tensors. An illustration of reconstructing a third-order tensor by three orthogonal bases is shown in Figure 2.

The orthogonal matrices  $\mathbf{U}^{(n)}$  can be determined by minimizing the cost function as

$$C = \sum_{i=1}^M \left\| \mathcal{A}_i - \mathcal{B}_i \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times \dots \times_n \mathbf{U}^{(n)} \times \dots \times_N \mathbf{U}^{(N)} \right\|^2. \quad (6)$$

Supposing that the rank of the  $N$  matrices  $\mathbf{U}^{(n)}$  is known, we use an iteration algorithm to obtain the  $N$  optimal matrices,  $\mathbf{U}_{\text{Opt}}^{(1)}, \mathbf{U}_{\text{Opt}}^{(2)}, \dots, \mathbf{U}_{\text{Opt}}^{(N)}$ , which are able to minimize the cost function  $C$ .

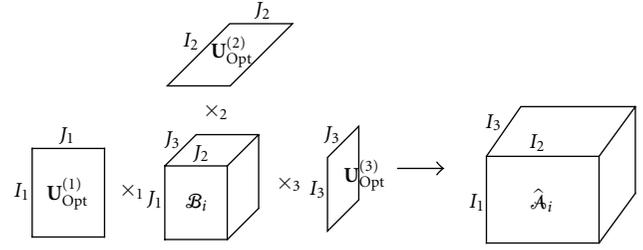


FIGURE 2: Reconstruction of a three-order tensor by the three orthogonal bases of mode subspace.

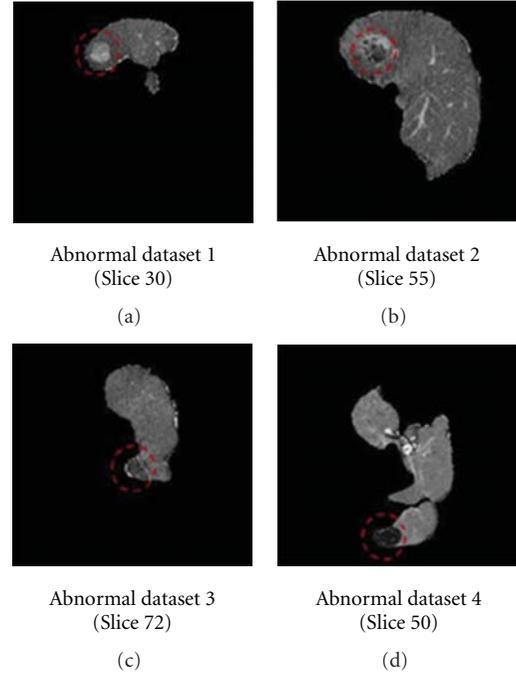


FIGURE 3: Slices of abnormal datasets.

Here, each matrix  $\mathbf{U}^{(n)}$  contains a set of basis vectors. An input sample can be calculated as a core tensor with the benefit of  $\mathbf{U}^{(n)}$ . This core tensor is the feature of the input sample.

Details about GND-PCA can be found in [13].

### 3. Experimental Results

*3.1. Datasets and Preprocessing Step.* The dataset we used to test the proposed method contains 23 abdominal CT scans collected from 23 patients, taken under similar conditions of illuminations and scanner setting. Each dataset obeys these conditions: slice thickness 2.5 mm, pitch 1.25 mm,  $256 \times 256$  matrix, and 79 slices. This dataset contains 19 cases with no radiologic finding (noted as *normal*) and 4 cases with radiologic finding (noted as *abnormal*). Figure 3 illustrates slices of abnormal datasets with tumors (red circles label the tumor positions).

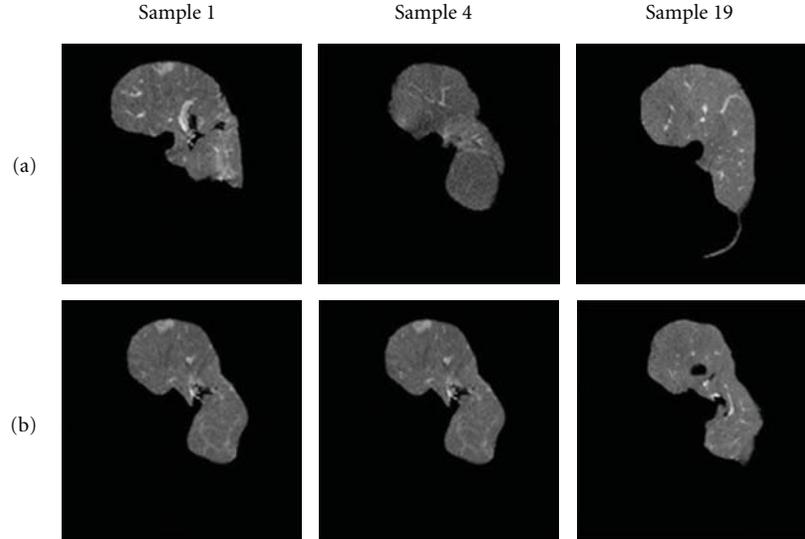


FIGURE 4: Example of datasets (some slices). (a) Original datasets. (b) 3D shape normalized datasets.

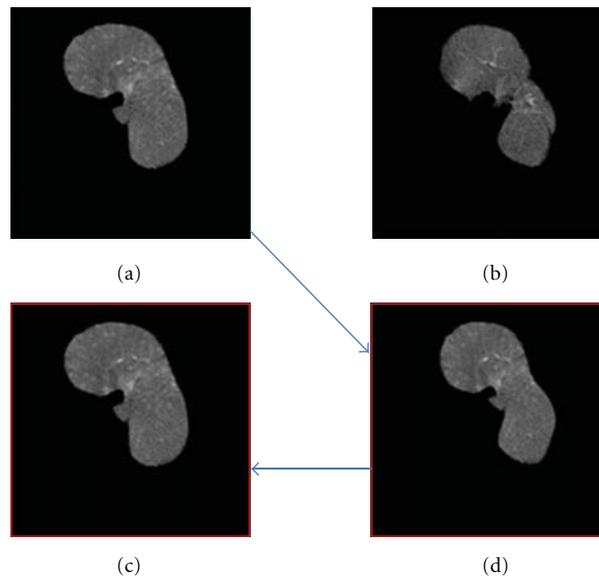


FIGURE 5: 3D shape normalization processing (one slice from volume data). (a) Moved image. (b) Fixed image. (c) Shape normalized moved image. (d) Inverse of 3D shape normalization.

The dimension of each sample is  $256 \times 256 \times 79$ . Initially livers are segmented manually from the datasets. Then we apply a rigid registration [9] for position normalization. Such pretreated datasets are noted as *original datasets*. As we mentioned in the previous section, we also apply a nonrigid registration to the dataset for both position and shape normalization to remove shape variations. The shape-normalized volumes are noted as *3D shape-normalized datasets*. Some original datasets and their 3D shape-normalized data are shown in Figure 4.

**3.2. 3D Shape Normalization Step.** We show the effectiveness of shape normalization in Figure 5. Here, Figure 5(a) is one slice of moved-volume dataset, and Figure 5(b) is the

corresponding slice of fixed-volume dataset. Figure 5(c) is the normalized slice of the moved volume dataset. In order to show that 3D shape-normalization processing causes little loss of texture information while interpolate the pixel values for shape deformation, we apply 3D shape normalization to the normalized moved-volume dataset again to transform it back to the original shape. Comparison of the inverse slice (Figure 5(d)) with the original slice (Figure 5(a)) shows that 3D shape normalization processing keeps almost all texture information. Thus, it is reasonable to apply 3D shape normalization as a preprocessing step to remove shape variations.

In our experiment, we chose the B-spline grid dimensions as (26 26 8), and we randomly chose one dataset as

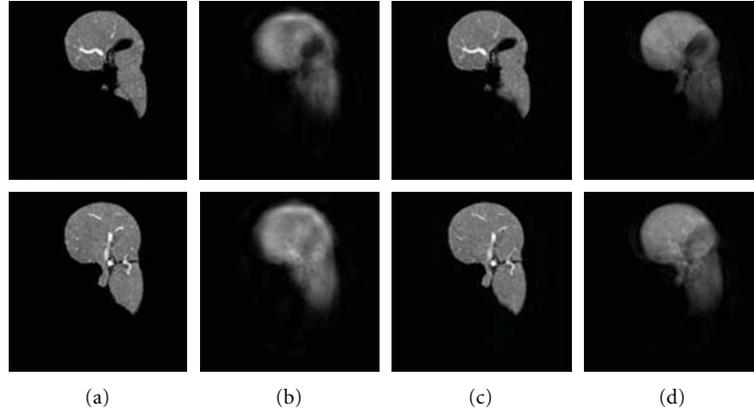


FIGURE 6: Reconstructed results for two slices of one test sample without shape normalization. (a) Original images. (b) Images reconstructed with  $10 \times 10 \times 4$  basis by GND-PCA. (c) Images reconstructed with  $100 \times 100 \times 40$  basis by GND-PCA. (d) Images reconstructed by PCA (eigenface method); all 15 available bases are used in the reconstruction.

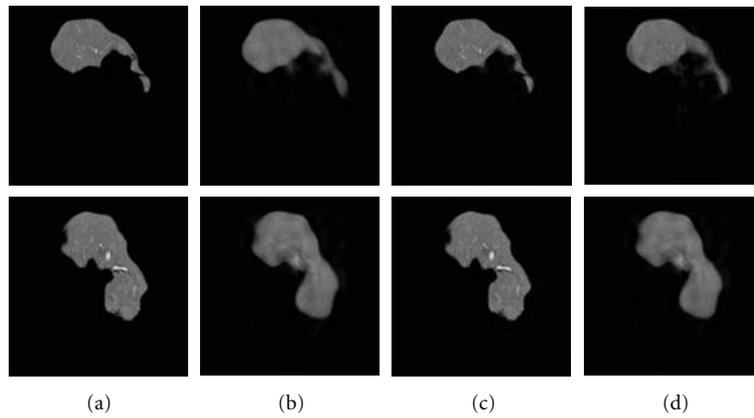


FIGURE 7: Reconstructed results for two slices of one test sample after shape normalization. (a) 3D shape-normalized images. (b) Images reconstructed with  $10 \times 10 \times 4$  basis by GND-PCA. (c) Images reconstructed with  $100 \times 100 \times 40$  basis by GND-PCA. (d) Images reconstructed by PCA (eigenface method); all 15 available bases are used in the reconstruction.

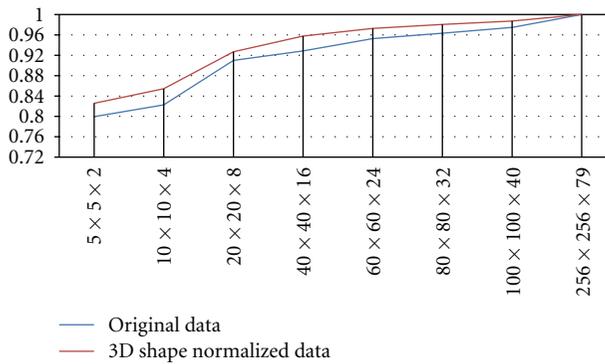


FIGURE 8: Normalized correlations while basis increasing.

the fixed volume and normalized the other dataset to the same shape.

**3.3. Modeling for Generalization.** The proposed GND-PCA is applied to both original and shape-normalized datasets. The

leave-one-out experiment is done to test the generalization ability of GND-PCA. As a small number of abnormal datasets of the liver, we randomly used 15 datasets to learn the optimal subspaces, and of the others the one left untrained was used as an input. Typical results are shown in Figures 6 and 7. The test volume was reconstructed from  $10 \times 10 \times 4$  and  $100 \times 100 \times 40$  mode-subspace bases by GND-PCA, respectively. Figure 8 illustrated that the reconstructed images were improved by an increase in the subspace basis. In spite of having very few samples, we still could obtain an almost perfect reconstruction with  $100 \times 100 \times 40$  basis. In order to make a comparison, we also show the reconstructed results by the conversational PCA (eigenface) method in Figures 6(d) and 7(d), which show that the quality of the reconstructed results are not satisfied even though the entire 15 available bases are used for reconstruction because of overfitting.

The normalized correlations between the original volume and the reconstructed volume are shown in Figure 8. Compared with in the case of the original dataset, the datasets can be represented by a small number of bases in

TABLE 1: Euclidean distances of features.

	Original data		3D shape-normalized data	
	ED	ED-LDT	ED	ED-LDT
LDT	26984	0	13817	0
Normal dataSet_1	29101	2117.5	13339	-477.55
Normal dataSet_2	20394	-6589.4	8474.4	-5342.3
Normal dataSet_3	16811	-10173	11872	-1944.7
Normal dataSet_4	21584	-5399.6	9432.6	-4384
Abnormal dataSet_1	25896	-1087.8	18400	4583.8
Abnormal dataSet_2	29633	2649.2	17314	3497.1
Abnormal dataSet_3	23303	-3680.3	16502	2685.6
Abnormal dataSet_4	30405	3421.6	19241	5424

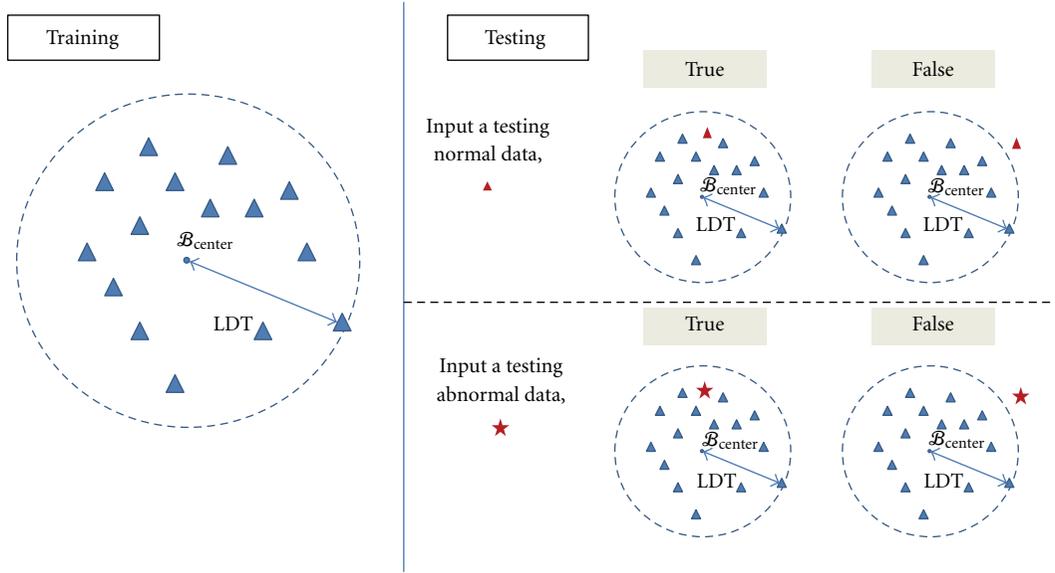


FIGURE 9: The rule for judging a test sample as normal or abnormal (2D case as example).

the case of shape-normalized dataset because the subspace contains only texture variations.

**3.4. Modeling for Discrimination.** Next, we introduce a simple experiment to show that the features extracted by our methods have the capability for discrimination. We used only 15 normal datasets for training and left the other 8 datasets for testing. The testing samples included 4 normal datasets and 4 abnormal datasets. After we obtained the optimal subspace by the GND-PCA method, each sample was represented by a core tensor. The core tensor is a feature of the sample and is noted as  $\mathcal{B}_i$ . We also calculate the mean feature of all of the training datasets and noted it as  $\mathcal{B}_{\text{center}}$ . Here, the dimension of the core tensor is  $100 \times 100 \times 40$ .

The Euclidean distance (ED) is applied to the calculation of the distance between  $\mathcal{B}_i$  and  $\mathcal{B}_{\text{center}}$ . Table 1 shows the ED for all the testing samples. Compared with those in the original datasets, the distances decreased in the shape-normalized dataset's experiments. We demonstrated that shape variations are removed by 3D shape normalization.

Next, we showed how to identify the normal datasets and abnormal datasets. The features captured by our method

are tensor formed; they can be flattened as high-dimensional vectors. In order to separate the features into two classes: normal and abnormal, we need to find a high-dimensional hyperplane. It is difficult to describe the hyperplane in high-dimensional space; we use Figure 9 as a 2D case to show how to find a hyperplane. Compared with the normal datasets, abnormal datasets have some significant parts in texture. If we do not consider the effect of shape, the significant parts caused a higher value of ED for abnormal datasets because we only used normal samples for training. We used the largest ED of the training sample (LDT), which is also shown in Table 1, as a boundary of normal and abnormal for classification. Table 2 gives the classified results for two kinds of dataset experiments. It demonstrates that the features extracted by our method have better performance for discriminations between the normal and abnormal classes.

## 4. Conclusion

In this paper, we developed a statistical texture modeling method for medical volumetric images based on 3D shape normalization and GND-PCA. We first propose to

TABLE 2: Classification result.

	Class	Test sample number	Correct classified number	Accuracy
Original data experiment	Normal	4	3	75%
	Abnormal	4	2	50%
3D shape normalized data experiment	Normal	4	4	100%
	Abnormal	4	4	100%

use a 3D shape normalization technique to normalize all volume datasets to the same shape to obtain the 3D shape-normalized datasets, which can be considered to contain only the texture variations. Then we trained them to construct the statistical model only for texture by GND-PCA method for application to liver volumes. Reconstruction results show a good performance on generalization by using our proposed method. We also designed a simple experiment to identify different types of data with corresponding features, such as normal and abnormal, which proved that the proposed model can be used for computer-assisted diagnostics of liver disease. In the future, we will test our method with more datasets for classification and use our method in practical applications.

## Acknowledgments

The authors are grateful to Kenji Suzuki, Ph.D., for his valuable suggestions and to E. F. Lanzl for improving the manuscript. This work was supported by the Japan Society for the Promotion of Science (JSPS). This work was also supported in part by a Grant-in-Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports under Grants no. 21300070 and 22103513, and in part by a Research fund from the Ritsumeikan Global Innovation Research Organization (R-GIRO).

## References

- [1] F. H. Netter, *Atlas of Human Anatomy*, WB Saunders, 4th edition, 2006.
- [2] K. H. Hohne, B. Pflesser, A. Pommert et al., "A new representation of knowledge concerning human anatomy and function," *Nature Medicine*, vol. 1, no. 6, pp. 506–511, 1995.
- [3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [4] H. C. Van Assen, M. G. Danilouchkine, F. Behloul et al., "Cardiac LV segmentation using a 3D active shape model driven by fuzzy inference," in *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, vol. 2878 of *Lecture Notes in Computer Science*, pp. 533–540, Montréal, Canada, November 2003.
- [5] M. R. Kaus, J. von Berg, J. Weese, W. Niessen, and V. Pekar, "Automated segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 8, no. 3, pp. 245–254, 2004.
- [6] T. Okada, R. Shimada, M. Hori et al., "Automated segmentation of the liver from 3D CT images using probabilistic atlas and multi-level statistical shape model," *Academic Radiology*, vol. 15, no. 11, pp. 1390–1403, 2008.
- [7] H. Huang, F. Makedon, and R. McColl, "High dimensional statistical shape model for medical image analysis," in *the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '08)*, pp. 1541–1544, May 2008.
- [8] M. R. Kaus, J. von Berg, J. Weese, W. Niessen, and V. Pekar, "Automated segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 8, no. 3, pp. 245–254, 2004.
- [9] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. Van der Geest, J. H. C. Reiber, and M. Sonka, "3-D active appearance models: segmentation of cardiac MR and ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1167–1178, 2002.
- [10] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [11] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 1014–1025, 2003.
- [12] S. Kohara, A. H. Foruzan, T. Tateyama et al., "Evaluation of statistical shape model based classification performance for liver disease of cirrhosis," Tech. Rep. PRMU 2010-154, IEICE, 2011.
- [13] R. Xu and Y. W. Chen, "Generalized N-dimensional principal component analysis (GND-PCA) and its application on construction of statistical appearance models for medical volumes with fewer samples," *Neurocomputing*, vol. 72, no. 10–12, pp. 2276–2287, 2009.
- [14] W. M. Hsu, J. F. Hughes, and H. Kaufman, "Direct manipulation of free-form deformations," *Computer Graphics (ACM)*, vol. 26, no. 2, pp. 177–184, 1992.
- [15] D. Rueckert, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [16] L. Ibanez et al., "The ITK software guide," <http://www.itk.org/>.
- [17] D. J. Kroon, "Non-Rigid B-Spline Grid Image Registration," <http://www.mathworks.co.jp/matlabcentral/fileexchange>.
- [18] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2002.
- [19] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [20] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.
- [21] H. Kong, L. Wang, E. K. Teoh, X. Li, J. G. Wang, and R. Venkateswarlu, "Generalized 2D principal component analysis for face image representation and recognition," *Neural Networks*, vol. 18, no. 5-6, pp. 585–594, 2005.
- [22] H. Yu and M. Bennamoun, "1D-PCA, 2D-PCA to nD-PCA," in *the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 181–184, August 2006.

- [23] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multi-linear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [24] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-( $R_1, R_2, \dots, R_N$ ) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

## Research Article

# Sparse Representation of Deformable 3D Organs with Spherical Harmonics and Structured Dictionary

Dan Wang,<sup>1</sup> Ahmed H. Tewfik,<sup>1</sup> Yingchun Zhang,<sup>2</sup> and Yunhe Shen<sup>2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA

<sup>2</sup>Department of Urology Surgery, University of Minnesota, Minneapolis, MN 55455, USA

Correspondence should be addressed to Dan Wang, wang0889@mail.utexas.edu

Received 3 March 2011; Revised 28 June 2011; Accepted 29 June 2011

Academic Editor: Kenji Suzuki

Copyright © 2011 Dan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposed a novel algorithm to sparsely represent a deformable surface (SRDS) with low dimensionality based on spherical harmonic decomposition (SHD) and orthogonal subspace pursuit (OSP). The key idea in SRDS method is to identify the subspaces from a training data set in the transformed spherical harmonic domain and then cluster each deformation into the best-fit subspace for fast and accurate representation. This algorithm is also generalized into applications of organs with both interior and exterior surfaces. To test the feasibility, we first use the computer models to demonstrate that the proposed approach matches the accuracy of complex mathematical modeling techniques and then both *ex vivo* and *in vivo* experiments are conducted using 3D magnetic resonance imaging (MRI) scans for verification in practical settings. All results demonstrated that the proposed algorithm features sparse representation of deformable surfaces with low dimensionality and high accuracy. Specifically, the precision evaluated as maximum error distance between the reconstructed surface and the MRI ground truth is better than 3 mm in real MRI experiments.

## 1. Introduction

Organ deformation during operations has imposed substantial challenges for performing precise diagnosis and surgery in minimum invasive surgery (MIS), such as natural orifice transluminal endoscopic (NOTES) [1]. The deformation markedly decreases the precision of the prior surgical plan that is based on the preoperative images (e.g. computed tomography (CT) or magnetic resonance imaging (MRI)), so it must be effectively compensated to lower surgical risks. However, this is not a trivial task due to the high degree of freedom of the 3D nonrigid deformation and limited field of view [1] for observation during MIS. To recovery, the 3D deformation with high resolution in real time, a critical issue is to seek an efficient representation of deformable surface, according to which the sampling and surface recovery strategy can be designed for updating the 3D visualization. This paper focuses on the topic of block sparse representation of deformable surfaces. The later topic of real time tracking an deformable organ with limited access to the organ is explored further in [2–4].

Various techniques have been proposed for surface description, and each has its own advantage and disadvantage according to the application requirements. Broadly speaking, there are two major categories of surface representation methods: local feature-based models and global or parametric models. The work in [5] based on geometric partial differential equations (PDE) belongs to the former category which derives Euler-Lagrange equation and then a geometric evolution equation (or geometric flow) to describe the surfaces. Similarly, the method in [6] treats the whole surface as a union of localized patches. Global surface representation methods [7–10], particularly, decomposing surfaces into other primitive shapes, are more appropriate for shape analysis and classification due to the lower dimensionality of the parameter space. This paper falls into the category of parametric global surface description.

Parametric surface representation describes a surface in a single functional form, such that the surface is fully characterized by a set of parameters in a particular domain. Surface harmonics such as spheroidal harmonics, cylindrical harmonics, and spherical harmonics (SH) [7] are widely

used as building blocks for global surface description. Each harmonic does not bear localized features but contributes to the entire shape description. Among those different types of harmonics, a well-known approach is the spherical harmonic decomposition (SHD), which has advantages of smoothness and high accuracy [7, 11]. With proper parameterization [11, 12], any genus-0 surface (The genus of a connected surface is an integer representing the maximum number of cuttings along nonintersecting closed simple curves without rendering the resultant manifold disconnected. It is equal to the number of handles on it, so a sphere is genus 0 and a torus is genus 1.) can be analyzed in the harmonic domain with reduced data dimensions. SHD has been widely used in applications related to surface description, including static modeling of kidney [8] and brain [9, 10], as well as spatial-temporal modeling of left-ventricular with known motion period. In [13], hemisphere is also applied to open surfaces.

Sparse signal representation has steadily gained attention over the years in the signal processing community. The aim is to find a representation which is sparse, or compact, such that most of the energy of a signal can be captured with only a few nonzero coefficients in a given dictionary. The first widely applied methods to seek sparse representation are greedy approaches, including matching pursuit (MP) [14], orthogonal matching pursuit (OMP) [15], and orthogonal least square (OLS) [16]. Those methods iteratively first select the most correlated element from dictionary and then remove the contribution of that element with decorrelation, before finding the next atom. Iteration terminates when any stopping criteria is met. The second type is global optimization algorithms, such as Basis Pursuit (BP) [17], FOCUSS [18] and Iterative Thresholding [19]. Global Optimization, in the approximate sense, relaxes the sparseness constraint, and its sparsity is a side-effect of the optimization. For example, the basis pursuit (BP) method approximates the  $l_0$ -norm sparsity constraint with an  $l_1$ -norm criteria, which effectively converts the problem into a convex optimization one, solved globally with linear programming. The orthogonal subspace pursuit (OSP) method [20] used in our paper belongs to the greedy category, which does not require prior knowledge of the dimension of the subspaces and combines the learned subspaces to produce a data-driven dictionary with good sparseness and generalizability.

Besides sparse decomposition algorithms as mentioned above, an equally important issue for sparse representation is how to select a dictionary for an application. The two main groups for dictionary design methods are structured dictionaries built out of common bases, and trained dictionaries that are inferred from the training data. For the common bases, it is well known that the wavelet transform can be used to generate sparse multiscale representations of images, the short-time Fourier transform (STFT) generates sparse time-frequency representation of speech signals, and the DCT is another transform that has been used for compression in audio coding algorithms due to its good compaction property. For dictionary learning, the applicable approaches include maximum likelihood estimation (MLE) [21], method of optimal directions (MOD) [22], maximum a-posteriori (MAP) [23], and so forth. Those methods

attempt to generalize the type of considered signal with the basis identified from a representative training data set. The proposed approach is based on trained dictionary, since, to the best of our knowledge, there is no common basis in which random surfaces can be sparsely represented.

Although sparse representation has been widely applied in the fields of signal compression, image denoising, blind source separation, and compressed sensing, there is still very limited application in 3D surfaces [24]. In fact, the main statistically motivated surface modeling methods are based on principle component analysis which is not sparse [25–29]. Those methods first compute the mean shape and then build the model by establishing legal variations learned from a set of training data for a given type of images, such as bone [25]. With PCA, the major variations of the shape populations are described by the first few basis vectors, such that any surface of that shape population can be projected into an orthogonal subspace spanned by the retained vectors. More advanced techniques, such as multiresolution deformable model [27], are provided to improve the accuracy considering limited sample size.

Most of the previous surface modeling/representation works are designed for either static models [8–10], or particular deformable organs with known physical properties (such as motion cycle) [30]. Further, computation bottleneck caused by large spherical harmonics basis hampered the applicability for real-time application. For PCA-based modeling methods, the resulting space that captures the variation in the population is either a super subspace including all training data or a truncated subspace with sacrificed generalization. In addition, PCA tends to be computational expensive when performing eigenvector decomposition as training data dimension increases, and it does not lead to any structure in the representation.

To bring the demonstrated merits of sparse coding to 3D surface representation, we propose a generally applicable algorithm of parametric sparse representation of deformable surfaces based on SHD and orthogonal subspace pursuit. The main contributions of this paper include the following.

- (i) Propose an algorithm of sparse representation of deformable surfaces.
- (ii) Generalize the representation approach for organs involving both interior and exterior surfaces.
- (iii) Present evaluation results conducted using computer models, ex vivo experiments based on 3D MRI scans of freshly excised porcine kidneys, and in vivo cardiac MRI scans of real patients.

This paper is organized as follows: in Section 2, we describe the proposed algorithm of sparse representation of deformable surface, denoted as SRDS thenceforth. Section 3 presents some experimental results using finite element model (FEM) data, ex vivo and in vivo experimental data. Finally, in Section 4, we finish with a few conclusions.

## 2. The SRDS Algorithm

The SRDS algorithm consists of three main steps to achieve sparse representations of deformable surfaces, as outlined in Figure 1. Initially, SHD is performed to depict the deformable surfaces in the training set in the harmonic domain. Then OSP is applied in the transformed domain to identify the subspaces in which the SH coefficient vectors of the deformations can be linearly represented. Finally, each deformation is clustered to the proper subspace and represented with the corresponding coefficient vector with block sparsity. This representation method is also extended to organ deformations occurred on both interior and exterior layers as described in Section 2.4. Furthermore, as a practical issue, pixel-wise surface alignment among all the 3D surfaces is also addressed in Section 2.5.

*2.1. Step 1: SHD.* Spherical harmonics are solutions to Laplace's equation expressed in the spherical coordinate system, defined as

$$Y_{lm}(\theta, \varphi) = (-1)^m \sqrt{\frac{2l+1}{4\pi}} \sqrt{\frac{(l-m)!}{(l+m)!}} P_{lm}(\cos \theta) e^{im\varphi}, \quad (1)$$

where  $\theta$  is the polar angle within  $[0, \pi]$ ,  $\varphi$  is the azimuthal angle within  $[0, 2\pi]$ ,  $l$  is the harmonic degree within  $[0, +\infty]$ , and  $m$  is the harmonic order varying in  $[-l, l]$ .  $P_{lm}$  is the associated Legendre function. After proper parameterization [11, 12], a 3D surface  $\mathbf{x}$  with finite energy can be expanded with SH series as

$$\mathbf{x}(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} f_{lm} Y_{lm}(\theta, \varphi). \quad (2)$$

Each harmonic coefficient  $f_{lm}$  is calculated using the inner product of the function  $\mathbf{x}(\theta, \varphi)$  and basis  $Y_{lm}(\theta, \varphi)$

$$f_{lm} = \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} \mathbf{x}(\theta, \varphi) Y_{lm}(\theta, \varphi) \sin \theta \, d\varphi \, d\theta. \quad (3)$$

Assume that harmonics up to level  $L$  ( $l \leq L$ ) are involved in the transformation. Let  $\mathbf{Y}$  denote the matrix composed of all  $(L+1)^2$  discretized harmonics, so  $\mathbf{Y}$  has the following formation:

$$\mathbf{Y} = \left[ \vec{Y}_{0,0} \quad \vec{Y}_{1,-1} \quad \vec{Y}_{1,0} \quad \vec{Y}_{1,1} \quad \cdots \quad \vec{Y}_{L,L} \right]_{N \times (L+1)^2}. \quad (4)$$

Then a surface can be represented in the matrix format as

$$\mathbf{x} = \mathbf{Y} \vec{f}, \quad (5)$$

where  $\mathbf{x}$  stands for a surface with  $N$  samples and  $\vec{f} = [f_{0,0} \quad f_{1,-1} \quad f_{1,0} \quad f_{1,1} \quad \cdots \quad f_{L,L}]^T$  is the harmonic coefficient vector. Notice that this equation is not exactly equal but approximately. For simplicity, we still use equal sign with least square estimation in this paper. The linear problem in (5) can be solved with the least square (LS) constraints outputting  $\vec{f}$

$$\vec{f} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{x}. \quad (6)$$

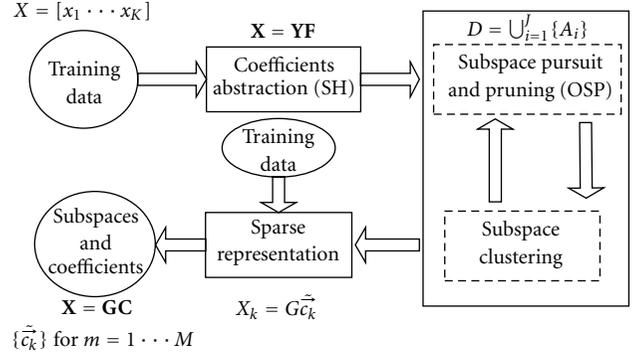


FIGURE 1: Flowchart of SRDS algorithm consists of three steps specified in the solid rectangulars; ellipsoids denote data input and output.

Perform SHD for each of the  $K$  training deformed surfaces in  $\mathbf{X} = \{\mathbf{x}_k\}_{k=1}^K$ , so the group of deformations can be described by matrix  $\mathbf{F}$

$$\mathbf{F} = \left[ \vec{f}_1 \quad \vec{f}_2 \quad \cdots \quad \vec{f}_K \right]_{(L+1)^2 \times K}. \quad (7)$$

Consequently, the training set of deformations can be totally characterized by columns in  $\mathbf{F}$

$$\mathbf{X} = \mathbf{Y}\mathbf{F}. \quad (8)$$

*2.2. Step 2: Subspace Identification with OSP.* The aim of the second step is to explore the structures in those training deformations in the transformed harmonic domain and recognize the inherent subspaces in which the SHD coefficient vectors of the training deformations can be projected with high accuracy. The newly developed OSP algorithm [20] is adopted since it features better generalization and less computational cost compared to the gold standard K-SVD algorithm [31]. OSP is an iterative process that terminates when one of the predefined criteria is met. In this paper, we specify the following two stopping criteria: (1) an error threshold for  $\varepsilon$  subspaces selection and (2) a maximum number of iterations  $E_{\max}$  for both controlling the subspace dimensions and avoiding deadlock searching. Further, the threshold for vector clustering is denoted as  $\eta$ , that is, we declare that a vector lives in a subspace if it can be projected to that subspace with error ( $l_2$  distance) no larger than  $\eta$ .

*A. Subspace Pursuit.* Initially, each vector  $\vec{f}_k$  of length  $(L+1)^2$  in  $\mathbf{F}$  is normalized by  $l_2$  norm. For convenience, we still use  $\mathbf{F}$  to denote SH coefficient matrix even after normalization. The algorithm first identifies a subspace from  $\mathbf{F}$  based on the stop criteria, then finds all the vectors in  $\mathbf{F}$  that can be represented in that subspace with error level below  $\eta$  and remove those vectors from  $\mathbf{F}$  to prepare for the next subspace pursuit. The process can be generalized as follows, in which  $A \otimes B$  means that elements from  $B$  are excluded from  $A$ , and  $A \oplus B$  stands for inclusion.

- (1) Initialization:  $i = 0$ ,  $\mathbf{D} = \emptyset$ ,  $\mathbf{F}^0 = \mathbf{F}$ ,
- (2) Subspace searching and clustering
  - (i)  $i = i + 1$ ; choose a vector  $\vec{f}_i$  from  $\mathbf{F}^i$  (e.g. first column of  $\mathbf{F}^i$ ) and let  $\mathbf{F}^i = \mathbf{F}^{i-1} \otimes \vec{f}_i$  to remove  $\vec{f}_i$ .
  - (ii) Find  $n_i$  vectors from  $\mathbf{F}^i = \mathbf{F} \otimes \vec{f}_i$  for representing  $\vec{f}_i$  with error no larger than  $\varepsilon$  within  $E_{\max}$  iterations and the  $n_i$  vectors form  $S_i$ .
  - (iii) Perform SVD decomposition on  $S_i$ :  $U\Sigma V^T = S_i$ ; let  $\mathbf{A}_i$  contain the first  $n_i$  vectors of  $U$ ; update  $\mathbf{D} = \mathbf{D} \oplus \mathbf{A}_i$ .
  - (iv) Select vectors from  $\mathbf{F}^i$  that can be represented by  $\mathbf{A}_i$  with error no larger than  $\eta$ , and then remove them from  $\mathbf{F}^i$ .
  - (v) Repeat above steps until all the vectors are clustered.

*B. Subspace Pruning.* One disadvantage of the traditional OSP algorithm is the presence of “spurious” or redundant subspaces especially as the dimension of the training data set increases. Those subspaces identified in the earlier iterations actually can be better represented by the later identified subspaces. Therefore, a postprocessing step is used to identify and then discard the redundancy among the subspaces without decreasing the performance. This is implemented by repartitioning the training data among the initial subspaces and then eliminating subspaces in which very few or no vector is clustered. In some cases, where the subspace size is limited to some constraint, an optimization step can be applied in conjunction with the pruning step. The details of the subspace optimization design is described in [20].

*C. Matrix  $\mathbf{F}$  Factorization.* After identifying the inherent subspaces, the coefficient matrix  $\mathbf{F}$  of training set can be partitioned into two-part union of subspaces and corresponding coefficients via the following procedures.

Since each vector  $\vec{f}_k$  has been clustered into the belonging subspace during the subspace identification process, the corresponding coefficients for each  $\vec{f}_k$  can be obtained accordingly. Suppose that  $\vec{f}_k$  lives in subspace  $\mathbf{A}_i$  which is spanned by  $n_i$  orthogonal basis, so its corresponding coefficients  $\vec{c}_k$  can be calculated using LS estimator

$$\vec{c}_k = (\mathbf{A}_i^T \mathbf{A}_i)^{-1} \mathbf{A}_i^T \vec{f}_k. \quad (9)$$

Then  $\vec{f}_k$  can be characterized by  $\vec{c}_k$  in its subspace

$$\vec{f}_k = \mathbf{A}_i \vec{c}_k. \quad (10)$$

If there are totally  $J$  subspaces identified from  $\mathbf{F}$ , a structured dictionary constructed by concatenating all deformation subspaces is established as  $\mathbf{D} = \bigcup_{i=1}^J \{\mathbf{A}_i\}$ , with dimension  $I = \sum_{i=1}^J n_i$ . Since each vector  $\vec{f}_k$  lies in one of the subspaces,  $\vec{f}_k$  can also be represented in the structured

dictionary with a block sparse vector  $\{\vec{c}_k\}$ , which is obtained via extending the coefficients  $\{\vec{c}_k\}_{k=1}^K$  by zero padding in positions corresponding to other subspaces in  $\mathbf{D}$ . Figure 2 provides an example of 3 subspaces to illustrate the sparsity of coefficient vector  $\vec{c}_1$ . If  $\vec{f}_1$  lies in subspace  $\mathbf{A}_2$  which are spanned by the 5th, 6th, and 7th columns in  $\mathbf{D}$ , then  $\vec{c}_1$  has nonzero values only at index of 5, 6, and 7. Consequently,  $\mathbf{F}$  can be factorized as

$$\mathbf{F} = \mathbf{D}\mathbf{C}, \quad (11)$$

where  $\mathbf{C} = |\vec{c}_1 \ \vec{c}_2 \ \cdots \ \vec{c}_K|_{I \times K}$  is the corresponding coefficient matrix with block sparsity.

### 2.3. Step 3: Structured Sparse Surface Representation

*A. Sparse Representation of Training Surfaces.* Integrating the subspace pursuing results in the harmonic domain in (11) with the initial SHD process in (8), the training deformations  $\mathbf{X}$  can be sparsely represented in the original spatial domain as

$$\begin{aligned} \mathbf{X} &= \mathbf{Y}\mathbf{D}\mathbf{C} \\ &= \mathbf{G}\mathbf{C}, \end{aligned} \quad (12)$$

where  $\mathbf{G} = \mathbf{Y}\mathbf{D}$  with size of  $N \times I$  is the desired structured dictionary in the spatial domain. Since  $\mathbf{D} = \bigcup_{i=1}^J \{\mathbf{A}_i\}$ ,  $\mathbf{G}$  is inherently structured by subspaces of  $\mathbf{G} = \bigcup_{i=1}^J \{\mathbf{G}_i\}$  with  $\mathbf{G}_i = \mathbf{Y}\mathbf{A}_i$  of size  $N \times n_i$ .

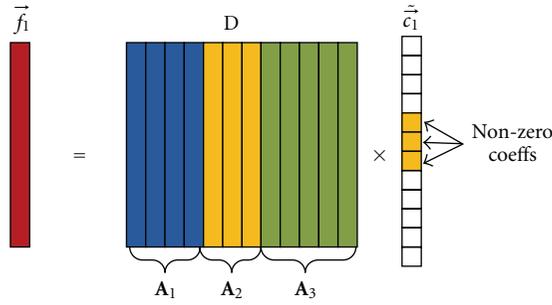
Up to this point, with matrix  $\mathbf{G}$  that captures the deformation features in the considered population, each training deformation  $\mathbf{x}_k$  in  $\mathbf{X}$  can be fully characterized by a sparse coefficient  $\vec{c}_k$  as

$$\mathbf{x}_k = \mathbf{G} \vec{c}_k. \quad (13)$$

The sparsity of  $\vec{c}_k$  has already been illustrated in Figure 2.

*B. Sparse Representation of Testing Surfaces.* For the testing deformations beyond the training set, we utilized the fact that the dictionary identified from an extensive training data features good generalization such that any deformation of that particular population can be represented in the subspaces with high accuracy. This is justified because organs only deform in limited ways due to their mechanical properties, so the deformation variations can be fully learned from a training data set. This applied structure allows fast deformation representation in subspaces of low dimensionality.

The testing set is denoted as  $\mathbf{H} = \{\mathbf{h}_m\}_{m=1}^M$ , where  $M$  is the number of deformations to be represented. The straightforward strategy is to find a best-fit subspace for  $\mathbf{h}_m$  by projecting it to every subspace  $\{\mathbf{G}_i\}_{i=1}^J$  and choose the subspace with minimal projection error. Since the number of subspaces  $J$  and the dimension of each subspace  $\{n_i\}_{i=1}^J$  are both small with the postprocessing of subspace pruning, this best-fit strategy still results in low computational cost. However, when the number of subspaces is too large, an alternative threshold approach can be applied by finding

FIGURE 2: Sparsity of coefficient vector  $\tilde{c}_1$ .

a subspace  $\mathbf{G}_i$  in which  $\mathbf{h}_m$  can be represented with an error level smaller than  $\eta$ . The former best-fit method is implemented in Section 3 for performance validation.

Suppose that  $\mathbf{G}_{i^*}$  is the chosen subspace, so coefficient vector  $\tilde{c}_m$  can be estimated with LS as

$$\tilde{c}_m = (\mathbf{G}_{i^*}^T \mathbf{G}_{i^*})^{-1} \mathbf{G}_{i^*}^T \mathbf{h}_m. \quad (14)$$

Then block sparse vector

$$\tilde{c}_m = [0 \ 0 \ \cdots \ \tilde{c}_m \ \cdots \ 0 \ 0 \ 0] \quad (15)$$

is obtained according to the rules described in Section 2.2. Further, the sparsity of  $\tilde{c}_k$  or  $\tilde{c}_m$  can be increased by trimming off nonzero elements with absolute value lower than a given threshold  $\delta$ .

It is worth noting that, different from the traditional learning approaches relying on orthogonal least square (OLS) [32] or matching pursuit (MP) [33] algorithms which select atoms from the training set and recombine them for representing each surface in the testing set sparsely, our SRDS algorithm avoids this heavy overload caused by reshuffling all the atoms. Instead, we apply the block structure of the dictionary learned from a representative training data set. This essentially enables the representation of each deformable surface compactly and sparsely with high accuracy and low computational cost.

**2.4. Extended Sparse Surface Representation.** For an organ with both interior and exterior surfaces, such as bladder, deformations can take place on both layers. The above theory can be extended to achieve sparse surface representation for deformations occurred on both interior and exterior wall of the object.

Initially, spherical parameterization is conducted on interior and exterior parts separately. We denote  $\tilde{x}_k^{\text{in}}$  and  $\tilde{x}_k^{\text{ex}}$  ( $1 \leq k \leq K$ ) as the corresponding interior and exterior of each training surface  $\tilde{x}_k$ . Then each pair of  $\tilde{x}_k^{\text{in}}$  (with  $N_1$  vertices) and  $\tilde{x}_k^{\text{ex}}$  (with  $N_2$  vertices) can be approximated by spherical harmonic basis as

$$\tilde{x}_k = \begin{bmatrix} \widehat{\tilde{x}_k^{\text{in}}} \\ \widehat{\tilde{x}_k^{\text{ex}}} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}^{\text{in}} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}^{\text{ex}} \end{bmatrix} \cdot \begin{bmatrix} \tilde{f}_k^{\text{in}} \\ \tilde{f}_k^{\text{ex}} \end{bmatrix} \quad (16)$$

where  $\mathbf{Y}^{\text{in}}$  of size  $N_1 \times (L+1)^2$  and  $\mathbf{Y}^{\text{ex}}$  of size  $N_2 \times (L+1)^2$  denote the spherical harmonic basis for inner and outer surfaces, respectively.  $L$  is the highest degree of harmonics included.  $\tilde{f}_k^{\text{in}}$  and  $\tilde{f}_k^{\text{ex}}$  are the corresponding harmonic coefficient vectors. Therefore, each deformation is represented by vector  $\tilde{f}_k = \begin{bmatrix} \tilde{f}_k^{\text{in}} \\ \tilde{f}_k^{\text{ex}} \end{bmatrix}$ , and all  $K$  training frames can be characterized by  $F = \cup \{\tilde{f}_k\}_{k=1}^K$  as

$$\mathbf{X} = \mathbf{YF} = \begin{bmatrix} \mathbf{Y}^{\text{in}} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}^{\text{ex}} \end{bmatrix} \begin{bmatrix} \tilde{f}_1^{\text{in}} & \cdots & \tilde{f}_K^{\text{in}} \\ \tilde{f}_1^{\text{ex}} & \cdots & \tilde{f}_K^{\text{ex}} \end{bmatrix}. \quad (17)$$

The following procedures of subspace identification and sparse surface representation as described in Sections 2.2 and 2.3 can be applied straightforwardly. After identifying  $J$  subspaces  $\mathbf{D} = \cup_{i=1}^J \{\mathbf{A}_i\}$  from SH coefficient matrix  $\mathbf{F}$ , each training deformation can be sparsely represented with block sparse coefficient vector  $\tilde{c}_k$  as:

$$\begin{aligned} \tilde{x}_k &= \begin{bmatrix} \widehat{\tilde{x}_k^{\text{in}}} \\ \widehat{\tilde{x}_k^{\text{ex}}} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}^{\text{in}} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}^{\text{ex}} \end{bmatrix} \mathbf{D} \tilde{c}_k \\ &= \mathbf{G} \tilde{c}_k = \mathbf{G}_i \tilde{c}_k, \end{aligned} \quad (18)$$

where  $\mathbf{G}_i = \begin{bmatrix} \mathbf{Y}^{\text{in}} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}^{\text{ex}} \end{bmatrix} \mathbf{A}_i$  is the subspace with size of  $(N_1 + N_2) \times n_i$ , and  $\mathbf{G} = \cup_{i=1}^J \{\mathbf{G}_i\}$  is the desired structured dictionary. Accordingly,  $\tilde{c}_k$  is the block sparse vector, and  $\tilde{c}_k$  is the nonzero coefficient values in the selected subspace.

**2.5. Surface Correspondence.** Similar to other surface modeling methods [25, 27, 28], the proposed approach requires point-wise correspondence across difference surfaces besides rigid registration [34]. Specifically, this point-to-point alignment is established, such that an identical spherical parameterization can be applied in the SHD procedure. Figure 3 illustrates the goal of surface correspondence. Same colored vertices on deformation 1 and 2 indicate a matched pair. After established correspondence of the point pairs over the two deformations, vertices on deformation 2 can be numbered in the same order as deformation 1.

Different correspondence methods have been proposed, such as minimum description length [35], SH-coefficient alignment [36], and so forth. We applied the SH based method [36] in this paper as well as ray-casting method for simple surfaces. The former SH-based method is based on the underlying fact that two points with the same parameter pair when mapped to a sphere are considered to be a corresponding pair. Therefore, it fixes parameterization of the template and rotate the other to optimize the surface correspondence by minimizing the root mean squared distance of the two SH coefficient vectors. The latter ray-casting method is introduced in the following section.

**2.5.1. Point Correspondence with Ray Casting.** For surfaces, if unique intersection exists between a ray starting from its

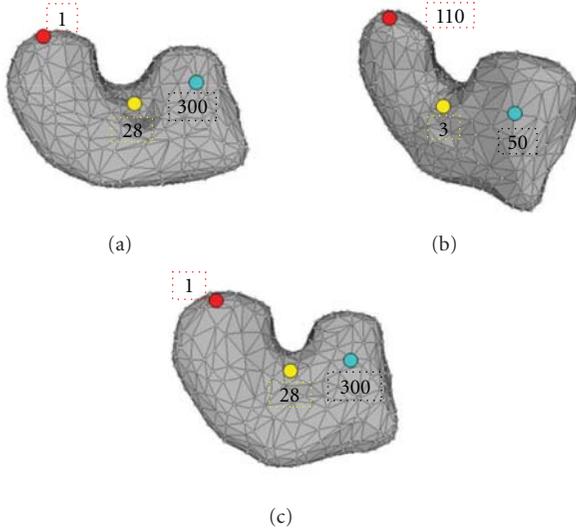


FIGURE 3: Surface correspondence: (a) vertex index on deformation 1, (b) vertex index on deformation 2 before correspondence, (c) corresponding vertex index on deformation 2.

object center and the surface, a ray-casting method can be applied to obtain sample pairs across all the deformations. For illustration, Figures 4(a) and 4(b) depict two different cases of ray-surface intersection in the simplified 2D space. In Figure 4(a), there is only one intersected point ( $p_1$  for ray  $\vec{r}_1$ , and  $p_2$  for ray  $\vec{r}_2$ ) between each ray and the deformed surface  $S_1$ . By contrast, Figure 4(b) gives an example when multiple intersections ( $p_1, p_1', p_1''$  for ray  $\vec{r}_1$  and  $p_2$  for ray  $\vec{r}_2$ ) are involved between rays and the surface  $S_2$ .

If the condition of single ray-surface intersection applies, deformations can be resampled through the following steps to achieve point correspondence.

- (i) Construct an icosahedron of  $W$  vertices with radius large enough to embrace the largest deformation volume among those under consideration; larger  $W$  results in denser samples to maintain the local details but incurs more computational cost.
- (ii) Align the center of the 3D surfaces to the origin of icosahedron such that rays casting from the origin can intersect with the surface.
- (iii) For each ray segment originated from the center to a vertex on the icosahedron, find the triangle on the surface mesh that intersects with the segment and use that intersected point as new surface sample.

Figure 4(c) illustrates the desired sample pairs as  $(p_{1,1}, p_{2,1})$  and  $(p_{1,2}, p_{2,2})$  over two deformations  $S_1$  and  $S_2$ . This resampling process also establishes a one-to-one map between a point on the object and a point on the sphere (icosahedron), which naturally meets the purpose of spherical parameterization. As a result, point-wise correspondence can be achieved across all the resampled surfaces, and a uniform spherical harmonic matrix  $\mathbf{Y}$  can be applied. As an example, Figures 5(a) and 5(b) compare an original kidney

surface with the corresponding resampled surface. We can see that ray-casting procedure well maintains the shape.

### 3. Experiments

Three types of experiments are conducted to demonstrate the feasibility of the proposed SRDS algorithm. The computer-generated FEM data is first used to demonstrate that the SRDS approach matches the accuracy of complex mathematical modeling techniques, then an ex vivo experiment is conducted using 3D MRI scans of porcine kidneys for evaluation in practical settings, and finally in vivo experiment is carried over dynamic cardiac MRI scans for evaluation in real patients.

**3.1. Experiment with FEM Data.** Three representative organs are employed in this FEM experiment: 3D cortical mesh as an example of complicated shapes, gallbladder as an instance with geometrically simple shape, and bladder consisting of both interior and exterior walls.

**3.1.1. Computer Model Setup.** The initial 3D models of different organs are fed into a FEM-based surgical simulation tool to generate deformation data for testing. For instance, Figure 6 demonstrates two examples of shape distortions due to the endoscope poking and grasping one side of the gallbladder.

Table 1 lists the FEM experimental setup of the three organs including number of vertices  $N$ , SH level  $L$ , number of deformations for training  $K$ , and testing  $M$ . “GBL” stands for gallbladder in all the tables. The maximum SH level used for brain model is chosen according to [9], and the levels for gallbladder and bladder are determined when the SHD representation error is below 0.1% (EOF). The complex brain structure requires more vertices and higher SH level for surface representation to achieve sufficient accuracy. To evaluate the representation precision qualitatively, an evaluation parameter EOF is defined as the normalized Euclidean distance between the original surface and the reconstructed surface

$$\text{EOF} = \frac{\|\hat{\mathbf{x}}_k - \mathbf{x}_k\|_2}{\|\mathbf{x}_k\|_2}. \quad (19)$$

All surfaces are centered to the origin of the coordinate system so that EOF will not be heavily affected by the denominator.

**3.1.2. Results.** With the FEM data, the proposed SRDS algorithm is evaluated from three perspectives: (1) subspace dimensionality, (2) sparsity and accuracy of representations, and (3) the effect of subspace pursuit threshold  $\varepsilon$  and coefficient truncation threshold  $\delta$  on the performance. The sparsity is defined as the  $l_0$  norm of the coefficient vector  $\hat{\mathbf{c}}_m$ .

**A. Training Results.** During training stage, we set  $\varepsilon = 0.005$  for subspace detection,  $\eta = 0.01$  for clustering,  $E_{\max} = 50$  as the maximum iteration times, and  $\delta = 0.005$  for

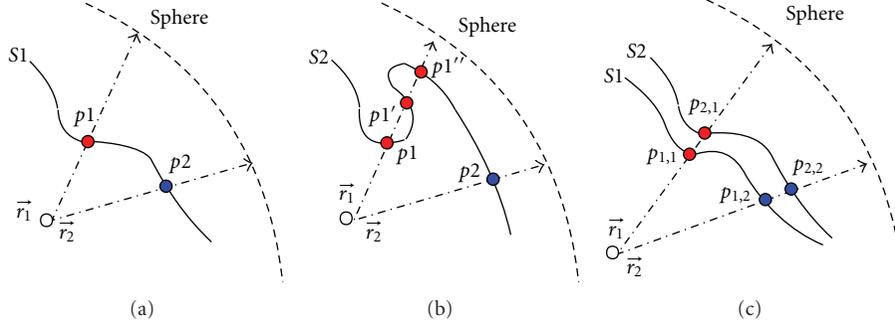


FIGURE 4: Ray casting sampling: (a) unique ray-surface intersection, (b) multiple ray-surface intersections, and (c) ray casting on two deformed surfaces.

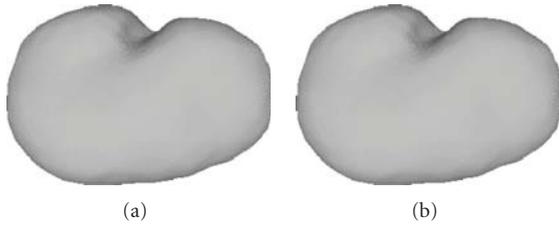


FIGURE 5: Resample with ray casting (a) original surface of kidney, (b) resampled surface of kidney.

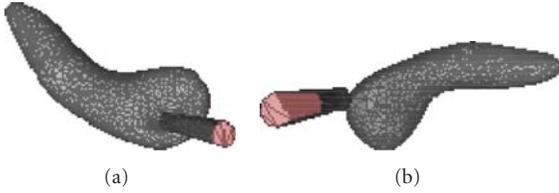


FIGURE 6: Gallbladder deformation generated by surgical simulator: (a) and (b) are distortions caused when the endoscope pokes and grasps the gallbladder.

coefficient truncation. Subspaces on  $X$ ,  $Y$ , and  $Z$  axis are identified separately. Table 2 shows that the subspace number  $J$  and dimensions of resulting dictionary ( $\dim(G) = I$ ) are markedly small relative to  $N$  or  $L^2$  in all three tests. We notice that the subspace dimensions of brain are relatively smaller than the other two. This is because of smaller training data size and minor extent of deformation considered in the brain experiment, which results in smaller dictionary size to capture the deformation features.

**B. Sparsity and Accuracy Evaluation.** Sparsity is examined in terms of  $(\mu/\sigma)$ , where  $\mu$  is the average  $l_0$  norm of the coefficient vector  $\tilde{c}_k$  (training) or  $\tilde{c}_m$  (testing) and  $\sigma$  is the corresponding standard deviation. To verify that whether our method achieves equivalent sparsity and precision when applying the structure of the dictionary, we also test the case without relying on any structure learned from training set, during which sparse representation of each deformation in the testing set is repursued from the training set using

TABLE 1: FEM Model setup.

	Vertices $N$	SH level $L$	Training $K$	Testing $M$
Brain	40962	80	35	35
GBL	3038	25	250	114
Bladder	$N1 = 4434,$ $N2 = 4274$	30	74	46

TABLE 2: Dimension of dictionary ( $J/I$ ).

	Brain	GBL	Bladder
Subspace on $X$ ( $J/I$ )	2/13	2/39	1/37
Subspace on $Y$ ( $J/I$ )	2/6	2/48	1/30
Subspace on $Z$ ( $J/I$ )	1/3	1/42	1/34

TABLE 3: Sparsity ( $\mu/\sigma$ ) and accuracy evaluation.

	$X$ ( $\mu/\sigma$ )	$Y$ ( $\mu/\sigma$ )	$Z$ ( $\mu/\sigma$ )	EOF (%)
Brain train (SRDS)	3.1/0.2	3.4/0.9	3.0/0.2	1.31
Brain test (SRDS)	3.5/0.5	3.9/0.4	3.0/0.0	1.32
Brain test (OSP)	4.1/0.6	4.5/0.5	3.9/0.8	1.30
GBL train (SRDS)	25.8/12.2	4.8/1.8	32.3/7.3	0.13
GBL test (SRDS)	33.2/1.0	44.1/1.1	40.8/1.3	0.15
GBL test (OSP)	44.4/4.0	48.3/2.3	43.9/4.5	0.13
Bladder train (SRDS)	24.1/3.7	20.8/2.9	24.8/4.0	0.076
Bladder test (SRDS)	24.6/2.7	21.0/2.5	25.6/3.2	0.073
Bladder test (OSP)	22.5/4.3	16.3/6.5	23.0/5.2	0.070

OSP approach. In the following tables, we use ‘‘OSP’’ to refer to the results obtained using such a repursuing process. Table 3 summarizes the sparsity of the SRDS representation of three organs for both training and testing set. It illustrates that the number of atoms needed for representing the complex deformations is much smaller than the dimension of spherical harmonic vectors  $((L + 1)^2)$ , and particularly the sparsity and accuracy via SRDS is very close to that from OSP repursuit for the testing deformations, which indicates the good generalization of the structured dictionary.

The reconstruction error in terms of EOF is further compared with that from standard SHD method, as shown in Figure 7. In general, the accuracy of SRDS is equivalent

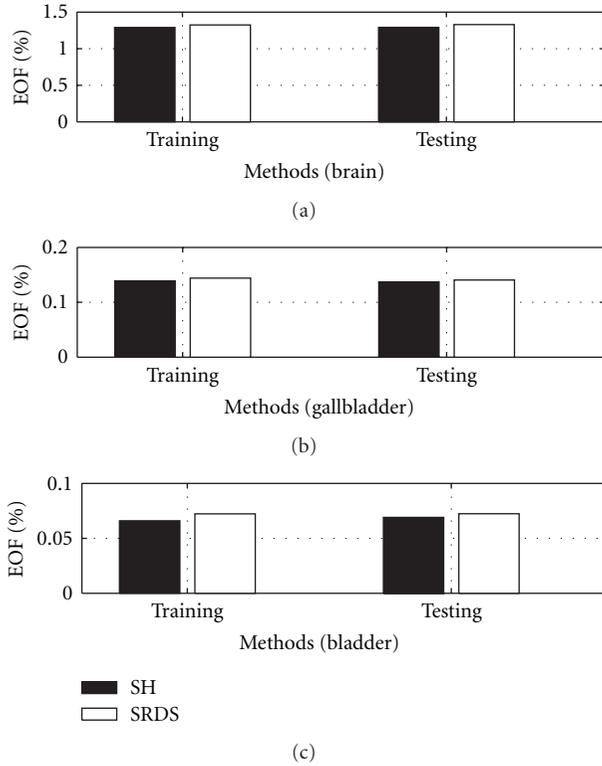


FIGURE 7: EOF of brain, gallbladder, and bladder reconstruction with SHD and SRDS methods. The left pair is for training set; the right pair is for testing set.

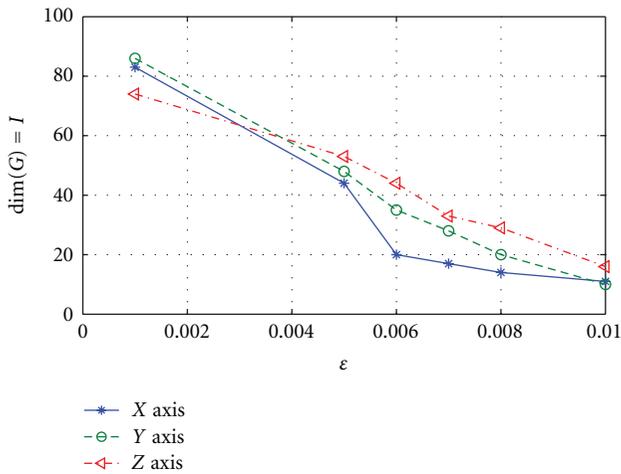


FIGURE 8: Dimension  $I$  of trained dictionary  $G$  decreases on all three axis as  $\epsilon$  increases.

to that of SHD method. Specifically, it shows that the SRDS method achieves average EOF of 1.32% (brain) and 0.14% (gallbladder) versus 1.29% (brain) and 0.13% (gallbladder) with SHD method. For bladder model with deformations on multiple layers, the overall representation error with SRDS is 0.07%, very close to 0.06% with SHD. Figures 11 and 12 show typical reconstructed deformations of the testing data for the three organs with SHD and SRDS methods.

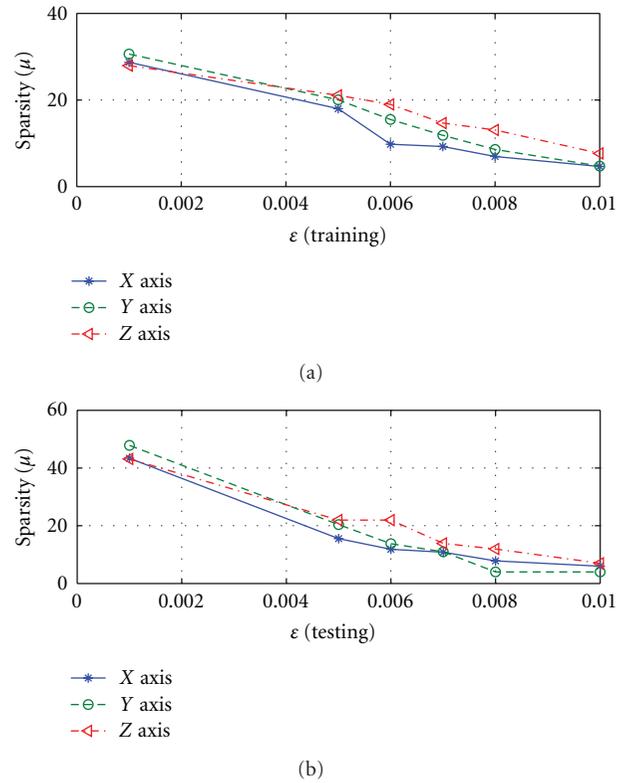


FIGURE 9: Sparsity ( $\mu$ ) of surface representation changes with different  $\epsilon$  on three axis: (a) is for training set; (b) is for testing set.

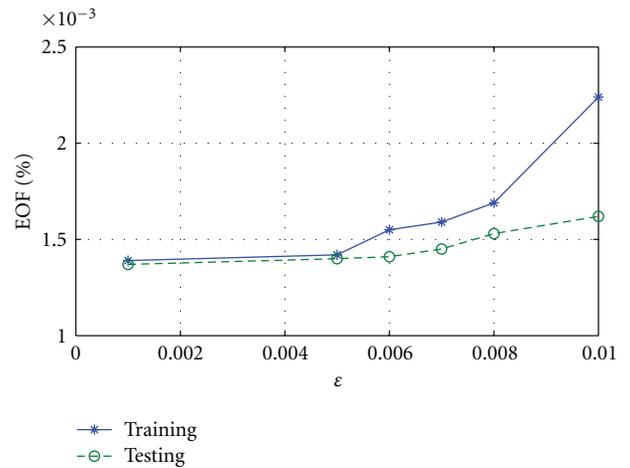


FIGURE 10: EOF of gallbladder reconstruction increases as  $\epsilon$  gets larger for both training and testing sets.

Figures 12(e) and 12(f) demonstrate the interior and exterior representation of the bladder at a same time instance. From those results, we can see that the SRDS algorithm achieves the accuracy equivalent to complex mathematical modeling techniques while significantly lowers the representation dimensionality.

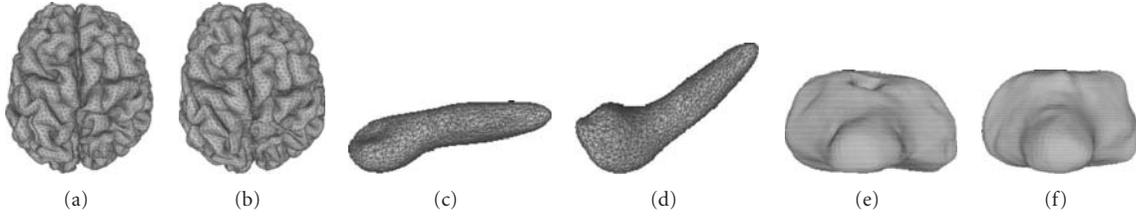


FIGURE 11: Reconstructed brain deformations using SHD approach: (a) is the initial brain shape and the circle marks one typical area under deformation, (b)–(e) are the reconstructed brain deformations.

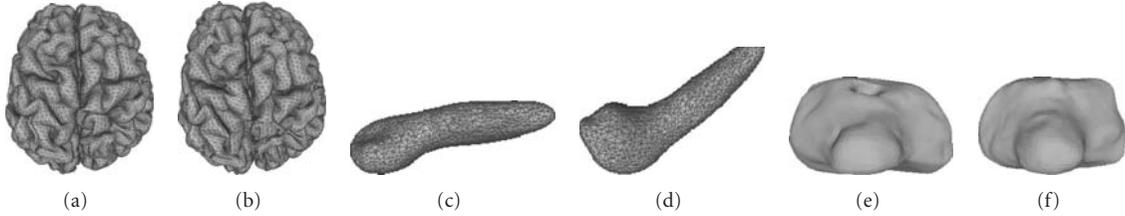


FIGURE 12: Reconstructed brain deformations using SRDS approach: (a) is the initial brain shape, (b)–(e) are the reconstructed brain deformations.

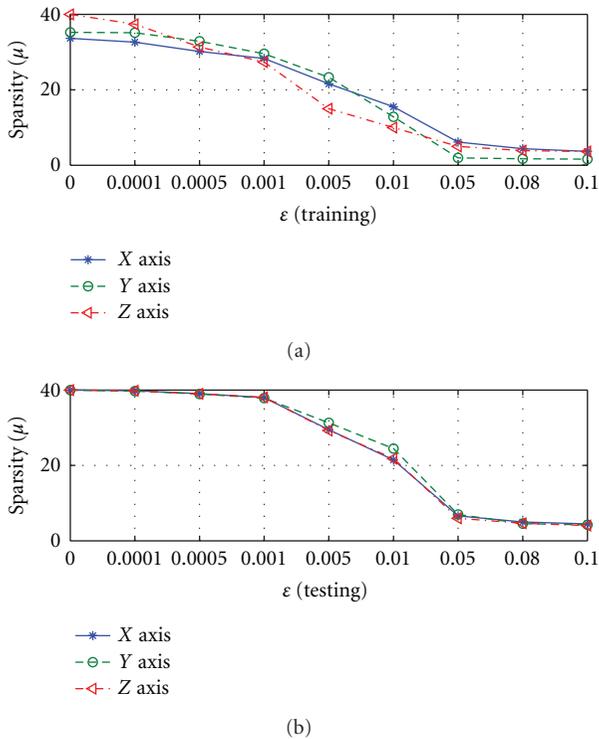


FIGURE 13: Sparsity ( $\mu$ ) of surface representation changes with different  $\delta$  on three axis: (a) is for training set; (b) is for testing set.

C. *Effect of  $\epsilon$ .* The performance of SRDS algorithm is examined as the subspace pursuit threshold  $\epsilon$  varies. Specifically, we study the effect of  $\epsilon$  on the dimensionality ( $I$ ) of the structured dictionary  $G$ , sparsity and accuracy of the surface representation. Figure 8 shows how the subspace dimensions on three axis change during the training stage as  $\epsilon$  increases

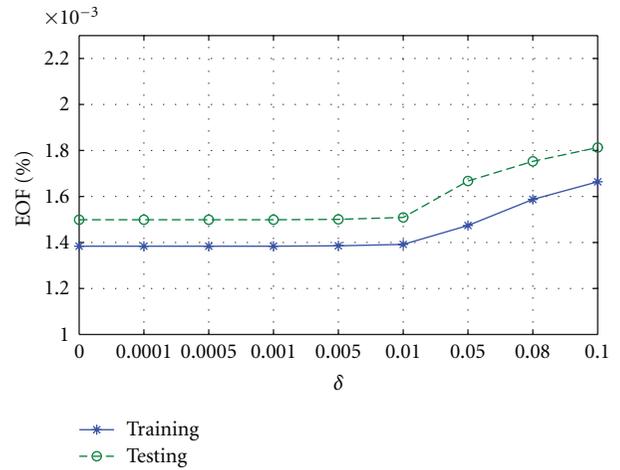


FIGURE 14: EOF of gallbladder reconstruction increases as  $\delta$  gets larger for both training and testing sets.

from 0.001 to 0.01. Figure 9 displays the influence of  $\epsilon$  on the average sparsity  $\mu$  of the surface representation results for both training and testing data sets. In general, smaller  $\epsilon$  leads to larger subspace size and less description sparsity, since lower  $\epsilon$  usually leads to more recruited atoms to meet the desired representation accuracy. Therefore, there is a tradeoff between representation accuracy and desired sparsity. Figure 10 reveals the representation EOF as a function of  $\epsilon$ . Not surprisingly, the reconstruction error is increased as  $\epsilon$  becomes larger. An empirical point can be chosen according to the training curve when space dimension  $I$  expands significantly but only trivial EOF improvement is gained, that is,  $\epsilon = 0.005$  is a preferred value in this test according to Figure 10.

TABLE 4: Subspace dimension and sparsity ( $\mu/\sigma$ ) for intramodel experiment.

	X ( $\mu/\sigma$ )	Y ( $\mu/\sigma$ )	Z ( $\mu/\sigma$ )	EOF (%)	Haus (mm)
$J/I$	1/17	2/31	2/31		
Train (SRDS)	16.9/0.4	15.5/1.5	15.8/1.5	0.24	0.55
Test (SRDS)	17.0/0.0	16.9/0.3	16.7/0.9	0.64	0.87
Test (OSP)	16.6/1.5	16.3/0.9	17.1/0.8	0.60	0.85

TABLE 5: Subspace dimension ( $J/I$ ) in intermodel experiments.

	X axis	Y axis	Z axis
Ex1 ( $J/I$ )	1/36	1/37	1/37
Ex2 ( $J/I$ )	2/87	2/86	2/87
Ex3 ( $J/I$ )	3/79	2/88	3/83

*D. Effect of  $\delta$ .* The influence of coefficient truncation threshold  $\delta$  on the performance of SRDS algorithm is also tested while  $\delta$  is varied among [00.0001 0.0005 0.001 0.005 0.01 0.05 0.08 0.1]. Figure 13 shows the effect of  $\delta$  on the average sparsity  $\mu$  of the surface representation results. We can see that, as the truncation threshold  $\delta$  enlarges, the sparsity of the representation is increased for both training and testing data sets at the price of decreased representation error as shown in Figure 14, so there is tradeoff between sparsity and accuracy. Empirically, one can choose the  $\delta$  value when the representation precision remarkably deteriorates while the sparsity is still increasing. Therefore, according to Figures 13 and 14, an appropriate value for  $\delta$  is between 0.005 and 0.01.

*3.2. Ex vivo Experiment Using MRI.* To evaluate the proposed algorithm in real applications, an ex vivo experiment using three porcine kidneys were conducted at the Center for Interdisciplinary Applications in Magnetic Resonance (CIAMR) of University of Minnesota. Deformations imposed to each kidney were controlled and maintained still during imaging by a customized nonmagnetic mechanical device as shown in Figure 15. Each deformed kidney shape was scanned in 3D MRI mode with spatial resolution of 1.2 mm to generate both training set and testing set. The SH degree  $L$  of the organ representation is set to be 20, and each 3D kidney mesh after surface correspondence has  $N = 4002$  vertices. Different from computer-generated deformations where surface correspondence is intrinsically established, the shapes from MRI scans are rendered independently, so the method described in Section 2.5 is applied to achieve point-wise correspondence.

Both intramodel and intermodel experiments are conducted. The former uses training and testing deformations from the same kidney; the later utilizes two out of the three kidneys for training and the third one for testing in a cross-evaluation fashion. Besides sparsity and EOF, Hausdorff distance between the represented shape and corresponding MRI surface is also examined as a physical measurement of

error. The Hausdorff distance between surface  $x$  and  $x'$  is defined as

$$d(x, x') = \max_{p \in x} d(p, x'), \quad (20)$$

where  $d(p, x')$  is defined as the distance between a point  $p$  on surface  $x$  and the closest point on surface  $x'$ , that is,

$$d(p, x') = \min_{p' \in x'} \|p, p'\|_2 \quad (21)$$

with  $\|\cdot\|_2$  denoting the Euclidean norm.

*3.2.1. Intramodel Test.* In the intramodel experiment, 31 deformations of the same kidney were generated and scanned by the MRI machine, among which 20 frames were randomly selected as training set, and the other 11 were applied for testing the generalization of the learned subspaces.

Table 4 shows the trained subspace dimensions ( $J$  as number of subspace,  $I$  as dictionary size of  $\mathbf{G}$ ), the sparsity of the descriptors in each axis for both training set and testing set, and the corresponding errors in terms of EOF and Hausdorff distance. Similar to the FEM experiment, the sparsity is also evaluated with OSP repursuit process in the testing set for comparison. The table shows that the sparsity and the accuracy achieved with SRDS is very close to that from OSP repursuing process. However, the SRDS method features delay-free surface representation by applying the structure in the identified dictionary. Further results about computational efficiency are shown in Section 3.4. One may notice that the size of training data in the MRI experiment is smaller than that in FEM test due to the less availability of 3D MRI images. As a rule of thumb, larger training set carries richer deformation information and thus leads to better generalization of the dictionary. However, given the size of training data and extent of deformation involved in the ex vivo experiment, high representation precision is still achieved.

Figure 16 illustrates the accuracy of the surface representation in the intramodel test. The average EOF in Figure 16(a) for training set is 0.24% and 0.64% for testing set, with maximum rate less than 1%. Further, error as Hausdorff distance (shown in Figure 16(b)) is  $0.55 \pm 0.23$  mm with 95th percentile error of 0.86 mm for the training set, and  $0.87 \pm 0.10$  mm with 95th percentile error of 0.96 mm for the testing set. This intramodel experiment demonstrated that the SRDS algorithm identifies subspaces generalizable enough to accurately represent deformations beyond the training set for the same object.

Figure 17 visualizes the color-coded error distribution at all vertices on the represented surface with SRDS relative to the actual MRI scans. Figure 17(a) illustrates the error range for different colors. Figures 17(b) and 17(d) show the error distribution for a typical reconstruction in the training and testing set, respectively. Figures 17(c) and 17(e) show maximum 90% level reconstruction errors, that is, 90% of all deformations in the training or testing set have representation point errors less than the values shown in the figures. Consistent with the EOF and Hausdorff distance results, the color diffusion in Figure 17 indicates that the

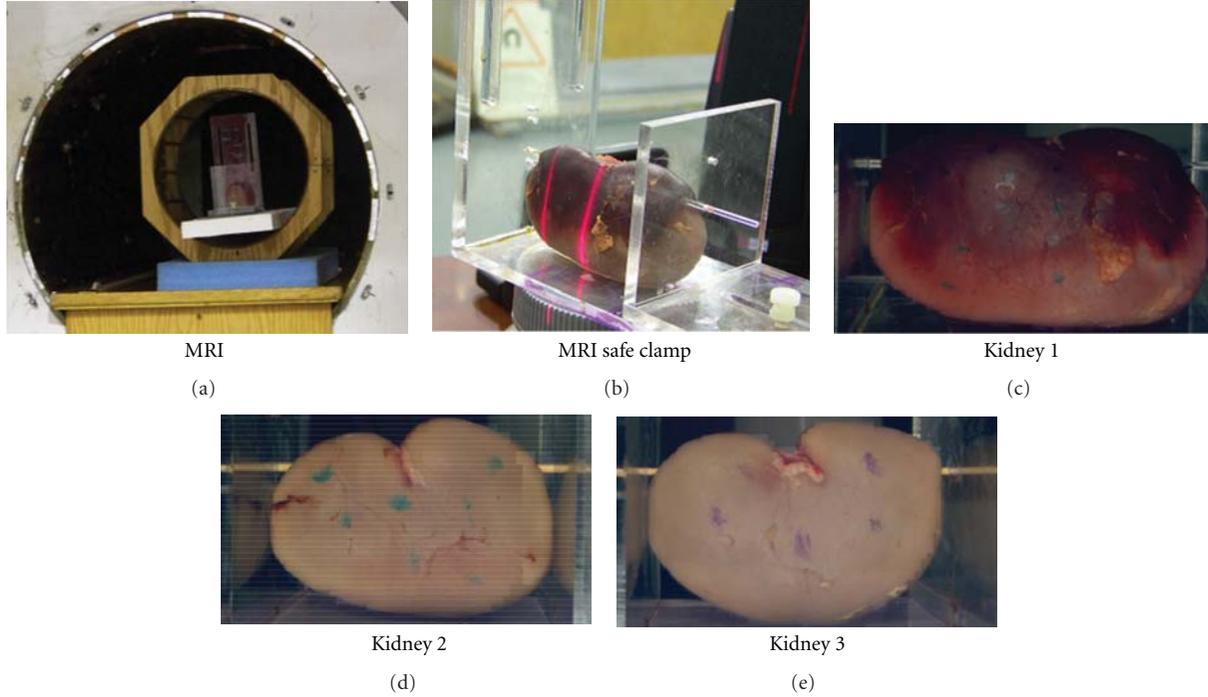


FIGURE 15: Images of three porcine kidneys for experiment.

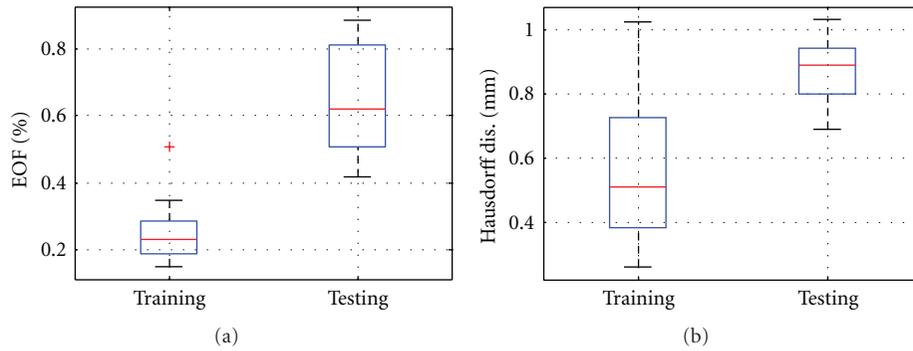


FIGURE 16: Boxplots of representation error in intramodel experiment: (a) EOF of training and testing set, (b) Hausdorff distance of training and testing sets.

precision in the testing group is relatively lower than that in the training group. However, among all the pixel-wise errors shown in Figure 17(e), less than 3% of all the surface points have error distance larger than 0.5 mm.

**3.2.2. Intermodel Test.** Three intermodel experiments are performed to further validate the proposed SRDS method applied to organs from different subjects. In the following context, “Ex1” stands for the experiment training on Kidney 2 and 3 plus one initial shape of Kidney 1 while testing on deformations of Kidney 1, and the like for “Ex2” and “Ex3”. In each experiment, both sparsity and accuracy are examined.

The number of subspaces ( $J$ ) and dimensions ( $I$ ) of the identified dictionary are listed in Table 5. The training

results vary among the three experiments but all features low subspace dimensions. Table 6 shows the sparsity of the intermodel experiments using the SRDS algorithm, and the error level is evaluated in terms of EOF and Hausdorff distance. Each testing deformation is also sparsely retrained using OSP for comparison. We can see that the sparsity and representation error resulting from SRDS method is very close to that using OSP.

Figure 18 shows the representation accuracy using SRDS algorithm in training and testing sets for the three tests. In general, the error in testing set is larger than that in the training set. Particularly, as for EOF evaluation, “Ex1” leads to the largest EOF error relative to “Ex2” and “Ex3”, but the average error rate is still as low as 0.3% for training set and 2.0% for testing set. Table 7 lists the specific Hausdorff measurements corresponding to boxplots in Figures 18(c) and 18(d),

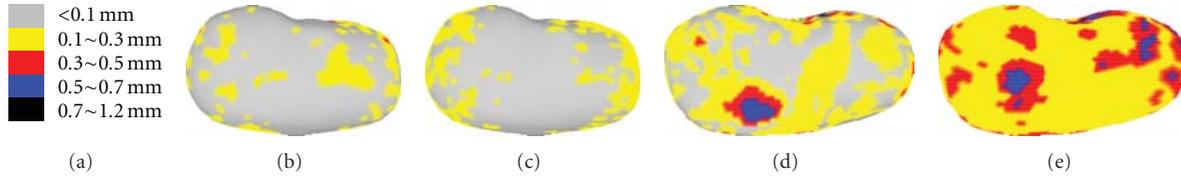


FIGURE 17: Representation error in intramodel experiment: (a) color-coded scales, (b) example representation error in training set, (c) 90% representation error in training set, (d) example of representation error in testing set, and (e) 90% representation error of testing set.

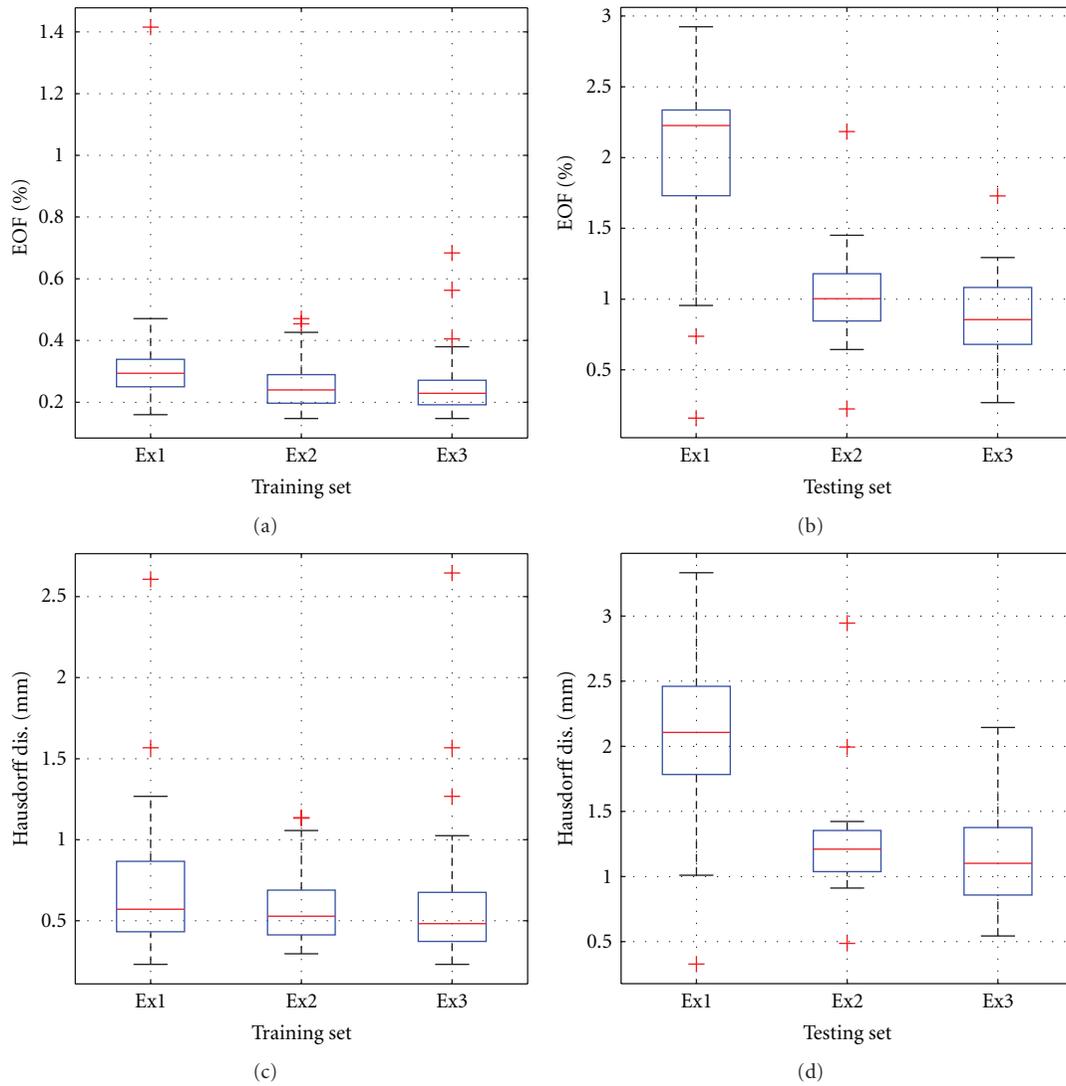


FIGURE 18: Boxplots of representation error in intermodel experiments: (a) EOF of training set, (b) EOF of testing set, (c) Hausdorff distance of training set, (d) Hausdorff distance of testing set.

including minimum, 95th percentile and mean. We can see that the 95th percentile Hausdorff distance across all experiments is below 3 mm, and the mean is below 2.1 mm. Comparing those error levels with the intramodel test, one can see that the homology existing among the training and testing deformations contributes to better dictionary generalization and, thus, leads to higher representation accuracy.

Figures 19, 20 and 21 show the color-coded error fields of a typical representation and at the maximum 90% level for the three intermodel experiments. In either training set or testing set, it is observed that large errors are mostly distributed around the edge area where local details are rich. Consistent with the previous EOF and Hausdorff distance measurements, the color diffusion in Figures 19–21 indicates that errors in testing set is larger than that in training set and

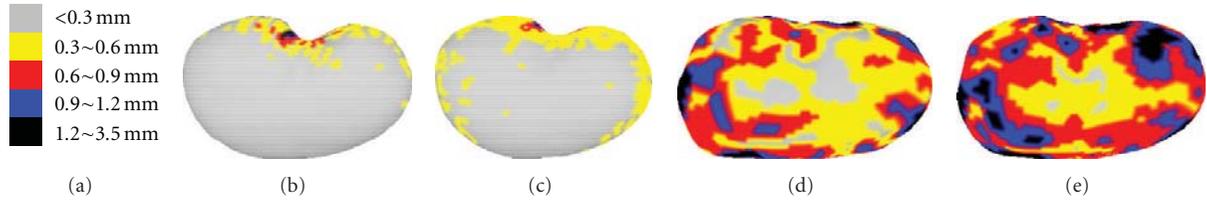


FIGURE 19: Representation error in intermodel Ex1: (a) color-coded scales, (b) example of color-coded point error in training set, (c) 90% color-coded point error in training set, (d) example of color-coded point error in testing set, and (e) 90% color-coded point error of testing set.

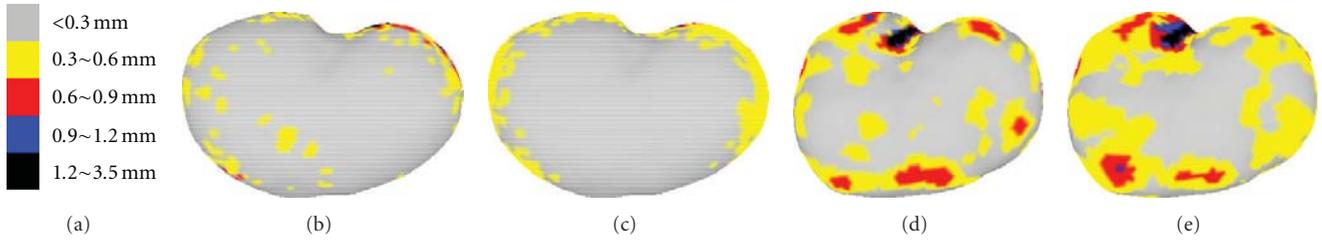


FIGURE 20: Representation error in intermodel Ex2: (a) color-coded scales, (b) example of color-coded point error in training set, (c) 90% color-coded point error in training set, (d) example of color-coded point error in testing set, and (e) 90% color-coded point error of testing set.

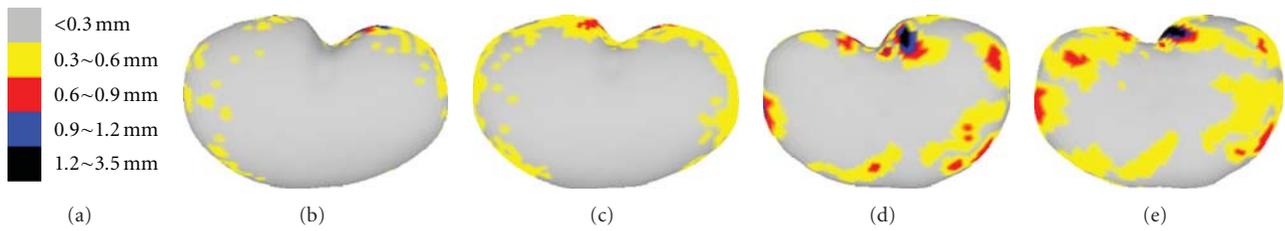


FIGURE 21: Representation error in intermodel Ex3: (a) color-coded scales, (b) example of color-coded point error in training set, (c) 90% color-coded point error in training set, (d) example of color-coded point error in testing set, and (e) 90% color-coded point error of testing set.

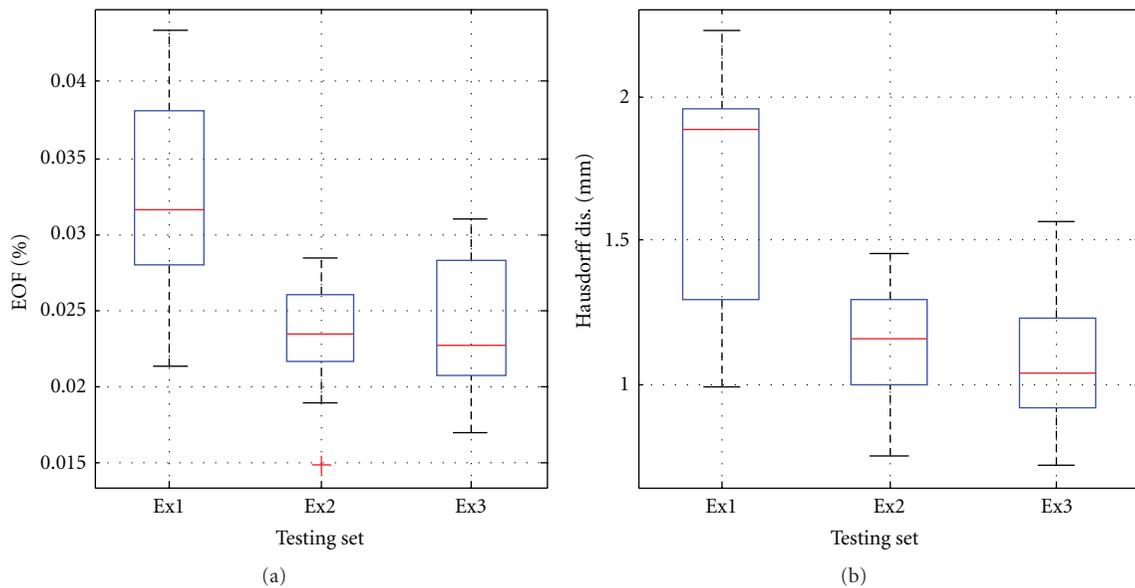


FIGURE 22: Boxplots of representation error in LV interpatient experiments: (a) EOF, (b) Hausdorff distance.

TABLE 6: Sparsity ( $\mu/\sigma$ ) and accuracy evaluation for intermodel experiment with kidneys.

	$X$ ( $\mu/\sigma$ )	$Y$ ( $\mu/\sigma$ )	$Z$ ( $\mu/\sigma$ )	EOF (%)	Haus (mm)
Ex1 train (SRDS)	34.7/5.5	35.6/5.5	36.4/1.6	0.32	0.69
Ex1 test (SRDS)	35.7/1.3	36.7/0.9	36.7/1.8	2.01	2.08
Ex1 test (OSP)	36.7/2.4	36.0/2.4	35.9/2.6	1.94	1.94
Ex2 train (SRDS)	42.8/4.5	42.8/2.2	43.5/3.7	0.26	0.59
Ex2 test (SRDS)	47.0/3.1	44.5/1.8	46.3/2.4	1.02	1.27
Ex2 test (OSP)	44.0/2.2	41.9/12.1	45.9/2.0	0.95	1.22
Ex3 train (SRDS)	37.1/9.1	43.3/2.2	40.8/1.3	0.25	0.59
Ex3 test (SRDS)	41.4/6.5	45.1/1.7	40.4/1.3	0.90	1.18
Ex3 test (OSP)	41.2/8.4	42.7/9.7	39.3/10.6	0.83	1.16

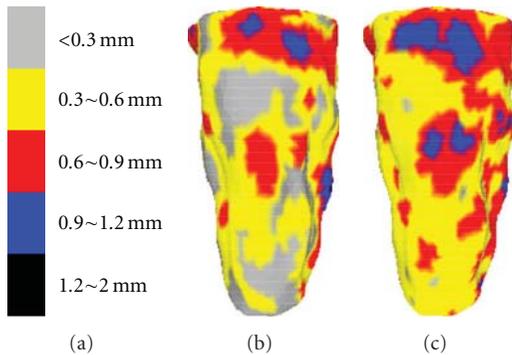


FIGURE 23: Representation error in in vivo experiment: (a) color-coded scales, (b) example representation error in testing set, (c) 90% representation error of testing set.

“Ex1” generates relatively larger error comparing to “Ex2” or “Ex3.”

**3.3. In Vivo Experiment Using MRI.** The proposed approach is also tested over the in vivo cardiac MR images [37], consisting of automatically segmented images from volumetric MRI scans of a diastole-systole-diastole cycle. For each patient, there are around 22 phases in a cardiac cycle. Surface correspondence of LV shapes within and across patients are accomplished using the approach described in Section 2.5. Since the generated surfaces from automatic segmentation software are quite rough, we use the spherical harmonic representation as a filter to smooth out those surface noises and then apply the smoothed surfaces as training and testing data. Therefore, the demonstrated error in this section is relative to the SHD surfaces, not to the original raw surfaces. The iter-patient results are reported as follows.

Similar to the ex vivo test, we use the segmented left ventricles (LV) of 2 different patients plus an initial LV surface for the third patient as training data, and the remaining LV shapes in a beating cycle of the third patient are used to test the generalization of the identified subspaces. The formulated tests are noted as “Ex1,” “Ex2,” and “Ex3.” Table 8 lists the sparsity test results of the three cross validations for both training and testing sets. We can see that the sparsity in the training set is close to that in the

TABLE 7: Hausdorff distance for intermodel experiment with kidneys.

	Min (mm)	95th (mm)	Mean (mm)
Ex1 train	0.23	1.14	$0.69 \pm 0.42$
Ex1 test	0.33	2.81	$2.08 \pm 0.61$
Ex2 train	0.30	1.03	$0.59 \pm 0.23$
Ex2 test	0.49	1.99	$1.27 \pm 0.49$
Ex3 train	0.23	1.03	$0.59 \pm 0.39$
Ex3 test	0.54	1.74	$1.18 \pm 0.44$

testing set, but the former achieves much higher accuracy. This is because that the identified subspaces generalize perfectly for those elected atoms among the training set after spherical harmonic smoothing. Consistent with the previous experiments, the representation of testing surfaces using SRDS is also compared with that using repursuing OSP. According to the results, SRDS achieves performance slightly worse than but close to that of OSP. However, as demonstrated in Section 3.4, without relying on the structured dictionary learned from the training population, OSP is a computational expensive task, since for each testing surface, it requires to research for atoms from the training set to achieve sparse representation.

Figure 22 provides boxplots for the representation accuracy of the testing set in terms of EOF and Hausdorff distance. Table 9 provides the minimum, 95th percentile, and mean Hausdorff measurements corresponding to Figure 22(b). In coincidence, “Ex1” leads to slightly larger errors than the other two tests, with average EOF of 3.2% (“Ex1”), and mean Hausdorff distance of  $1.67 \pm 0.39$  mm. The 95th percentile Hausdorff distance across all experiments is below 2.2 mm. Figures 23(b) and 23(c) show the color-coded error field of a typical representation and at the maximum 90% level for the testing set in one interpatient experiment (c). As indicated by the color distribution, majority of the point errors are below 0.9 mm. Particularly, in the 90th percentile evaluation in Figure 23, only 3% of all the point-wise errors are above 0.9 mm.

**3.4. Efficiency.** To examine the efficiency of the proposed SRDS method quantitatively, the computational time to represent each surface in the testing set using SRDS method is

TABLE 8: Sparsity ( $\mu/\sigma$ ) along  $X$ ,  $Y$ ,  $Z$  axis and accuracy of in vivo LV tests.

	$X$ ( $\mu/\sigma$ )	$Y$ ( $\mu/\sigma$ )	$Z$ ( $\mu/\sigma$ )	EOF (%)	Haus (mm)
Ex1 train (SRDS)	37.0/0.0	35.4/0.5	31.7/9.0	0.13	0.08
Ex1 test (SRDS)	37.0/0.0	36.0/0.0	34.6/1.5	3.21	1.07
Ex1 test (OSP)	33.3/10.5	36.8/2.6	33.8/9.2	2.92	0.98
Ex2 train (SRDS)	35.4/7.4	32.8/14.1	32.4/3.3	0.17	0.10
Ex2 test (SRDS)	38.0/0.0	37.0/0.0	34.0/0.0	2.31	1.15
Ex2 test (OSP)	32.1/6.3	24.0/10.5	32.3/6.5	1.93	1.07
Ex3 train (SRDS)	36.7/3.2	38.8/1.5	38.4/0.5	0.12	0.07
Ex3 test (SRDS)	39.6/1.2	40.0/0.0	38.8/0.4	2.33	1.09
Ex3 test (OSP)	32.7/11.5	33.4/11.3	30.1/12.2	1.98	0.99

TABLE 9: Hausdorff distance for in vivo experiment with LV.

	Min (mm)	95th (mm)	Mean (mm)
Ex1 train	0	0.50	$0.08 \pm 0.18$
Ex1 test	0.99	2.16	$1.67 \pm 0.39$
Ex2 train	0	0.63	$0.10 \pm 0.24$
Ex2 test	0.75	1.37	$1.14 \pm 0.18$
Ex3 train	0	0.60	$0.07 \pm 0.21$
Ex3 test	0.72	1.46	$1.09 \pm 0.24$

TABLE 10: Computational time of SRDS and OSP.

	Training $K$	SH level $L$	SRDS (sec) $t_1$	OSP (sec) $t_2$	$t_2/t_1$
Brain	35	80	0.3	51.2	170.7
GBL	350	25	0.6	39.1	65.2
Bladder	74	30	2.5	57.2	22.9
Kidney	52	20	0.8	9.2	11.5
LV	51	25	0.5	6.9	13.8

compared with that resulting from OSP repursuing approach for the above five organs. The results are summarized in Table 10, including training set size  $K$ , maximum SH level  $L$ , average time (in seconds) required with SRDS ( $t_1$ ) and OSP ( $t_2$ ), respectively, and the ratio between the two. As shown in Table 10, the time consumption for seeking sparse representation of the testing surfaces using the SRDS is at least 10 times lower than that using the original OSP method which does not rely on the dictionary structure learned from the training data set. The advantage is more pronounced when the training data size  $K$  or the SH level  $L$  is large. For example, in the brain model, the high SH level  $L$  leads to substantial computational delay during the search for proper atoms for representation, such that the SRDS achieves a speed orders of magnitude faster than the OSP method without training. On the other hand, for the case of gallbladder, the large training size also increases the time used by repursuing OSP, so it runs 65 times slower than SRDS.

To summarize, considering the test results for sparsity, accuracy, and efficiency given in this experiment section, we can see that the proposed SRDS method achieves sparse

surface representation with high computational efficiency and accuracy.

## 4. Conclusions and Discussion

This paper introduced a new algorithm for block sparse representation of deformable organ surfaces with high accuracy. The proposed SRDS design first identifies the deformation subspaces from the training data set in the transformed spherical harmonic domain and then represents each deformed surface with a block sparse vector in the structured dictionary. SRDS is generalized to applications involving organs with multiple surface layers, such as bladder. The algorithm has been validated with FEM data and real 3D MRI scans under both ex vivo and in vivo conditions. The FEM test results demonstrate that SRDS achieves accuracy matching that of complex mathematical modeling techniques. Further, the maximum representation error in ex vivo experiment is below 1 mm for intramodel test and below 3 mm for intermodel test. For the in vivo experiment, the SRDS achieves an accuracy of better than 2.5 mm.

SRDS algorithm has already been used in tracking organ deformations in minimum invasive surgery [2–4]. The structure introduced in the dictionary enables efficient surface recovery from limited samples. In addition, the merits of block sparse surface representation presented here can be applied to various medical organ modeling, shape classification, and similarity retrieval where reduced parameter dimension can potentially speed up the implementations.

## References

- [1] M. S. Wagh and C. C. Thompson, “Surgery insight: natural orifice transluminal endoscopic surgery—an analysis of work to date,” *Nature Clinical Practice Gastroenterology & Hepatology*, vol. 4, no. 7, pp. 386–392, 2007.
- [2] D. Wang and A. H. Tewfik, “In vivo tracking of 3D organs using spherical harmonics and subspace clustering,” in *Proceedings of the IEEE International Conference on Image Processing, (ICIP ’09)*, pp. 817–820, November 2009.
- [3] D. Wang, Y. C. Zhang, and A. H. Tewfik, “Real time tracking of exterior and interior organ surfaces using sparse sampling of the exterior surfaces,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP ’10)*, pp. 1022–1025, March 2010.

- [4] D. Wang and A. H. Tewfik, "Real time tracking of 3D organ surfaces using single MR image and limited optical viewing," in *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging, (ISBI '10)*, pp. 620–623, April 2010.
- [5] G. L. Xu and Q. Zhang, "A general framework for surface modeling using geometric partial differential equations," *Computer Aided Geometric Design*, vol. 25, no. 3, pp. 181–202, 2008.
- [6] G. Zeng, S. Paris, L. Quan, and F. Sillion, "Accurate and scalable surface representation and reconstruction from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 141–158, 2007.
- [7] A. Matheny and D. B. Goldgof, "The use of three- and four-dimensional surface harmonics for rigid and nonrigid shape recovery and representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, no. 10, pp. 967–981, 1995.
- [8] J. L. Dillenseger, H. Guillaume, and J. J. Patard, "Spherical harmonics based intrasubject 3-D kidney modeling/registration technique applied on partial information," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2185–2193, 2006.
- [9] M. K. Chung, K. M. Dalton, L. Shen, A. C. Evans, and R. J. Davidson, "Weighted Fourier series representation and its application to quantifying the amount of gray matter," *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 566–581, 2007.
- [10] M. K. Chung, K. M. Dalton, and R. J. Davidson, "Tensor-based cortical surface morphometry via weighted spherical harmonic representation," *IEEE Transactions on Medical Imaging*, vol. 27, no. 8, Article ID 4449087, pp. 1143–1151, 2008.
- [11] C. Brechbhlher, G. Gerig, and O. Kbler, "Parametrization of closed surfaces for 3-D shape description," *Computer Vision and Image Understanding*, vol. 61, no. 2, pp. 154–170, 1995.
- [12] L. Shen and F. Makedon, "Spherical mapping for processing of 3D closed surfaces," *Image and Vision Computing*, vol. 24, no. 7, pp. 743–761, 2006.
- [13] H. Huang, L. Zhang, D. Samaras et al., "Hemispherical harmonic surface description and applications to medical image analysis," in *Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization, and Transmission, (DPVT '06)*, pp. 381–388, June 2006.
- [14] P. J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985.
- [15] S. G. Mallat, G. Davis, and Z. Zhang, "Adaptive time-frequency decompositions," *Optical Engineering*, vol. 33, no. 7, pp. 2183–2191, 1994.
- [16] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.
- [17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [18] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.
- [19] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, 2008.
- [20] B. V. Gowreesunker and A. H. Tewfik, "Learning sparse representation using iterative subspace identification," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, Article ID 5419963, pp. 3055–3065, 2010.
- [21] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [22] K. Engan, S. O. Aase, and J. H. Husy, "Multi-frame compression: theory and design," *Signal Processing*, vol. 80, no. 10, pp. 2121–2140, 2000.
- [23] J. F. Murray and K. Kreutz-Delgado, "An improved FOCUSS-based learning algorithm for solving sparse linear inverse problems," in *Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers*, pp. 347–351, November 2001.
- [24] M. Mahmoudi and G. Sapiro, *Sparse Representations for Three-Dimensional Range Data Restoration*, IMA, Minneapolis, Minn, USA, 2009.
- [25] K. T. Rajamania, M. A. Stynerb, H. Taliba, G. Y. Zheng, L. P. Noltea, and M. A. Ballester, "Statistical deformable bone models for robust 3D surface extrapolation from sparse data," *Medical Image Analysis*, vol. 11, no. 2, pp. 99–109, 2007.
- [26] C. Basso and T. Vetter, "Statistically motivated 3D faces reconstruction," in *Proceedings of the 2nd International Conference on Reconstruction of Soft Facial Parts*, Remagen, Germany, March 2005.
- [27] J. Feng and H. S. Ip, "A multi-resolution statistical deformable model (MISTO) for soft-tissue organ reconstruction," *Journal of Pattern Recognition*, vol. 42, no. 7, pp. 1543–1558, 2009.
- [28] T. Albrecht, R. Knothe, and T. Vetter, "Modeling the remaining flexibility of partially fixed statistical shape models," in *Proceedings of the Workshop on the Mathematical Foundations of Computational Anatomy*, New York, NY, USA, March 2008.
- [29] G. Y. Zheng, S. Gollmerb, S. Schumanna, X. Dong, T. Feilkasb, and M. A. G. Ballester, "A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images," *Medical Image Analysis*, vol. 13, no. 6, pp. 883–899, 2009.
- [30] H. Huang, R. Zhang, F. Makedon, L. Shen, and J. Pearlman, "A spatio-temporal modeling method for shape representation," in *Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization, and Transmission, (DPVT '06)*, pp. 1034–1040, June 2006.
- [31] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [32] S. Chen, C. F. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.
- [33] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [34] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [35] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "A minimum description length approach to statistical shape modeling," *IEEE Transactions on Medical Imaging*, vol. 21, no. 5, pp. 525–537, 2002.
- [36] L. Shen, H. Farid, and M. A. McPeck, "Modeling three-dimensional morphological structures using spherical harmonics," *Evolution*, vol. 63, no. 4, pp. 1003–1016, 2009.

- [37] L. Najman, J. Cousty, M. Couprie et al., "An open, clinically validated database of 3d+t cine-MR images of the left ventricle with associated manual and automated segmentations," *Insight Journal*, special issue entitled ISC/NA-MIC Workshop on Open Science, 2007.

## Research Article

# Multiresolution Analysis Using Wavelet, Ridgelet, and Curvelet Transforms for Medical Image Segmentation

**Shadi AlZubi, Naveed Islam, and Maysam Abbod**

*Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University, West London UB8 3PH, UK*

Correspondence should be addressed to Shadi AlZubi, sh\_sheeh@yahoo.com

Received 10 October 2010; Revised 12 February 2011; Accepted 17 May 2011

Academic Editor: Kenji Suzuki

Copyright © 2011 Shadi AlZubi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The experimental study presented in this paper is aimed at the development of an automatic image segmentation system for classifying region of interest (ROI) in medical images which are obtained from different medical scanners such as PET, CT, or MRI. Multiresolution analysis (MRA) using wavelet, ridgelet, and curvelet transforms has been used in the proposed segmentation system. It is particularly a challenging task to classify cancers in human organs in scanners output using shape or gray-level information; organs shape changes throw different slices in medical stack and the gray-level intensity overlap in soft tissues. Curvelet transform is a new extension of wavelet and ridgelet transforms which aims to deal with interesting phenomena occurring along curves. Curvelet transforms has been tested on medical data sets, and results are compared with those obtained from the other transforms. Tests indicate that using curvelet significantly improves the classification of abnormal tissues in the scans and reduce the surrounding noise.

## 1. Introduction

In the last decade, the use of 3D image processing has been increased especially for medical applications; this leads to increase the qualified radiologists' number who navigate, view, analyse, segment, and interpret medical images. The analysis and visualization of the image stack received from the acquisition devices are difficult to evaluate due to the quantity of clinical data and the amount of noise existing in medical images due to the scanners itself. Computerized analysis and automated information systems can offer help dealing with the large amounts of data, and new image processing techniques may help to denoise those images.

Multiresolution analysis (MRA) [1–3] has been successfully used in image processing specially with image segmentation, wavelet-based features has been used in various applications including image compression [4], denoising [5], and classification [6]. Recently, the finite ridgelet and curvelet transforms have been introduced as a higher dimensional MRA tool [7, 8].

Image segmentation requires extracting specific features from an image by distinguishing objects from the background. The process involves classifying each pixel of an

image into a set of distinct classes, where the number of classes is much smaller. Medical image segmentation aims to separate known anatomical structures from the background such cancer diagnosis, quantification of tissue volumes, radiotherapy treatment planning, and study of anatomical structures.

Segmentation can be manually performed by a human expert who simply examines an image, determines borders between regions, and classifies each region. This is perhaps the most reliable and accurate method of image segmentation, because the human visual system is immensely complex and well suited to the task. But the limitation starts in volumetric images due to the quantity of clinical data.

Curvelet transform is a new extension of wavelet transform which aims to deal with interesting phenomena occurring along curved edges in 2D images [9]. It is a high-dimensional generalization of the wavelet transform designed to represent images at different scales and different orientations (angles). It is viewed as a multiscale pyramid with frame elements indexed by location, scale, and orientation parameters with needle-shaped elements at fine scales. Curvelets have time-frequency localization properties of wavelets but also shows a very high degree of

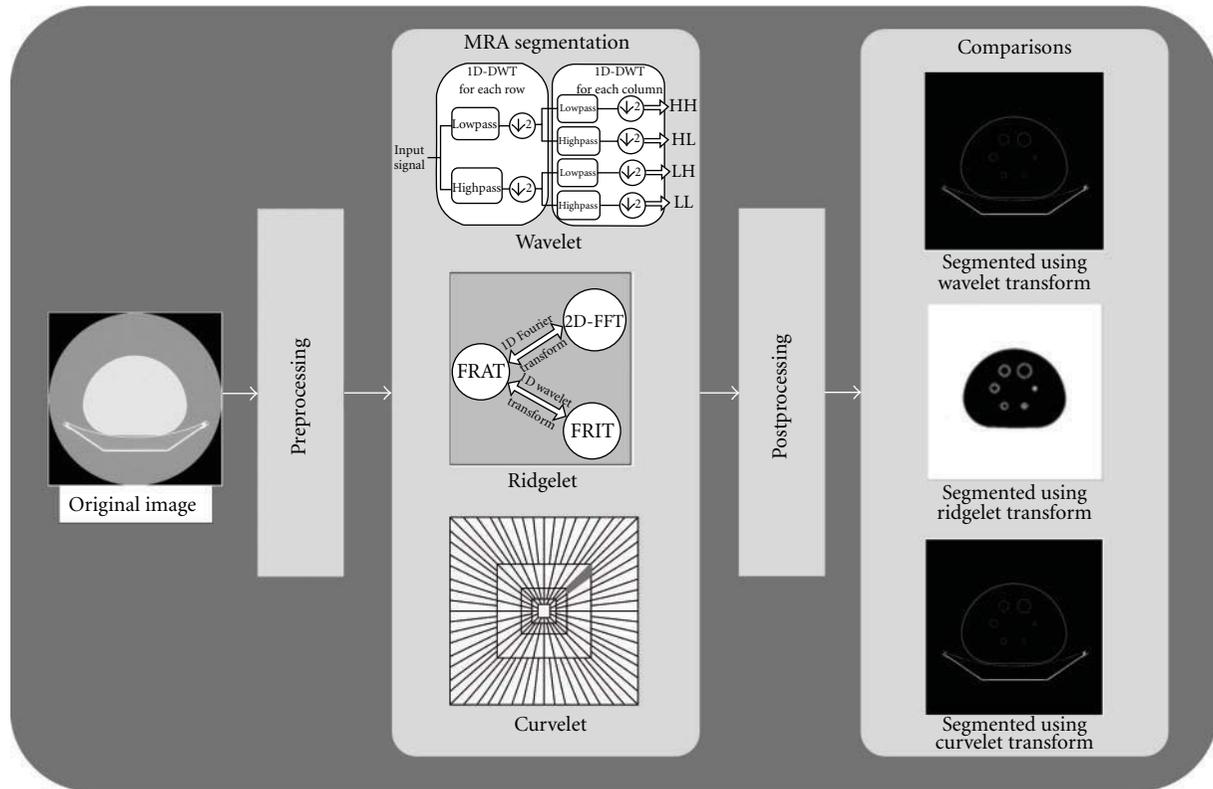


FIGURE 1: Proposed segmentation system for medical images.

directionality and anisotropy, and its singularities can be well approximated with very few coefficients.

This paper is focusing on a robust implementation of MRA techniques for segmenting medical volumes using features derived from the wavelet, ridgelet, and curvelet transforms of medical images obtained from a CT scanner. The rest of this paper is organised as follow: Section 2 illustrates the proposed medical image segmentation system using MRA techniques. The mathematical background and the methodology for the proposed MRA techniques have been explained in Section 3. The results and analysis of the implemented wavelet, ridgelet, and curvelet transforms for medical image segmentation are illustrated in Section 4. Finally, Section 5 includes the conclusions and future work of this research.

## 2. Proposed Medical Image Segmentation System

The main aim of this research is to facilitate the process of highlighting ROI in medical images, which may be encapsulated within other objects or surrounded by noise that make the segmentation process not easy. Figure 1 illustrates the proposed medical image segmentation system using MRA. Wavelet, ridgelet, and curvelet transforms are applied on medical images with other pre- and postprocessing techniques to present segmented outputs and detected ROI in an easier and more accurate way.

## 3. Methodology—Multiresolution Analysis

Image segmentation using MRA such as wavelets has been widely used in recent years and provides better accuracy in segmenting different types of images. Many recent developments in MRA have taken place, while wavelets are suitable for dealing with objects with point singularities. Wavelets can only capture limited directional information due to its poor orientation selectivity. By decomposing the image into a series of high-pass and low-pass filter bands, the wavelet transform extracts directional details that capture horizontal, vertical, and diagonal activity. However, these three linear directions are limiting and might not capture enough directional information in noisy images, such as medical CT scans, which do not have strong horizontal, vertical, or diagonal directional elements. Ridgelet improves MRA segmentation; however, they capture structural information of an image based on multiple radial directions in the frequency domain. Line singularities in ridgelet transform provides better edge detection than its wavelet counterpart. One limitation to use ridgelet in image segmentation is that ridgelet is most effective in detecting linear radial structures, which are not dominant in medical images. The curvelet transform is a recent extension of ridgelet transform that overcome ridgelet weaknesses in medical image segmentation. Curvelet is proven to be particularly effective at detecting image activity along curves instead of radial directions which are the most comprising objects of medical images.

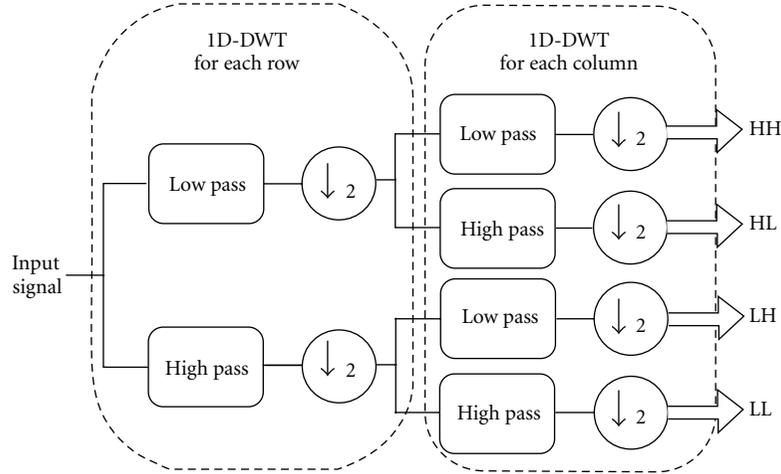


FIGURE 2: 2D DWT filter structure.

**3.1. Wavelet Transform.** In the last decade, wavelet transform has been recognized as a powerful tool in a wide range of applications, including image/video processing, numerical analysis, and telecommunication. The advantage of wavelet is that wavelet performs an MRA of a signal with localization in both time and frequency [10, 11]. In addition to this, functions with discontinuities and functions with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sine cosine basis vectors to achieve a comparable approximation. Wavelet operates by convolving the target function with wavelet kernels to obtain wavelet coefficients representing the contributions in the function at different scales and orientations. Wavelet or multiresolution theory can be used alongside segmentation approaches, creating new systems which can provide a segmentation of superior quality to those segmentation approaches computed exclusively within the spatial domain [12].

Discrete wavelet transform (DWT) can be implemented as a set of high-pass and low-pass filter banks. In standard wavelet decomposition, the output from the low-pass filter can be then decomposed further, with the process continuing recursively in this manner. According to [13], DWT can be mathematically expressed by

$$\begin{aligned} a^j(n) &= \sum_{i=0}^{L-1} l(i) \cdot a^{j-1}(2n-i), \quad 0 \leq n < N_j, \\ d^j(n) &= \sum_{i=0}^{L-1} h(i) \cdot d^{j-1}(2n-i), \quad 0 \leq n < N_j. \end{aligned} \quad (1)$$

The coefficients  $a^j(n)$  and  $d^j(n)$  refer to approximation and detailed components in the signal at decomposition level  $j$ , respectively. The  $l(i)$  and  $h(i)$  represent the coefficients of low-pass and high-pass filters, respectively.

DWT decomposes the signal into a set of resolution-related views. The wavelet decomposition of an image creates at each scale  $j$  a set of coefficient values  $w_j$  with an overall mean of zero.  $w_j$  contains the same number of voxels as the

original image; therefore, this wavelet transform is redundant [14, 15].

For images, 1D-DWT can be readily extended into 2D. In standard 2D wavelet decomposition, the image rows are fully decomposed, with the output being fully decomposed columnwise. In nonstandard wavelet decomposition, all the rows are decomposed by one decomposition level followed by one decomposition level of the columns. Figure 2 illustrates the filter structure of 2D-DWT.

Wavelet uses a set of filters to decompose images depending on filter coefficients and the number of those coefficients. The most popular wavelet filter is Haar wavelet filter (HWF) which takes the averages and differences from the low- and high-pass filters, respectively. Figure 3 illustrates an example of applying 2D-DWT using HWF on an image for 2 levels of decompositions.

**3.2. Ridgelet Transform.** In 1998, Donoho introduced the ridgelet transform [16] continuous ridgelet transform (CRT) can be defined from a 1D wavelet function oriented at constant lines and radial directions. Ridgelet transform [17–19] has been generating a lot of interest due to their superior performance over wavelets. While wavelets have been very successful in applications such as denoising and compact approximations of images containing zero dimensional or point singularities. Wavelets do not isolate the smoothness along edges that occurs in images [20], and they are thus more appropriate for the reconstruction of sharp point singularities than lines or edges. These shortcomings of wavelet are well addressed by the ridgelet transform; the functionality of wavelet has been extended to higher dimensional singularities and becomes an effective tool to perform sparse directional analysis [3, 21]. Generally speaking, wavelets detect objects with point singularities, while ridgelets are able to represent objects with line singularities.

The finite ridgelet transform (FRIT) was computed in two steps: a calculation of discrete radon transform and an application of a wavelet transform. The finite radon transform (FRAT) is computed in two steps: a calculation

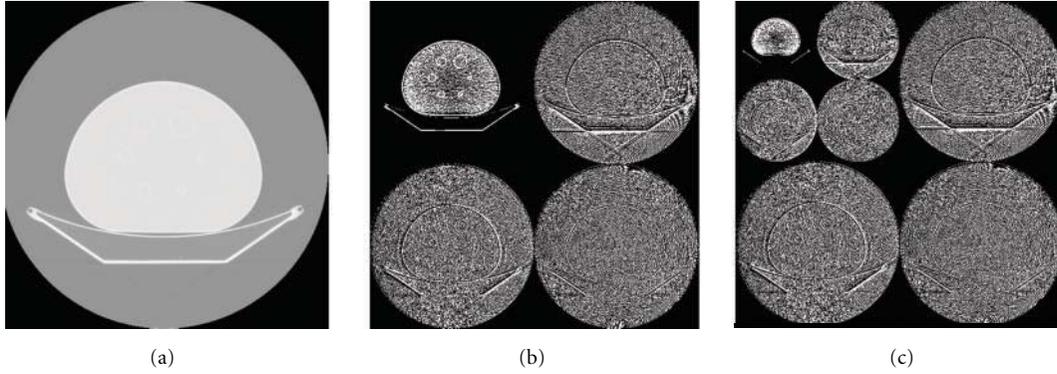


FIGURE 3: 2D-DWT. Original image (a), first decomposition level (b), and second decomposition level (c).

of 2D Fast Fourier Transform (FFT) for the image and an application of a 1D inverse fast Fourier transform (iFFT) on each of the 32 radial directions of the radon projection. 1D wavelet is applied restricted to radial directions going through the origin for three levels of decompositions.

Applying FRAT on image can be presented as a set of projections of the image taken at different angles to map the image space to projection space. Its computation is important in image processing and computer vision for problems such as pattern recognition and the reconstruction of medical images. For discrete images, a projection is computed by summation of all data points that lie within specified unit-width strips; those lines are defined in a finite geometry [22].

Depending on [23], FRAT of a real function on the finite grid  $Z_p^2$  is defined in

$$r_k[l] = \text{FRAT}_f(k, l) = \frac{1}{\sqrt{P}} \sum_{(i,j) \in L(k,l)} f(i, j). \quad (2)$$

Here,  $L(k, l)$  denotes the set of points that make up a line on the lattice  $Z_p^2$  as in

$$L(k, l) = \{(i, j) : j = k_i + l(\text{mod } p), i \in Z_p\}, \quad 0 \leq k < p, \\ L(p, l) = \{(l, j) : j \in Z_p\}. \quad (3)$$

To compute the  $K$ th radon projection (i.e., the  $K$ th row in the array), all pixels of the original image need to be passed once and use  $P$  histogrammers: one for every pixel in the row [12]. At the end, all  $P$  histogrammed values are divided by  $K$  to get the average values.

According to Alzu'bi and Amira in [3], once the wavelet and radon transforms have been implemented, the ridgelet transform is straightforward. Each output of the radon projection is simply passed through the wavelet transform before it reaches the output multiplier. As shown in Figure 4, ridgelets use FRAT as a basic building block, where FRAT maps a line singularity into point singularity, and the wavelet transform has used to effectively detect and segment the point singularity in radon domain.

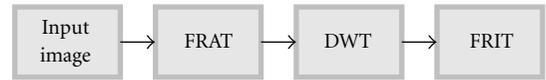


FIGURE 4: FRIT block diagram.

Figure 5 shows a clinical chest slice from a CT scanner [24] in the last step of ridgelet transform before image reconstruction at different block sizes.

Continuous ridgelet transform is similar to the continuous wavelet transform except that point parameters  $(x, y)$  in the cartesian grid (Figure 6(a)) which perform pixels in the image or an entry in a 2D matrix are now replaced by line parameters  $(\beta, \theta)$ , where  $\beta$  is the intercept and  $\theta$  is the angle. Figure 6(b) illustrates the radial grid in ridgelet transform; however, straight lines evaluate the image in the frequency domain [3].

The segmentation result achieved using ridgelet transformation on medical images was not promising. Medical images comprised from curves which are still not singularity points after applying radon transform. Wavelet transform cannot detect those singularities properly, since it still not singularity points [3], resulting that ridgelet transformation is not suitable for segmenting these images.

Ridgelet transform can be used in other applications, where images contain edges and straight lines. Curvelet transform has been introduced to solve this problem; it deals with higher singularities compared to wavelet and ridgelet transforms.

**3.3. Curvelet Transform.** The curvelet transform has gone through two major revisions. It was first introduced in [25, 26] by Candés and Donoho in 2000, which used a complex series of steps involving the ridgelet analysis of the radon transform of an image. Their performance was very slow; hence, researchers developed a new version which is easier to use and understand. In this new method, the use of the ridgelet transform as a preprocessing step of curvelet was discarded, thus reducing the amount of redundancy in the transform and increasing the speed considerably [3].

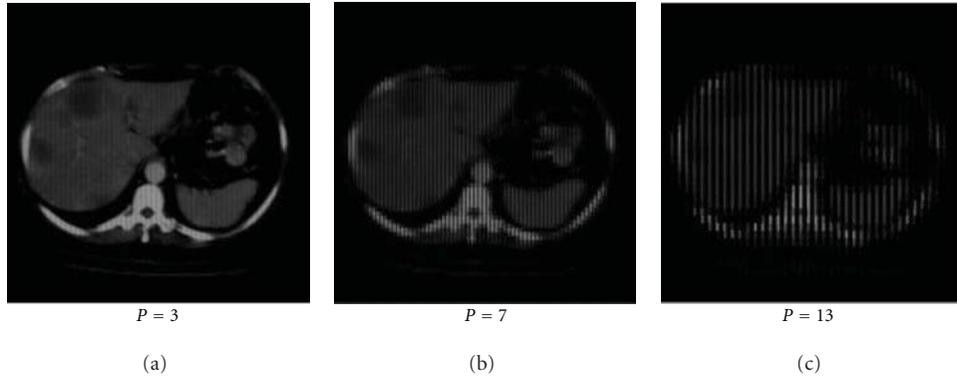


FIGURE 5: Ridgelet transform for real CT images at block sizes (3, 7, and 13).

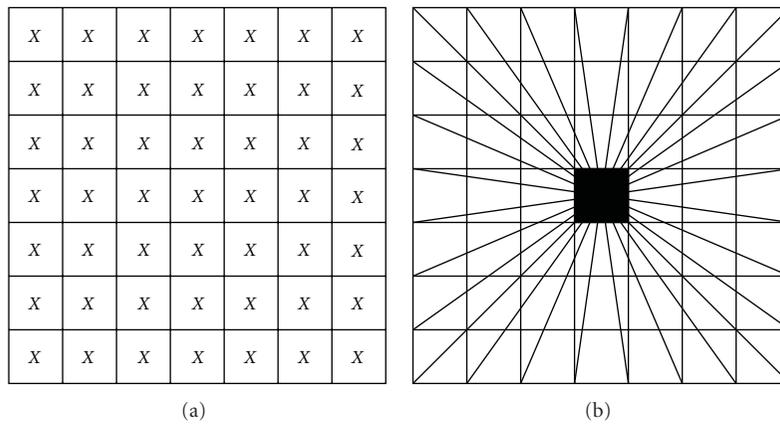


FIGURE 6: Wavelet and ridgelet parameters.

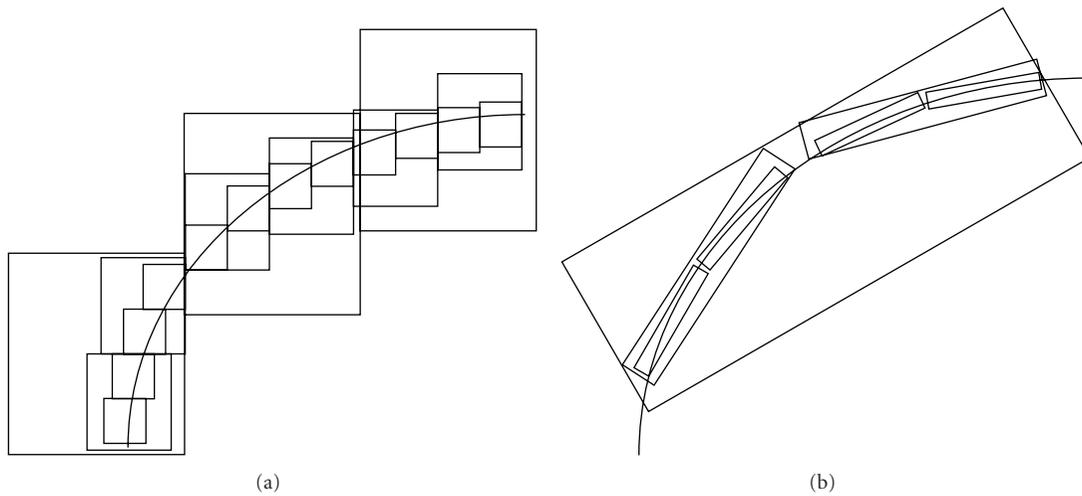


FIGURE 7: An approximating comparison between wavelet (a) and curvelet (b).

Curvelet aims to deal with interesting phenomena occurring along curved edges in a 2D image. As illustrated in Figure 7, curvelet needs fewer coefficients for representation, and the edge produced from curvelet is smoother than wavelet edge [27].

The newly constructed and improved version of curvelet transform is known as Fast Discrete Curvelet Transform

(FDCT). This new technique is simpler, faster and less redundant than the original curvelet transform which based on ridgelets. According to Candes et al. in [9], two implementations of FDCT are proposed:

- (i) unequally spaced Fast Fourier transforms (USFFT),
- (ii) wrapping function.

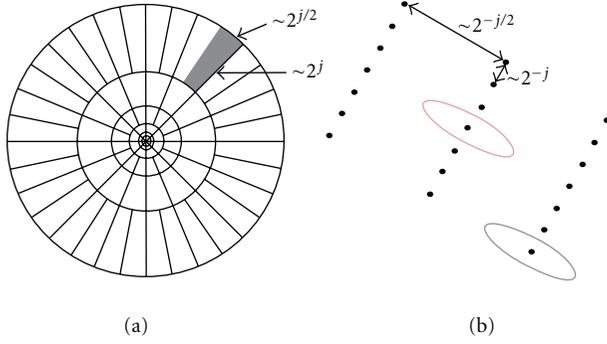


FIGURE 8: Curvelet tiling of space and frequency. The induced tiling of the frequency plane (a). The spatial Cartesian grid associated with a given scale and orientation (b).

Both implementations of FDCT differ mainly by the choice of spatial grid that used to translate curvelets at each scale and angle. Both digital transformations return a table of digital curvelet coefficients indexed by a scale parameter, an orientation parameter, and a spatial location parameter. Wrapping-based transform is based on wrapping a specially selected Fourier samples, and it is easier to implement and understand.

**3.3.1. Continuous-Time Curvelet Transform.** Curvelet transform works in two dimensions with spatial variable  $x$ , frequency domain variable  $\omega$ , and the frequency-domain polar coordinates  $r$  and  $\theta$ . Curvelet transform can be defined by a pair of windows, radial window  $\{W(r)\}$ , and angular window  $\{V(t)\}$  [9]. As illustrated in (4), these windows will always obey the admissibility conditions.

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in \left(\frac{3}{4}, \frac{3}{2}\right), \quad (4)$$

$$\sum_{j=-\infty}^{\infty} V^2(t - l) = 1, \quad t \in \left(-\frac{1}{2}, \frac{1}{2}\right).$$

A polar “wedge” represented by  $U_j$  is supported by the radial window  $\{W(r)\}$  and angular window  $\{V(r)\}$ . Equation (5) defines  $U_j$  in the Fourier domain

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor} \theta}{2\pi}\right). \quad (5)$$

Equation (6) defines the curvelet transform as a function of  $\{x = (x_1, x_2)\}$  at scale  $2^{-j}$ , orientation  $\theta_l$  and position  $x_k(j, l)$ , where  $R_\theta$  is the rotation in radians. Figure 8 illustrates the induced tiling of the frequency plane and the spatial Cartesian grid associated with a given scale and orientation [9], and shaded area represents the polar wedge by  $U_j$

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}\left(x - x_k^{(j,l)}\right)\right). \quad (6)$$

**3.3.2. Fast Discrete Curvelet Transform via Wrapping.** The new implementation of curvelet transform based on Wrapping of Fourier samples takes a 2D image as an input in the

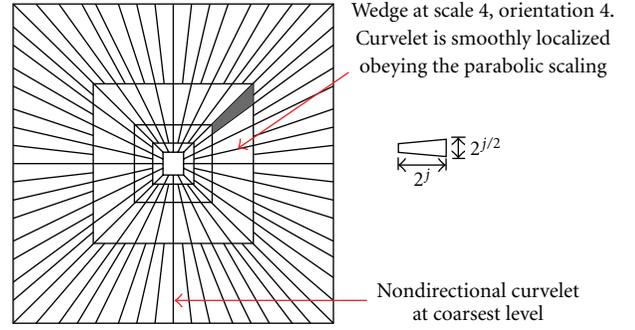


FIGURE 9: 5-level curvelet digital tiling of an image.

form of a Cartesian array  $f[m, n]$ , where  $0 \leq m < M$ ,  $0 \leq n < N$  where  $M$  and  $N$  are the dimensions of the array. As illustrated in (7), the outputs will be a collection of curvelet coefficients  $c^D(j, l, k_1 k_2)$  indexed by a scale  $j$ , an orientation  $l$  and spatial location parameters  $k_1$  and  $k_2$ .

$$c^D(j, l, k_1 k_2) = \sum_{\substack{0 \leq m < M \\ 0 \leq n < N}} f[m, n] \varphi_{j,l,k_1 k_2}^D[m, n]. \quad (7)$$

Each  $\varphi_{j,l,k_1 k_2}^D$  is a digital curvelet waveform, superscript  $D$  stands for “digital.” These approach implementations are the effective parabolic scaling law on the subbands in the frequency domain to capture curved edges within an image in more effective way. As mentioned earlier, wrapping based curvelet transform is a multiscale pyramid which consists of several subbands at different scales consisting of different orientations and positions in the frequency domain. At a high frequency level, curvelets are so fine and looks like a needle shaped element and they are non-directional coarse elements at low frequency level.

Figure 9 illustrates the whole image represented in spectral domain in the form of rectangular frequency tiling by combining all frequency responses of curvelets at different scales and orientations. It can be seen that curvelets are needle like elements at higher scale.

It can be seen from Figure 9 that curvelet becomes finer and smaller in the spatial domain and shows more sensitivity to curved edges as the resolution level is increased, thus allowing to effectively capturing the curves in an image, and curved singularities can be well-approximated with fewer coefficients.

In order to achieve a higher level of efficiency, curvelet transform is usually implemented in the frequency domain. This means that a 2D FFT is applied to the image. For each scale and orientation, a product of  $U_{jl}$  “wedge” is obtained; the result is then wrapped around the origin, and 2D IFFT is then applied resulting in discrete curvelet coefficients. Candes et al. describe the discrete curvelet transform in [9] as illustrated in

$$\text{Curvelet transform} = \text{IFFT}[\text{FFT}(\text{Curvelet}) \times \text{FFT}(\text{Image})]. \quad (8)$$

The difficulty behind this is that trapezoidal wedge does not fit in a rectangle of size  $2^j \times 2^{j/2}$  aligned with the axes in

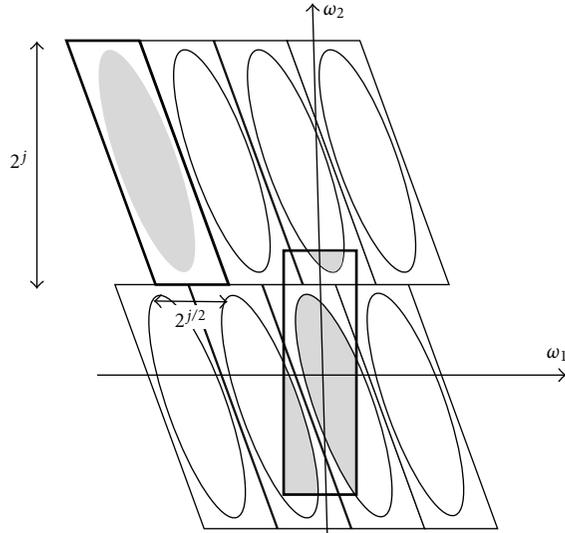


FIGURE 10: Wrapping wedge data.

the frequency plane in which the 2D IFFT could be applied to collect curvelet coefficients. Wedge wrapping procedure proposed in [9] uses a parallelogram with sides  $2^j$  and  $2^{j/2}$  to support the wedge data. The wrapping is done by periodic tiling of the spectrum inside the wedge and then collecting the rectangular coefficient area in the centre. The centre rectangle of size  $2^j \times 2^{j/2}$  successfully collects all the information in that parallelogram [28]. Figure 10 illustrates the process of wrapping wedge where the angle  $\theta$  is in the range  $(\pi/4, 3\pi/4)$  and the rectangles have the same width and length as the parallelogram is centred at the origin [9].

The following are the steps of applying wrapping based FDCT algorithm [9].

*Step 1.* Apply the 2D FFT to an image to obtain Fourier samples

$$\hat{f}[m, n], \quad -\frac{n}{2} \leq m, n < \frac{n}{2}. \quad (9)$$

*Step 2.* For each scale  $j$  and angle  $l$ , form the product

$$\tilde{U}_{j,l}[m, n] \hat{f}[m, n]. \quad (10)$$

*Step 3.* Wrap this product around the origin and obtain

$$\tilde{f}_{j,l}[m, n] = W(\tilde{U}_{j,l} \hat{f})[m, n], \quad (11)$$

where the range for  $m, n$ , and  $\theta$  is now  $0 \leq m < 2^j$ ,  $0 \leq n < 2^{j/2}$ , and  $-\pi/4 \leq \theta < \pi/4$ .

*Step 4.* Apply IFFT to each  $\tilde{f}_{j,l}$ , hence collecting the discrete coefficients  $c^D(j, l, k_1, k_2)$ .

The curvelet transform is a multiscale transform such as wavelet, with frame elements indexed by scale and location parameters. Wavelets are only suitable for objects with point singularities, Ridgelets are only suitable for objects with line singularities, while curvelets have directional parameters and

its pyramid contains elements with a very high degree of directional specificity. The elements obey a special scaling law, where the length and the width of frame elements support are linked using

$$\text{width} \approx \text{length}^2 \quad (12)$$

Discrete curvelet transform in the spectral domain utilizes the advantages of FFT. During FFT, both image and curvelet at a given scale and orientation are transformed into the Fourier domain. The convolution of the curvelet with the image in the spatial domain then becomes their product in the Fourier domain. At the end of this computation process, a set of curvelet coefficients are obtained by applying IFFT to the spectral product. This set contains curvelet coefficients in ascending order of the scales and orientations.

Curvelets are superior to the other transforms as in the following.

(a) *Optimally Sparse Representation of Objects with Edges.* Curvelets provide optimally sparse representation of objects which display curve-punctuated smoothness except for discontinuity along a general curve with a bounded curvature. Such representations are nearly as sparse as if the object were not singular and turn out to be far sparser than other transforms decomposition of the object.

(b) *Optimal Image Reconstruction in Severely Ill-Posed Problems.* Curvelets also have special microlocal features which make them especially adapted to certain reconstruction problems with missing data. For example, in many important medical applications, one wishes to reconstruct an object  $f(x_1, x_2)$  from noisy and incomplete tomographic data [28]. Because of its relevance in biomedical imaging, this problem has been extensively studied, yet curvelets offer surprisingly new quantitative insights [18]. For example, an application of the phase-space localization of the curvelet transform allows a very precise description of those features of the object of function ( $f$ ) which can be reconstructed accurately from such data and how well, and of those features which cannot be recovered.

As illustrated in (8), the data acquisition geometry separates the curvelet expansion of the object into two pieces as illustrated in

$$f = \sum_{n \in \text{Good}} \langle f, \varphi_n \rangle \varphi_n + \sum_{n \notin \text{Good}} \langle f, \varphi_n \rangle \varphi_n. \quad (13)$$

The first part of (13) can be recovered accurately, while the second part cannot. What is interesting here is that one can probably reconstruct the recoverable part with an accuracy similar to that one would achieve even if one had complete data. There is indeed a quantitative theory showing that for some statistical models which allow for discontinuities in the object to be recovered, there are simple algorithms based on the shrinkage of curvelet biorthogonal decompositions, which achieve optimal statistical rates of convergence [18].

Figure 11 illustrates the frequency response of curvelets at different scales and orientations for some test images using

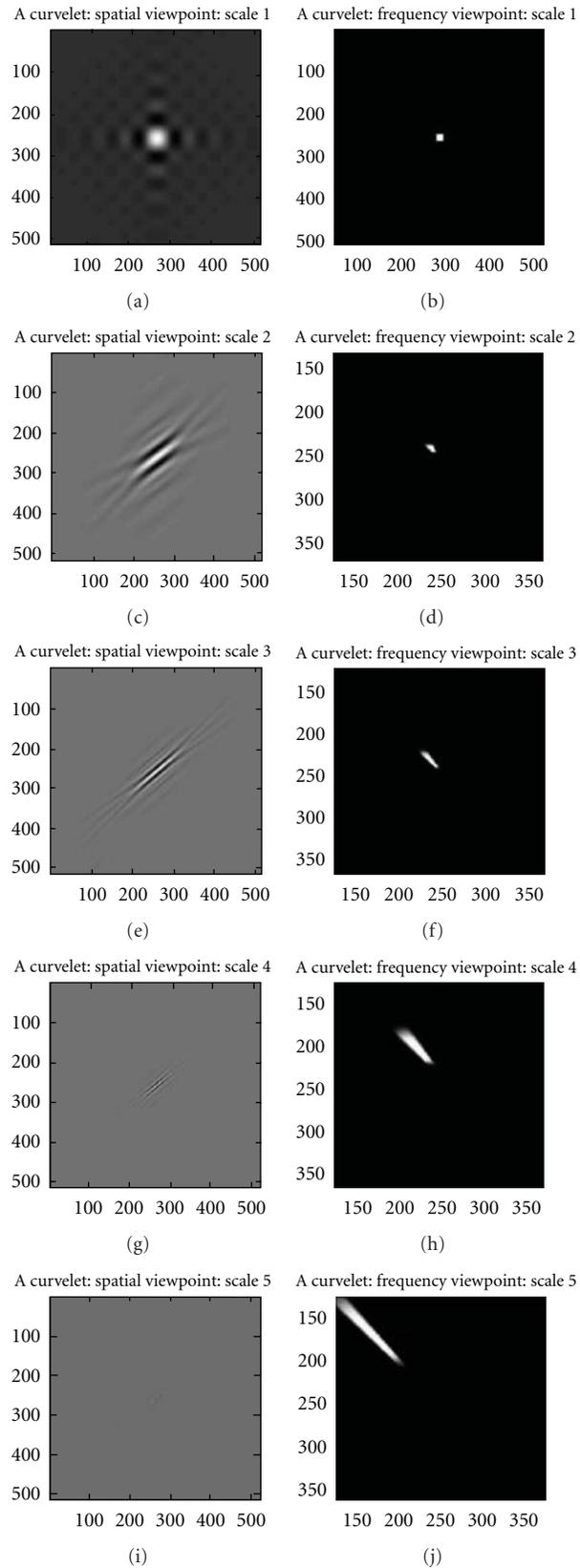


FIGURE 11: Curvelets at increasingly fine scales from 1 to 5. Spatial domain (a, c, e, g, i). Frequency domain (b, d, f, h, j) [29].

TABLE 1: Wavelet and ridgelet comparisons depending on SNR and processing time.

Domain	Wavelet			Ridgelet			Spatial
	Level 1	Level 2	Level 3	$P = 5$	$P = 11$	$P = 31$	
SNR (dB)	10.63	11.14	10.95	10.37	11.43	11.88	7.17
Time (sec)	0.23	0.24	0.50	71.5	29.91	10.01	1.18

“Curvelab (version: 2.1.2)” [29] in both spatial and frequency domain.

It can be seen from Figure 11 that curvelets are nondirectional at coarsest level. Figures 11(a), 11(c), 11(e), 11(g), and 11(i) are the spatial representation of curvelet at scales 1 to 5. And Figures 11(b), 11(d), 11(f), 11(h), and 11(j) are the frequency domain representation of curvelet that is modulus of FFT. Figure 12 illustrates a clinical data for human chest from CT scanner in spatial domain and its curvelet coefficients.

In Figure 12, the low frequency coefficients (coarse scale) are stored at the centre of the display. The concentric coronae (formed by black strips) show coefficients at different scales and the outer coronae correspond to higher frequencies. Each corona has four strips further subdivided in angular panels; each panel represents coefficients at a specified scale and orientation suggested by the position of the panel.

Wedge wrapping is done for all the wedges at each scale in the frequency domain to obtain a set of subbands or wedges at each curvelet decomposition level, and these subbands are the collection of discrete curvelet coefficients.

The aim is to identify the most effective texture descriptor for medical images to capture edge information more accurately. The discrete curvelet transform can be calculated to various resolutions or scales and angles; the maximum number of resolution depends on the original image size and the angles. Number of angles at the second coarsest level must be at least eight and a multiple of four; that is,  $512 \times 512$  image has five maximum possible resolution levels containing structural information of the image. Figure 13 illustrates how curvelet-based edge reconstruction in medical imaging differs from other transforms methods.

#### 4. Results and Analysis

The end users of the proposed system are the radiologists and specialists who analyse medical images for cancer diagnosis. After several meetings with those people in the radiology departments in some hospitals, the main goal that they are working is to detect the accurate cancer size in medical images with the least error. This process may be affected by the noise surrounding ROI, which make the process of measuring the exact dimensions of the lesion so hard.

Different datasets have been carried out with the proposed system to validate it for clinical applications. The first one is NEMA IEC body phantom which consists of an elliptical water filled cavity with six spherical inserts suspended by plastic rods of inner diameters: 10, 13, 17, 22, 28, and 37 mm [25, 26]. Real clinical human images acquired by a CT scanner [24] have also been used to experiment the

proposed approaches, this data has been previously analysed by the radiologists and the provided reports explains that the patients are diagnosed by cancer. Table 1 illustrates the SNR values of extracted features from NEMA IEC DATA SET in spatial domain, different levels of decomposition of wavelet domain and at different block sizes in ridgelet domain.

It can be seen from Table 1 that small values of SNR have been obtained for all techniques; this is due to the noise from the acquisition systems itself. This noise will be a part of the medical image after the reconstruction of all slices. Relatively, better SNR values can be achieved with the second level of wavelet decomposition and as the block size ( $p$ ) is getting bigger with the ridgelet transform, where the transformed image is getting more similar to the original image. This can be assigned to the major limitation of using ridgelet transformation in medical image segmentation, where ridges rarely exist in such data.

MRA transforms have been used with thresholding technique to segment the experimental data. Thresholding technique has been applied as a preprocessing step on the original images at threshold value ( $t = 35$ ) to remove as much artificial spam sequel produced from the scanners. The transform then applied to effectively represent objects with edges which are the contours of the medical images followed by another thresholding at ( $t = 7$ ) to remove most of the remaining noise and facilitate the measurement process.

Figure 14 illustrates the segmentation using curvelet transform. Figures 14(a) and 14(c) illustrate the original images from a CT scanner, and Figures 14(b) and 14(d) illustrate the segmented phantom image and real chest image, respectively, using curvelet transform. As illustrated in Figure 15, results of the proposed segmentation technique are vary in terms of smooth reconstruction of the spheres. Curvelet transform segments the input image and removes artifacts from the image to exhibit smooth and optimal segmentation of NEMA phantom. Ridgelet transform detect ROI but does not give promising segmentation results due to the lack of ridges or straight lines in the tested data set. Wavelet quadrants are varying also in their quality; relatively, the best results have been achieved with the LL-filter output.

Table 2 illustrates NEMA spheres diameters error percentages measured using different multiresolution analysis techniques and compared to previously implemented techniques. ED has been used to measure the spheres diameters and calculate the error percentages for each technique, and sphere diameter error percentages have been calculated as follows:

$$\text{error \%} = \frac{\text{Measured Diameter} - \text{Actual Diameter}}{\text{Actual Diameter}} \times 100\%. \quad (14)$$

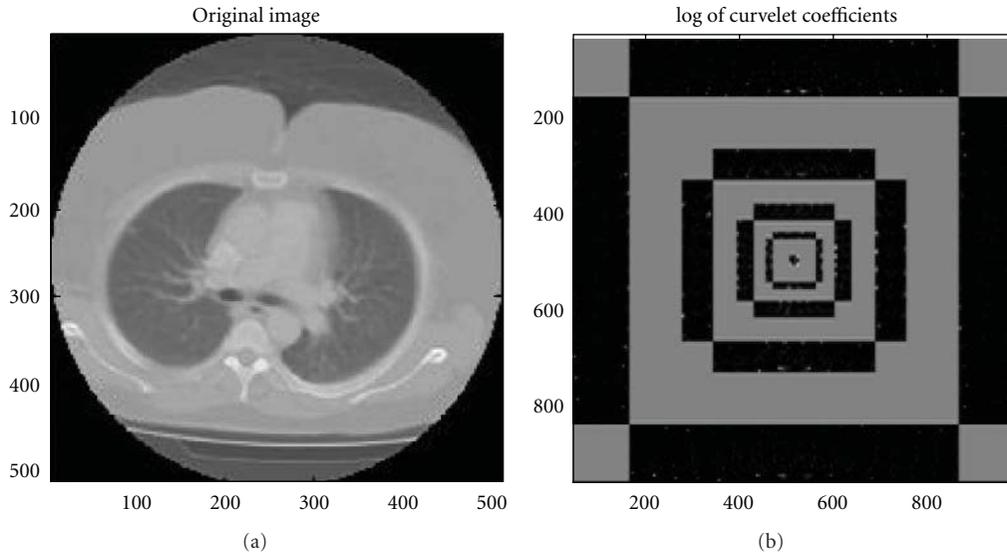


FIGURE 12: Clinical slice for the human chest from a CT scanner in spatial (a) and curvelet coefficients (b).

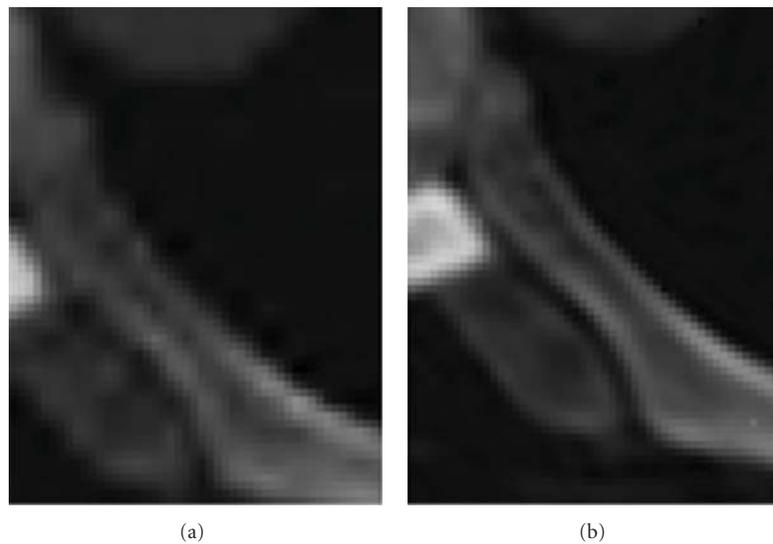


FIGURE 13: Reconstruction of tomographic data. Wavelet domain (a), and curvelet domain (b).

From Table 2, in the case of K-means clustering, tumor volumes are underestimated by approximately 5%-6% in most cases; however, for the two smaller spherical inserts, with diameter of 10 mm and 13 mm, respectively, these underestimations are significantly greater. For the smallest sphere, more than a 13% volume discrepancy is recorded, with the K-means algorithm finding it difficult to quantify the tumor accurately. Sphere 2 similarly is massively underestimated (11.5%). Unlike K-means clustering, MRFM tends to overestimate the volumes of the spherical inserts, with the exception of Spheres 1 and 2.

Spheres diameters are reduced to the half with each decomposition level of wavelet transform. Three decomposition levels of DWT have been applied on NEMA phantom [25, 26] using two different filters (Haar, Daubechies),

and the measured diameters were doubled at each level to produce a fair comparison with the other available techniques. It can be seen that most of the error percentages were decreasing while the spheres diameter increasing; it is worth mentioning that there is no upper bound of the spheres diameters to keep the errors decreasing, because the ROI becomes clearer and easier to be detected and measured properly. But tumours in real life are usually very small in the early stage cancer, and the problem is to detect those turnouts' tumours as soon as possible.

The two smallest spherical inserts are still underestimated in most of the techniques and got the largest error percentages. The large volumetric errors encountered using this acquisition exist as a consequence of the poor slice thickness setting selected for the scan. The 4.25 mm slice thickness

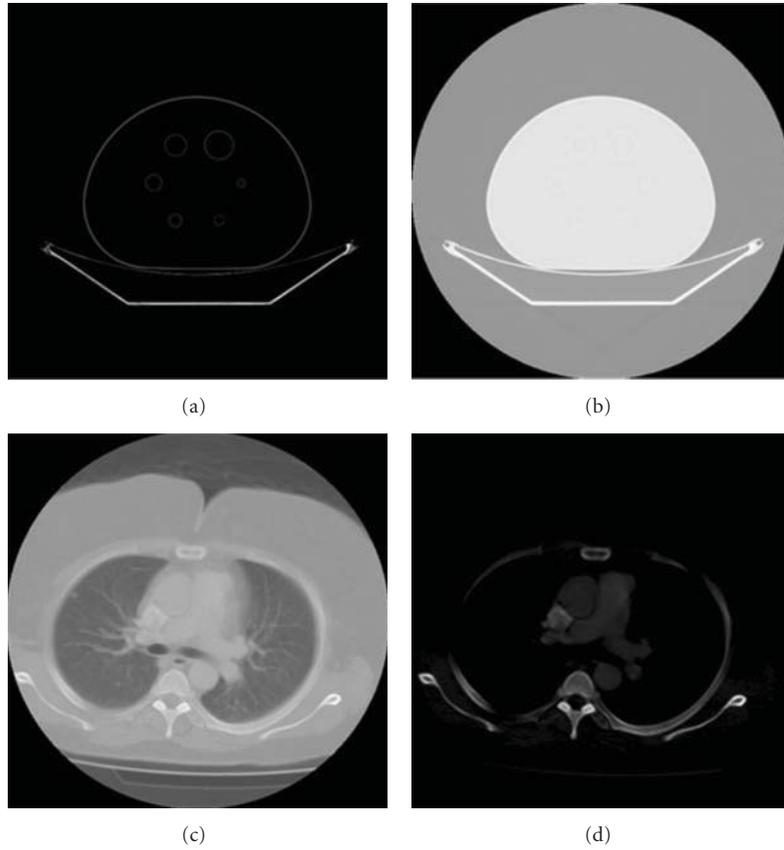


FIGURE 14: Curvelet transform for segmentation. (a) NEMA IEC body phantom, (b) segmented phantom slice, (c) original real chest slice, and (d) segmented real chest slice.

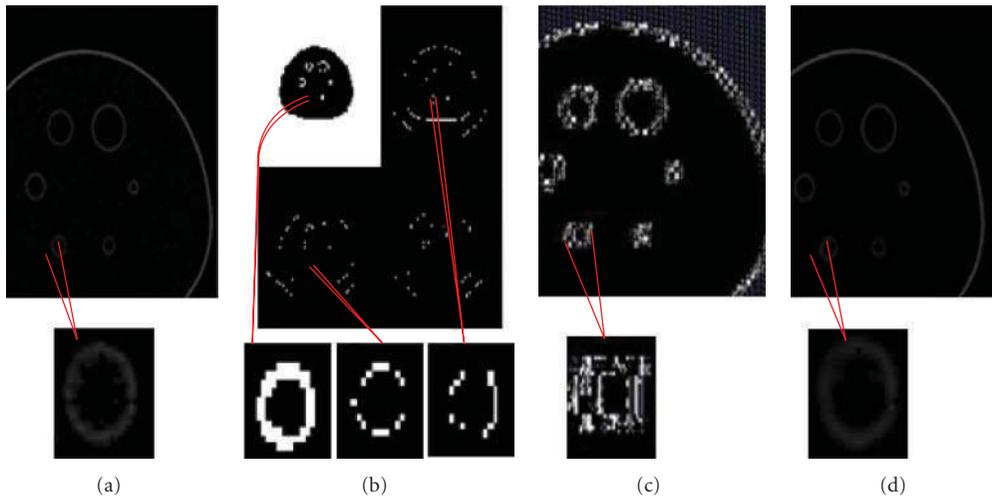


FIGURE 15: Segmentation using conventional hard thresholding and curvelet-based segmentation. (a) Denoised spatial domain. (b) First level in wavelet domain. (c) Ridgelet domain. (d) Curvelet domain.

causes large fluctuations in transaxial tumour areas to occur between image slices. This problematic characteristic occurs most notably with the smallest spherical inserts, where single voxel reallocation causes a large deviation in percentage error. In Figure 16, the percentage error computed between the actual sphere volume and the volumes obtained using all

methodologies for each of the six tumours inserts is plotted. It can be seen that all techniques are settled down according to the error percentages as the sphere diameters increased.

It can be also seen from Table 2 that acceptable error percentages have been achieved using ridgelet transform for the big spheres (22 mm, 28 mm, and 37 mm), where the

TABLE 2: The error percentages of spheres diameters measurements for NEMA IEC body phantom.

Spheres (mm)	Error % for measured diameters					
	10	13	17	22	28	37
K-means [3]	-13.6	-11.5	-5.77	-5.51	-5.1	-5.01
MRFM [3]	-7.41	-8.69	4.28	4.06	3.9	3.89
Clustering [3]	18.6	16.0	9.0	7.5	5.5	1.1
Iterative Thresholding [3]	3.0	3.1	0.6	0.9	1.1	1.8
Wavelet						
Haar						
Level 1	-2.9	-2.46	1.35	0.82	0.29	0.05
Level 2	-10.9	-6.67	3.88	-1.3	-0.76	-1.95
Level 3	—	—	5.65	-18.2	2.57	-3.24
Daubechies						
Level 1	-7.43	-2.69	0.12	2.0	2.17	1.81
Level 2	-5.2	0.15	-4.24	0.73	0.62	-0.11
Ridgelet	-10.93	-6.67	3.88	-1.30	-0.76	-1.95
Curvelet	2.65	1.62	1.07	-0.82	-0.33	-0.09

TABLE 3: Comparison of curvelet, ridgelet, and wavelet denoising in terms of PSNR and MSE.

Image name	Curvelet denoising		Ridgelet denoising		Wavelet denoising	
	MSE	PSNR (dB)	MSE	PSNR (dB)	MSE	PSNR (dB)
NEMA	41.67	31.93	108.78	26.14	101.12	28.08
Chest	58.8	30.44	152.45	23.55	147.63	26.44

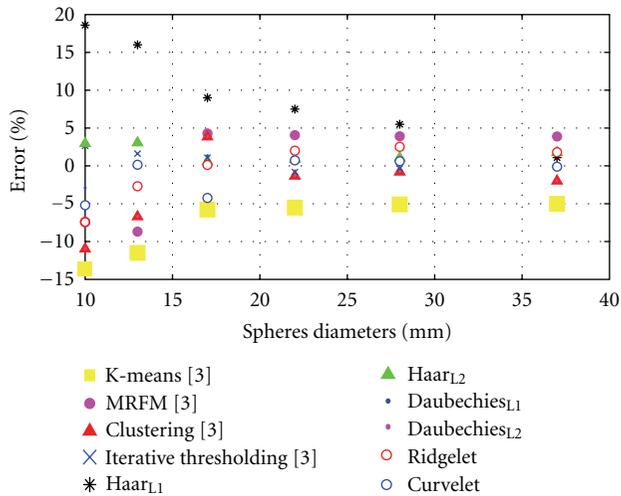


FIGURE 16: Visual comparison for error percentages in Table 2.

curves are not sharp and ridgelet detect it accurately as it close to be ridges. But for the small spheres, ridgelet weakness for medical image segmentation start appears clearly.

Curvelet transform overcomes the weakness of wavelet for segmenting sharp curves and detect the small spheres accurately with error percentages (0.82%–2.65%). For the big spheres, errors achieved using wavelet transform are still better than those achieved using curvelet transform due to the sharpness of that spheres. But still very good results using curvelet transform and acceptable for clinical applications.

PSNR and MSE have been also used to evaluate the quality of the proposed techniques. The original image has been contaminated with Gaussian white noise at  $\sigma = 20\%$  of the maximum intensity. Table 3 illustrates a comparison study of curvelet transform with the other traditional transforms, and comparison terms PSNR and MSE have been used to test the quality of the transformed image.

From Table 3, it can be seen that the best results according to both PSNR and MSE have been achieved using curvelet transform. Wavelet transform performs better results compared to ridgelet transform in both validation metrics. Figure 17 illustrates two noisy images and the denoised outputs using both wavelet and curvelet.

According to a study done by Dettori and Semler [23], the ridgelet-based descriptors had significantly higher performance measures in comparison to wavelet-based descriptors, with accuracy rates higher than any other wavelet-based feature set for all individual organs. This is not surprising given the fact that the ridgelet transform is able to capture multidirectional features, as opposed to the wavelet transform which focus mainly on horizontal, vertical, and diagonal features. This can be generalized to most of the images except for medical scanners, where the weakness of wavelet is not dominant in such images.

Curvelet-based descriptors had an even higher performance in comparison to both the wavelet and ridgelet, with accuracy rates higher, respectively. The accuracy rate using curvelet transform is better; this is expected, since the curvelet transform is able to capture multidirectional

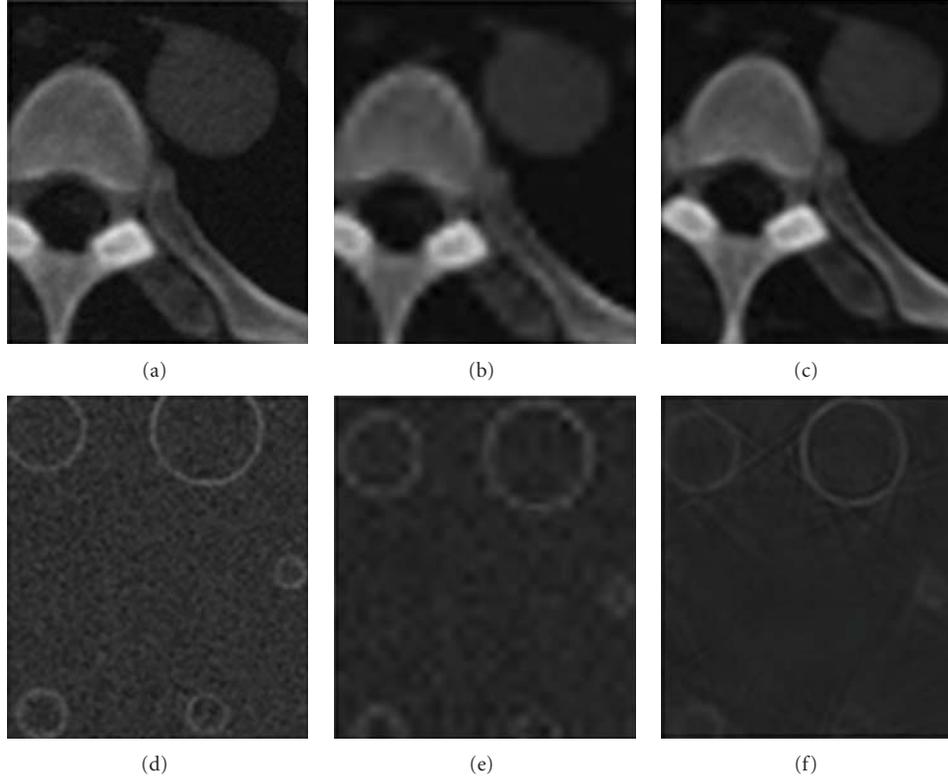


FIGURE 17: MRA for image denoising. (a, d) Noisy images. (b, e) Wavelet. (c, f) Curvelet.

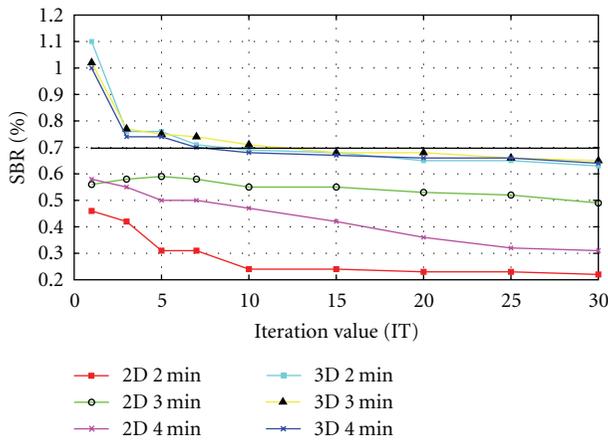


FIGURE 18: Scanner variables effects on the segmented image.

features in wedges, as opposed to lines or points as in the ridgelet or wavelet transform. The multidirectional features in curvelets are very effective in extracting the important features from medical images and then segmented accurately.

As illustrated in the previous tables and figures, it can be seen that more efficient and smooth image reconstruction is achieved using curvelet transform. In terms of optimal reconstruction of the objects with edges and curves, curvelet-based techniques outperform the traditional wavelet and ridgelet transforms.

The algorithm presented in this chapter is able to classify normal tissues in CT scans with high accuracy rates. These hypotheses will be further tested and validated on different predefined clinical data sets in chapter 8 of this thesis.

Segmentation using curvelet transform has been chosen for experimenting the PET scanner sensitivity variables, curvelet was applied in parallel with multithresholding and classification techniques to classify the spheres in a separate class from the other comprising objects at least noise included. The experiment was evaluated based on the ratio between the spheres area to the other area of the scanned slice. The actual spheres area can be calculated according to (15), given that the spheres diameters are 10, 13, 17, 22, 28, 37 mm

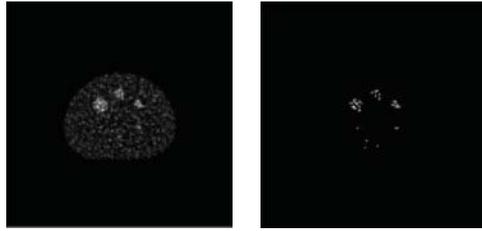
$$S_{\text{Original}} = \frac{1}{2}\pi \sum_{r \in \{a\}} r^2, \quad \{a\} = \{10, 13, 17, 22, 28, 37\}, \quad (15)$$

where  $S_{\text{Original}}$  is the actual area of all six spheres together. The scan resolution can be acquired from *Amide* software where each pixel size is  $4.6875 \times 4.6875$  mm and each slice size is  $128 \times 128$  pixels. The overall slice area and the ratio between both areas can be calculated according to (16), respectively, where SBR is the spheres to background ratio

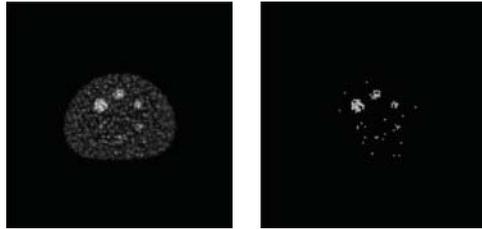
$$A_{\text{Original}} = 128 \times 128 \times 4.6875 \times 4.6875 = 360000 \text{ mm}^2,$$

$$\text{SBR} = \frac{S_{\text{Original}}}{A_{\text{Original}} - S_{\text{Original}}} \times 100\% = 0.702\%.$$

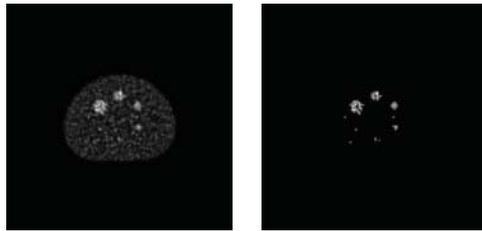
(16)



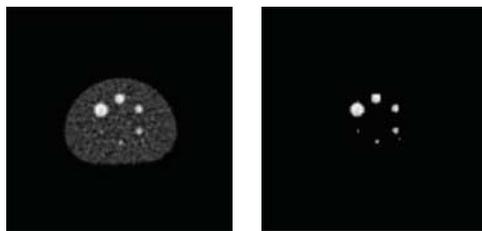
(a) IT = 10, time = 2 min, collimator: 2D (left), segmented (right)



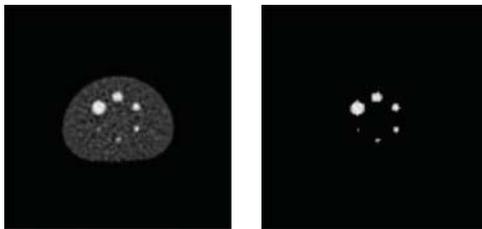
(b) IT = 10, time = 3 min, collimator: 2D (left), segmented (right)



(c) IT = 10, time = 4 min, collimator: 2D (left), segmented (right)

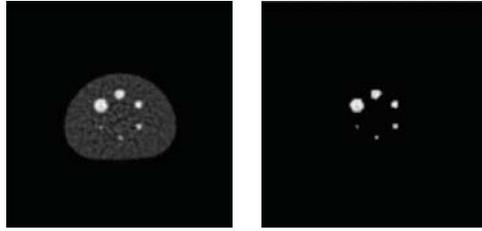


(d) IT = 10, time = 2 min, collimator: 3D (left), segmented (right)



(e) IT = 10, time = 3 min, collimator: 3D (left), segmented (right)

FIGURE 19: Continued.



(f) IT = 10, time = 4 min, collimator: 3D (left), segmented (right)

FIGURE 19: Segmented results achieved at IT value (10), where the best results detected.

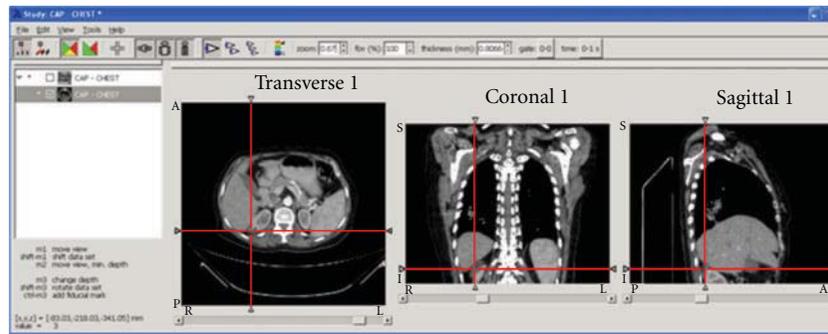


FIGURE 20: AMIDE snapshot locating the kidney cancer.



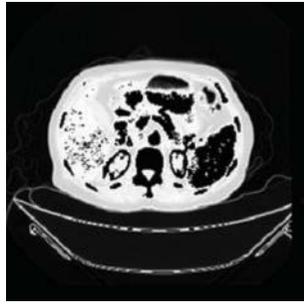
FIGURE 21: ROI highlighted in the original image (kidney cancer).

TABLE 4: Spheres to background ratio (SBR) for different variable samples.

2D/3D	Time/bed section	Iteration	SUB	SBR (%)
2D	2 min	1	30	0.46
2D	2 min	5	30	0.31
2D	2 min	10	30	0.24
2D	2 min	30	30	0.22
2D	3 min	1	30	0.56
2D	3 min	5	30	0.59
2D	3 min	10	30	0.55
2D	3 min	20	30	0.53
2D	4 min	3	30	0.55
2D	4 min	15	30	0.42
2D	4 min	20	30	0.36
2D	4 min	30	30	0.31
3D	2 min	1	32	1.1
3D	2 min	3	32	0.76
3D	2 min	7	32	0.71
3D	2 min	10	32	0.69
3D	3 min	1	32	1.02
3D	3 min	3	32	0.77
3D	3 min	10	32	0.71
3D	3 min	15	32	0.68
3D	4 min	3	32	0.74
3D	4 min	10	32	0.68
3D	4 min	30	32	0.64

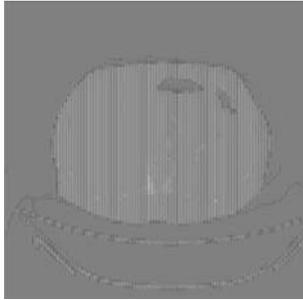
Table 4 illustrates the results for a sample data provided from the collaborator for different scanner variables. It can be seen that the SBR percentages varies based on the scanner variables used. To explain the effects of those variables on the output image, Figure 18 illustrates the changes in the quality of the segmented image based on the scanners variables.

It can be seen from Figure 18 that 3D scans produce closer SBR percentages than 2D for all iterations except at IT = 1. It can be noticed that the area evaluating the spheres decreases as the IT value increases for both 2D and 3D and for all bed section scanning times. These results match the expectation of the radiologists at Paul Strickland Scanner



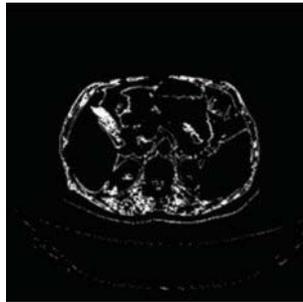
Wavelet transform

(a)



Ridgelet transform

(b)



Curvelet transform

(c)

FIGURE 22: MRA for real clinical data segmentation.

Centre. The segmented results achieved at  $IT = 10$  are illustrated in Figure 19.

A predefined clinical dataset comprised of 217 slices, with slice thickness of 3.0 mm has been tested on the proposed system. Based on the provided report, the patient is affected by multiple bilateral renal cortical cysts; the largest one is seen in the lower pole of the right kidney, measuring about  $47 \times 45$  mm. A snapshot taken for a DICOM viewer window and the ROI has been located by the red lines in three different orientations of the patient's body scan (Figure 20). It can be seen that ROI appears more clearly in its biggest illustration in slice 198; this slice is illustrated in Figure 21, and the ROI (kidney cancer) has been highlighted by red colour.

MRA have been applied on the medical image to segment it and detect ROI. Figure 22 illustrates the outputs of applying those techniques.

TABLE 5: Segmentation techniques' performance based on patient data (kidney cancer data).

Segmentation technique	Cancer area accuracy (%)	MSE	PSNR (dB)	Data loss
DWT				
Haar	91.0	102.7	35.2	Normal
Daubechies	89.5	104.5	34.3	Normal
Wavelet Packet	83.2	111.2	30.9	Normal
Ridgelet	—	109.9	30.3	High
Curvelet	96.2	88.2	29.5	Normal

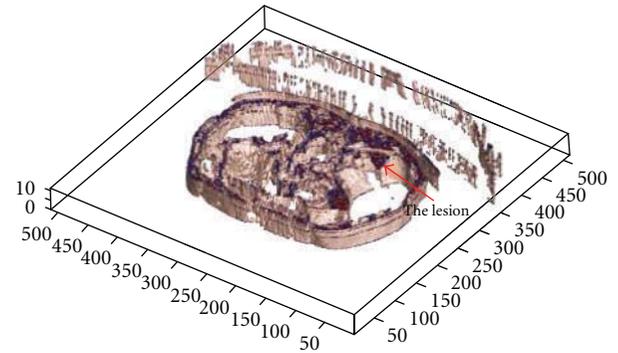


FIGURE 23: Segmenting patient volume data affected by the kidney cancer.

The performance of the proposed techniques for segmenting the illustrated slice in Figure 21 is explained in Table 5. The area of the cancer has been measured and compared to the provided report and then used to qualify the performance of each technique as well as MSE, data loss, and PSNR.

The clinical datasets have been segmented also using 3D segmentation techniques, and the lesions were detected accurately. Curvelet transform has been used before 3D segmentation to achieve a denoised CT output and ensure smoother edges. Patient data which includes lesions in liver, kidney and lung has been segmented and visualized in Figures 23, 24, and 25, where the OOIs are located.

## 5. Conclusion

Due to the changing shapes of organs in medical images, segmentation process using multiresolution analysis combined with thresholding as pre- and postprocessing step allows accurate detection of ROIs. Multiresolution analysis such as wavelet transform is extensively used in medical image segmentation and provides better accuracy in results. Curvelet and ridgelet transforms are new extension of the wavelet transform that aims to deal with interesting phenomena occurring along higher dimensional singularities. Though wavelets are well suited to point singularities, they have limitations with orientation selectivity hence do not represent changing geometric features along edges effectively.

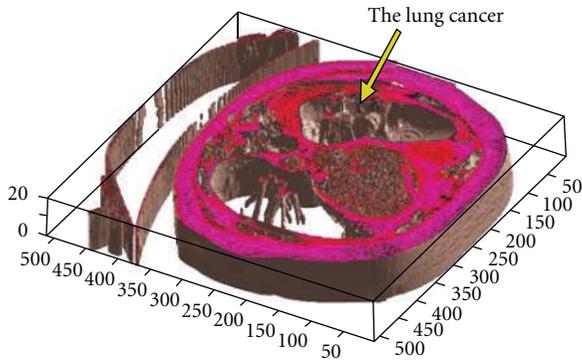


FIGURE 24: Segmenting patient volume data affected by the lung cancer.

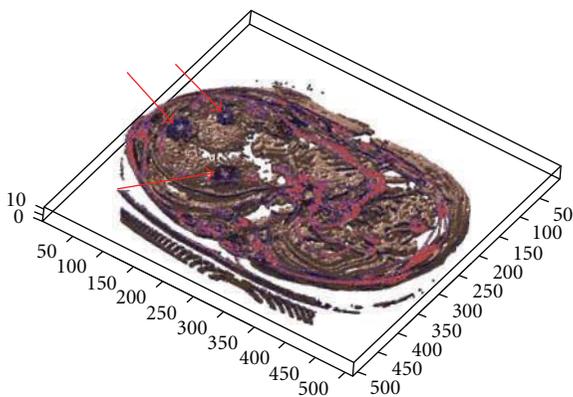


FIGURE 25: Segmenting patient volume data affected by the liver cancer (located by red arrows).

Curvelet transform exhibits good reconstruction of the edge data by incorporating a directional component to the traditional wavelet transform. Experimental study in this report has shown that curvelet-based segmentation of the medical images not only provide good-quality reconstruction of detected ROI, promising results are also achieved in terms of accurately detecting ROI and denoising process. Curvelet transform is a new tool and utilization of this technique; it is far from sufficient in the medical image processing area. The future work related to this is the implementation of 3D MRA transform which can be applied directly on medical volumes to detect obstacle and objects of interest.

## References

- [1] J. L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component separation," *Advances in Imaging and Electron Physics*, vol. 132, pp. 287–348, 2004.
- [2] A. Mojsilovic, M. Popovic, S. Markovic, and M. Krstic, "Characterization of visually similar diffuse diseases from B-scan liver images using nonseparable wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 541–549, 1998.
- [3] S. Alzu'bi and A. Amira, "3D medical volume segmentation using hybrid multiresolution statistical approaches," *Advances in Artificial Intelligence*, vol. 2010, Article ID 520427, 15 pages, 2010.
- [4] C. Mulcahy, "Image compression using the Haar wavelet transform," *Spelman Science and Mathematics Journal*, vol. 1, pp. 22–31, 1997.
- [5] W. Fourati, F. Kammoun, and M. S. Bouhleh, "Medical image denoising using wavelet thresholding," *Journal of Testing and Evaluation*, vol. 33, no. 5, pp. 364–369, 2005.
- [6] B. Kara and N. Watsuji, "Using wavelets for texture classification," in *IJCI Proceedings of International Conference on Signal Processing*, pp. 920–924, September 2003.
- [7] M. N. Do and M. Vetterli, "The finite ridgelet transform for image representation," *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 16–28, 2003.
- [8] M. Do and M. Vetterli, "Image denoising using orthonormal finite ridgelet transform," in *Wavelet Applications in Signal and Image Processing*, vol. 4119 of *Proceedings of SPIE*, pp. 831–842, 2003.
- [9] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transform," *SIAM: Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899, 2006.
- [10] I. Daubechies, "Wavelet transforms and orthonormal wavelet bases, different perspectives on wavelets," in *Proceedings of the Symposia in Applied Mathematics*, vol. 47, pp. 1–33, American Mathematical Society, San Antonio, Tex, USA, 1993.
- [11] S. G. Mallat, "Theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [12] K. Rajpoot and N. Rajpoot, "Hyperspectral colon tissue cell classification," in *Medical Imaging, Proceedings of SPIE*, 2004.
- [13] I. S. Uzun and A. Amira, "Design and fpga implementation of finite ridgelet transform," in *Proceedings of the IEEE International Symposium on Circuits and Systems, (ISCAS '05)*, vol. 6, pp. 5826–5829, May 2005.
- [14] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, "Wavelets for computer graphics: a primer, part 1," *IEEE Computer Graphics and Applications*, vol. 15, no. 3, pp. 76–84, 2002.
- [15] A. Haar, "Theorie der Orthogonalen Funkt Ionensysteme," *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, 1910.
- [16] D. L. Donoho, "Ridge functions and orthonormal ridgelets," *Journal of Approximation Theory*, vol. 111, no. 2, pp. 143–179, 2001.
- [17] M. N. Do and M. Vetterli, "Orthonormal finite ridgelet transform for image compression," in *Proceedings of the International Conference on Image Processing, (ICIP '00)*, pp. 367–370, September 2000.
- [18] E. Cand'es and D. Donoho, *A Surprisingly Effective Non adaptive Representation for Objects With Edges, Curves and Surfaces*, Vanderbilt University Press, Nashville, Tenn, USA, 2000.
- [19] E. Candes, *Ridgelets: theory and application*, Ph.D. thesis, Department of Statistics, Stanford University, Stanford, Calif, USA.
- [20] E. J. Candes and D. L. Donoho, "Ridgelets: a key to higher-dimensional intermittency?" *Philosophical Transactions of the Royal Society A*, vol. 357, no. 1760, pp. 2495–2509, 1999.
- [21] J. L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.
- [22] J. He, "A characterization of inverse Radon transform on the Laguerre hypergroup," *Journal of Mathematical Analysis and Applications*, vol. 318, no. 1, pp. 387–395, 2006.

- [23] L. Dettori and L. Semler, "A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 486–498, 2007.
- [24] Computed Tomography Scanner, King Abdullah University Hospital, Ramtha, Jordan, 2009.
- [25] International Electrotechnical Commission (IEC), Tech. Rep. 61675-1, Geneva, Switzerland, 1998.
- [26] National Electrical Manufacturers Association (NEMA), Standards Publication NU2, Washington, DC, USA, 2001.
- [27] L. Boubchir and J. Fadili, "Multivariate statistical modelling of images with the curvelet transform," in *Proceedings of the 8th International Conference on Signal Processing, Pattern Recognition, and Applications*, pp. 747–750, 2005.
- [28] I. Sumana, *Image retrieval using discrete curvelet transform*, M.S. thesis, Monash University, Australia, 2008.
- [29] L. Demanet, "The curvelet Organization," <http://www.curvelet.org/software.html>.

## Research Article

# Biomedical Imaging Modality Classification Using Combined Visual Features and Textual Terms

Xian-Hua Han<sup>1</sup> and Yen-Wei Chen<sup>1,2</sup>

<sup>1</sup> College of Information Science and Engineering, Ritsumeikan University, Kusatsu-Shi, 525-8577, Japan

<sup>2</sup> College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA

Correspondence should be addressed to Xian-Hua Han, hanxhua@fc.ritsumei.ac.jp

Received 22 February 2011; Revised 12 May 2011; Accepted 6 July 2011

Academic Editor: Fei Wang

Copyright © 2011 X.-H. Han and Y.-W. Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We describe an approach for the automatic modality classification in medical image retrieval task of the 2010 CLEF cross-language image retrieval campaign (ImageCLEF). This paper is focused on the process of feature extraction from medical images and fuses the different extracted visual features and textual feature for modality classification. To extract visual features from the images, we used histogram descriptor of edge, gray, or color intensity and block-based variation as global features and SIFT histogram as local feature. For textual feature of image representation, the binary histogram of some predefined vocabulary words from image captions is used. Then, we combine the different features using normalized kernel functions for SVM classification. Furthermore, for some easy misclassified modality pairs such as CT and MR or PET and NM modalities, a local classifier is used for distinguishing samples in the pair modality to improve performance. The proposed strategy is evaluated with the provided modality dataset by ImageCLEF 2010.

## 1. Introduction

Imaging modality is an important aspect of the image for medical retrieval [1–6]. “In user studies, clinicians have indicated that modality is one of the most important filters that they would like to be able to limit their search by. Many image retrieval websites (Goldminer, Yottalook) allow users to limit the search results to a particular modality. However, this modality is typically extracted from the caption and is often not correct or present” [7]. Some works have shown that image modality can be extracted from the image itself using visual features [8–10]. Therefore, in this paper, we propose to use both visual and textual features for medical image representation, and combine the different features using normalized kernel function in SVM.

In computer vision, studies have shown that the simple global features such as histogram of edge, gray or color intensity, can represent images, and give the acceptable performance in image retrieval or recognition research fields. Based on the success of the above-mentioned visual features for general image recognition, we also use them as medical

image representation for modality classification. Recently, using local visual feature for image representation has become very popular, and been proved to be very effective for image categorization or retrieval [11]. The most famous approach for image representation using local visual feature is bag of keypoints [12, 13]. The basic idea of bag of keypoints is that a set of local image patches is sampled using some method (e.g., densely, randomly, or using a keypoint detector) and a vector of visual descriptors is evaluated on each patch independently (e.g., SIFT descriptor, normalized pixel values). The resulting distribution of descriptors in descriptor space is then quantified in some way (e.g., by using vector quantization against a prespecified codebook to convert it to a histogram of votes for (i.e., patches assigned to codebook centres) and the resulting global descriptor vector is used as a characterization of the image (e.g., as feature vector on which to learn an image classification rule based on an SVM classifier). Furthermore, according to the visual properties of medical images, we also calculate a histogram of small-block variance as visual feature for image representation. For textual feature, we predefine 90

vocabulary words somewhat according to the statistical properties of training samples' captions and our knowledge about medical modality, and calculate a binary histogram for any medical image using their captions. After obtaining the different features for image representation, we combine them together using kernel function for SVM classifier. Because different features maybe have deferent scale and dimension, in order to allow each individual feature to contribute equally for modality classification, we normalize the distance between two samples using mean distance of all training samples, and then, obtain the kernel function for each individual feature. The final kernel for SVM classification is the mean of individual kernel, which can be called Joint Kernel Equal Contribution (JKEC). Furthermore, for some easy misclassified modalities such as CT and MR or PET and NM, a global classifier, which deals with all modalities in the used database, may not be effective in distinguishing the local modalities from each other. Therefore, after the global classification, a local classifier is used in the easy misclassified modality pairs to refine the classification results. Finally, the proposed algorithm is evaluated on the modality dataset of ImageCLEF 2010, and almost achieve the expected accuracy rate expected by the modality classification task of ImageCLEF 2010, which is about 97% classification rate.

## 2. Feature Extraction for Image Representation

In this section we describe how we extract a feature representation, which is somewhat robust to the high variability inherent in medical images and includes enough discriminative information for modality category. Some previous studies showed that it is difficult to correctly classify image categorization with only one type of image feature [14, 15]. So in this paper, we represent images with different images features including gray and color intensity histogram, block-based edge and variance histogram, popular bag-of-words model as visual feature, and a binary histogram of the predefined vocabulary words from image captions as textual feature. Then we merge them together for modality classification. Next, we simply introduce the used features for medical image representation.

### 2.1. Visual Features

*2.1.1. Gray and Color Intensity Histogram.* Intensity histograms are widely used to capture the distribution information in an image. They are easy to compute and tend to be robust against small changes of camera viewpoints. For Gray intensity histogram, we can calculate the number of each intensity (0–255) for all image pixel, and normalize it using pixel number. Given an image  $\mathbf{I}$  in some color space (e.g., red, green, and blue), to calculate color histogram the color channels are quantized into a coarser space with  $k$  bins for red,  $m$  bins for green, and  $l$  bins for blue. Therefore the color histogram is a vector  $\mathbf{h} = (h_1, h_2, \dots, h_n)^T$ , where  $n = kml$ , and each element  $h_i$  represents the number of pixels of the discretized color in the image. We assume that all images

have been scaled to the same size. Otherwise, we normalize histogram elements as

$$h'_j = \frac{y_j}{\sum_{j=0}^n y_j}. \quad (1)$$

*2.1.2. Block-Based Edge Histogram.* We firstly segment the image into several blocks, and calculate edge histogram weighted by gradient intensity in each block [16]. In experiment, we grid-segment an image into 4-by-4 block, and calculate a 20-bin edge histogram in each block. So we have 320-(20 \* 16-)dimensional edge histogram feature for medical image representation.

*2.1.3. Block-Based Variance Histogram.* For each pixel in an image, a small patch centered by the specific pixel are used for calculating the local variation of the pixel. After obtaining the local variation of all pixels in the image, a histogram of variation intensity is calculated for the image representation.

*2.1.4. Bag-of-Words Feature.* In computer vision, local descriptors (i.e., features computed over limited spatial support) have proved well-adapted to matching and recognition tasks, as they are robust to partial visibility and clutter. In this paper, we use grid-sampling patches, and then compute appearance-based descriptors on the patches. In contrast to the interest points from the detector, these points can also fall onto very homogeneous areas of the image. After the patches are extracted, the SIFT [11] descriptor is applied to represent the local features. The SIFT descriptor computes a gradient orientation histogram within the support region. For each of 8 orientation planes, the gradient image is sampled over a 4-by-4 grid of locations, thus resulting in a 128-dimensional feature vector for each region. A Gaussian window function is used to assign a weight to the magnitude of each sample point. This makes the descriptor less sensitive to small changes in the position of the support region and puts more emphasis on the gradients that are near the center of the region. To obtain robustness to illumination changes, the descriptors are made invariant to illumination transformations of the form  $aI(x) + b$  by scaling the norm of each descriptor to unity [11]. These SIFT features are then clustered with a  $k$ -means algorithm using the Euclidean distance. Then we discard all information for each patch except its corresponding closest cluster center identifier. For the test data, this identifier is determined by evaluating the Euclidean distance to all cluster centers for each patch. Thus, the clustering assigns a cluster  $c(x)$  ( $c = 1, \dots, C$ ) to each image patch  $x$  and allows us to create histograms of cluster frequencies by counting how many of the extracted patches belong to each of the clusters. The histogram representation  $h_c(X)$  with  $C$  bins is then determined by counting and normalization such that

$$h_c(X) = \frac{1}{L_X} \sum_{l=1}^{L_X} \delta(c, c(x_l)), \quad (2)$$

where  $\delta$  denotes the Kronecker delta function. Figure 1 shows the procedure of bag-of-words (BoW) feature extrac-

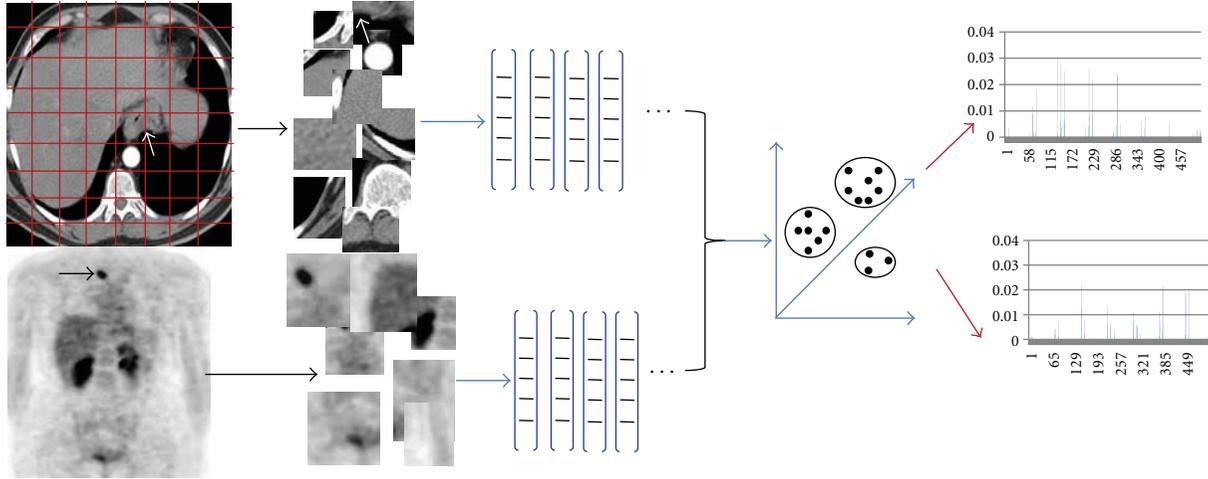


FIGURE 1: BOW feature extraction procedure.

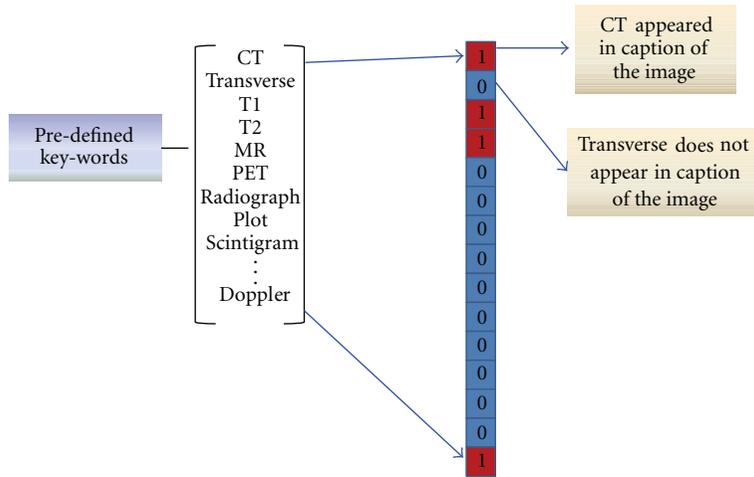


FIGURE 2: Textual feature extraction procedure.

tion and the extracted histogram feature of an example image. Obviously, there exist alternatives to algorithmic choices made in the proposed method. For example, different interest point detectors can be used. However, it does not manifest obvious merit for different background cluster of images. Furthermore, the geometrical relation between the extracted patches is completely neglected in the approach presented here. While this relation could be used to improve classification accuracy, it remains difficult to achieve an effective reduction of the error rate in various situations by doing so.

**2.2. Textual Features.** According to the statistical properties of word occurrence in each training modality image’s captions and our prior knowledge about the classifying modalities, we select 90 key-words, such as CT, curve, MR, urethrogram, and PET, as the vocabulary for forming a binary histogram for each medical image. The binary histogram for image representation is 90-dimension vector, where each dimension is correspond to one selected keyword.

If one keyword appeared one or more than one time in an image’s caption, the value of the corresponding dimension in its represented binary histogram will be 1, otherwise it will be 0. The textual feature extraction procedure is illustrated in Figure 2.

### 3. Feature Fusion

Given a training set  $(x_i, y_i)_{i=1,2,\dots,N}$  of  $N$  instances consisting of an image  $x_i \in \chi$  and a class label  $y_i \in 1, 2, \dots, C$ , and given a set of  $F$  image features  $f_m : \chi \rightarrow \mathcal{R}^{d_m}$ ,  $m = 1, 2, \dots, F$ , where  $d_m$  denotes the dimensionality of the  $m$ th feature, the problem of learning a classification function  $y : \chi \rightarrow 1, 2, \dots, C$  from the features and training set is called feature combination problem. In computer vision, the problem of learning a multiclass classifier from training data is often addressed by means of kernel methods. Kernel methods make use of kernel functions defining a measure of similarity between pairs of instances. In the context of feature combination it is useful to associate a kernel to each

image feature as the following equation (3), and combine the kernels of different features together. For a kernel function  $K$  of each feature between real vectors we define the short-hand notation:

$$K_m(\mathbf{I}_i, \mathbf{I}_j) = K(f_m(\mathbf{I}_i), f_m(\mathbf{I}_j)) = K(S(f_m(\mathbf{I}_i), f_m(\mathbf{I}_j))), \quad (3)$$

where  $\mathbf{I}_i$  and  $\mathbf{I}_j$  are two samples,  $f_m(\mathbf{I}_i)$  is the  $m$ th extracted feature from the sample  $\mathbf{I}_i$ , and  $S(f_m(\mathbf{I}_i), f_m(\mathbf{I}_j))$  is the similarity measure between the  $m$ th features of the samples  $\mathbf{I}_i$  and  $\mathbf{I}_j$ . Then the image kernel  $K_m: \chi \times \chi \in \mathfrak{R}$  only considers similarity with respect to image feature  $f_m$ . If the image feature is specific to a certain aspect, say, it only considers color information, then the kernel measures similarity only with regard to this aspect. The subscript  $m$  of the kernel can then be understood as indexing into the set of features. Because different features maybe have different scale and dimension, in order to allow each individual feature to contribute equally for modality classification, we normalize the distance between two samples using mean distance of all training samples, and then, obtain the kernel function for each individual feature  $f_m$ . The final kernel for SVM classification is the mean of individual kernel, which can be called Joint Kernel Equal Contribution (JKEC). For the feature similarity calculation of two samples, we use several distances: Euclidean distance ( $L_2$  distance),  $L_1$  distance, and  $\chi^2$  distance, for evaluating the classification performance. The  $\chi^2$  distance for two samples can be calculated as follows:

$$S_m^{i,j} = S(f_m(\mathbf{I}_i), f_m(\mathbf{I}_j)) = \sum_1^L \frac{(x_l - y_l)^2}{x_l + y_l}, \quad (4)$$

where  $x$  and  $y$  represent the  $m$ th features  $f_m(\mathbf{I}_i)$ ,  $f_m(\mathbf{I}_j)$  of samples  $i$  and  $j$ , respectively, and  $x_l$  is the  $l$ th element of the vector  $x$ .  $S_m^{i,j}$  is the similarity measure of the  $m$ th feature between the  $i$ th and  $j$ th training samples. Then, the RBF function is used for calculating the kernel:

$$K_m^{i,j} = K_m(\mathbf{I}_i, \mathbf{I}_j) = \exp\left(\frac{-S(f_m(\mathbf{I}_i), f_m(\mathbf{I}_j))}{\gamma_m}\right), \quad (5)$$

where  $\gamma_m$  is the normalized item for kernel function of the  $m$ th feature. Here, we use the distance mean of all training samples as  $\gamma_m = 1/N^2(\sum_i^N \sum_j^N S_m^{i,j})$  ( $N$  is the training sample number), which will lead to similar contribution of each feature to kernel. Then the final combined kernel function can be obtained by

$$K^{i,j} = \frac{1}{M} \sum_i^M K_m^{i,j}, \quad (6)$$

where  $M$  is the feature number for image representation. The proposed algorithm is evaluated on the modality training dataset of ImageCLEF 2010, which expects about 97% classification rate on the released evaluated and test datasets. Because the ground-truths of the evaluated and test dataset are not released, we cross-validate our proposed strategy with the released training dataset firstly. The classification rate with our experiment on training dataset almost approximated the required goal of the modality classification task.

## 4. Refinement Procedure for Easy Misclassified Modalities

In the released medical database by ImageCLEF 2011, some modalities have a lot of visual similarity such as PET and NM modality. Therefore, it is difficult to distinguish them in the global modality classification, which deals with all modalities in the database. In this section, after the global conventional classification, we design local classifiers to refine the classification results in easy-misclassified modalities. Next, we firstly explain the used dataset, and then, introduce how to design the local classifier according to evaluation results.

**4.1. Image Data.** The database released for the ImageCLEF-2010 Medical modality classification in medical retrieval task includes 2390 annotated modality images (CT: 314; GX: 355; MR: 299; NM: 204; PET: 285; PX: 330; US: 307; XR: 296) for training and a separate evaluated set consisting of 2620 images. The aim is to automatically classify the evaluated set using 8 different modality label sets including CT, MR, and PET. Some example images are shown in Figure 3. A more detailed explanation of the database and the tasks can be found in [17].

**4.2. Local Classifier Designing.** For validating the discriminant properties of different modalities, we randomly select 180 samples from each medical modality in ImageCLEF 2010 training dataset, and the remainder for testing. We combine all visual and textual features using the JKEC fusion strategy introduced in Section 3, for modality classification. The confusion matrix of one run is shown in Table 1. From Table 1, it can be seen that 92.537% CT sample images are correctly recognized as CT modality, and 3.9851% and 2.2388% are classified as MR and XR modalities, respectively. On the other hand, about 2-3% MR or XR sample images are also misrecognized as CT modality. At the same time, it is obvious that NM and PET or GX and PX modalities are also easily misclassified from each other. Therefore, we design three local classifiers for the limited easy-misclassified modalities, which are CT, MR, and XR group, NM and PET group, GX and PX group, to refine the classification results in local regions. The refinement procedure with the local classifiers are shown in Figure 4. The compared experimental results with or without refinement procedure are shown in Figure 5. From Figure 5, it can be seen that the recognition rates for CT, MR, PET, PX, and XR modalities with local classifier refinement can be improved more than 1% compared to those without refinement.

## 5. Experiments

In this section, we validate the recognition rates of different features with three types of similarity measures: Euclidean distance ( $L_2$  distance),  $L_1$  distance and  $\chi^2$  distance, and do the cross-validation experiments using the combined visual and textual features on ImageCLEF 2010 training dataset. Then, the submitted runs to medical modality classification of ImageCLEF 2010 and the released results will be introduced.

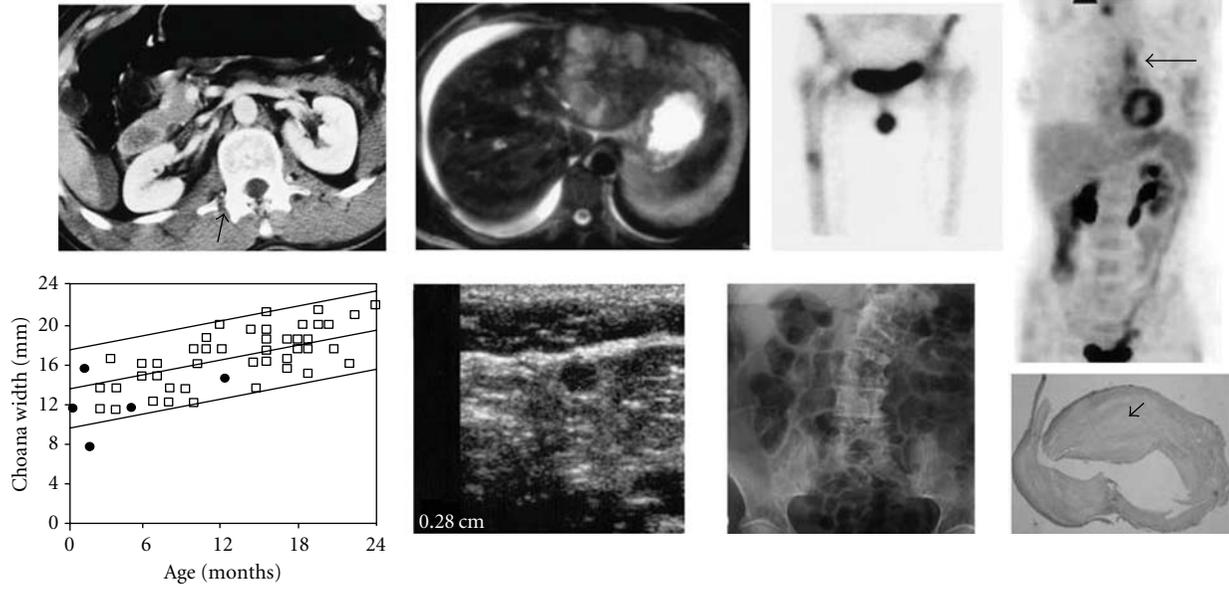


FIGURE 3: Sample images of 8 medical modalities. From left to right and top to bottom, the images are CT, MR, NM, PET, GX, US, XR, and PX, respectively.

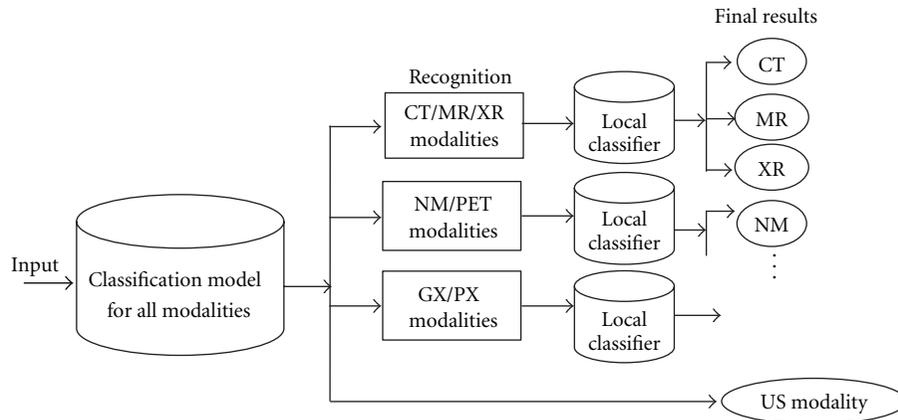


FIGURE 4: The refinement procedure with the local classifiers.

TABLE 1: Confusion matrix of one run on medical evaluated dataset using combined visual and texture features.

Modality	CT	GX	MR	NM	PET	PX	US	XR
CT (%)	92.537	0	3.9851	0	0.4925	0.7463	0	2.2388
GX (%)	0	97.714	0	0	0	0.2857	0	0
MR (%)	3.3613	0	94.118	0	0	0.8403	0	1.6807
NM (%)	0	0	0	98.23	1.77	0	0	0
PET (%)	0	0	0.4224	2.53	97.048	0	0	0
PX (%)	0	1.333	0	0.6667	0	96.667	0	1.3333
PX (%)	0	0	0.7874	0	0.7874	98.425	0	0
XR (%)	1.7241	0	2.5862	0.8621	0	0.8621	0	93.966

(1) The recognition rates of different features with three types of similarity measures: in order to validate what kind of distance measure is adaptive to each extracted feature for image representation, we apply three types of similarity measures: Euclidean distance ( $L_2$

distance),  $L_1$  distance, and  $\chi^2$  distance, for calculating the kernel function as in (5) of SVM classifier. In the experiments, with the ImageCLEF 2010 training dataset, 180 images are randomly selected for training from each modality, the remainder are for test. The

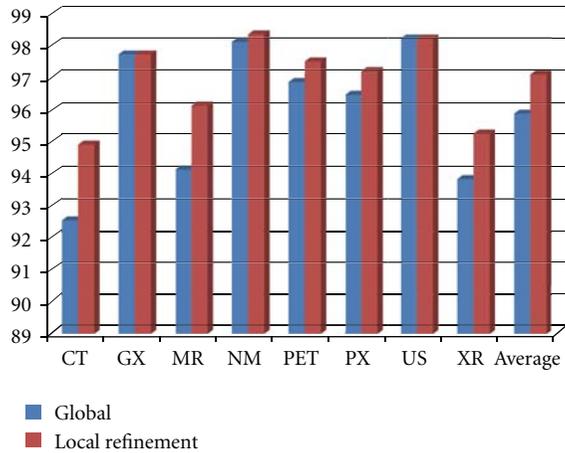


FIGURE 5: The compared recognition rates for all modalities. Blue bar: without refinement using local classifiers; Red bar: with refinement.

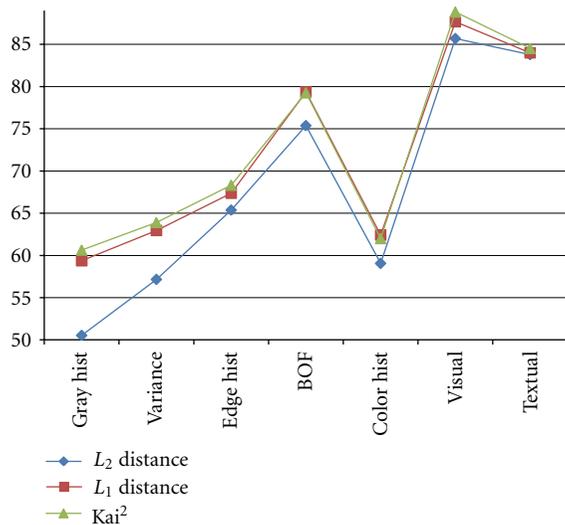


FIGURE 6: The compared recognition rates with different features using three types of similarity measures: Euclidean distance ( $L_2$  distance),  $L_1$  distance, and  $\chi^2$  distance. Gray hist, variance, edge hist and color hist represent the histogram of gray, variance, edge, and color, respectively; BOF and textual mean the BOF and textual features.

compared recognition rates are shown in Figure 6, where “Kai2” means  $\chi^2$  distance. From Figure 6, it can be seen that  $L_1$  and  $\chi^2$  distance can obtain much better performance than  $L_2$  distance for most features, and  $\chi^2$  distance can achieve a little better than or similar results to  $L_1$  distance. Then, in the next experiments, we utilize  $\chi^2$  distance for a similarity measure of all features to calculate SVM kernel functions.

- (2) Cross-validation experiments: in the experiments, we firstly divide the training dataset of ImageCLEF 2010 into 5 groups, and use 4 groups as training and 1 group as test to cross-validate the performance

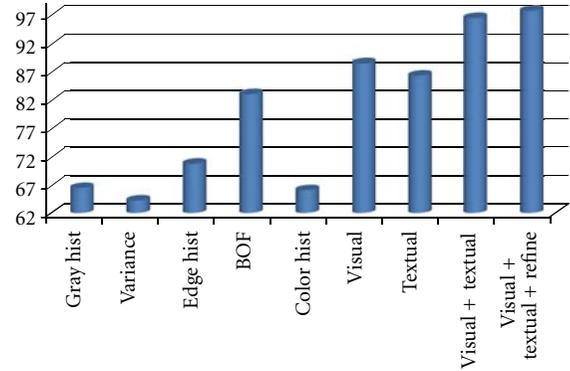


FIGURE 7: The recognition rates of cross-validation strategy using different features, where visual means the recognition rate using the combined feature of all visual ones, visual + textual means those using the combined textual and visual features, and visual + textual + refine means using refinement procedure after classification with the combined textual and visual features.

of different features with  $\chi^2$  distance as similarity measure. The recognition rates are shown, Figure 7, where visual means with the combined features of all visual ones, Visual + Textual means with the combined features of all visual and textual ones, visual + textual + refine means using refinement procedure after classification with all combined features. From Figure 7, it can be seen that the average recognition rate can be improved about 1% after the refinement procedure introduced in Section 4.

- (3) Submitted runs: as Medical Image Processing Group (MIPG) of our Intelligent Image Processing Laboratory (IIPL) in Ritsumeikan University, we prepared four runs for evaluation image set, which used combine visual feature, textual feature, both visual and textual features, and weighted visual and textual features. The recognition results are shown in Table 2. We submitted two runs using textual, combined textual and visual features by on-line-system, respectively. Our results are ranked second among 6 participating teams, and the result of one run is also ranked second among about 50 runs [18]. At the same time, the recognition rates (submitted run: 93.36%, unsubmitted run: 93.89%) of our methods using mixed feature (Visual plus textual) are similar to the first ranking results 94% by Xerox Research Centre Europe.

## 6. Conclusions

In this paper, we proposed to extract different visual and textual features for medical image representation, and use JKEC strategy to fusion them for modality classification. To extract visual features from the images, we used histogram descriptor of edge, gray, or color intensity and block-based variation as global features and SIFT histogram as local feature, and the binary histogram of some predefined vocabulary words for image captions is used for textual

TABLE 2: Overall classification rates on medical evaluated dataset using combination of different features.

Features	Visual	Textual	Visual + texture	Weighted visual + textual
Classification rate (%)	87.07	84.58	93.36	93.89

feature. Because different features maybe have different scale and dimension, in order to allow each individual feature to contribute equally for modality classification, we proposed to use joint kernel equal contribution (JKEC) for kernel fusion of different features. Furthermore, for some easy misclassified modality pairs such as CT and MR or PET and NM modalities, a local classifier is used for distinguishing samples in the pair modality to improve performance. The proposed algorithm is evaluated by the provided modality dataset by ImageCLEF 2010.

## Acknowledgments

This paper was supported in part by the Grand-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture, and Sports under the Grand no. 21300070 and 22103513, and in part by the Research fund from Ritsumeikan Global Innovation Research Organization (R-GIRO).

## References

- [1] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medicine clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, 2004.
- [2] H. Muller, T. Deselaers, T. Lehmann, P. Clough, and W. Hersh, "Overview of the ImageCLEFmed 2006 medical retrieval annotation tasks, evaluation of multilingual and multimodal information retrieval," in *Proceedings of the Seventh Workshop of the Cross-Language Evaluation Forum, (CLEF 2006)*, C. Peters, P. Clough, F. Gey et al., Eds., Lecture Notes in Computer Science, Alicante, Spain, 2006.
- [3] W. Hersh, J. Kalpathy-Cramer, and J. Jensen, "Medical image retrieval and automated annotation: OHSU at ImageCLEF 2006," in *Proceedings of the Working Notes for the CLEF 2006 Workshop*, Alicante, Spain, 2006.
- [4] M. O. Güld, M. Kohnen, D. Keysers et al., "Quality of DICOM header information for image categorization," in *Proceedings of the International Society of Optics and Photonics (SPIE)*, vol. 4685, pp. 280–287, February 2002.
- [5] W. Hersh, H. Muller, J. Jensen, J. Yang, P. Gorman, and P. Ruch, "Advancing biomedical image retrieval: development and analysis of a test collection," *Journal of the American Medical Informatics Association*, vol. 13, no. 5, pp. 488–496, 2006.
- [6] J. Kalpathy-Cramer and W. Hersh, "Automatic image modality based classification and annotation to improve medical image retrieval," *Student Health Technology Information*, vol. 129, part 2, pp. 1334–1338, 2007.
- [7] "Medical retrieval task," <http://www.imageclef.org/node/104/>.
- [8] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image-databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1996.
- [9] A. Lakdashti and M. S. Moin, "A new content-based image retrieval approach based on pattern orientation histogram," *Computer Vision/Computer Graphics Collaboration Techniques*, vol. 4418, pp. 587–595, 2007.
- [10] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233–1244, 1996.
- [11] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [12] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proceedings of the ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1–16.
- [13] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR 2006)*, pp. 2169–2178, June 2006.
- [14] X. H. Han, Y. W. Chen, and X. Ruan, "Image recognition by learned linear subspace of combined bag-of-features and low-level features," in *Proceedings of the 20th International Conference on Pattern Recognition, (ICPR 2010)*, pp. 1049–1052, August 2010.
- [15] X. H. Han, Y. W. Chen, and X. Ruan, "Image categorization by learned nonlinear subspace of combined visual-words and low-level features," in *Proceedings of the 20th International Conference on Pattern Recognition, (ICPR 2010)*, pp. 3037–3040, August 2010.
- [16] C. S. Won, D. K. Park, and S. J. Park, "Efficient use of MPEG-7 edge histogram descriptor," *Electronics and Telecommunications Research Institute Journal*, vol. 24, no. 1, pp. 23–30, 2002.
- [17] H. Muller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, C. E. Kahn Jr., and W. Hersh, "Overview of the CLEF 2010 medical image retrieval track," in *Proceedings of the Working Notes of Cross-Language Evaluation Forum, (CLEF 2010)*, Padova, Italy, 2010.
- [18] <http://imageclef.org/2010/medical>.

## Research Article

# Automatic Segmentation of Dermoscopic Images by Iterative Classification

Maciel Zortea,<sup>1</sup> Stein Olav Skrøvseth,<sup>2</sup> Thomas R. Schopf,<sup>2</sup>  
Herbert M. Kirchesch,<sup>3</sup> and Fred Godtliebsen<sup>1</sup>

<sup>1</sup>Department of Mathematics and Statistics, University of Tromsø, 9037 Tromsø, Norway

<sup>2</sup>Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, 9038 Tromsø, Norway

<sup>3</sup>Venloer Straße 107, 50259 Pulheim, Germany

Correspondence should be addressed to Maciel Zortea, maciel.zortea@uit.no

Received 23 March 2011; Accepted 4 May 2011

Academic Editor: Fei Wang

Copyright © 2011 Maciel Zortea et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate detection of the borders of skin lesions is a vital first step for computer aided diagnostic systems. This paper presents a novel automatic approach to segmentation of skin lesions that is particularly suitable for analysis of dermoscopic images. Assumptions about the image acquisition, in particular, the approximate location and color, are used to derive an automatic rule to select small seed regions, likely to correspond to samples of skin and the lesion of interest. The seed regions are used as initial training samples, and the lesion segmentation problem is treated as binary classification problem. An iterative hybrid classification strategy, based on a weighted combination of estimated posteriors of a linear and quadratic classifier, is used to update both the automatically selected training samples and the segmentation, increasing reliability and final accuracy, especially for those challenging images, where the contrast between the background skin and lesion is low.

## 1. Introduction

Melanoma is a common cancer in the adult population, and accounts for a considerable number of deaths in fair skinned people worldwide. It may arise within preexistent moles or *de novo* in unaffected skin. When diagnosed at an early-stage prognosis is excellent, the melanoma can be cured by simple excision. However, as melanomas can be hard to distinguish from common moles, even for experienced dermatologists, early diagnosis is a challenge, especially for general practitioners.

A recent advance in diagnosis of melanomas is the emergence of dermoscopy, also known as epiluminescence microscopy. Dermoscopy is a noninvasive diagnostic technique that consists in the examination of skin lesions with a dermoscope, which is a hand held optical device that typically consists of a magnifying lens and a light source, used to illuminate the skin. Usually, a transparent plate and liquid medium is placed between the instrument and the skin, allowing the inspection of a lesion, in the upper layers of

the skin, in great detail. This clear view of the skin can help with diagnosing skin cancers [1].

There is a considerable variation in the appearance of common moles and melanomas. They both appear as spots on the skin, with diameters ranging from a few millimeters up to several centimeters. Color typically is brown or black, but red and blue also exists. In most situations the pigmented lesion can easily be identified from the normal surrounding skin which appears lighter.

Delineation of the contour of pigmented skin lesions (segmentation) plays a relevant role for automatic feature extraction, where the purpose is automatic diagnosis of melanomas. Image segmentation is the process of adequately grouping pixels into a few regions, whose pixels share some similar characteristic, like color, texture, or shape [2]. Automated analysis of edges, colors, and shape of the lesion relies upon an accurate segmentation and is an important first step in any computer aided diagnosis system. Indeed, most commercially available systems show a great variability in reliability and specificity in the diagnostic process, and

the image segmentation also varies greatly [3]. Irregular shape, nonuniform color, and ambiguous structures make the problem challenging [4]. The presence of hairs and skin flakes are additional undesirable features that may interfere with segmentation. An additional complication arises in the validation of any technique, as there is no gold standard to refer to. Even trained dermatologists differ significantly when delineating the same lesion in separate incidents [5], so validation of any technique has to be treated with care. Indeed, an important feature of any segmentation procedure must be reproducibility. Even under the best effort to counter this, different images of the same lesion will differ slightly in illumination, rotation, and shear due to the flexibility of the skin.

Supervised segmentation methods are those methods that require input from the analyst, such as examples of skin and lesion pixels, a rough approximation of the lesion borders to be optimized, or a final refinement of a proposed solution [6–8]. Generally, in such settings, the user needs to provide a priori input for each particular image being analyzed. This task relies on the experience and knowledge of the user. Although this kind of approach may be very effective, the process may be particularly time consuming for health care professionals. As an alternative, automatic segmentation methods (also called unsupervised methods) attempt to find the lesion borders without any input from the analyst and can, therefore, be applied even by persons who are not trained in dermatology. Several approaches have been proposed in this direction. Most common automatic segmentation algorithms rely on techniques based on histogram thresholding [7, 9–12], where most commonly RGB information is mapped to a one or two-dimensional color space through choice of one of the channels, luminance, or principal component analysis. Other approaches include clustering [4, 13, 14], region-based techniques [7, 15, 16], contour-based approaches [7, 17–20], genetic programming [21], and segmentation fusion techniques [22]. A recent overview of methods applied to segmentation of skin lesions in dermoscopy images suggest that clustering is the most popular segmentation method, probably due to the availability of robust algorithms [6].

The algorithm proposed in Celebi et al. (2008) [16] uses a region growing and merging technique called statistical region merging (SRM) [23] to segment the image. SRM was proven to be a robust segmentation algorithm for segmentation of color images. In [16], SRM achieved better results on segmentation of skin lesions when compared to five automated methods: the dermatologist-like tumor extraction algorithm [15], the JSEG segmentation algorithm [24], the mean-shift clustering [4], and the orientation-sensitive fuzzy c-means [13]. The recently proposed automatic adaptive thresholding (AT) by Silveira et al. (2009) [7] segments the image by comparing the color of each pixel with a threshold. The algorithm uses entropy to find the most suitable RGB channel for discrimination, and a pixel is classified as lesion if it is darker than the threshold. Distinct rules are proposed to compute the threshold for the cases of bimodal or single component histograms.

Currently, several instruments designed for computer aided diagnosis (CAD) of skin lesions are commercially

available. Despite the use of powerful and dedicated video cameras, the cost related to the acquisition material [25] or the actual usefulness of these instruments for dermatologists practising [3] may be the reasons that prevent their wide diffusion to physicians.

At the other hand, current limitations of state-of-art CAD instruments motivates the development of new algorithms for analysis of skin lesions and simple data acquisition options. Following an approach that might be practical and intuitive for dermatologists, the images considered in this study are taken by a consumer digital camera with a dermatoscope attached. This simple image acquisition setup was considered for instance in [26–28]. As a usual procedure for dermatologists, an alcohol-based contact fluid liquid between the skin and the dermoscope is used during the acquisition of the images. This minimizes the formation of air bubbles, reducing artifacts.

Although several algorithms for unsupervised image segmentation have been proposed in the literature, they do not necessarily perform well for the specific problem of skin lesion segmentation, and it appears unlikely that one particular method could outperform all other segmentation methods for any lesion. Our idea is to exploit (a) the fact that skin lesions are approximately located in the center of the dermoscope during the acquisition process and (b) the color characteristics of typical lesions and integrate such information into a reliable framework for automated segmentation of skin lesions. This results in an algorithm for segmentation based on classification, where the initial training data set is selected as portions of the image based on these initial assumptions, and the training set is iteratively expanded. This training and classification occurs within one image only, the training set being small portions of the image called seed regions. Thus, we arrive at our proposed algorithm, which we will denote iterative classification Segmentation (ICS).

The proposed segmentation method introduces novelties in the current state-of-art of methods designed specifically for segmentation of dermoscopic skin lesions.

- (1) An automated method is proposed to select small seed regions to represent both the background skin and lesion. In the proposed approach, seed regions are used for initialization of an iterative classification framework. This differs from previous studies, in some of which seed regions are manually placed for modeling the background skin [4], or others, where it is assumed that the corners of the lesion images are good estimates for the background skin [16]. We believe our approach provides more flexibility in the search for suitable regions despite its simplicity.
- (2) The lesion segmentation problem is treated as a binary classification problem. A hybrid strategy that combines classification posterior probability estimates from two distinct classifiers is presented. For each particular lesion, the best choice between a linear or quadratic classifier (or a weighted combination of both) is automatically set by optimization of the classification accuracy. The proposed classifier proved to be valuable in the context of iterative classification.

- (3) An iterative approach is used to update the parameters of the classifiers and the final segmentation. In contrast to semisupervised learning techniques that make use of both labeled and unlabeled data for training [29], where typically high confident samples are iteratively added to the initial set of ground truth manually labeled samples; here, the proposed method attempts to use only automatically selected samples with no supervision involved. It appears of interest to understand if such kind of approach would be valuable to solve the problem of segmentation of skin lesions.
  - (4) The proposed method makes specific hypothesis for the segmentation problem of interest, allowing the design of safeguard criteria for driving the unsupervised update of training samples. We believe that the idea might also be successfully extended to segmentation of other kind of medical images, provided that the hypothesis, notably the a priori location of and relative color of elements of interests, is properly defined.
  - (5) As with most of methods dealing with image processing, there are inevitably parameters that must be set by the user. Throughout this paper, we attempt to use parameters defined in length units as much as possible, thus using the magnitude of a physical quantity. Since in practical applications, the resolution of the equipment used for image acquisition can be measured, the parameters can be easily converted to pixel unities for image analysis. Surprisingly, we observe in the literature that most papers dealing with segmentation of skin lesions cite the relevant algorithm parameters in pixels, or a value relative to the image size, often without reference to the image resolution. This complicates unbiased implementation of alternative techniques for comparison purposes. Hopefully, our approach might help other researchers using different equipment.
- (2) Classification: two distinct base classifiers are made available to the classification procedure, specifically a linear and quadratic classifier. The decision of which classifier to use for each particular lesion, or possibly a combination of both classifiers (weighting), is automatically decided using optimization of the classification accuracy. This classification strategy can be seen as a hybrid classifier, here defined in terms of the combination of the posterior classification probabilities. The classification is iterated, facilitating robust selection of new training samples and segmentation.
  - (3) Iteration: the automatically selected training samples are updated and the classification is repeated iteratively until convergence.

A flowchart of the proposed ICS algorithm is depicted in Figure 1. A detailed description of the assumptions and the relevant processing stages is provided in Sections 2.2–2.5. The postprocessing that leads to the final segmentation mask is presented in Section 3.

**2.1. Preprocessing.** The RGB image is first processed to the perceptual uniformity CIE  $L^*u^*v$  color space. The  $L^*u^*v$  color space attempts perceptual uniformity, and it is extensively used for applications such as computer graphics and was previously used for segmentation of skin lesions [13].

**2.2. Assumptions.** Specific assumptions about the dermoscopic image acquisition and the colors of skin and lesion are used for the initial automatic selection of seed regions, as described below.

**2.2.1. Image Acquisition.** We assume that at least part of the lesion is located inside the circular spatial domain  $d_1 = \{r \leq 5 \text{ mm}\}$ , where  $r$  is the radius of a disk centered in the image ( $r = 5 \text{ mm}$  corresponds to the red circular ring in Figure 2). In practice, this assumes that some care was taken during the acquisition of the image. The assumption is not very stringent: as it will be described in Section 2.3, only a small part of the lesion needs to be located inside this spatial domain. When the lesion is relatively small, the domain  $d_1$  might enclose the entire lesion.

In a similar way, we assume that in the remaining part of the imaged area, the spatial domain  $d_2 = \{5 \text{ mm} < r \leq \ell\}$ , where  $\ell$  is the radius of the entire circular imaged area acquired using the dermatoscope (with our equipment,  $\ell \approx 8.7 \text{ mm}$ ), a few small skin regions, possibly spatially homogeneous, are available. In general, this should be true, except for big lesions covering the entire image. For skin lesions covering the full dermoscopic area, (i.e.,  $\approx 240 \text{ mm}^2$  with the current dermatoscope), segmentation between skin and mole is not applicable. From a medical point of view, the size of a lesion is itself indicative of further medical attention, but since segmentation is meaningless in such cases, we will ignore these in the remainder.

**2.2.2. Skin Color.** An a priori assumption regarding the color of both skin and lesion areas is necessary to initialize

The remaining of the paper is organized as follows; in Section 2, we introduce the ICS framework. Section 3 outlines the postprocessing steps allowing the final segmentation. Section 4 presents experimental evidence of the competitiveness of the ICS algorithm for segmentation of skin lesions using dermoscopy images. We conclude with final remarks in Section 5.

## 2. Methodology

The proposed ICS framework is characterized by three main processing stages.

- (1) Initialization: automatic search for seed regions under assumptions on the approximate location of the lesion and the usual lighter nature of the skin color compared to the lesion. The seed regions provide training samples for the binary classification between background skin and lesion.

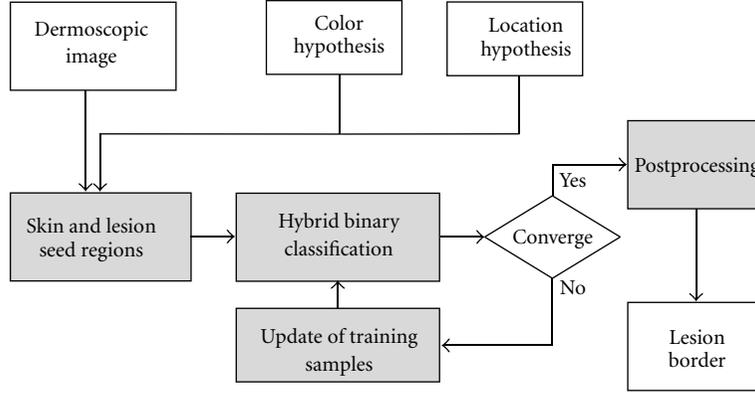


FIGURE 1: A flowchart of the proposed ICS algorithm showing the main processing modules.

the unsupervised (automatic) selection of training samples. We assume that the lighter areas in  $d_2$  are likely to be skin and those areas in  $d_1$  that differs most, in a statistical sense, from the skin correspond to lesion.

The statistical difference is here computed measuring the Euclidian distances in the Luminance component of the calibrated RGB image processed to the CIE  $L * u * v$  color space. As will be shown later in experiments (in Figure 3), the location of the seed regions using only the Luminance component proved to be robust enough to the great majority of lesions analyzed.

### 2.3. Stage1: Initial Unsupervised Selection of Training Samples

**2.3.1. Skin Samples.** The peripheral spatial domain  $d_2$  is divided into four quadrants. In each quadrant, we search for a small seed, a box of size  $w_{\text{skin}} \times w_{\text{skin}}$  pixels corresponding to the lighter colored areas by looking for the regions with maximum average value of the pixel's luminance component. According to the assumption in Section 2.2.2, these regions are likely to be skin portions. We suggest  $w_{\text{skin}} = 1.0$  mm. The skin seed regions corresponds to the green boxes depicted in green in Figure 2.

Once the four skin seeds are positioned, in order to further increase the robustness of the process, we select and merge only the three seeds that correspond to the regions that are statistically most similar by testing all three vs one combination of regions (again, measuring the Euclidian distance of the luminance component in the two sets). The pixels located in the seed regions are the initial skin training samples. Excluding one out of the four seeds reduces the risk of incorporating regions that are contaminated by artifacts.

**2.3.2. Lesion Samples.** Once representative initial skin samples are identified (Section 2.3.1), we search in the central spatial domain  $d_1$  for the box of size  $w_{\text{mole}} \times w_{\text{mole}}$  whose average pixel values in the luminance component has the largest Euclidian distance from the initial skin training samples average value.

Depending on the size of the box compared to the size of the lesion, skin pixels may be present inside the box.

Intuitively, a bigger box over the lesion would better capture the variability of the lesion, but it increases the probability of including skin pixels, so a tradeoff must be made. We suggest  $w_{\text{mole}} = 1.5$  mm. Examples of this initial location for the lesion seed are the blue boxes shown in Figure 2.

For illustrative purposes, Figure 2(b) includes the case of a challenging lesion characterized by a brighter color compared to the background skin. In this particular example, the proposed method was able to accurately position both seed regions.

**2.4. Stage 2: Classification.** Once the initial training samples are automatically selected using the procedure described in the previous section, the samples can be used for “supervised” classification. Note that the initial training samples were obtained in an automated way using only assumptions about the images. Notice, however, that since the initial training samples were obtained from an automatic strategy, a mechanism aiming at improving the segmentation accuracy and the confidence of the automatic training samples selection would be desirable.

From a classification perspective, it is worth mentioning that the initial seed regions (the automatically selected skin and mole box regions in Figure 2 using the luminance  $L$  component) are likely to be perfectly separable. This is because the seeds are found by looking for bright pixels on the periphery and the most dissimilar pixels in the center. Therefore, it is likely that the decision boundary will be poorly specified and hence unlikely that a quadratic classifier could initially outperform a simple linear classifier. However, as new training samples are iteratively added to the model, and using all the  $L * u * v$  color components, one could expect that the decision boundaries between both skin and lesion would become progressively better defined in the feature space and potentially allowing a quadratic classifier to benefit from the higher flexibility in placing the discriminating boundaries.

Because of their simplicity and fast calculation, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) [30] are initially considered as base classifiers for the binary classification problem. QDA fits multivariate normal densities with covariance estimates stratified by

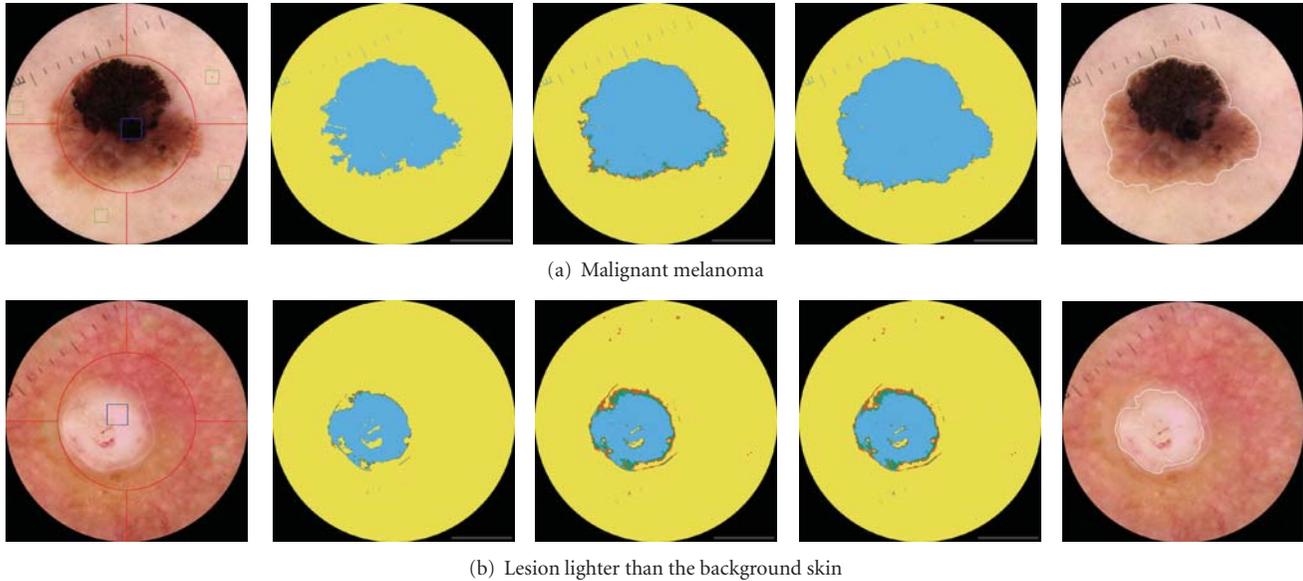


FIGURE 2: (Left) The small automatically selected green boxes correspond to the skin seed regions used as initial training. The location is based on the brightest luminance in each quadrant. The blue box corresponds to the seed region for the lesion, selected as the region statistically most different from the skin seeds. The three middle images show intermediate iterative steps at iterations  $j = \{1, 2\}$  and the final iteration which is  $j = 7$  in (a),  $j = 3$  in (b). In these maps, yellow and blue pixels are those classified with high confidence as skin and lesion, respectively. The maps include also red and green pixels, corresponding to those pixels classified with low confidence as skin and lesion, respectively. The white contour in the rightmost figure is the final border obtained after postprocessing.

lesion and skin groups, when LDA uses a pooled covariance matrix estimation. We also considered using a nonparametric classifier (specifically, the nearest-neighbor classifier), but due to the computational burden it was discarded. LDA can be seen as a particular case of the QDA classifier, raising the question if there is some added value in considering the use of a linear classifier in the proposed method. In preliminary experiments, we found scenarios where the simple linear choice appears to be preferable, for instance, in cases where the seed regions corresponding to skin are placed on very homogenous areas. This areas typically exhibit very low variance, compromising the estimation of the covariance matrix for the skin class in QDA. The low variance may result also from the placement of the seed in regions that does not provide an accurate representation of the statistics of all the background skin (poor location). If the quadratic classifier is used, despite the perceptual small differences of color between the seed region for skin and the whole background skin, it appeared that skin portions would be wrongly classified as high confident skin ( $p_k^\lambda(y = 1 | x) > 0.98$ ). This undesired behavior would compromise the iterative update of confident training samples, especially for the lesion. Especially at the beginning of the iterative classification, it appears to be more prudent to use of linear classifier. As more samples are added, a better estimate of the class statistics would be a benefit for the quadratic classifier.

The above considerations suggests that the use of an iterative procedure would be a reasonable choice for update of the initially selected training samples, and ideally, the choice between LDA and QDA should be automatic in the iterative process.

Our proposed methodology for selection between LDA and QDA for classification is very simple in nature. Assume that the classification output for a given pixel  $x$  is  $y = 1$  for pixels classified as skin and  $y = 2$  for lesion. We consider a class of hybrid posterior estimates

$$p_k(y = \Omega | x) = \lambda_k p_k^{\text{QDA}}(y = \Omega | x) + (1 - \lambda_k) p_k^{\text{LDA}}(y = \Omega | x), \quad (1)$$

where  $0 \leq \lambda_k \leq 1$ . Given the training data, our aim is to choose a value of  $\lambda_k$  that maximizes classification accuracy of classes  $\Omega = \{1, 2\}$  at each iteration  $k = \{1, 2, 3, \dots, n\}$ .

Without ground-truth samples, estimating classification accuracy is a challenging task. We rely on the training samples from the automatically placed seed regions. Such samples are divided in two spatially disjoint subsets. One subset is used to train LDA and the other QDA. In practice, one and a half of the three seed regions corresponding to skin, and half of the mole seed region are used in each classifier. This reduces the spatial correlation between the samples used in both classifiers. For an easy implementation of (1), the interval  $[0 - 1]$  for  $\lambda_k$  is finely partitioned and the value of  $\lambda_k$  that provides the best classification is chosen. For each value of  $\lambda_k$ , we compute the classification accuracy for both the skin and the lesion classes and retain the minimum value. At least, half of the samples remained unseen by each classifier, acting like independent test samples. An important point is that in case of identical classification accuracies for distinct values of  $\lambda_k$ , the lowest value of  $\lambda_k$  is selected, privileging the simpler LDA. Typically, this is the case in the first iterations.

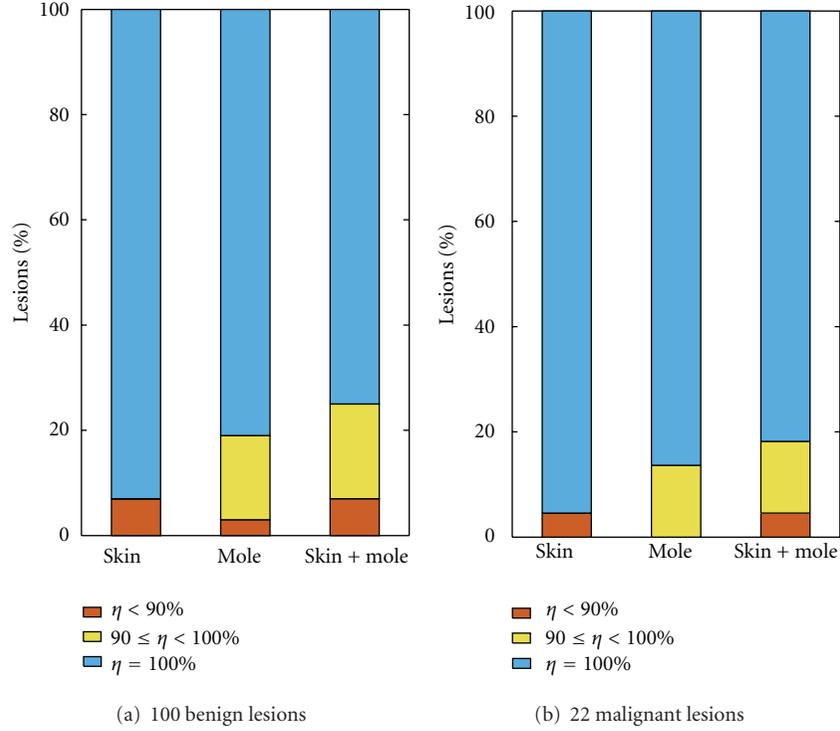


FIGURE 3: Average accuracy ( $\eta$ ) of the detected seed regions for the images in the test set, according to the diagnosis. For visualization purposes, instead of the number of lesions, the vertical axis shows the corresponding percentage out of 100 benign and 22 malignant lesions.

**2.5. Stage 3: Iterative Update of Training Samples and Classification.** After the initial binary classification, new additional samples are added iteratively to the classification process to better estimate the parameters of the classification model. In order to keep classification stable, it is desirable to focus on those skin and mole pixels classified with high confidence: those pixels classified with an estimated a posteriori probability higher than the predefined threshold  $p_k^\lambda(y = \Omega | x) > \tau$ . In our experiments, we set  $\tau = 0.98$ , which seems a reasonable empirical choice. In the case of the background skin, any position in the image classified as skin with high confidence is eligible for providing additional training samples. In the case of the lesion, an additional safeguard constraint is imposed. The idea is to eliminate artifacts and lesion pixel candidates from small isolated regions that are possibly not representative. For this purpose, mathematical morphological opening [31] is applied to regions classified as mole in the input binary mask. In our experiments, opening is applied using a disk of diameter  $d = 1$  mm as structural element. In the sequence, the continuous subregion resulting from opening that best encloses the initial lesion seed box is selected. Lesion pixels classified with high confidence inside this subregion are eligible for update. This restriction is not applied for the segmentation itself but only for the selection of training samples.

The classifiers are, thus, iteratively trained using those samples from the original seed regions plus an additional set of training samples randomly selected at each iteration. For practical computational, the maximum number of added samples for each class is limited to the corresponding number

of samples contained in the original seed regions. Thus, at each iteration, half of samples are the original, and half are updated with high confident ones.

The classification and training sample selection is repeated until convergence. The default stop criteria of the iterative update of training samples is defined as

$$\left| 1 - \frac{m_{k+1}}{m_k} \right| < 0.01, \quad (2)$$

where  $m_k$  is the number of pixels classified as lesion with high confidence at iteration  $k$ .

### 3. Postprocessing

The final binary segmentation mask might contain a certain number of disjoint regions classified as lesion. Ideally, the segmentation procedure is expected to produce two independent regions: lesion and background. Since these regions are rarely homogeneous, segmentation can classify multiple isolated objects as lesion. To obtain a single lesion object, a set of postprocessing steps are applied.

First, mathematical morphological erosion [31] is applied to the binary mask obtained from classification. Erosion is commonly used in image processing to eliminate small isolated regions and artifacts. In our experiments, erosion is applied using a disk of diameter  $d = 0.12$  mm as structural element. The size of this element seems to be a safe choice [32] when the intention is to eliminate potential isolated hairs classified as mole pixels over the background

skin. Morphological dilation [31] is applied to the resulting “cleaned” image. In order to better preserve the fine details of the border, otherwise eliminated by the morphological operations, a slightly bigger disk element of size  $d = 0.3$  mm is used for the dilation, and the resulting image is multiplied pixelwise by the input binary mask obtained from classification.

In sequence, the remaining objects in the generated binary mask that are spatially closer than a given threshold are connected. The idea is that small scattered objects classified as mole could be part of a continuous subregion, that eventually could be connected to a big segment, if close together. Connecting regions classified as mole might also be useful for those patients with light-colored hairs imaged over the lesion, likely to be classified as background skin, thus “breaking” the main lesion into subregions during the classification. Again, this can be easily done using a classical spatial image filtering procedure. A Gaussian filter is selected for our experiments. The bandwidth is set identical to the radius used in the above erosion procedure, and the empirical selected value 0.1 is used as threshold for generating the final binary mask containing a certain number of subregions.

Each contiguous subregion  $i$  labeled as lesion is given a score  $n_i$ . If each region  $i$  is labeled  $\mathcal{R}_i$  each region’s score is given by

$$n_i = \sum_{u,v \in \mathcal{R}_i} f(u,v), \quad (3)$$

where  $f(u,v)$  is the two-dimensional Gaussian function

$$f(u,v) = \exp \left\{ - \left[ \frac{(u - u_o)^2}{2\sigma_u^2} + \frac{(v - v_o)^2}{2\sigma_v^2} \right] \right\}, \quad (4)$$

where  $(u_o, v_o)$  are the coordinates of the center of the disk image, and  $\sigma$  is the Gaussian bandwidth. This gives each region a score that increases with size and proximity to the centre, and penalizes regions that are small or off centre. In our experiments, we set  $\sigma_u = \sigma_v = 2.5$  mm. This particular choice of  $\sigma$  gives over 95% of the total weight to regions (at least partially) centered in a disk of diameter 10 mm. In practice this postprocessing step allows to eliminate potentially disconnected peripheral objects classified as lesion and serves as a guide allowing the doctor to select the lesion to be analyzed by framing the lesion at the center of the picture.

Except for the region with largest score  $n_i$ , all remaining objects labeled as lesion are discarded. Any holes in this region are filled. It is worth mentioning that although “single” lesions are more common, there exists also multifocal lesions. For such particular cases, the current default preferred behaviour of our algorithm is to outline the larger individual lesion.

Finally, the border of the lesion is drawn around the subregion with the highest score. The Gaussian filter used in the postprocessing renders the contour visually smooth, as usually is the output of hand-drawn borders by dermatologists. See the examples in Figure 2. However, when the border line is intended to be used for computing features for diagnosis in a computer aided diagnosis system, smoothing should be used carefully, since it may be removing information about

the contour irregularity, which is an important feature, for example, for the ABCD Rule of Dermatoscopy [33].

## 4. Experiments

**4.1. Data Acquisition.** The data used in this study are dermoscopic images acquired by a portable dermoscope (Dermlite Pro II HR) attached to a digital camera (Ricoh GR, Ricoh Company Ltd, Tokyo, Japan). The equipment acquires a circular imaged area of diameter about 17.4 mm (1650 pixels), spatial resolution of about 2400 dpi and 8 bits per channel color depth. The images have been corrected for nonuniform illumination using calibration color standards. A  $5 \times 5$  median filter was applied for noise reduction, but no further preprocessing was used for removing additional artifacts. As mentioned in the introductory section, the images considered in this study were acquired using an alcohol-based contact fluid liquid between the skin and the dermoscope, which minimizes formation of air bubbles in the images. If bubbles were present, it might call for an ad hoc preprocessing algorithm to remove light-colored areas due to reflection in the image that might compromise the skin modeling and location of the seed lesions, with loss of accuracy [6].

A set of 122 images of pigmented lesions divided between 100 benign and 22 malignant lesions (i.e., melanomas) were used for clinical evaluation of the segmentation. These images were randomly selected from a larger database of dermoscopic images and not used during development of the ICS framework. Copies of the images were printed in  $178 \times 178$  mm paper format and independently given to three dermatologists, who were asked to manually draw the contours of the skin lesions. No additional information, like the histopathological diagnosis, was given to the dermatologists. After careful digitalization, the contour obtained from dermatologists was stored for reference purposes.

It is worth noticing that delineation of the borders of lesions is challenging for dermatologists and is not part of the daily routine, and results can greatly vary between doctors. Essentially, doctors are trained to differentiate between benign and malignant lesions, not necessarily in the specific task of border location.

**4.2. Measures for Border Detection Evaluation.** Qualitative and quantitative approaches are the most common strategies used in literature for the purpose of evaluation of the performance of border detection in dermoscopic images.

Qualitative evaluation of lesion segmentation is a passive strategy in the sense that a candidate border is shown to a dermatologist, who is asked to provide a score or grade to the solution (e.g., good, acceptable, poor, and bad) based on visual assessment. Examples include [19, 34].

In quantitative evaluation, the role of the dermatologist is reversed. In this context, the dermatologist is asked to manually draw the border around the lesion, and the manually drawn border is used as ground truth. Assessing accuracy of an alternative segmentation requires definition of a similarity score between ground truth and a candidate border. Among

the many scores are the overlap-based agreement ratio in [35] that uses the logical operation exclusive disjunction (symbolized XOR), the sensitivity and specificity, precision and recall, true positive rate, false positive rate, pixel misclassification probability [36], the Hausdorff and the Hammoude distances in [7], the weighted performance index [37], among others. A common fact with these scores is that they are computed from paired comparisons of borders. Recently an extension considering simultaneous comparison with multiple reference ground-truths (obtained by several dermatologists) was examined in [38].

In this paper, we will focus on three of the previous scores. In addition, an alternative score derived from the Hausdorff distance computation will be introduced. The first two scores considered are the sensitivity and specificity. Sensitivity and Specificity are statistical measures of the performance of a binary classification test, commonly used in medical studies. In the context of segmentation of skin lesions, sensitivity measures the proportion of actual lesion pixels which are correctly identified as such. Specificity measures the proportion of background skin pixels which are correctly identified. Given the following definitions:

TP: true positive, lesion pixels correctly classified as lesion,

FP: false positive, skin pixels incorrectly identified as lesion,

TN: true negative, skin pixels correctly identified as skin,

FN: false negative, lesion pixels incorrectly identified as skin,

and the number of pixels (#) in each of the above categories, the sensitivity and specificity are given by

$$\begin{aligned} \text{sensitivity} &= \frac{\#TP}{\#TP + \#FN}, \\ \text{specificity} &= \frac{\#TN}{\#TN + \#FP}. \end{aligned} \quad (5)$$

The third score considered is the Hausdorff distance defined as follows. Let  $\mathcal{M} = \{m_1, m_2, \dots, m_z\}$  and  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$  denote the set of points belonging to the manually and automatically drawn contours  $\mathcal{M}$  and  $\mathcal{A}$ , respectively. The distance from  $m_i$  to its closest point in  $\mathcal{A}$  is given by

$$d(m_i, \mathcal{A}) = \min_j \|m_i - a_j\|. \quad (6)$$

The Hausdorff distance is the maximum of the distance to the closest points between the two curves,

$$d_H = \max \left\{ \max_i d(m_i, \mathcal{A}), \max_j d(a_j, \mathcal{M}) \right\}. \quad (7)$$

We convert  $d_H$  from pixels to millimeters for easier interpretation.

The forth and last score that will be considered is also based on the distance between the ground truth and

candidate contour. Instead of taking the overall maximum value of the distance between contour points, as done in (7), we will compute the fraction of the contour pixels with an error lower than a predefined threshold  $\tau$

$$e(\tau) = \frac{1}{n+z} \left( \sum_{i=1}^z I(d(m_i, \mathcal{A}) \leq \tau) + \sum_{j=1}^n I(d(a_j, \mathcal{M}) \leq \tau) \right), \quad (8)$$

where the indicator function  $I(\cdot) = 1$  when the condition is satisfied, 0 otherwise.

The metric  $e(\tau)$  is convenient in the sense that it allows the analyst to set an error tolerance during the comparison of a candidate border and the ground-truth reference. What is computed in (8) is a ratio of contour pixels matched below the error threshold tolerance.  $e(\tau)$  is robust to the presence of local high disagreement between contours (outlier distances). For  $\tau \geq d_H$ ,  $e(\tau) = 1$ , and we should therefore set  $\tau < d_H$ . Since the value of  $d_H$  is different for each lesion, and the disagreement between contours usually depends on how easy or difficult is the lesion to be segmented, a tradeoff value for  $\tau$  should be set. In our experiments, we set  $\tau = 0.5$  mm. This applies for all lesions, and we believe it to be a reasonable tolerance for assessing the accuracy of the location of the contour of lesions seen at the high magnification provided by dermoscopes. In contrast to  $d_H$ , a higher  $e(\tau)$  score, indicates a better agreement of contours.

To conclude this section, it is worth noticing that when the sensitivity and specificity are computed, all pixels in the binary segmentation masks contribute to the final result. The Hausdorff distance, and the ratio of border samples with an error lower than the above threshold  $\tau = 0.5$  mm used in (8), are complementary scores to sensitivity and specificity, that are more oriented towards the quantification using the magnitude of a physical quantity as a measure of accuracy and tolerance of segmentation.

**4.3. Algorithm Settings and Benchmark.** During the classification, equal priors were assumed for skin and lesion. In order to speed up computation, all images were downsampled to  $826 \times 826$  pixels using bilinear interpolation [39].

For benchmarking, both unsupervised (automatic) and supervised segmentation methods are considered.

**4.3.1. Fully Automatic Segmentation Methods.** The first algorithm considered for benchmark was proposed in Celebi et al. (2008) [16]. The approach uses a region growing and merging technique called statistical region merging (SRM) [23] to segment the image. The SRM-based approach requires estimation of the color of the background skin. In our implementation, instead of placing four boxes at the corners of the image, as originality proposed, we used the location of the seed regions found by ICS to model the background skin. In practice, this avoids ambiguity, since the results would otherwise depend on how a square is cropped from the circular sector provided by dermoscopes, specially for lesions covering a large portion of the dermoscope.

The light-colored regions, that is, the regions whose mean color has a distance  $<60$  to the background skin color are eliminated. The final result is obtained by removing the isolated regions and then merging the remaining regions, followed by a postprocessing stage [16]. SRM has an internal parameter  $Q$  that makes it possible to control the coarseness of the segmentation. Higher values of  $Q$  result in finer segmentation and thus the generation of more regions [23]. Specific details about the setting of the parameter  $Q$  are not reported by Celebi et al. in [16]. Using a trial and error procedure, a few candidate values for the parameter  $Q$  (as suggested in [23]) were applied to images from an independent training set.  $Q = 32$  was found to be a reasonable choice. Clearly, a more in depth search using a more objective criterion for setting the optimal value of  $Q$  might lead to slightly different SRM results.

The second automatic algorithm considered for benchmark purposes is the adaptive thresholding (AT) recently proposed by Silveira et al. (2009) in [7]. AT performs segmentation by comparing the color of each pixel with a threshold  $\tau$ . A pixel is classified as lesion if it is darker than  $\tau$ . First the algorithm search for the RGB channel which allows the best discrimination. It is assumed that the entropy provides the answer, and the channel with maximum entropy value is selected. In most dermoscopic images, the blue channel is selected. The histogram of the automatically selected color component is computed. For bimodal histograms, the threshold is automatically computed as the local minima between the maxima, plus a small offset to account for quantization issues. When the histogram has a single component, the threshold is obtained from the 5% percentile color of a squared region located at the center of the image, plus a constant offset. The default offset values presented in [7] are used in our implementation.

**4.3.2. Supervised Segmentation Methods.** As mentioned in the introductory section, supervised methods implies user interaction during the segmentation process. The way supervision can be used varies a lot. Supervision could be used in the initialization of the algorithm, or alternatively, at the end of the process, allowing for instance user interaction for manual correction of a proposed solution. Different algorithms are available in literature. For instance, in [7], the authors consider three edge-based methods that require only two mouse clicks for initialization. Very promising results are reported. In our study, however, we will prefer to deal with approaches more oriented to pixelwise classification along the lines of the proposed ICS method.

It is important to stress that in general a direct comparison between supervised and unsupervised methods is unfair. Supervised methods are expected to outperform unsupervised methods, since they are given additional input information. But the reason why the proposed automatic ICS is compared with supervised methods in this paper has an important motivation: we would like to estimate how well a classification algorithm could perform if representative ground-truth (manually labeled) skin and lesion samples were used for training purposes.

In order to make the supervised analysis more trackable, a few practical experimental simplifications are done. First, we consider an estimated ground truth, given by the pixelwise majority voting of the independent segmentation masks provided by the three dermatologists (Section 4.1). In addition, the image is randomly sampled with a ratio corresponding to 1% of the total amount of pixels. Although low, this ratio provide a visually dense sampling of the image ( $\approx 22$  pixels/mm<sup>2</sup>) that, given the high spatial resolution of the dermoscopic image, typically allows the selection of thousand of samples for both the background skin and the mole class.

Three supervised classifiers will be used for the segmentation (binary classification) task: the classic linear and quadratic discriminant analysis, LDA and QDA, respectively, and the support vector machines classifier (SVM) [40]. The SVM kernel used is the Gaussian radial basis function (RBF). The regularization term in the formulation of SVM and the kernel bandwidth are set using a traditional grid-search procedure with maximization of the 10-fold cross validation accuracy. Both LDA and QDA are also available in the proposed ICS algorithm. SVM is a nonparametric method based on a mathematical framework and presents several advantages compared to other pattern recognition methods, including the ability to handle large numbers or predictors with relatively small sample. Diagnosis of skin lesions is one among the many examples of use of SVM [41].

**4.3.3. Postprocessing.** In order to minimize differences in the final border location due to differences introduced by specific postprocessing choices, mainly related to the criteria of exclusion of isolated regions and smoothing applied to draw the final border, the postprocessing discussed in Section 3 was applied to all the automatic and supervised segmentation methods, allowing a fair comparison.

**4.4. Experiment 1: Evaluation of the Accuracy of the Automatic Selection of the Seed Regions.** The location hypothesis that allows the initial sample selection for the proposed iterative segmentation algorithm plays a key role in the methodology. Intuitively, a good initialization increases the chance of a good final segmentation. In this first experiment, the interest is to experimentally verify if the proposed algorithm is able to accurately locate the skin and mole seed regions. For location accuracy assessment, here, the segmentation masks provided by the three dermatologists are combined to generate a reference map. For each image, we focus on the regions of agreement, where pixels are labeled as mole or skin by at least two of the three dermatologists. Ideally, we would like the algorithm to automatically place the initial small seed regions inside this “safe regions”. How it deviates from this ideal situation is then investigated in Figure 3. This figure shows the percentage of cases in the test set, where the seed regions are correctly located according with distinct accuracy levels. For assessment purposes, the lesions are grouped by diagnosis. For each lesion, the individual accuracies of both skin and mole seed regions are computed as the ratio

TABLE 1: Average accuracy of segmentation, quantified by the sensitivity (%), showing the performance of the proposed ICS method against alternative solutions. Borders manually drawn by three dermatologists, and their agreement, are considered as the ground-truth reference for accuracy assessment.

Ref.	Manual by doctors			Fully automatic			Supervised		
	D1	D2	D3	SRM	AT	ICS	LDA	QDA	SVM
D1	—	86.9	85.4	79.2	87.3	91.8	93.4	95.3	92.2
D2	89.5	—	87.1	79.6	87.1	93.4	94.3	96.8	93.6
D3	92.7	91.0	—	82.1	90.0	94.0	96.1	97.7	95.1
All	95.9	93.5	93.2	81.2	89.2	94.1	97.0	98.5	96.0

TABLE 2: Average accuracy of segmentation, quantified by the specificity (%), showing the performance of the proposed ICS method against alternative solutions. Borders manually drawn by three dermatologists, and their agreement, are considered as the ground-truth reference for accuracy assessment.

Ref.	Manual by doctors			Fully automatic			Supervised		
	D1	D2	D3	SRM	AT	ICS	LDA	QDA	SVM
D1	—	98.3	98.4	98.0	91.6	96.7	97.6	96.9	97.7
D2	96.1	—	97.6	98.3	90.3	96.3	96.9	96.2	96.8
D3	96.4	97.9	—	98.0	90.7	96.2	96.8	96.2	96.8
All	97.4	98.9	99.1	98.2	90.8	96.4	97.3	96.7	97.4

TABLE 3: Average accuracy of segmentation, quantified by the percentage of contour with an error  $\leq 0.5$  mm, showing the performance of the proposed ICS method against alternative solutions. Borders manually drawn by three dermatologists, and their agreement, are considered as the ground-truth reference for accuracy assessment.

Ref.	Manual by doctors			Fully automatic			Supervised		
	D1	D2	D3	SRM	AT	ICS	LDA	QDA	SVM
D1	—	84.5	87.1	75.7	75.0	81.1	86.3	81.8	88.8
D2	84.5	—	86.8	76.8	74.8	84.5	88.2	84.8	90.1
D3	87.1	86.8	—	78.5	77.1	82.7	88.5	83.3	90.6
All	92.6	91.7	95.5	78.7	77.1	84.8	91.5	87.5	94.4

between pixels correctly classified and the respective size of the seed. The “skin + mole” label in Figure 3 refers to the average value. Results in Figure 3 shows that for over 90% of the 122 dermoscopic lesions analyzed the seed regions are positioned in locations where at least two out of the three dermatologists would agree as accurate background skin or mole regions.

*4.5. Experiment 2: Clinical Evaluation.* When it comes to computing performance scores for assessment of the accuracy of the segmentation of skin lesions, like the scores presented in Section 4.2, the reference segmentation to be used as ground truth must be established.

According to previous studies, the use of the borders provided by a single dermatologist as ground truth should be avoided, since the solution for the same lesion by other dermatologists would exhibit a natural variability. A detailed discussion about this important point is beyond the scope of this work, but the reader is referred to [36, 42] for more details.

For the evaluation of the proposed method, the borders provided by three dermatologists are considered individually. An additional segmentation mask, generated by the pixelwise agreement in terms of the majority voting of the solutions

provided by the three dermatologists is computed. The use of this additional segmentation is a simple attempt to obtain a more accurate estimation of the underline ground truth. Majority voting attributes the same weight for each dermatologist, which seems a reasonable hypothesis given their similar professional training. The segmentation mask generated by combination of the single solutions provided by dermatologists is expected to be more accurate than each of the individual solutions, because each manual solution is drawn independently, with an expected accuracy better than 50% (random guess). This is one of the key points usually exploited in the design of the so called Ensemble methods in statistics and machine learning [43]. We believe majority voting would be a good attempt to estimate the underline ground truth, ideally considering manual borders from many independent dermatologists.

The average accuracy of segmentation is shown in Tables 1, 2, 3, and 4. The performance scores are computed using different references (marked “Ref.”) as ground truth: each of the manually drawn border by the three dermatologists, marked as “D1”, “D2”, and “D3”, plus an additional mask generated by the pixelwise majority voting of the three dermatologists’s segmentations, marked as “All”. The accuracies are grouped in three blocks according to the nature of

TABLE 4: Average accuracy of segmentation, quantified by the Hausdorff distance (mm), showing the performance of the proposed ICS method against alternative solutions. Borders manually drawn by three dermatologists, and their agreement, are considered as the ground-truth reference for accuracy assessment.

Ref.	Manual by doctors			SRM	Fully automatic			Supervised		
	D1	D2	D3		AT	ICS	LDA	QDA	SVM	
D1	—	1.50	1.35	2.50	1.91	1.68	1.40	1.47	1.17	
D2	1.50	—	1.31	2.42	2.14	1.56	1.22	1.33	1.14	
D3	1.35	1.31	—	2.43	1.93	1.61	1.25	1.37	1.13	
All	1.04	1.01	0.75	2.39	1.92	1.52	1.09	1.20	0.86	

the segmentation solution: automatic methods and supervised methods. For reference purposes, scores corresponding to manual segmentation are included. For 2 out of the 100 benign lesions present in the test set, SRM did not provide a segmentation. These two lesions were excluded in this experiment.

Concerning the automatic methods, the main focus in this paper, Tables 1–4 shows that the proposed ICS method provide very competitive results when compared to SRM and AT, all the reference segmentations taken as ground truth, and for the different accuracy scores considered. In terms of sensitivity ICS scores above 91.8%, while SRM remains under 82.1%, with intermediate values for AT. On the other hand, the lower sensitivity scores of SRM are balanced by the higher specificities that lies above 98.0%, with ICS performing slightly lower at 96.7%. Surprisingly, the sensitivities and specificities of the automatic methods, in particular the proposed ICS, are very close to those of the dermatologists. However, the better performance of human segmentations when compared to the automatic methods is much clearer when the other two scores, the Hausdorff distance and the percentage of border pixels with an error lower of 0.5 mm, are considered. For these two scores, the ICS algorithm performs better than the SRM and AT alternatives.

Tables 1–4 clearly evidence the very good segmentation performance by the three supervised methods. Notice, however, that the supervised results were obtained after a dense sampling of training pixels, selected using the skin and lesion portions given by the majority voting solution (see details in Section 4.3.2). Thus, it is natural that in most of cases the performance of LDA, QDA, and SVM are even better than those of the single doctors. On the other hand, the results are interesting in the sense that they provide an idea where is the limit of the accuracy of a computer generated segmentation would be in the current test set, provided that individual training for each image based on the estimated underline ground truth was available and considering the current postprocessing employed. Somehow surprising is the poor performance of QDA compared to LDA. In any event, results show clear evidence that the nonlinear SVM classifier performs the best.

The above experimental results are based on the average values of the scores used for accuracy assessment. The dispersion of the scores is also an important factor to be analyzed. This is shown in Figure 4. Here, to reduce the amount of

plots otherwise required for visualization of all data used to compute the averages in Tables 1–4, only the case corresponding to the use of the reference ground truth generated by majority voting is depicted (the scores corresponding to the row “all” in the previous tables). To understand the traditional box plots depicted in Figure 4, it is necessary to visualize its distinctive features. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints the algorithm considers to be not outliers, and the outliers are plotted individually. Not surprisingly, the automatic segmentation methods show higher variability when compared to the borders manually drawn by dermatologists and the supervised methods tested. Overall, we once again appreciate the very competitive results provided by the proposed ICS algorithm, when compared to the alternative automatic methods considered, for all the four scores considered. Except for the Hausdorff distance, higher values indicate a more accurate segmentation. For the test set, ICS proved to be more stable as shown by the lower degree of dispersion of the scores, in the plots shown by the shorter distance between the lower and upper quartiles summarizing the data.

The contrast between the background skin and the lesion is expected to influence the accuracy of the delineation of the borders of the lesion. When evaluating the reasons why state-of-art CAD systems rejected analysing some lesions that would be of interest for dermatologists, Perrinaud et al. [3] observed that all instruments required the presence of “adequate contrast” between the lesion and surrounding nonlesional skin. Not surprisingly, borders manually drawn by different dermatologists tends to agree less when the delineation of the contours of the lesion is unclear. The same behavior is expected in automatic segmentation methods, impacting CAD systems.

In order to get additional insights about the performance of the different segmentation methods for distinct levels of contrast, Figure 5 shows the average values of the accuracy of segmentations, quantified by the four distinct scores, focusing on the automatic segmentation methods analyzed. For reference purposes, the scores obtained from manual segmentation are included. In this complementary analysis, the set of 120 test lesions is grouped in three disjoint subsets (categories), here labeled as “low”, “intermediate”, and “high”, referring to increasing levels of contrast between background skin and lesion. In this study, the luminance component of

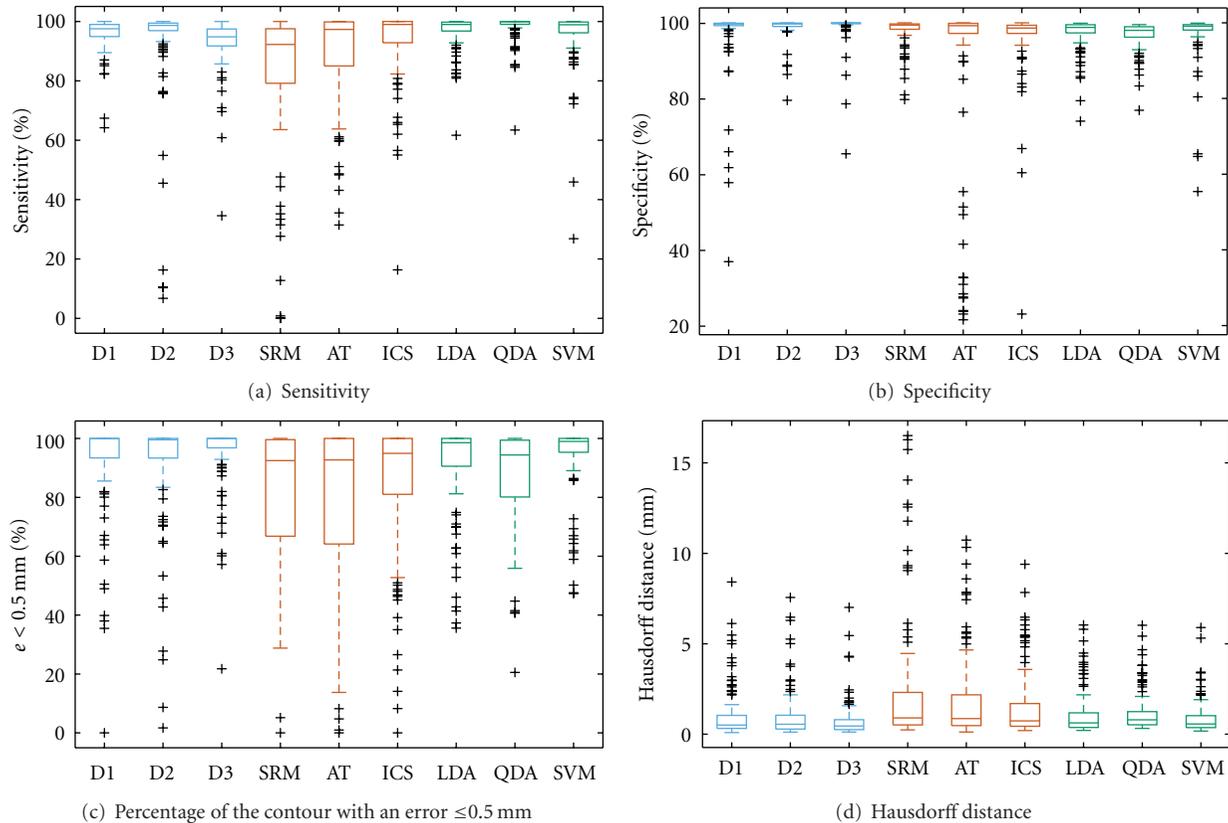


FIGURE 4: Dispersion of the accuracy of segmentation, quantified by the four distinct border measures, showing the performance of the proposed ICS method against the different segmentation methods analyzed. D1, D2, and D3 corresponds to the manual segmentations provided by dermatologists. SRM, AT, and ICS are automatic methods. LDA, QDA, and SVM are supervised methods that here include only for reference purposes. The accuracy scores are computed using the mask generated by majority voting of the manual segmentations provided by three dermatologists as ground-truth reference.

the dermoscopic image is used to measure the contrast, quantified by the absolute value of the difference between the median of the lesion and skin pixels. For the purpose of computing the luminance difference between background skin and lesion pixels, the mask generated by majority voting of the manual segmentation of the three dermatologists was used.

The three subsets of lesions were automatically partitioned by minimizing the within-cluster sum of squares of the luminance values, according to the Euclidian distance criteria. The resulting subsets contained 49, 56, and 15 lesions, corresponding to the subsets labeled as low, intermediate, and high contrast, respectively. For indicative purposes, the respective values obtained for the contrast at the center of the clusters were 9.0, 21.9, 40.4 (from  $3 \times 8$  bits RGB images converted to  $L * u * v$  space). The high number of lesions with very low contrast stress the challenge of accurately segment the current test set, composed by many lesions of very low contrast (accounting for 41% of the 120 lesions), randomly selected, and previously unseen by the segmentation methods tested.

Figure 5 confirms that the accuracy of all the segmentation methods decreases when the contrast between lesion and skin is lower. In terms of segmentation scores, we observe

that the proposed ICS presents very competitive results when compared to the SRM and AT algorithms, specially in presence of low contrast lesions. For instance, in terms of average sensitivity and specificity for this particular “low contrast” subset of lesions, ICS scores 88.6% and 93.7%, whereas AT scores 84.4% and 80.7%, when SRM scores 61.7% and 98.3%, respectively. Concerning these performance scores, we conclude that the best tradeoff is provided by ICS, followed by AT and SRM. The competitive performance of ICS is also confirmed when compared to the best performance by alternative algorithms, for both the Hausdorff distance (2.4 mm for ICS against 3.5 mm for AT) and the percentage of contour with an error  $\leq 0.5$  mm (74.5% for ICS against 60.5% for SRM). The good news is that when the level of contrast between skin and lesion increases, suggesting that the moles are easier to be segmented, the performance of the three methods increases considerably, remaining similar for all algorithms.

We conclude this experiment examining the accuracy of the segmentation methods in terms of histopathological diagnosis. The lesions are grouped in two subsets according to benignity (98 lesions) or malignancy (22 lesions). Figure 6 once again confirms the very competitive results provided by ICS. In the current test, set we observe that ICS performed

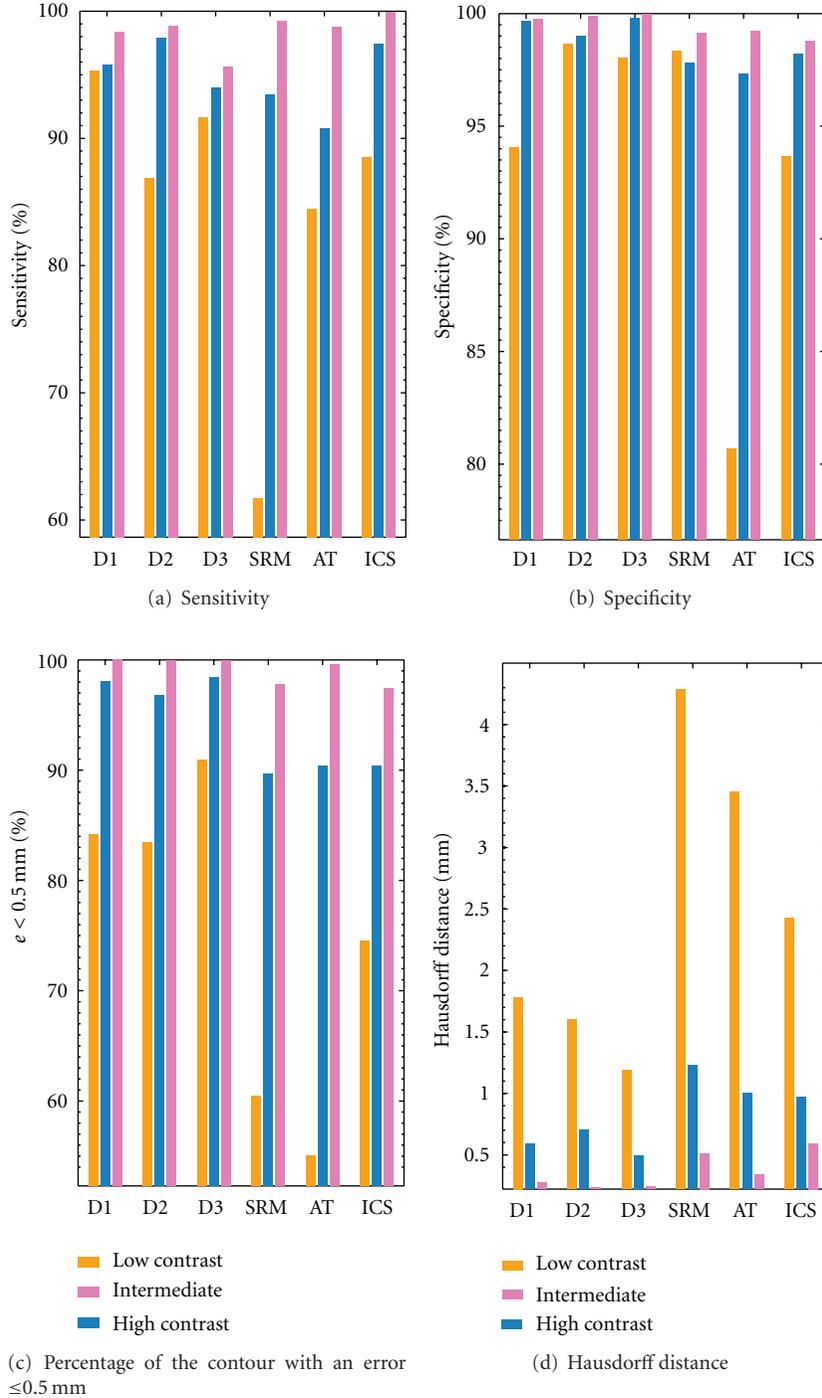


FIGURE 5: Average values of the accuracy of segmentation, quantified by the four distinct border measures, showing the performance of the proposed ICS method against the different segmentation methods analyzed. The lesions are grouped in three disjoint sets: low, intermediate, and high contrast between skin and lesion. The mask generate by majority voting of the manual segmentation by three dermatologists is used as ground-truth reference. The bars refer to low (orange), medium (blue), and high (pink) contrast lesions.

similarly well for both benign and malignant lesions. This behavior of ICS was found consistent for the different accuracy scores. The differences in terms of performance for benign and malignant lesions is more remarkable for both the AT and SRM algorithms, that in general tended to score

better on the malignant set, compared to their respective scores in the benign set.

4.6. *Experiment 3: Convergence and Posteriors in ICS.* The aim of the experiment is to provide additional details about

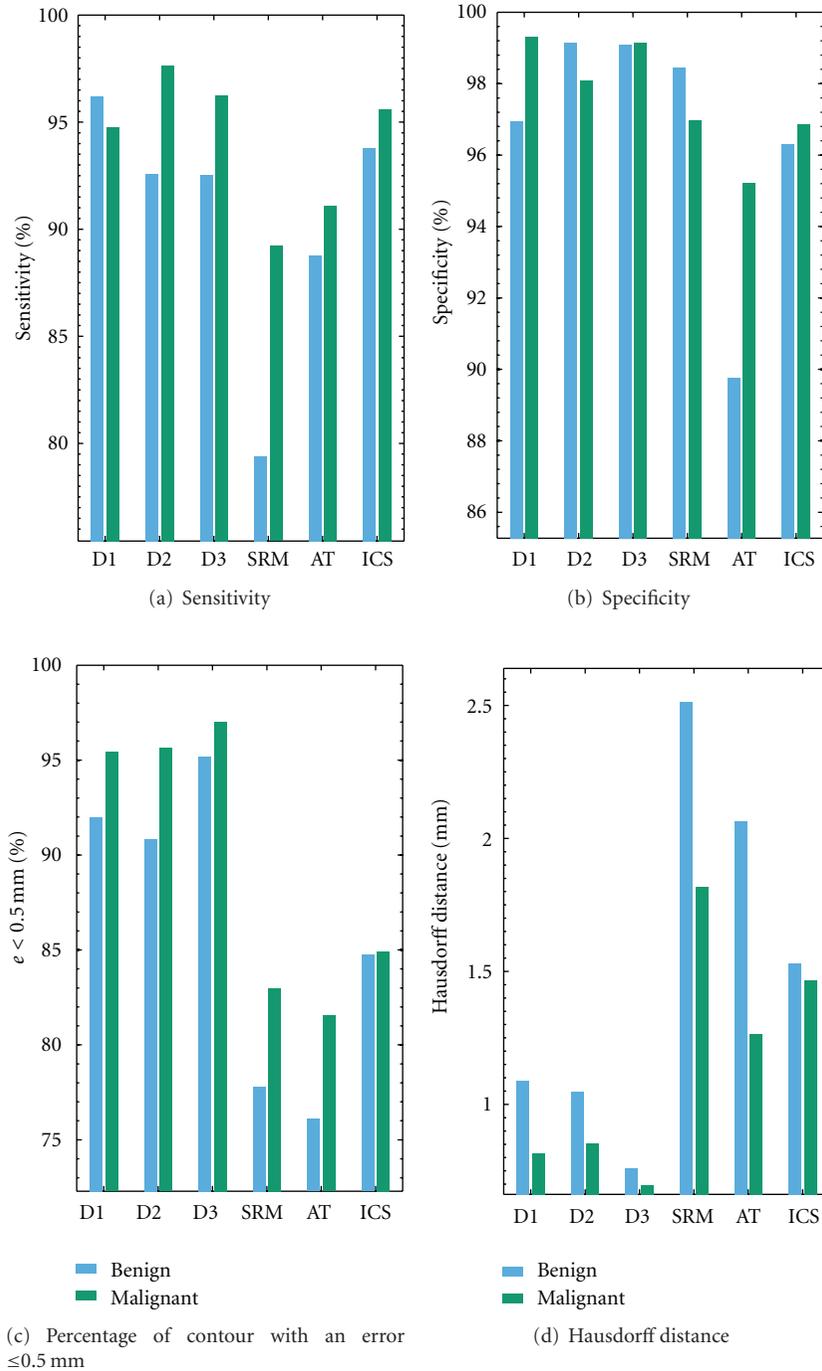


FIGURE 6: Average values of the accuracy of segmentations, quantified by the four distinct border measures, showing the performance of the proposed ICS method against the different segmentation methods analyzed. The lesions are grouped by histopathological diagnosis. The mask generate by majority voting of the manual segmentation of the three dermatologists is used as reference. The bars refer to benign (blue) and malignant (green) lesions.

the behavior of the proposed ICS framework during the classification. The two main aspects, (i) the number of iterations, and (ii) how the posteriors of the LDA and QDA classifiers are combined, are experimentally investigated.

Figure 7(a) provides details about the number of iterations. On average, the method converged in 4.7 iterations.

This suggests that just a few iterations are enough for modeling the statistics of the background skin and lesion. During experiments, we set the maximum number of iterations  $n = 30$ . If not converged at iteration  $n = 30$ , the iterative classification was stopped, and the last result used. As shown in Figure 7(a), this occurred only for one lesion, suggesting

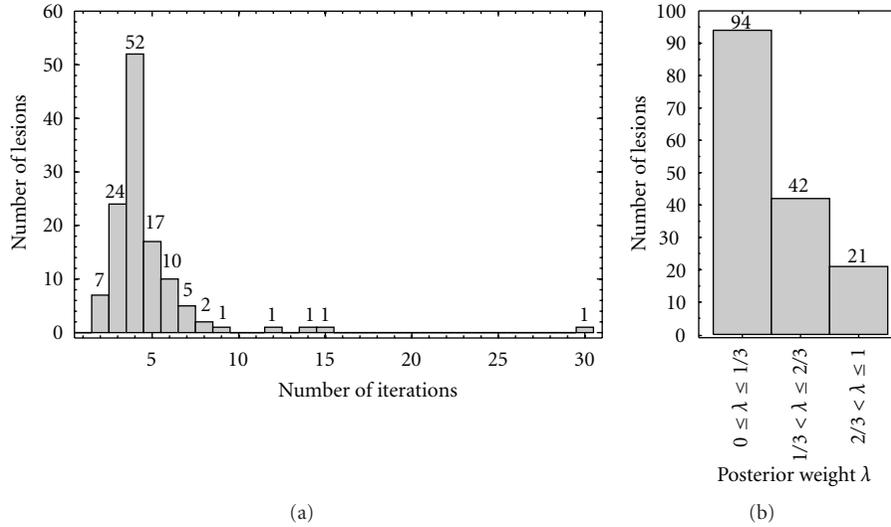


FIGURE 7: (a) Number of iterations for convergence and the (b) frequencies of the automatically selected value for the weight  $\lambda$  showing how the LDA ( $\lambda = 0$ ) and QDA ( $\lambda = 1$ ) posteriors were combined according to (1) in the proposed ICS algorithm. Results are summarized according to three intervals of possible  $\lambda$  values.

that the proposed method is very likely to converge for the task of segmentation of dermoscopic images.

How the posteriors are combined in the proposed ICS framework is analyzed in Figure 7(b). For summarization purposes, we group the values of  $\lambda_k$  in (1) in three intervals (i)  $0 \leq \lambda_k \leq 1/3$  (ii)  $1/3 < \lambda_k \leq 2/3$ , and (iii)  $2/3 < \lambda_k \leq 1$  and count the number of lesions that had, for some of the iterations  $k = \{1, \dots, n\}$ , a  $\lambda_k$  value inside the intervals. As shown in Figure 7(b), for most of the cases (94 lesions) the proposed ICS framework selected low values of  $\lambda_k$ , suggesting the preference for LDA. This fact is not surprising in the current implementation. Since the separability between the initial samples from the skin and lesion seeds is usually good, the accuracy of the seed regions is very high in most cases ( $\approx 100\%$ ), and in case of identical accuracy for both LDA and QDA,  $\lambda_k$  in (1) is taken as the lowest value possible, thus intentionally privileging the simpler LDA (for the reasons discussed in Section 2.4). It is also worth noticing that when LDA was used in the context of supervised segmentation, LDA tended to perform slightly better than QDA (Tables 1–4). But selection of low values for  $\lambda_k$  is not always the case, in particular, for other 42 and 21 lesions in the two remaining intervals shown in Figure 7(b), indicating intermediate preferences between LDA and QDA, and a stronger preference for QDA, respectively. In the current analysis, the total number of lesions contained in the three intervals is bigger than the number of lesions in the test set. This is due to the fact that  $\lambda_k$  is optimized for each iteration, so in this analysis, a given lesion may account in more than one of the intervals considered until reaching convergence.

We conclude our analysis showing in Figures 8 and 9 a few examples of pigmented skin lesions from the test set, and the manual border drawn by the dermatologists, the proposed ICS, and the alternative AT and SRM methods. Note that none of the methods performs better in all cases.

## 5. Conclusions and Final Remarks

We have presented an automatic algorithm for segmentation of pigmented skin lesions. The technique is primarily developed for analysis of images acquired by a portable dermatoscope attached to a consumer-grade digital camera.

In contrast to other automatic segmentation techniques, the proposed ICS algorithms relies on specific assumptions about the image acquisition, in particular the approximate location and color of the skin and lesion. The assumptions are simple in nature and are designed for the specific problem of segmentation of dermoscopic images. The main purpose is a safe selection of initial small seed regions corresponding to skin and lesion portions that through iterative classification leads to the final segmentation mask.

The clinical accuracy assessment using 122 dermoscopic images, randomly selected, with ground-truth lesion borders manually drawn by three dermatologists suggests competitive segmentation results when the proposed ICS algorithm is compared to alternative automatic segmentation methods. The improvements are particularly remarkable for lesions with low contrast between background skin and lesion. In addition, in the current test set the proposed algorithm was found to perform similarly well for both benign and malignant lesions.

Overall, the proposed framework is simple and flexible enough to allow testing with different classifiers. Compared to a traditional 1-D histogram-based segmentation, the proposed approach uses all the RGB color information available. In addition, since the proposed segmentation framework is essentially classification based, it could eventually accommodate additional input features such as contextual information. Usually, convergence was reached in a few iterations. The algorithm is relatively fast (takes about 1 min.), and the processing time depends essentially on

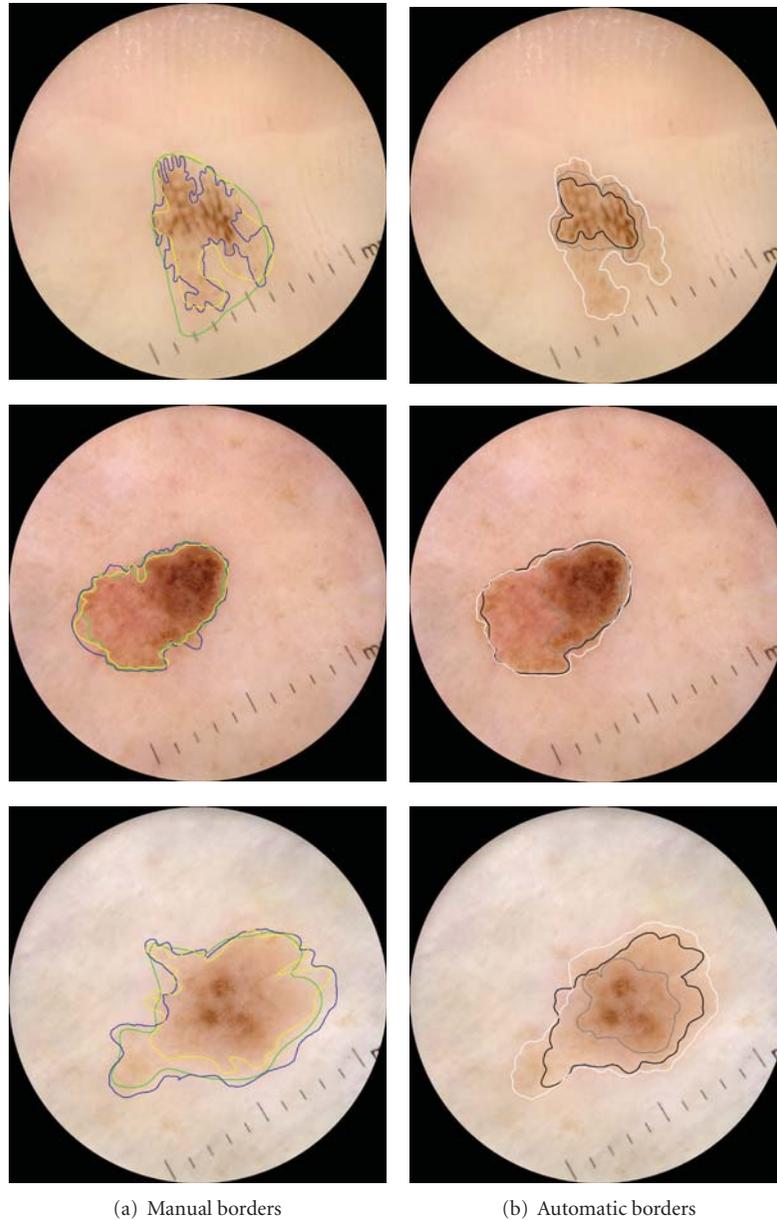


FIGURE 8: Example of lesions in the test set and the border provided by (a) the three dermatologists and (b) the automatic borders by AT (black border), SRM (gray border), and the proposed ICS algorithm (white border). None of the methods performed better in all cases.

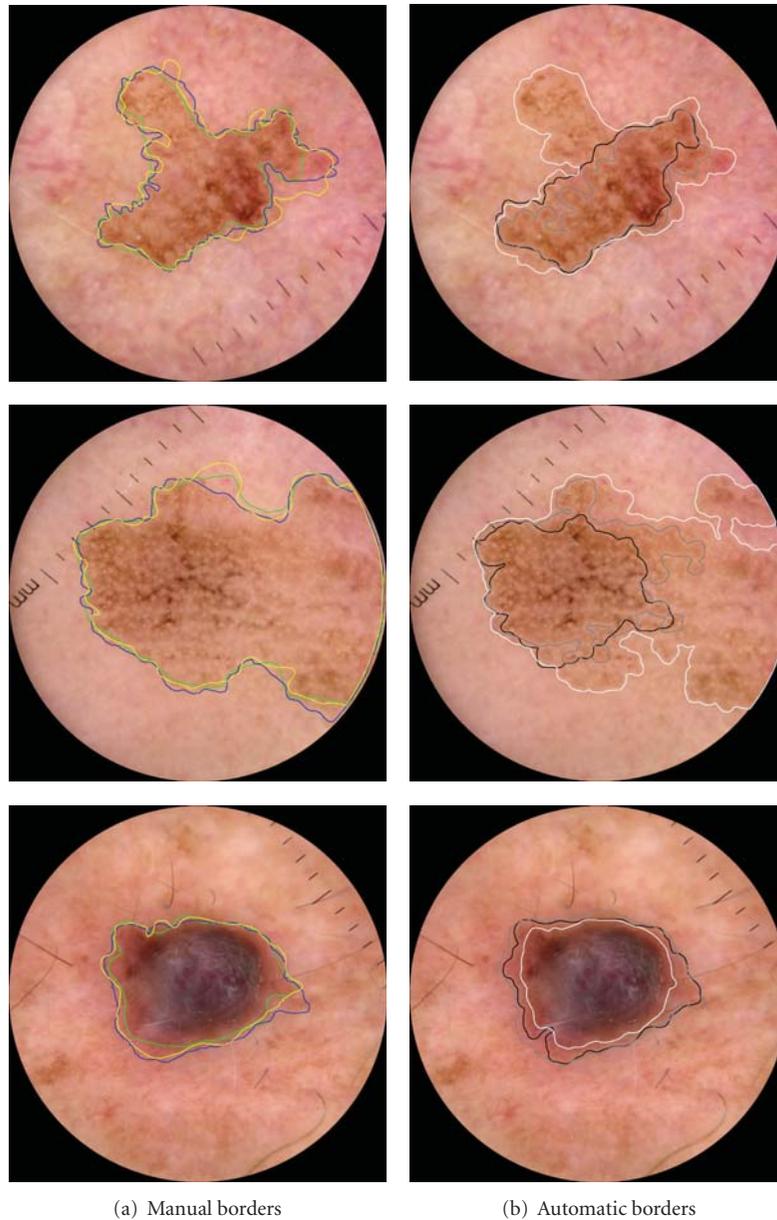
the choice of the classifiers, whose posteriors are combined automatically.

We believe that the suggested framework is general enough to be useful for analysis of other kind of images acquired by different equipments, adapting the initial assumptions about the geometry of acquisition and color of the lesion of interest to the specific problem at hand. In addition, despite not directly envisaged in this paper, it appears also that the proposed method could be easily modified to accommodate user iteration, for instance, by manual placement of seed regions, rendering the proposed method even more robust.

We conclude reminding that it is important to keep in mind that the effect of border detection error upon the accuracy of a computer aided diagnosis system can only be validated when used as a part of a final diagnostic system.

### Acknowledgments

The authors gratefully acknowledge the help they received from dermatologists Dr. M. Lorentzen, G. M. Traaen, and K. Bremnes at the University Hospital of North Norway for the delineation of the borders of the the pigmented skin



(a) Manual borders

(b) Automatic borders

FIGURE 9: Example of lesions in the test set and the border provided by (a) the three dermatologists and (b) the automatic borders by AT (black border), SRM (gray border), and the proposed ICS algorithm (white border). None of the methods performed better in all cases.

lesions. The authors would like to thank the Mathematics and Statistics Group at the Tromsø Telemicine Laboratory, in particular Marc Geilhufe and Kevin Thon for help and assistance. The authors would also like to thank anonymous reviewers for their helpful comments on this paper.

## References

- [1] G. Argenziano and H. Soyer, "Dermoscopy of pigmented skin lesions—a valuable tool for early diagnosis of melanoma," *The Lancet Oncology*, vol. 2, no. 7, pp. 443–449, 2001.
- [2] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [3] A. Perrinaud, O. Gaide, L. E. French, J. H. Saurat, A. A. Marghoob, and R. P. Braun, "Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? a study comparing the results of three systems," *British Journal of Dermatology*, vol. 157, no. 5, pp. 926–933, 2007.
- [4] R. Melli, C. Grana, and R. Cucchiara, "Comparison of color clustering algorithms for segmentation of dermatological images," in *Proceedings of the SPIE Medical Imaging: Image Processing*, vol. 6144, San Diego, Calif, USA, February 2006.
- [5] P. Wighton, M. Sadeghi, T. K. Lee et al., "A fully automatic random walker segmentation for skin lesions in a supervised setting," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI, 09)*, G.-Z. Yang, D. Hawkes, D. Rueckert et al., Eds., vol. 5762 of *Lecture Notes in Computer Science*, pp. 1108–1115, Springer, London, UK, 2009.

- [6] M. Celebi, H. Iyatomi, G. Schaefer, and W. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 148–153, 2009.
- [7] M. Silveira, J. Nascimento, J. Marques et al., "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, 2009.
- [8] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV '01)*, vol. 1, pp. 105–112, Vancouver, Canada, July 2001.
- [9] A. Green, N. Martin, J. Pfitzner, M. O'Rourke, and N. Knight, "Computer image analysis in the diagnosis of melanoma," *Journal of the American Academy of Dermatology*, vol. 31, no. 6, pp. 958–964, 1994.
- [10] C. Grana, G. Pellacani, R. Cucchiara, and S. Seidenari, "A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 959–964, 2003.
- [11] D. D. Gómez, C. Butakoff, B. K. Ersbøll, and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 157–161, 2008.
- [12] K. Møllersen, H. M. Kirchesch, T. G. Schopf, and F. Godtliessen, "Unsupervised segmentation for digital dermoscopic images," *Skin Research and Technology*, vol. 16, no. 4, pp. 401–407, 2010.
- [13] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.
- [14] T. Lee, M. Atkins, M. King, S. Lau, and D. McLean, "Counting moles automatically from back images," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1966–1969, 2005.
- [15] H. Iyatomi, H. Oka, M. Saito et al., "Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system," *Melanoma Research*, vol. 16, no. 2, pp. 183–190, 2006.
- [16] M. Celebi, H. Kingravi, H. Iyatomi et al., "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, no. 3, pp. 347–353, 2008.
- [17] E. Claridge and A. Orun, "Modelling of edge profiles in pigmented skin lesions," in *Proceedings of Medical Image Understanding and Analysis*, pp. 53–56, 2002.
- [18] B. Erkol, R. Moss, R. J. Stanley, W. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," *Skin Research and Technology*, vol. 11, no. 1, pp. 17–26, 2005.
- [19] T. Mendonca, A. Marcal, A. Vieira et al., "Comparison of segmentation methods for automatic diagnosis of dermoscopy images," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6573–6576, IEEE Engineering in Medicine and Biology Society, 2007.
- [20] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [21] M. Roberts and E. Claridge, "An artificially evolved vision system for segmenting skin lesion images," *Lecture Notes in Computer Science*, vol. 2878, pp. 655–662, 2003.
- [22] H. Ganster, A. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 233–239, 2001.
- [23] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1452–1458, 2004.
- [24] M. Celebi, Y. Aslandogan, W. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Research and Technology*, vol. 13, no. 4, pp. 454–462, 2007.
- [25] A. Tenenhaus, A. Nkengne, J. Horn, C. Serruys, A. Giron, and B. Fertil, "Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions," *Skin Research and Technology*, vol. 16, no. 1, pp. 85–97, 2010.
- [26] D. Eedy and R. Wootton, "Teledermatology: a review," *British Journal of Dermatology*, vol. 144, no. 4, pp. 696–707, 2001.
- [27] A. Gewirtzman and R. Braun, "Computerized digital dermoscopy," *Journal of Cosmetic Dermatology*, vol. 2, no. 1, pp. 14–20, 2003.
- [28] A. Baldi, R. Murace, E. Dragonetti et al., "Definition of an automated content-based image retrieval (CBIR) system for the comparison of dermoscopic images of pigmented skin lesions," *BioMedical Engineering Online*, vol. 8, no. 1, p. 18, 2009.
- [29] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, MIT Press, Cambridge, Mass, USA, 2006.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 2009.
- [31] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer, New York, NY, USA, 2003.
- [32] N. Otberg, H. Richter, H. Schaefer, U. Blume-Peytavi, W. Sterry, and J. Lademann, "Variations of hair follicle size and distribution in different body sites," *Journal of Investigative Dermatology*, vol. 122, no. 1, pp. 14–19, 2004.
- [33] W. Stolz, A. Riemann, A. Cagnetta et al., "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," *European Journal of Dermatology*, vol. 4, no. 7, pp. 521–527, 1994.
- [34] F. Ercal and W. Stoecker, "Detection of skin tumor boundaries in color images," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 624–627, 1993.
- [35] G. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Unsupervised color image segmentation: with application to skin tumor borders," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, no. 1, pp. 104–111, 1996.
- [36] G. Joel, S. Philippe, G. David et al., "Validation of segmentation techniques for digital dermoscopy," *Skin Research and Technology*, vol. 8, no. 4, pp. 240–249, 2002.
- [37] R. Garnavi, M. Aldeen, and M. E. Celebi, "Weighted performance index for objective evaluation of border detection methods in dermoscopy images," *Skin Research and Technology*, vol. 17, no. 1, pp. 35–44, 2011.
- [38] M. E. Celebi, G. Schaefer, and H. Iyatomi, "Objective evaluation of methods for border detection in dermoscopy images," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '08)*, pp. 3056–3059, Vancouver, Canada, August 2008.
- [39] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, New York, NY, USA, 2nd edition, 2007.
- [40] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2000.
- [41] A. Sboner and C. F. Aliferis, "Modeling clinical judgment and implicit guideline compliance in the diagnosis of melanomas using machine learning," in *Proceedings of the AMIA Annual Symposium*, pp. 664–668, American Medical Informatics Association, 2005.

- [42] E. Claridge, J. Smith, and P. Hall, "Evaluation of border irregularity in pigmented skin lesions against a consensus of expert clinicians," in *Proceedings of Medical Image Understanding and Analysis*, pp. 85–88, 1998.
- [43] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, New York, NY, USA, 2004.

## Research Article

# Multiclass Sparse Bayesian Regression for fMRI-Based Prediction

Vincent Michel,<sup>1,2,3</sup> Evelyn Eger,<sup>3,4</sup> Christine Keribin,<sup>2,5</sup> and Bertrand Thirion<sup>1,3</sup>

<sup>1</sup> PARIETAL Team, INRIA Saclay-Île-de-France, 91191 Saclay, France

<sup>2</sup> Laboratoire de Mathématiques, Université Paris-Sud 11, 91400 Orsay, France

<sup>3</sup> CEA, DSV, I2BM, NeuroSpin, 91191 Gif-sur-Yvette, France

<sup>4</sup> CEA, DSV, I2BM, INSERM U562, 91191 Gif-sur-Yvette, France

<sup>5</sup> SELECT Team, INRIA Saclay-Île-de-France, 91400, France

Correspondence should be addressed to Bertrand Thirion, bertrand.thirion@inria.fr

Received 23 December 2010; Revised 3 March 2011; Accepted 7 April 2011

Academic Editor: Kenji Suzuki

Copyright © 2011 Vincent Michel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Inverse inference* has recently become a popular approach for analyzing neuroimaging data, by quantifying the amount of information contained in brain images on perceptual, cognitive, and behavioral parameters. As it outlines brain regions that convey information for an accurate prediction of the parameter of interest, it allows to understand how the corresponding information is encoded in the brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, as there are far more features (voxels) than samples (images), and dimension reduction is thus a mandatory step. We introduce in this paper a new model, called *Multiclass Sparse Bayesian Regression (MCBR)*, that, unlike classical alternatives, automatically adapts the amount of regularization to the available data. MCBR consists in grouping features into several classes and then regularizing each class differently in order to apply an adaptive and efficient regularization. We detail these framework and validate our algorithm on simulated and real neuroimaging data sets, showing that it performs better than reference methods while yielding interpretable clusters of features.

## 1. Introduction

In the context of neuroimaging, machine learning approaches have been used so far to address diagnostic problems, where patients were classified into different groups based on anatomical or functional data. By contrast, in cognitive studies, the standard framework for functional or anatomical brain mapping was based on mass univariate inference procedures [1]. Recently, a new way of analyzing functional neuroimaging data has emerged [2, 3], and it consists in assessing how well behavioral information or cognitive states can be predicted from brain activation images such as those obtained with functional magnetic resonance imaging (fMRI). This approach opens new ways for understanding the mental representation of various perceptual and cognitive parameters, which can be regarded as the study of the corresponding *neural code*, albeit at a relatively low spatial resolution. The accuracy of the prediction of the behavioral or cognitive target variable, as well as the spatial layout of predictive regions, can provide valuable information about

functional brain organization; in short, it helps to *decode* the brain system [4].

Many different pattern recognition and machine learning methods have been used to extract information from brain images and compare it to the corresponding target. Among them, *Linear Discriminant Analysis (LDA)* [3, 5], *Support Vector Machine (SVM)* [6–9], or regularized prediction [10, 11] has been particularly used. The major bottleneck in this kind of analytical framework is that there are far more features than samples, so that the problem is plagued by the curse of dimensionality, leading to overfitting. Dimension reduction can be used to extract relevant information from the data, the standard approach in functional neuroimaging being feature selection (e.g., *Anova*) [3, 6, 11, 12]. However, by performing feature selection and parameter estimation separately, such approach is not optimal. Thus, a popular combined selection/estimation scheme, *Recursive Feature Elimination* [13], may be used. However, this approach relies on a specific heuristic, which does not guarantee the optimality of the solution and is particularly costly.

By contrast, there is great interest in sparsity-inducing regularizations, which optimize both simultaneously.

In this paper, we assume that the code under investigation is about some scalar parameter that characterizes the stimuli, such as a scale/shape parameters but possibly also position, speed (assuming a 1-D space), or cardinality. Thus, we focus on regression problems and defer the generalization to classification to future work. Let us introduce the following predictive linear model:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + b, \quad (1)$$

where  $y$  represents the behavioral variable and  $(\mathbf{w}, b)$  are the parameters to be estimated on a training set. A vector  $\mathbf{w} \in \mathbb{R}^p$  can be seen as an image;  $p$  is the number of features (or voxels), and  $b \in \mathbb{R}$  is called the *intercept*. The matrix  $\mathbf{X} \in \mathbb{R}^{n \times p}$  is the design matrix. Each row is a  $p$ -dimensional sample, that is, an activation map related to the observation. With  $n \ll p$ , the estimation of  $\mathbf{w}$  is ill posed.

To cope with the high dimensionality of the data, one can penalize the estimation of  $\mathbf{w}$ , for example, based on the  $\ell_2$  norm of the weights. Classical regularization schemes have been used in functional neuroimaging, such as the Ridge regression [14], Lasso [15], or Elastic Net regression [16]. However, these approaches require the amount of penalization to be fixed beforehand and possibly optimized by cross-validation. To deal with the choice of the amount of penalization, one can use the Bayesian regression techniques, which include the estimation of regularization parameters in the whole estimation procedure. Standard Bayesian regularization schemes are based on the fact that a penalization by weighted  $\ell_2$  norm is equivalent to setting the Gaussian priors on the weights  $\mathbf{w}$ :

$$\begin{aligned} \mathbf{w} &\sim \mathcal{N}(0, A^{-1}), \quad A = \text{diag}(\alpha_1, \dots, \alpha_p), \\ \forall i \in [1, \dots, p], \quad \alpha_i &\in \mathbb{R}^+, \end{aligned} \quad (2)$$

where  $\mathcal{N}$  is the Gaussian distribution and  $\alpha_i$  the precision of the  $i$ th feature. The model in (2) defines two classical Bayesian regression schemes. The first one is *Bayesian Ridge Regression (BRR)* [17], which corresponds to the particular case  $\alpha_1 = \dots = \alpha_m$ . By regularizing all the features identically, BRR is not well suited when only few features are relevant. The second classical scheme is *Automatic Relevance Determination (ARD)* [18], which corresponds to the case  $\alpha_i \neq \alpha_j$  if  $i \neq j$ . The regularization performed by ARD is very adaptive, as all the weights are regularized differently. However, by regularizing each feature separately, ARD is prone to underfitting when the model contains too many regressors [19] and also suffers from convergence issues [20].

These classical Bayesian regularization schemes have been used in fMRI inverse inference studies [10, 14, 21]. However, these studies used only sparsity as built-in feature selection and do not consider neuroscientific assumptions for improving the regularization (i.e., within the design of the matrix  $A$ ). Indeed, due to the intrinsic smoothness of functional neuroimaging data [22], predictive information is rather encoded in different groups of features sharing similar information. A potentially more adapted approach

is the Bayesian regression scheme presented in [23], which regularizes patterns of voxels differently. The weights of the model are defined by  $\mathbf{w} = U\eta$ , where  $U$  is a matrix defined as set of spatial patterns (one pattern by column) and  $\eta$  are the parameters of the decomposition of  $\mathbf{w}$  in the basis defined by  $U$ . The regularization is controlled through the covariance of  $\eta$ , which is assumed to be diagonal with only  $m$  possible different values  $\text{cov}(\eta) = \exp(\lambda_1)\mathbf{I}^{(1)} + \dots + \exp(\lambda_m)\mathbf{I}^{(m)}$ .

The matrices  $\mathbf{I}^{(i)}$  are diagonal and defined subsets of columns of  $U$  sharing similar variance  $\exp(\lambda_i)$ . Due to its class-based model, this approach is similar to the one proposed in this paper, but the construction of  $I$  relies on ad hoc voxel selection steps, so that there is no proof that the solution is correct. A contrario, the proposed approach jointly optimizes, within the same framework, the construction of the pattern of voxels and the regularization parameter of each pattern.

In this paper, we detail a model for the Bayesian regression in which features are grouped into  $K$  different classes that are subject to different regularization penalties. The estimation of the penalty is performed in each class separately, leading to a stable and adaptive regularization. The construction of the group of features and the estimation of the predictive function are performed jointly. This approach, called *Multiclass Sparse Bayesian Regression (MCBR)*, is thus an intermediate solution between BRR and ARD. It requires less parameters to estimate than ARD and is far more adaptive than BRR. Another asset of the proposed approach in fMRI inverse inference is that it creates a clustering of the features and thus yields useful maps for brain mapping. After introducing our model and giving some details on the parameter estimation algorithms (the variational Bayes or Gibbs sampling procedures), we show that the proposed algorithm yields better accuracy than reference methods, while providing more interpretable models.

## 2. Multiclass Sparse Bayesian Regression

We first detail the notations of the problem and describe the priors and parameters of the model. Then, we detail the two different algorithms used for model inference.

*2.1. Model and Priors.* We recall the linear model for regression:

$$\mathbf{y} = f(\mathbf{X}, \mathbf{w}, b) = \mathbf{X}\mathbf{w} + b. \quad (3)$$

We denote by  $\mathbf{y} \in \mathbb{R}^n$  the targets to be predicted and  $\mathbf{X} \in \mathbb{R}^{n \times p}$  the set of activation images related to the presentation of different stimuli. The integer  $p$  is the number of voxels and  $n$  the number of samples (images). Typically,  $p \sim 10^3$  to  $10^5$  (for a whole volume), while  $n \sim 10$  to  $10^2$ .

*Priors on the Noise.* We use classical priors for regression, and we model the noise on  $y$  as an *i.i.d.* Gaussian variable:

$$\begin{aligned} \epsilon &\sim \mathcal{N}(0, \alpha^{-1}\mathbf{I}_n), \\ \alpha &\sim \Gamma(\alpha; \alpha_1, \alpha_2), \end{aligned} \quad (4)$$

where  $\alpha$  is the precision parameter and  $\Gamma$  stands for the *gamma density* with two hyperparameters  $\alpha_1, \alpha_2$ :

$$\Gamma(x; \alpha_1, \alpha_2) = \alpha_2^{\alpha_1} x^{\alpha_1 - 1} \frac{\exp^{-x\alpha_2}}{\Gamma(\alpha_1)}. \quad (5)$$

*Priors on the Class Assignment.* In order to combine the sparsity of *ARD* with the stability of *BRR*, we introduce an intermediate representation, in which each feature  $j$  belongs to one class among  $K$  indexed by a discrete variable  $z_j$  ( $\mathbf{z} = \{z_1, \dots, z_p\}$ ). All the features within a class  $k \in \{1, \dots, K\}$  share the same precision parameter  $\lambda_k$ , and we use the following prior on  $\mathbf{z}$ :

$$\mathbf{z} \sim \prod_{j=1}^p \prod_{k=1}^K \pi_k^{\delta_{jk}}, \quad (6)$$

where  $\delta$  is *Kronecker's*  $\delta$ , defined as

$$\delta_{jk} = \begin{cases} 0 & \text{if } z_j \neq k, \\ 1 & \text{if } z_j = k. \end{cases} \quad (7)$$

We finally introduce an additional Dirichlet prior [24] on  $\pi$ :

$$\pi \sim \text{Dir}(\eta) \quad (8)$$

with a hyperparameter  $\eta$ . By updating at each step the probability  $\pi_k$  of each class, it is possible to prune classes. This model has no spatial constraints and thus is not spatially regularized.

*Priors on the Weights.* As in *ARD*, we make use of an independent Gaussian prior for the weights:

$$\mathbf{w} \sim \mathcal{N}(0, \mathbf{A}^{-1}) \quad \text{with } \text{diag}(\mathbf{A}) = \{\lambda_{z_1}, \dots, \lambda_{z_p}\}, \quad (9)$$

where  $\lambda_{z_j}$  is the precision parameter of the  $j$ th feature, with  $z_j \in \{1, \dots, K\}$ . We introduce the following prior on  $\lambda_k$ :

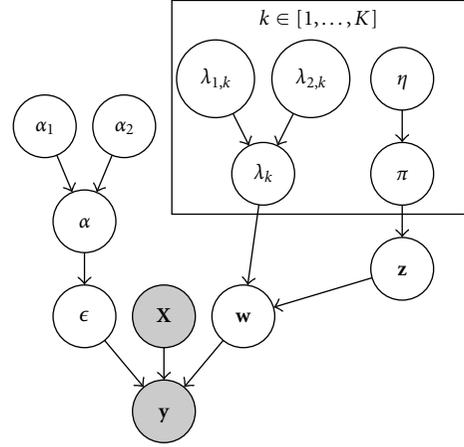
$$\lambda_k \sim \Gamma(\lambda_k; \lambda_{1,k}, \lambda_{2,k}) \quad (10)$$

with hyperparameters  $\lambda_{1,k}, \lambda_{2,k}$ . The complete generative model is summarized in Figure 1.

*2.1.1. Link with Other Bayesian Regularization Schemes.* The link between the proposed MCBR model and the other regularization methods, Bayesian Ridge Regression and Automatic Relevance Determination, is obvious.

- (i) With  $K = 1$ , that is,  $\lambda_{z_1} = \dots = \lambda_{z_p}$ , we retrieve the *BRR* model,
- (ii) With  $K = p$ , that is,  $\lambda_{z_i} \neq \lambda_{z_j}$  if  $i \neq j$ , and assigning each feature to a singleton class (i.e.,  $z_j = j$ ), we retrieve the *ARD* model.

Moreover, the proposed approach is related to the one developed in [25]. In this paper, the authors proposed, for the distribution of weights of the features, a binary mixture of Gaussians with small and large precisions. This model is used for variable selection and estimated by the *Gibbs sampling*. Our work can be viewed as a generalization of this model to a number of classes  $K \geq 2$ .



$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{w} + \epsilon + b & \pi &\sim \text{Dir}(\eta) \\ \epsilon &\sim \mathcal{N}(0, \alpha^{-1}\mathbf{I}_n) & \mathbf{w} &\sim \mathcal{N}(0, \mathbf{A}^{-1}) \\ \alpha &\sim \Gamma(\alpha; \alpha_1, \alpha_2) & &\text{with } \text{diag}(\mathbf{A}) = \{\lambda_{z_1}, \dots, \lambda_{z_p}\} \\ \mathbf{z} &\sim \prod_{j=1}^p \prod_{k=1}^K \pi_k^{\delta_{jk}} & \lambda_k &\sim \Gamma(\lambda_k; \lambda_{1,k}, \lambda_{2,k}) \end{aligned}$$

FIGURE 1: Graphical model of *Multiclass Sparse Bayesian Regression (MCBR)*. We denote by  $\mathbf{y} \in \mathbb{R}^n$  the targets to be predicted and by  $\mathbf{X} \in \mathbb{R}^{n \times p}$  the set of activation images. both the weights of the model  $\mathbf{w}$  depend on a discrete variable  $\mathbf{z}$  that assigns each feature to a class  $k$  among  $K$ . Both the noise  $\epsilon$  and the weights  $\mathbf{w}$  have a Gamma prior on their precisions. The variable  $\mathbf{z}$  follows a Dirichlet prior  $\pi$ .

*2.2. Model Inference.* For models with latent variables, such as MCBR, some singularities can exist. For instance in a mixture of components, a singularity is a component with one single sample and thus zero variance. In such cases, maximizing the *log likelihood* yields flawed solutions, and one can use the posterior distribution of the latent variables  $p(\mathbf{z} | \mathbf{X}, \mathbf{y})$  for this maximization. However, the posterior distribution of the latent variables given the data does not have a closed-form expression, and some specific estimation methods, such as *variational Bayes* or *Gibbs sampling*, have to be used.

We propose two different algorithms for inferring the parameters of the MCBR model. We first estimate the model by the *variational Bayes*, and the resulting algorithm is thus called *VB-MCBR*. We also detail an algorithm, called *Gibbs-MCBR*, based on a *Gibbs sampling* procedure.

*2.2.1. Estimation by Variational Bayes: VB-MCBR.* The *variational Bayes* (or *VB*) approach provides an approximation  $q(\Theta)$  of  $p(\Theta | \mathbf{y})$ , where  $q(\Theta)$  is taken in a given family of distributions and  $\Theta = [\mathbf{w}, \lambda, \alpha, \mathbf{z}, \pi]$ . Additionally, the *variational Bayes* approach often uses the following *mean field approximation*, which allows the factorization between the approximate distribution of the latent variables and the approximate distributions of the parameters:

$$q(\Theta) = q(\mathbf{w})q(\lambda)q(\alpha)q(\mathbf{z})q(\pi). \quad (11)$$

We introduce the *Kullback-Leibler* divergence  $\mathcal{D}(q(\Theta))$  that measures the similarity between the true posterior

$p(\Theta | \mathbf{y})$  and the variational approximation  $q(\Theta)$ . One can decompose the *marginal log-likelihood*  $\log p(\mathbf{y})$  as

$$\log p(\mathbf{y} | \Theta) = \mathcal{F}(q(\Theta)) + \mathcal{D}(q(\Theta)) \quad (12)$$

with

$$\begin{aligned} \mathcal{F}(q(\Theta)) &= \int d\Theta q(\Theta) \log \frac{p(\mathbf{y}, \Theta)}{q(\Theta)}, \\ \mathcal{D}(q(\Theta)) &= \int d\Theta q(\Theta) \log \frac{q(\Theta)}{p(\Theta | \mathbf{y})}, \end{aligned} \quad (13)$$

where  $\mathcal{F}(q(\Theta))$  is called *free energy* and can be seen as the measure of the quality of the model. As  $\mathcal{D}(q(\Theta)) \geq 0$ , the free energy is a lower bound on  $\log p(\mathbf{y})$  with equality if and only if  $q(\Theta) = p(\Theta | \mathbf{y})$ . So, inferring the density  $q(\Theta)$  of the parameters corresponds to maximizing  $\mathcal{F}$  with respect to the free distribution  $q(\Theta)$ . In practice, the VB approach consists in maximizing the free energy  $\mathcal{F}$  iteratively with respect to the approximate distribution  $q(\mathbf{z})$  of the latent variables and with respect to the approximate distributions of the parameters of the model  $q(\mathbf{w})$ ,  $q(\lambda)$ ,  $q(\alpha)$ , and  $q(\pi)$ .

The variational distributions and the pseudocode of the VB-MCBR algorithm are provided in Appendix A. This algorithm maximizes the free energy  $\mathcal{F}$ . In practice, iterations are performed until convergence to a local maximum of  $\mathcal{F}$ . With an ARD prior (i.e.,  $K = p$  and fixing  $z_j = j$ ), we retrieve the same formulas as the ones found for *Variational ARD* [18].

**2.2.2. Estimation by Gibbs Sampling: Gibbs-MCBR.** We develop here an estimation of the MCBR model using Gibbs sampling [26]. The resulting algorithm is called *Gibbs-MCBR*; the pseudocode of the algorithm and the candidate distributions are provided in Appendix B. The Gibbs sampling algorithm is used for generating a sequence of samples from the joint distribution to approximate marginal distributions. The main idea is to use conditional distributions that should be known and possibly easy to sample from, instead of directly computing the marginals from the joint law by integration (the joint law may not be known or hard to sample from). The sampling is done iteratively among the different parameters, and the final estimation of parameters is obtained by averaging the values of the different parameters across the different iterations (one may not consider the first iterations, this is called the *burn in*).

**2.2.3. Initialization and Priors on the Model Parameters.** Our model needs few hyperparameters; we choose here to use slightly informative and class-specific hyperparameters in order to reflect a wide range of possible behaviors for the weight distribution. This choice of priors is equivalent to setting heavy-tailed centered *Student's t*-distributions with variance at different scales, as priors on the weight parameters. We set  $K = 9$ , with weakly informative priors  $\lambda_{1,k} = 10^{k-4}$ ,  $k \in [1, \dots, K]$  and  $\lambda_{2,k} = 10^{-2}$ ,  $k \in [1, \dots, K]$ . Moreover, we set  $\alpha_1 = \alpha_2 = 1$ . Starting with a given number of classes and letting the model automatically prune the classes can be seen as a means of avoiding costly model

selection procedures. The choice of class-specific priors is also useful to avoid label switching issues and thus speeds up convergence. Crucially, the priors used here can be used in any regression problem, provided that the target data is approximately scaled to the range of values used in our experiments. In that sense, the present choice of priors can be considered as *universal*. We also randomly initialize  $q(\mathbf{z})$  for VB-MCBR (or  $\mathbf{z}$  for Gibbs-MCBR).

### 2.3. Validation and Model Evaluation

**2.3.1. Performance Evaluation.** Our method is evaluated with a cross-validation procedure that splits the available data into training and validation sets. In the following,  $(\mathbf{X}^t, \mathbf{y}^t)$  are a learning set  $(\mathbf{X}^t, \mathbf{y}^t)$  is a test set, and  $\hat{\mathbf{y}}^t = F(\mathbf{X}^t \hat{\mathbf{w}})$  refers to the predicted target, where  $\hat{\mathbf{w}}$  is estimated from the training set. The performance of the different models is evaluated using  $\zeta$ , the ratio of explained variance:

$$\zeta(\mathbf{y}^t, \hat{\mathbf{y}}^t) = \frac{\text{var}(\mathbf{y}^t) - \text{var}(\mathbf{y}^t - \hat{\mathbf{y}}^t)}{\text{var}(\mathbf{y}^t)}. \quad (14)$$

This is the amount of variability in the response that can be explained by the model (perfect prediction yields  $\zeta = 1$ , while  $\zeta < 0$  if prediction is worse than chance).

**2.3.2. Competing Methods.** In our experiments, the proposed algorithms are compared to different state-of-the-art regularization methods.

- (i) *Elastic Net Regression* [27], which requires setting two parameters  $\lambda_1$  and  $\lambda_2$ . In our analyzes, a cross-validation procedure within the training set is used to optimize these parameters. Here, we use  $\lambda_1 \in \{0.2\tilde{\lambda}, 0.1\tilde{\lambda}, 0.05\tilde{\lambda}, 0.01\tilde{\lambda}\}$ , where  $\tilde{\lambda} = \|\mathbf{X}^T \mathbf{y}\|_\infty$ , and  $\lambda_2 \in \{0.1, 0.5, 1., 10., 100.\}$ . Note that  $\lambda_1$  and  $\lambda_2$  parametrize heterogeneous norms.
- (ii) *Support Vector Regression (SVR)* with a linear kernel [28], which is the reference method in neuroimaging. The  $C$  parameter is optimized by cross-validation in the range of  $10^{-3}$  to  $10^1$  in multiplicative steps of 10.
- (iii) *Bayesian Ridge Regression (BRR)*, which is equivalent to MCBR with  $K = 1$  and  $\lambda_1 = \lambda_2 = \alpha_1 = \alpha_2 = 10^{-6}$ , that is, weakly informative priors.
- (iv) *Automatic Relevance Determination (ARD)*, which is equivalent to MCBR with  $K = p$  and  $\lambda_1 = \lambda_2 = \alpha_1 = \alpha_2 = 10^{-6}$ , that is, weakly informative priors.

All these methods are used after an *Anova*-based feature selection as this maximizes their performance. Indeed, irrelevant features and redundant information can decrease the accuracy of a predictor [29]. The optimal number of voxels is selected within the range  $\{50, 100, 250, 500\}$ , using a nested cross-validation within the training set. We do not directly select a threshold on  $P$  value or cluster size, but rather a predefined number of features. The estimation of the parameters of the learning function is also performed using a nested cross-validation within the training set, to ensure

TABLE 1: *Simulated regression data.* Explained variance  $\zeta$  for different methods (average of 15 different trials). The  $P$ -values are computed using a paired  $t$ -test.

Methods	Mean $\zeta$	Std $\zeta$	$P$ -value to Gibbs-MCBR
SVR	0.11	0.1	.0**
Elastic net	0.77	0.11	.0004**
BRR	0.19	0.14	.0**
ARD	0.79	0.06	.0**
Gibbs-MCBR	0.89	0.04	—
VB-MCBR	0.04	0.05	.0**

\*\* Level of significance of the  $P$ -values between 0.01 and 0.05.

a correct validation and an unbiased comparison of the methods. All methods are developed in *C* and used in *Python*. The implementation of elastic net is based on *coordinate descent* [30], while SVR is based on LibSVM [31]. Methods are used from *Python* via the *Scikit-learn* open source package [32].

For VB-MCBR and Gibbs-MCBR, in order to avoid a costly *internal cross-validation*, we select 500 voxels, and this selection is performed on the training set. The number of iterations used is fixed to 5000 (*burn in* of 4000 iterations) for Gibbs-MCBR and 500 for VB-MCBR. Preliminary results on both simulated and real data showed that these values are sufficient enough for an accurate inference of the model. As explained previously, we set  $K = 9$ , with weakly informative priors  $\lambda_{1,k} = 10^{k-4}$ ,  $k \in [1, \dots, K]$  and  $\lambda_{2,k} = 10^{-2}$ ,  $k \in [1, \dots, K]$ . Moreover, we set  $\alpha_1 = \alpha_2 = 1$ , and we randomly initialize  $q(\mathbf{z})$  for VB-MCBR (or  $\mathbf{z}$  for Gibbs-MCBR).

### 3. Experiments and Results

3.1. *Experiments on Simulated Data.* We now evaluate and illustrate MCBR on two different sets of simulated data.

3.1.1. *Details on Simulated Regression Data.* We first test MCBR on a simulated data set, designed for the study of ill-posed regression problem, that is,  $n \ll p$ . Data are simulated as follows:

$$\begin{aligned} \mathbf{X} &\sim \mathcal{N}(0, 1) \quad \text{with } \epsilon \sim \mathcal{N}(0, 1), \\ \mathbf{y} &= 2(\mathbf{X}_1 + \mathbf{X}_2 - \mathbf{X}_3 - \mathbf{X}_4) + 0.5(\mathbf{X}_5 + \mathbf{X}_6 - \mathbf{X}_7 - \mathbf{X}_8) + \epsilon. \end{aligned} \quad (15)$$

We have  $p = 200$  features,  $n^l = 50$  images for the training set, and  $n^t = 50$  images for the test set. We compare MCBR to the reference methods, but we do not use feature selection, as the number of features is not very high.

3.1.2. *Results on Simulated Regression Data.* We average the results of 15 different trials, and the average explained variance is shown in Table 1. Gibbs-MCBR outperforms the other approaches, yielding higher prediction accuracy than the reference elastic net and ARD methods. The prediction accuracy is also more stable than the other methods. VB-MCBR falls into the local maximum of  $\mathcal{F}$  and does not yield

an accurate prediction. BBR has a low prediction accuracy compared to other methods such as ARD. Indeed, it cannot finely adapt the weights of the relevant features, as these features are regularized similarly as the irrelevant ones. SVR has also low accuracy, due to the fact that we do not perform any feature selection. Thus, SVR suffers from the curse of dimensionality, unlike other methods such as ARD or elastic net, which performs feature selection and model estimation jointly.

In Figure 2, we represent the probability density function of the distributions of the weights obtained with BRR (a), Gibbs-MCBR (b), and ARD (c). With BRR, the weights are grouped in a monomodal density. ARD is far more adaptive and sets lots of weights to zero. The Gibbs-MCBR algorithm creates a multimodal distribution, lots of weights being highly regularized (pink distributions), and informative features are allowed to have higher weights (blue distributions).

With MCBR, weights are clustered into different groups, depending on their predictive power, which is interesting in application such as fMRI inverse inference, as it can yield more interpretable models. Indeed, the class to the features with higher weights ( $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4\}$ ) belong which is small (average size of 6 features) but has a high *purity* (percentage of relevant features in the class) of 74%.

3.1.3. *Comparison between VB-MCBR and Gibbs-MCBR.* We now look at the values of  $w_1$  and  $w_2$  for the different steps of the two algorithms (see Figure 3). We can see that VB-MCBR (b) quickly falls into a local maximum, while Gibbs-MCBR (a) visits the space and reaches the region of the correct set of parameters (red dot). VB-MCBR is not optimal in this case.

### 3.2. Simulated Neuroimaging Data

3.2.1. *Details on Simulated Neuroimaging Data.* The simulated data set  $\mathbf{X}$  consists of  $n = 100$  images (size  $12 \times 12 \times 12$  voxels) with a set of four square regions of interest (ROI) (size  $2 \times 2 \times 2$ ). We call  $\mathcal{R}$  the support of the ROI (i.e., the 32 resulting voxels of interest). Each of the four ROIs has a fixed weight in  $\{-0.5, 0.5, -0.5, 0.5\}$ . We call  $w_{i,j,k}$  the weight of the  $(i, j, k)$  voxel. The resulting images are smoothed with a Gaussian kernel with a standard deviation of 2 voxels, to mimic the correlation structure observed in real fMRI data. To simulate the spatial variability between images (intersubject variability, movement artifacts in intrasubject variability), we define a new support of the ROIs, called  $\tilde{\mathcal{R}}$  such that, for each image  $l$ th, 50% (randomly chosen) of the weights  $\mathbf{w}$  are set to zero. Thus, we have  $\tilde{\mathcal{R}} \subset \mathcal{R}$ . We simulate the target  $\mathbf{y}$  for the  $l$ th image as

$$y_l = \sum_{(i,j,k) \in \tilde{\mathcal{R}}} w_{i,j,k} X_{i,j,k,l} + \epsilon_l \quad (16)$$

with the signal in the  $(i, j, k)$  voxel of the  $l$ th image simulated as

$$X_{i,j,k,l} \sim \mathcal{N}(0, 1), \quad (17)$$

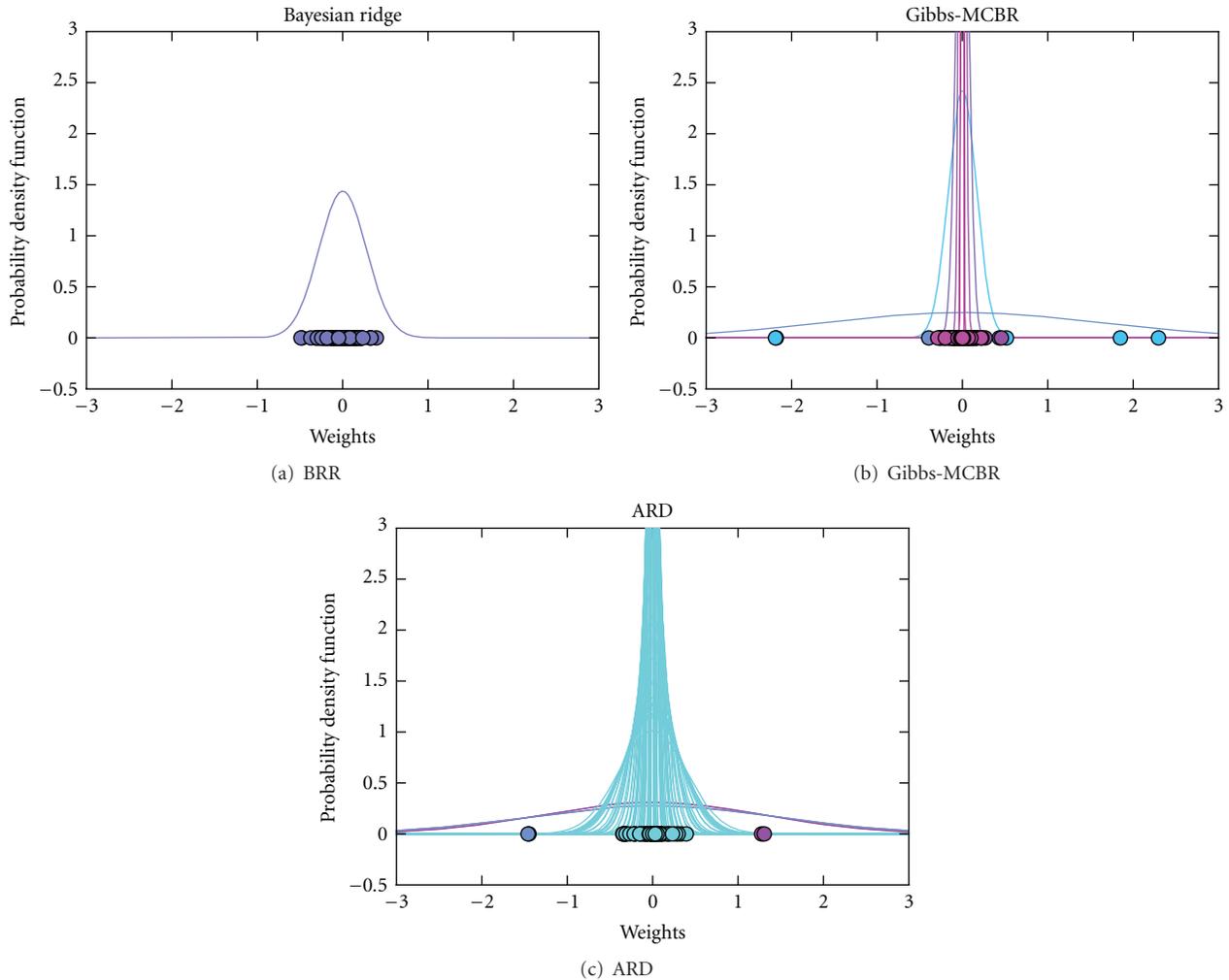


FIGURE 2: Results on simulated regression data. Probability density function of the weight distributions obtained with BRR (a), Gibbs-MCBR (b), and ARD (c). Each color represents a different component of the mixture model.

and  $\epsilon_l \sim \mathcal{N}(0, \gamma)$  is a Gaussian noise with standard deviation  $\gamma > 0$ . We choose  $\gamma$  in order to have a signal-to-noise ratio of 5 dB.

**3.2.2. Results on Simulated Neuroimaging Data.** We compare VB-MCBR and Gibbs-MCBR with the different competing algorithms. The resulting images of weights are given in Figure 4, with the true weights (a) and resulting Anova F-scores (b). The reference methods can detect the truly informative regions (ROIs), but elastic net (f) and ARD (h) retrieve only part of the support of the weights. Moreover, elastic net yields an overly sparse solution. BRR (g) also retrieves the ROIs but does not yield a sparse solution, as all the features are regularized in the same way. We note that the weights in the *feature space* estimated by SVR (e) are nonzero everywhere and do not outline the support of the ground truth. VB-MCBR (c) converges to a local maximum similar to the solution found by BRR (g); that is, it creates only one nonempty class, and thus regularizes all the features similarly. We can thus clearly see that, in this model, the variational

Bayes approach is very sensitive to the initialization and can fall into nonoptimal local maxima, for very sparse support of the weights. Finally, Gibbs-MCBR (d) retrieves most of the true support of the weights by performing an adapted regularization.

**3.3. Experiments and Results on Real fMRI Data.** In this section, we assess the performance of MCBR in an experiment on the *mental representation of object size*, where the aim is to predict the size of an object seen by the subject during the experiment, in both intrasubject and intersubject cases. The size (or scale parameter) of the object will be the target variable  $\mathbf{y}$ .

**3.3.1. Details on Real Data.** We apply the different methods on a real fMRI dataset related to an experiment studying the representation of objects, on ten subjects, as detailed in [33]. During this experiment, ten healthy volunteers viewed objects of 4 shapes in 3 different sizes (yielding 12 different experimental conditions), with 4 repetitions of each

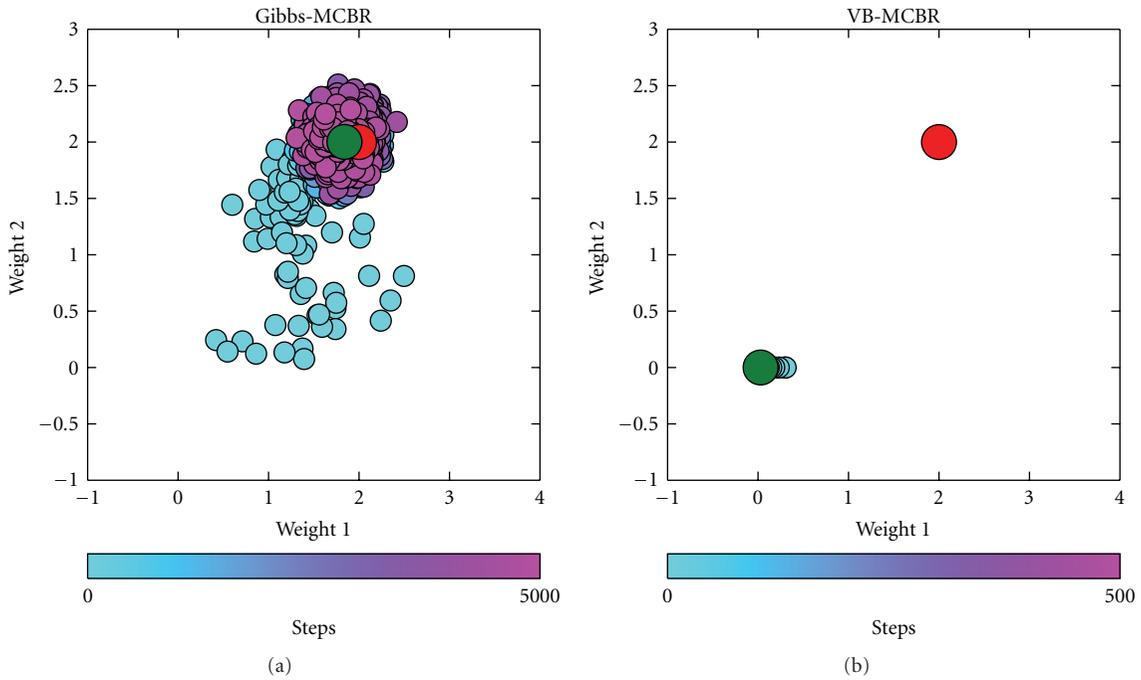


FIGURE 3: Results on simulated regression data. Weights of the first two features found for the different steps of Gibbs-MCBR (a) and VB-MCBR (b). The red dot represents the ground truth of both weights, and the green dot represents the final state found by the two algorithms. VB-MCBR is stuck in a local maximum, and Gibbs-MCBR finds the correct weights.

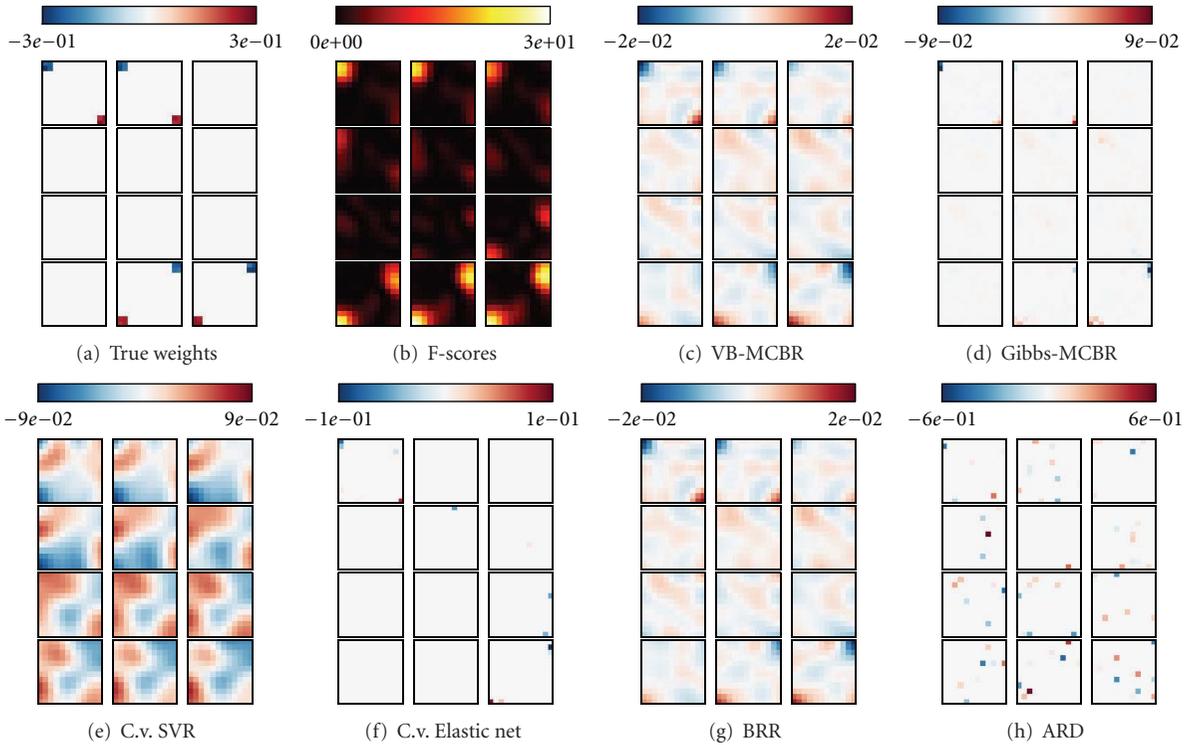


FIGURE 4: Two-dimensional slices of the three-dimensional volume of simulated data. Weights found by different methods, the true target (a) and F-score (b). The Gibbs-MCBR method (d) retrieves almost the whole spatial support for the weights. The sparsity-promoting reference methods, elastic net (f) and ARD (h), find an overly sparse support of the weights. VB-MCBR (c) converges to a local maximum similar to BRR (g) and thus does not yield a sparse solution. SVR (e) yields smooth maps that are not similar to the ground truth.

TABLE 2: *Intrasubject analysis*. Explained variance  $\zeta$  for the three different methods. The  $P$ -values are computed using a paired  $t$ -test. VB-MCBR yields the best prediction accuracy, while being more stable than the reference methods.

Methods	Mean $\zeta$	Std $\zeta$	$P$ -val/Gibbs-MCBR
SVR	0.82	0.07	.0006***
Elastic net	0.9	0.02	.001***
BRR	0.92	0.02	.0358**
ARD	0.89	0.03	.0015***
Gibbs-MCBR	0.93	0.01	—
VB-MCBR	0.94	0.01	.99

\*\* Level of significance of the  $P$ -values between 0.01 and 0.05.

\*\*\* Level of significance of the  $P$ -values below 0.01.

stimulus in each of the 6 sessions. We pooled data from the 4 repetitions, resulting in a total of  $n = 72$  images by subject (one image of each stimulus by session). Functional images were acquired on a 3-T MR system with an eight-channel head coil (Siemens Trio, Erlangen, Germany) as T2\*-weighted echo-planar image (EPI) volumes. Twenty transverse slices were obtained with a repetition time of 2 s (echo time: 30 ms; flip angle: 70°;  $2 \times 2 \times 2$ -mm voxels; 0.5 mm gap). Realignment, normalization to MNI space, and general linear model (GLM) fit were performed with the SPM5 software (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>). The normalization is the conventional method of SPM (implying affine and nonlinear transformations) and not the one using unified segmentation. The normalization parameters are estimated on the basis of a whole-head EPI acquired in addition and are then applied to the partial EPI volumes. The data are not smoothed. In the GLM, the effect of each of the 12 stimuli convolved with a standard hemodynamic response function was modeled separately, while accounting for serial autocorrelation with an AR(1) model and removing low-frequency drift terms using a high-pass filter with a cutoff of 128 s. The GLM is fitted separately in each session for each subject, and we used in the present work the resulting session-wise parameter estimate images (the  $\beta$ -maps are used as rows of  $\mathbf{X}$ ). The four different shapes of objects were pooled across for each one of the three sizes, and we are interested in finding discriminative information on sizes. This reduces to a regression problem, in which our goal is to predict a simple scalar factor (size of an object). All the analyzes are performed without any prior selection of regions of interest and use the whole acquired volume.

*Intrasubject Regression Analysis.* First, we perform an intrasubject regression analysis. Each subject is evaluated independently, in a 12-fold cross-validation. The dimensions of the real data set for one subject are  $p \sim 7 \times 10^4$  and  $n = 72$  (divided in 3 different sizes, 24 images per size). We evaluate the performance of the method by a leave-one-condition-out cross-validation (i.e., leave-6-image-out), and doing so the GLM is performed separately for the training and test sets. The parameters of the reference methods are optimized with a nested leave-one-condition-out cross-validation within the training set, in the ranges given before.

TABLE 3: *Intersubject analysis*. Explained variance  $\zeta$  for the different methods. The  $P$ -values are computed using a paired  $t$ -test. MCBR yields highest prediction accuracy than the two other Bayesian regularizations BRR and ARD.

Methods	Mean $\zeta$	Std $\zeta$	$P$ -val/Gibbs-MCBR
SVR	0.77	0.11	.14
Elastic net	0.78	0.1	.75
BRR	0.72	0.1	.01**
ARD	0.52	0.33	.02*
Gibbs-MCBR	0.79	0.1	—
VB-MCBR	0.78	0.1	0.4

\* Level of significance of the  $P$ -values.

\*\* Level of significance of the  $P$ -values between 0.01 and 0.05.

*Intersubject Regression Analysis.* Additionally, we perform an intersubject regression analysis on the sizes. The intersubject analysis relies on subject-specific fixed-effect activations that is, for each condition, the 6 activation maps corresponding to the 6 sessions are averaged together. This yields a total of 12 images per subject, one for each experimental condition. The dimensions of the real data set are  $p \sim 7 \times 10^4$  and  $n = 120$  (divided into 3 different sizes). We evaluate the performance of the method by cross-validation (leave-one-subject-out). The parameters of the reference methods are optimized with a nested leave-one-subject-out cross-validation within the training set, in the ranges given before.

### 3.3.2. Results on Real Data

*Intrasubject Regression Analysis.* The results obtained by the different methods are given in Table 2. The  $P$ -values are computed using a paired  $t$ -test across subjects. VB-MCBR outperforms the other methods. Compared to the results on simulated data, VB-MCBR still falls in a local maximum similar to the Bayesian ridge regression which performs well in this experiment. Moreover, both Gibbs-MCBR and VB-MCBR are more stable than the reference methods.

*Intersubject Regression Analysis.* The results obtained with the different methods are given in Table 3. As in the intrasubject analysis, both MCBR approaches outperform the reference methods, SVR, BRR, and ARD. However, the prediction accuracy is similar to that of elastic net. In this case, Gibbs-MCBR performs slightly better than VB-MCBR, but the difference is not significant.

One major asset of MCBR (and more particularly Gibbs-MCBR, as VB-MCBR often falls into a one-class local maximum) is that it creates a clustering of the features, based on the relevance of the features in the predictive model. This clustering can be accessed using the variable  $\mathbf{z}$ , which is implied in the regularization performed on the different features. In Figure 5, we give the histogram of the weights of Gibbs-MCBR for the intersubject analysis. We keep the weights and the values of  $\mathbf{z}$  of the last iteration; the different classes are represented as dots of different colors and are superimposed on the histogram. We can notice that the pink distribution represented at the bottom of the

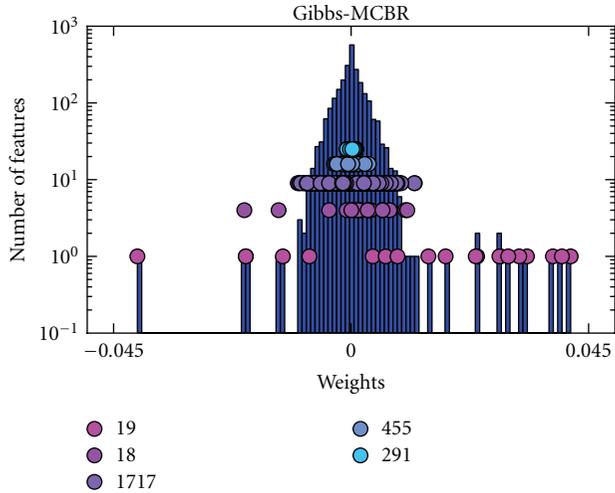


FIGURE 5: *Intersubject analysis*. Histogram of the weights found by Gibbs-MCBR and corresponding  $z$  values (each color of dots represents a different class), for the intersubject analyzes. We can see that Gibbs-MCBR creates clusters of informative and noninformative voxels and that the different classes are regularized differently, according to the relevance of the features in each of them.

histogram corresponds to relevant features. This cluster is very small (19 voxels), compared to the two blue classes represented at the top of the histogram that contain many voxels (746 voxels) which are highly regularized, as they are noninformative.

The maps of weights found by the different methods are detailed in Figure 6. The methods are used combined with an Anova-based *univariate feature selection* (2500 voxels selected, in order to have a good support of the weights). As elastic net, Gibbs-MCBR yields a sparse solution but extracts a few more voxels. The map found by elastic net is not easy to interpret, with very few informative voxels scattered in the whole occipital cortex. The map found by SVR is not sparse in the *feature space* and is thus difficult to interpret, as the spatial layout of the neural code is not clearly extracted. VB-MCBR does not yield a sparse map either, all the features having nonnull weights

#### 4. Discussion

It is well known that in high-dimensional problems, regularization of feature loadings significantly increases the generalization ability of the predictive model. However, this regularization has to be adapted to each particular dataset. In place of costly cross-validation procedures, we cast regularization in a Bayesian framework and treat the regularization weights as hyperparameters. The proposed approach yields an adaptive and efficient regularization and can be seen as a compromise between a global regularization (Bayesian Ridge Regression) that does not take into account the sparse or focal distribution of the information and automatic relevance determination. Additionally, MCBR

creates a clustering of the features based on their relevance and thus explicitly extracts groups of informative features.

Moreover, MCBR can cope with the different issues of ARD. ARD is subject to an underfitting in the hyperparameter space that corresponds to an underfitting in model selection (i.e., on the features to be pruned) [19]. Indeed, as ARD is estimated by maximizing evidence, models with less selected features are preferred, as the integration is done on less dimensions, and thus evidence is higher. ARD will choose the sparsest model across models with similar accuracy. A contrario, MCBR requires far less hyperparameter ( $2 \times K$ , with  $K \ll p$ ) and suffers less from this issue, as the sparsity of the model is defined by groups. Moreover, a full Bayesian framework for estimating ARD requires to set some priors on the *hyperparameters* (e.g.,  $\alpha_1$  and  $\alpha_2$ ), and it may be sensitive to specific choice of these hyperparameters. A solution is to use an *internal cross-validation* for optimizing these parameters, but this approach can be computationally expensive. In the case of MCBR, the distributions of the hyperparameters are bound to a class and not to each feature. Thus, the proposed approach is less sensitive to the choice of the hyperparameters. Indeed, the choice of good hyperparameters for the features is dealt with at the class level.

On simulated data, our approach performs better than other classical methods such as SVR, BRR, ARD, and elastic net and yields a more stable prediction accuracy. Moreover, by adapting the regularization to different groups of voxels, MCBR retrieves the true support of the weights and recovers a sparse solution. Results on real data show that MCBR yields more accurate predictions than other regularization methods. As it yields less sparse solution than elastic net, it gives access to more plausible loading maps which are necessary for understanding the spatial organization of brain activity, that is, retrieving the spatial layout of the neural coding. On real fMRI data, the explicit clustering of Gibbs-MCBR is also an interesting aspect of the model, as it can extract few groups of relevant features from many voxels.

In some experiments, the variational Bayes algorithm yields less accurate predictions than the Gibbs sampling approach, which can be explained by the difficulty of initializing the different variables (especially  $z$ ) when the support of the weight is overly sparse. Moreover, the VB-MCBR algorithm relies on a variational Bayes approach, which may not be optimal, due to strong approximations in model inference. A contrario Gibbs-MCBR is more time consuming but yields a better model inference. Finally, the variability in the results may be explained by the difficulty to estimate the model (optimality is not ensured).

The question of model selection (i.e., the number of classes  $K$ ) has not been addressed in this paper. One can use the free energy in order to select the best model, but due to the instability of VB-MCBR, this approach does not seem promising. A more interesting method is the one detailed in [34], which can be used with the Gibbs sampling algorithm. Here, model selection is performed implicitly by emptying classes that do not fit the data well. In that respect, the choice of heterogeneous priors for different classes is crucial: replacing our priors with class-independent priors (i.e.,  $\lambda_{1,k} = 10^{-2}$ ,  $k \in [1, \dots, K]$ ) in the intersubject analysis

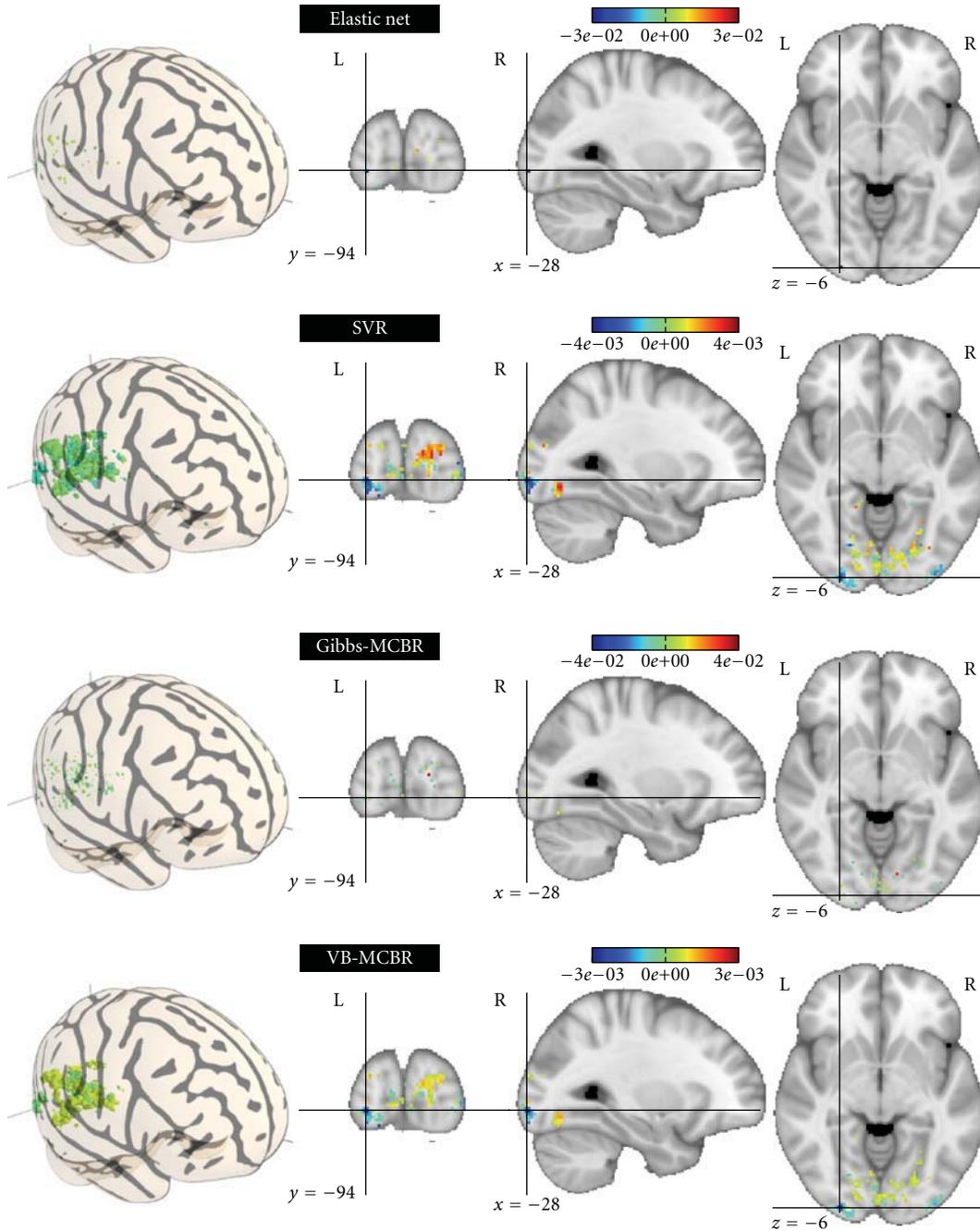


FIGURE 6: *Intersubject analysis.* Maps of weights found by the different methods on the 2500 most relevant features by Anova. The map found by elastic net is difficult to interpret as the very few relevant features are scattered within the whole brain. SVR and VB-MCBR do not yield a sparse solution. Gibbs-MCBR, by performing an adaptive regularization, draws a compromise between the other approaches and yields a sparse solution, but also extracts small groups of relevant features.

on size prediction leads Gibbs-MCBR to a local maximum similar to VB-MCBR.

Finally, this model is not restricted to the Bayesian regularization and can be used for classification, within a probit or logit model [35, 36]. The proposed model may thus be used for diagnosis in medical imaging, for the prediction of both continuous or discrete variables.

## 5. Conclusion

In this paper, we have proposed a model for adaptive regression, called *MCBR*. The proposed method integrates, in the same Bayesian framework, BRR and ARD and performs a different regularization for relevant and irrelevant features. It can tune the regularization to the possible different

Initialize  $a_1 = \alpha_1, a_2 = \alpha_2, l_1 = \lambda_1, l_2 = \lambda_2$  and  $d_k = \eta_k$   
 Randomly initialize  $q(z_j = k)$   
 Set a number of iterations *max steps*  
**repeat**  
   Compute  $A$  using (A.1),  $\Sigma$  using (A.2) and  $\mu$  using (A.3).  
   Compute  $l_1$  using (A.4) and  $l_2$  using (A.5).  
   Compute  $a_1$  using (A.6) and  $a_2$  using (A.7).  
   Compute  $\rho_{jk}$  using (A.8).  
   Compute  $\pi_k$  using (A.9) and  $d_k$  using (A.10).  
**until** *max steps*;  
**Return**  $\mu$ .

ALGORITHM 1: VB-MCBB algorithm.

Initialize  $\alpha_1, \alpha_2, \lambda_1, \lambda_2$  and  $\eta_k$   
 Randomly initialize  $z$   
 Set a number of iterations *burn number* for *burn-in*  
 Set a number of iterations *max steps*  
**Repeat**  
   Compute  $\Sigma$  using (B.1) and  $\mu$  using (B.2).  
   Sample  $\mathbf{w}$  in  $\mathcal{N}(\mathbf{w} | \mu, \Sigma)$ .  
   Compute  $l_1$  using (B.3) and  $l_2$  using (B.4).  
   Sample  $\lambda$  in  $\prod_{k=1}^K \Gamma(\lambda_k | l_{1,k}, l_{2,k})$ .  
   Compute  $a_1$  using (B.5) and  $a_2$  using (B.6).  
   Sample  $\alpha$  in  $\Gamma(a_1, a_2)$ .  
   Compute  $\rho_{jk}$  using (B.7).  
   Sample  $\mathbf{z}$  in  $\text{mult}(\exp \rho_{j,1}, \dots, \exp \rho_{j,K})$ .  
   Compute  $d_k$  using (B.8).  
   Sample  $\pi_k$  in  $\text{Dir}(d_k)$ .  
**until** *max steps*;  
**return** Average value of  $\mathbf{w}$  after *burn number* iterations.

ALGORITHM 2: Gibbs-MCBB algorithm.

level of sparsity encountered in fMRI data analysis, and it yields interpretable information for fMRI inverse inference, namely, the  $\mathbf{z}$  variable (latent class variable). Experiments on both simulated and real data show that our approach is well suited for neuroimaging, as it yields accurate and stable predictions compared to the state-of-the-art methods.

## Appendices

### A. VB-MCBB Algorithm

The *variational Bayes* approach yields the following variational distributions:

(i)  $q(\mathbf{w}) \sim \mathcal{N}(\mathbf{w} | \mu, \Sigma)$  with

$$\bar{\mathbf{A}} = \text{diag}(\bar{l}_1, \dots, \bar{l}_p) \text{ with} \quad (A.1)$$

$$\bar{l}_j = \sum_{k=1}^K q(z_j = k) \frac{l_{1,k}}{l_{2,k}} \quad \forall j \in \{1, \dots, p\},$$

$$\Sigma = \left( \frac{a_1}{a_2} \mathbf{X}^T \mathbf{X} + \bar{\mathbf{A}} \right)^{-1}, \quad (A.2)$$

$$\mu = \frac{a_1}{a_2} \Sigma \mathbf{X}^T \mathbf{y}; \quad (A.3)$$

(ii)  $q(\lambda_k) \sim \Gamma(l_{1,k}, l_{2,k})$  with

$$l_{1,k} = \lambda_{1,k} + \frac{1}{2} \sum_{j=1}^p q(z_j = k), \quad (A.4)$$

$$l_{2,k} = \lambda_{2,k} + \frac{1}{2} \sum_{j=1}^p (\mu_{jj}^2 + \Sigma_{jj}) q(z_j = k); \quad (A.5)$$

(iii)  $q(\alpha) \sim \Gamma(a_1, a_2)$  with

$$a_1 = \alpha_1 + \frac{n}{2}, \quad (A.6)$$

$$a_2 = \alpha_2 + \frac{1}{2} (\mathbf{y} - \mathbf{X}\mu)^T (\mathbf{y} - \mathbf{X}\mu) + \frac{1}{2} \text{Tr}(\Sigma \mathbf{X}^T \mathbf{X}); \quad (A.7)$$

(iv)  $q(z_j = k) \sim \exp(\rho_{jk})$  with

$$\rho_{jk} = -\frac{1}{2} (\mu_{jj}^2 + \Sigma_{jj}) \frac{l_{1,k}}{l_{2,k}} + \ln(\pi_k) + \frac{1}{2} (\Psi(l_{1,k}) - \log(l_{2,k})), \quad (A.8)$$

$$\pi_k = \exp^{\{\Psi(d_k) - \Psi(\sum_{k=1}^K d_k)\}}, \quad (A.9)$$

$$d_k = \eta_k + \sum_{j=1}^p q(z_j = k), \quad (A.10)$$

where  $\Psi$  is the digamma function  $\Psi(x) = \Gamma'(x)/\Gamma(x)$ . The VB-MCBB algorithm is provided in pseudo-code in Algorithm 1.

### B. Gibbs-MCBB Algorithm

With  $\Theta = [\mathbf{w}, \lambda, \alpha, \mathbf{z}, \pi]$ , we have the following candidate distributions (i.e., the distributions used for the sampling of the different parameters):

(i)  $p(\mathbf{w} | \Theta - \{\mathbf{w}\}) \propto \mathcal{N}(\mathbf{w} | \mu, \Sigma)$  with

$$\Sigma = (\mathbf{X}^T \mathbf{X} \alpha + \mathbf{A})^{-1} \quad \text{with } \mathbf{A} = \text{diag}(\lambda_{z_1}, \dots, \lambda_{z_p}), \quad (B.1)$$

$$\mu = \Sigma \alpha \mathbf{X}^T \mathbf{y}; \quad (B.2)$$

(ii)  $p(\lambda | \Theta - \{\lambda\}) \propto \prod_{k=1}^K \Gamma(\lambda_k | l_{1,k}, l_{2,k})$  with

$$l_{1,k} = \lambda_{1,k} + \frac{1}{2} \sum_{j=1}^p \delta(z_j = k), \quad (B.3)$$

$$l_{2,k} = \lambda_{2,k} + \frac{1}{2} \sum_{j=1}^p \delta(z_j = k) w_j^2; \quad (B.4)$$

(iii)  $p(\alpha \mid \Theta - \{\alpha\}) \propto \Gamma(a_1, a_2)$  with

$$a_1 = \alpha_1 + \frac{n}{2}, \quad (\text{B.5})$$

$$a_2 = \alpha_2 + \frac{1}{2}(\mathbf{y} - \mathbf{X}\mu)^T(\mathbf{y} - \mathbf{X}\mu); \quad (\text{B.6})$$

(iv)  $p(z_j \mid \Theta - \{z\}) \propto \text{mult}(\exp \rho_{j,1}, \dots, \exp \rho_{j,K})$  with

$$\rho_{jk} = -\frac{1}{2}w_j^2\lambda_k + \ln(\pi_k) + \frac{1}{2}\log \lambda_k; \quad (\text{B.7})$$

(v)  $p(\pi_k \mid \Theta - \{\pi\}) \propto \text{Dir}(d_k)$  with

$$d_k = \eta_k + \sum_{j=1}^p \delta(z_j = k). \quad (\text{B.8})$$

The algorithm is provided in pseudocode in Algorithm 2.

## Acknowledgment

The authors acknowledge support from the ANR Grant ViMAGINE ANR-08-BLAN-0250-02.

## References

- [1] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189–210, 1994.
- [2] S. Dehaene, G. Le Clec'h, L. Cohen, J. B. Poline, P. F. van de Moortele, and D. Le Bihan, "Inferring behavior from functional brain images," *Nature Neuroscience*, vol. 1, no. 7, pp. 549–550, 1998.
- [3] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex," *NeuroImage*, vol. 19, no. 2, pp. 261–270, 2003.
- [4] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, 2001.
- [5] J. D. Haynes and G. Rees, "Predicting the stream of consciousness from activity in human visual cortex," *Current Biology*, vol. 15, no. 14, pp. 1301–1307, 2005.
- [6] T. M. Mitchell, R. Hutchinson, R. S. Niculescu et al., "Learning to decode cognitive states from brain images," *Machine Learning*, vol. 57, no. 1-2, pp. 145–175, 2004.
- [7] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu, "Support vector machines for temporal classification of block design fMRI data," *NeuroImage*, vol. 26, no. 2, pp. 317–329, 2005.
- [8] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data," *NeuroImage*, vol. 28, no. 4, pp. 980–995, 2005.
- [9] S. J. Hanson and Y. O. Halchenko, "Brain reading using full brain support vector machines for object recognition: there is no face identification area," *Neural Computation*, vol. 20, no. 2, pp. 486–503, 2008.
- [10] O. Yamashita, M. A. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns," *NeuroImage*, vol. 42, no. 4, pp. 1414–1429, 2008.
- [11] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, "Sparse logistic regression for whole-brain classification of fMRI data," *NeuroImage*, vol. 51, no. 2, pp. 752–764, 2010.
- [12] F. De Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel, and E. Formisano, "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns," *NeuroImage*, vol. 43, no. 1, pp. 44–58, 2008.
- [13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [14] C. Chu, Y. Ni, G. Tan, C. J. Saunders, and J. Ashburner, "Kernel regression for fMRI pattern prediction," *NeuroImage*, vol. 56, no. 2, pp. 662–673, 2011.
- [15] H. Liu, M. Palatucci, and J. Zhang, "Blockwise coordinate descent procedures for the multi-task Lasso, with applications to neural semantic basis discovery," in *Proceedings of the 26th International Conference On Machine Learning (ICML '09)*, pp. 649–656, June 2009.
- [16] M. K. Carroll, G. A. Cecchi, I. Rish, R. Garg, and A. R. Rao, "Prediction and interpretation of distributed neural activity with sparse models," *NeuroImage*, vol. 44, no. 1, pp. 112–122, 2009.
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, Berlin, Germany, 1st edition, 2007.
- [18] M. Tipping, *The Relevance Vector Machine*, Morgan Kaufmann, 2000.
- [19] Y. Qi, T. P. Minka, R. W. Picard, and Z. Ghahramani, "Predictive automatic relevance determination by expectation propagation," in *Proceedings of the 21st International Conference on Machine Learning (ICML '04)*, ACM Press, 2004.
- [20] D. Wipf and S. Nagarajan, "A new view of automatic relevance determination," in *Advances in Neural Information Processing Systems*, vol. 20, pp. 1625–1632, MIT Press, 2008.
- [21] Y. Ni, C. Chu, C. J. Saunders, and J. Ashburner, "Kernel methods for fmri pattern prediction," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '08)*, pp. 692–697, 2008.
- [22] K. Uğurbil, L. Toth, and D. S. Kim, "How accurate is magnetic resonance imaging of brain function?" *Trends in Neurosciences*, vol. 26, no. 2, pp. 108–114, 2003.
- [23] K. Friston, C. Chu, J. Mourão-Miranda et al., "Bayesian decoding of brain images," *NeuroImage*, vol. 39, no. 1, pp. 181–205, 2008.
- [24] H. Steck and T. S. Jaakkola, "On the dirichlet prior and bayesian regularization," in *Advances in Neural Information Processing Systems*, vol. 15, pp. 697–704, 2002.
- [25] E. I. George and R. E. McCulloch, "Variable selection via gibbs sampling," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 881–889, 1993.
- [26] S. Geman and D. Geman, *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, Morgan Kaufmann, 1987.
- [27] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society. Series B*, vol. 67, no. 2, pp. 301–320, 2005.
- [28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

- [29] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.
- [30] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [31] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [32] scikit-learn, version 0.2, 2010, <http://scikit-learn.sourceforge.net/>.
- [33] E. Eger, C. A. Kell, and A. Kleinschmidt, "Graded size sensitivity of object-exemplar-evoked activity patterns within human LOC subregions," *Journal of Neurophysiology*, vol. 100, no. 4, pp. 2038–2047, 2008.
- [34] S. Chib and I. Jeliazkov, "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 270–281, 2001.
- [35] J. H. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 669–679, 1993.
- [36] R. E. McCulloch, N. G. Polson, and P. E. Rossi, "A Bayesian analysis of the multinomial probit model with fully identified parameters," *Journal of Econometrics*, vol. 99, no. 1, pp. 173–193, 2000.

## Research Article

# Development of a New Tool for 3D Modeling for Regenerative Medicine

**Filippo Mattoli,<sup>1</sup> Roberto Tiribuzi,<sup>1</sup> Francesco D'Angelo,<sup>1</sup> Ilaria di Girolamo,<sup>1</sup> Mattia Quattrocchi,<sup>1,2</sup> Simona Montesano,<sup>1</sup> Lucia Crispoltoni,<sup>1</sup> Vasileios Oikonomou,<sup>1</sup> Maria Gabriella Cusella De Angelis,<sup>3</sup> Peggy Marconi,<sup>4</sup> Antonio Orlacchio,<sup>5,6</sup> Maurilio Sampaolesi,<sup>2,3</sup> Sabata Martino,<sup>1</sup> and Aldo Orlacchio<sup>1</sup>**

<sup>1</sup> *Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Sezione di Biochimica e Biologia Molecolare, Università degli Studi di Perugia, Via del Giochetto, 06126 Perugia, Italy*

<sup>2</sup> *Translational Cardiology Laboratory, SCIL K.U. Leuven, 3000 Leuven, Belgium*

<sup>3</sup> *Dipartimento di Medicina Sperimentale, Sezione Anatomia Umana, Università di Pavia, 27100 Pavia, Italy*

<sup>4</sup> *Dipartimento di Medicina Sperimentale e Diagnostica, Sezione di Microbiologia, Università degli Studi di Ferrara, 44121 Ferrara, Italy*

<sup>5</sup> *Laboratorio di Neurogenetica, CERC-IRCCS Santa Lucia, Rome, Italy*

<sup>6</sup> *Dipartimento di Neuroscienze, Università di Roma "Tor Vergata", Rome, Italy*

Correspondence should be addressed to Sabata Martino, martinosa@unipg.it and Aldo Orlacchio, orly@unipg.it

Received 4 November 2010; Revised 4 March 2011; Accepted 6 April 2011

Academic Editor: Pingkun Yan

Copyright © 2011 Filippo Mattoli et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The effectiveness of therapeutic treatment based on regenerative medicine for degenerative diseases (i.e., neurodegenerative or cardiac diseases) requires tools allowing the visualization and analysis of the three-dimensional (3D) distribution of target drugs within the tissue. Here, we present a new computational procedure able to overcome the limitations of visual analysis emerging by the examination of a molecular signal within images of serial tissue/organ sections by using the conventional techniques. Together with the 3D anatomical reconstitution of the tissue/organ, our framework allows the detection of signals of different origins (e.g., marked generic molecules, colorimetric, or fluorimetric substrates for enzymes; microRNA; recombinant protein). Remarkably, the application does not require the employment of specific tracking reagents for the imaging analysis. We report two different representative applications: the first shows the reconstruction of a 3D model of mouse brain with the analysis of the distribution of the  $\beta$ -Galactosidase, the second shows the reconstruction of a 3D mouse heart with the measurement of the cardiac volume.

## 1. Introduction

Regenerative medicine-based applications represent a promising therapeutic approach for diseases with degenerative tissues as hallmarks (i.e., Alzheimer's disease, lysosomal storage disorders, and some cardiac pathologies). To this end, innovative combined gene transfer/stem cell implantation strategies are advanced in order to reestablish the genetic defect as well as the damaged tissues/organs [1–7]. Parallel effort is also made on the development of new tools able to investigate the tissues/organs architecture after treatment [8–13]. Thus, the three-dimensional (3D) computational recon-

struction of anatomical tissues/organs following surgery represents one of the necessary analytical instruments.

The most common technique for creating such models is based on the 3D reconstruction from serial cross-section images collected through several conventional techniques (Computerized Tomography, Positron Emission Tomography, Magnetic Resonance Imaging, 3D ultrasound, and X-ray) as well as by synchrotron radiation or diffraction-enhanced imaging [12–17]. However, these techniques do not supply information on the macromolecular composition in the image. Conversely, spectroscopy-based techniques (i.e., 3D IR-imaging) present the advantage of underlying

the image macromolecular composition directly [18–20], but, because of the lack of penetration of mid IR radiation into the tissue, this method precludes a real-time imaging of whole samples. Advancing in computational technology (open source library (e.g., <http://www.eecs.tufts.edu/~alauri02/install.htm>) [20, 21] as well as open source software (e.g., <http://www.fas.org/dh/>)) support the overall imaging analysis procedures and the development of new tools of imaging investigations. In this contest, progress will come from computational methods able to detect target molecules of various compositions within tissues/organs.

Addressing this issue, we present a simple and usable framework that does not require sophisticated or expensive apparatus. The procedure integrate 2D images obtained by collecting serial tissue/organ slices into a 3D computational model but is finalized to analyze the presence and distribution of target molecules as well as other tissue/organ parametric characteristics. The main advantage of our method is that it is applicable to postmortem analysis of samples processed in every experimental condition. The application resolves some limitations of the above described techniques since together with the anatomical structure information of the reconstructed tissue/organ the method allows the real-time detection of target macromolecules.

Here, we reported two different applications of our computational procedure. The first consists in the generation of a 3D brain model using C57/BL6 mice after *in vivo* gene transfer with HSV-T0Z herpes simplex viral vector. Mainly, we present the reconstruction of the 3D distribution of the transgene within the brain. The second consists in the generation a 3D model of C57/BL6 mouse heart in order to measure the cardiac volume values.

## 2. Materials and Methods

**2.1. Materials.** C57/BL6 mice were from Charles River, Italy. The 5-bromo-4-cloro-2-indlyl- $\beta$ -D-galactopiranoside (X-Gal) was obtained from Sigma Chemical Co. The medium for tissue culture was from Euroclone, Celbio Lab., fetal calf serum was from Mascia Brunelli, penicillin/streptomycin was from Gibco BRL. All reagents used in this study were of analytical grade.

**2.2. HSV-T0Z Viral Vector Direct Injection into the Mice Brain.** One group of 5-month-old animals was injected with a dose of herpes simplex viral vector encoding for the  $\beta$ -Galactosidase enzyme ( $5 \times 10^6$  PFU, HSV-T0Z) into the internal capsule of the left brain hemisphere of the mice as previously described [22]. Mice were anesthetized with 0.02 mL/g body weight of 2,2,2-tribromoethanol and 2-methyl-2-butanol and placed on the Styrofoam platform of a stereotaxic injection apparatus (David Kopf Instruments, Tujunga, Calif, USA). The skull was exposed following a 10 mm incision in the midline. The injection coordinates for the internal capsule were  $-0.34$  mm to bregma, 1.4 mm mediolateral, and 3.8 mm depth. These coordinates were chosen in order to minimize vector leakage into the ventricular space. Each injection was 5  $\mu$ L total, and the injection

speed was 0.1  $\mu$ L/min. The injections were carried out using a needle capillary (1.2 mm  $\times$  0.6 mm) attached to a Hamilton syringe. The injections were delivered at a rate of 0.1  $\mu$ L/min, and the needle was slowly withdrawn after an additional 5 minutes. The scalp was closed by suture.

All procedures were performed according to protocols approved by an internal animal care and use committee and were reported to the Ministry of Health, as per Italian law.

**2.3. Brain Serial Section Preparation.** One month after injection, mice were sacrificed by cardiac perfusion. The left ventricle was cannulated, an incision was made in the right atrium, and the animals were perfused with 2% paraformaldehyde in PBS until the outflow ran clear, then the brain was included in ornithine carbamoyl transferase (O.C.T. compound, Tissue-Tek, Sakamura, The Netherlands) after exposure to 5%–30% glucose gradient and finally sectioned on a cryostat into 15- $\mu$ m-thick serial sections.

We collected brain serial sections in four series of slides (A-B-C-D), so that: section 1 on slide B3 was collected immediately after section 1 on slide A3 and immediately before section 1 on slide C3. Thus, staining only 1/4 of the sections (A), we checked the beta-gal staining distribution (1 section every 60  $\mu$ m) along the whole brain extension.

After perfusion, whole spines were removed, and after decalcification in 3% Trifluoroacetic acid (Merck), spinal cords, surrounded by vertebrae and remains of skeletal muscles, were cut into blocks containing a known number of vertebrae (four or five). Each block was sectioned on a cryostat into 10  $\mu$ m serial sections.

Animal experimentation protocols were approved by the Italian ethical committee.

**2.4. Galactosidase Analysis.**  $\beta$ -Galactosidase activity was assayed through the histological substrate X-Gal as previously described [22].

**2.5. C57/BL6 Hearts Sectioning.** For histological heart analysis, C57/BL6 mice were sacrificed with an intravenous bolus injection of saturated potassium chloride, aiming at inducing cardiac arrest at systole, and the hearts were rapidly excised for fixation in buffered formaldehyde (4%), washed in PBS and quickly frozen in OCT. Serial heart sections were fixed with 4% PFA, and processed for hematoxylin-eosin staining according to standard procedures. Images were taken with a S100 TV microscope (Carl Zeiss MicroImaging Inc.) or with deconvolution microscope DeltaVision RT for the panelling image.

**2.6. Development of Routines.** The routines were developed in Visual Basic 6 and applied on images of format “.bmp” acquired by microscope.

**2.6.1. To Reduce Image Soils and Glares.** The routine clears images, transforms to black colour (RGB 0,0,0) any pixel with intensity of red/green/blue comprised between the values chosen by customers. Customer can select values by

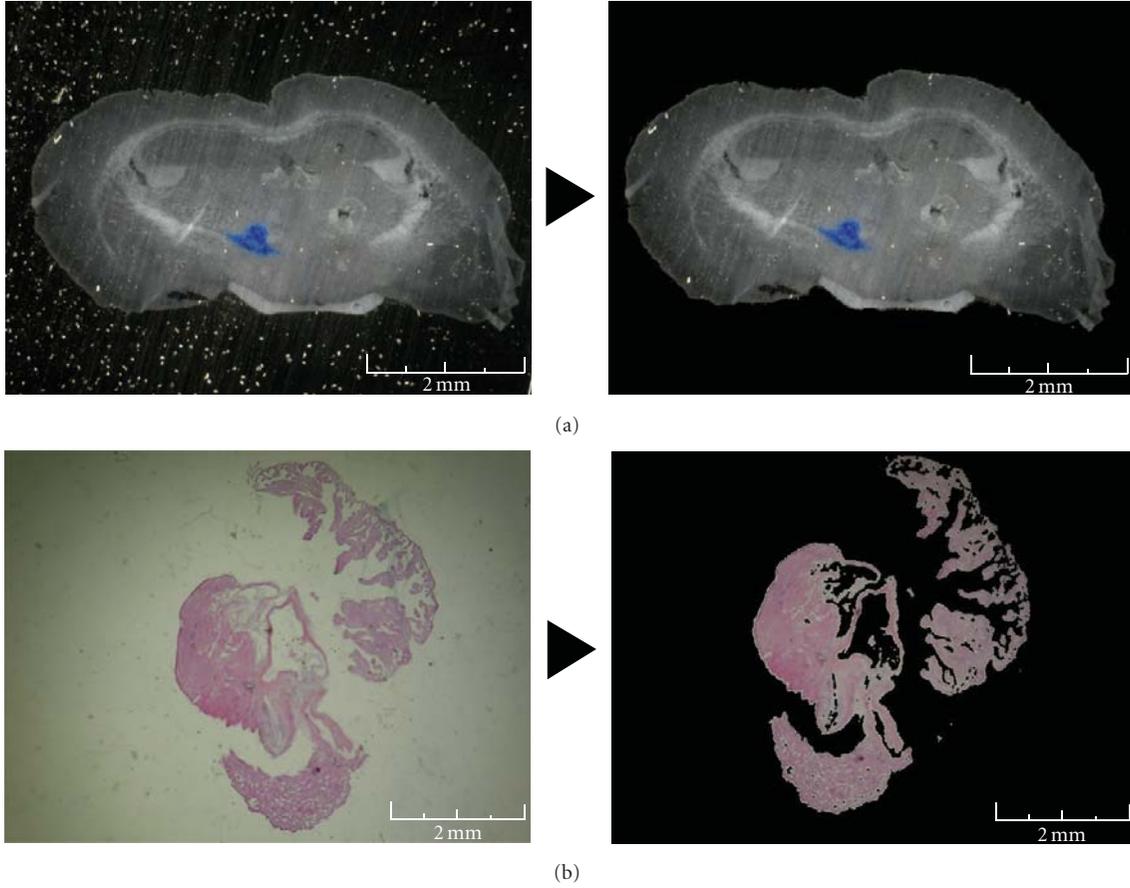


FIGURE 1: Soil and glare reduction. Soils and glares have been removed in order to have a clear vision of the final model. (a) We have on the left the original image of the section of brain tissue, and on the right the image after soil and glare reduction. (b) The same procedure has been used for heart section images.

dragging windows in tissue and background areas. The software allows separating foreground from background. In particular, salt/pepper noise may be eliminated through the selection of colour levels adequately. The output images will contain only the tissue colour range (Figure 1). We did not modify the images. We have always converted the background into black colour. This step is based on the selection of the colour tonality that is already part of the background.

**2.6.2. To Centre the Image.** The routine centres the section area in the image elaborating each file by moving the selected area from the total amount:

$$\frac{\sum_n^1 x_n}{n} - x_0, \quad \frac{\sum_n^1 y_n}{n} - y_0. \quad (1)$$

“ $n$ ” is the number of not black pixels (not RGB 0,0,0), “ $x_n$ ” and “ $y_n$ ” are coordinates of these pixels, “ $x_0$ ” and “ $y_0$ ” are the centre of bitmap coordinates.

The centering step does not require foreground center of each slice. In fact, even in the case where the slice section is deformed, the method guarantees a correct analysis (Figure 2).

**2.6.3. To Orientate the Image.** The routine orients the image of each section with respect to the adjacent section by rotating the tissue area of the bitmap  $360^\circ$ , and for every fraction of rotation calculates the difference “ $d$ ”:

$$d = \sum_n^1 (i_n - k_n). \quad (2)$$

“ $i$ ” is the intensity of the pixel “ $n$ ” of the first image and “ $k$ ” is the intensity of the pixel “ $n$ ” of the previous image. Notably, the section orientation may be validated by comparing the anatomic tissue structure of the slice section using specialized mouse atlas as reference (e.g., <http://www.bnl.gov/CTN/mouse/>).

The image chosen is the one with the smallest value of “ $d$ ” (Figure 3).

**2.6.4. To Reduce Deformation.** To reduce image deformation, we used an algorithm that extracts the contour of each section. This algorithm divides the image in an amount of slices depending on the image size. Each slice is a line of pixels from the centre to the edge of image with the concentric lines that can cover the whole  $360^\circ$  degrees of picture. The contour

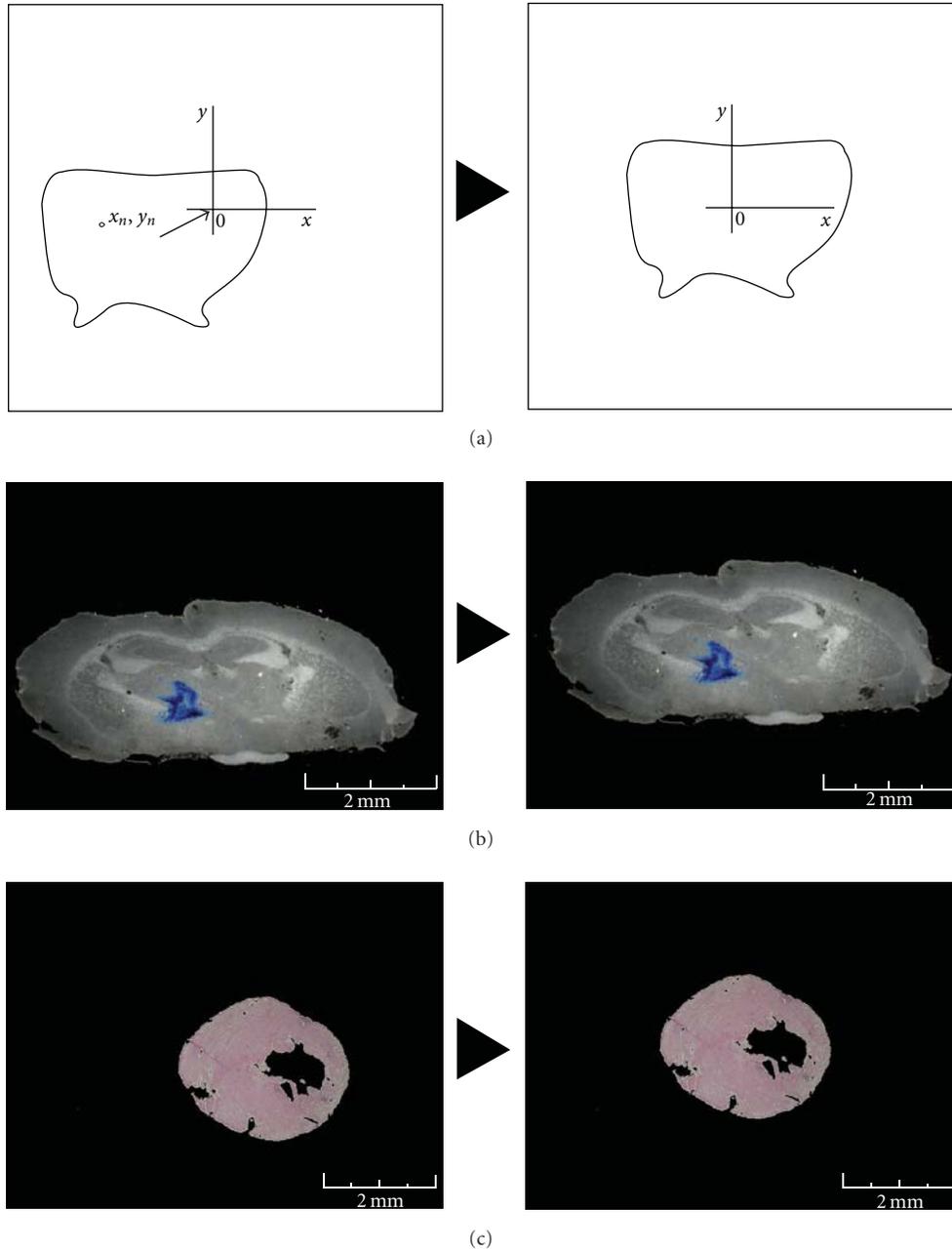


FIGURE 2: Woven image centre. (a) Outline (the reported image doesn't represent any specific tissue section) of the image centring routine (see routine 1).  $x_n$  and  $y_n$  are coordinates of not black pixels (not RGB 0,0,0),  $x_0$  and  $y_0$  are the centre of the bitmap coordinates. The program calculates a kind of barycentre of the woven area in the image and moves it to the centre of the bitmap. (b) Example of the image centring procedure: left panel is an original image of the coronal serial sections, right panel is the same image after the automatic woven centring. (c) Example of the image centring procedure for heart section images.

pixels are defined by checking each pixel for each line from the border of picture toward the centre. The first not black pixel is taken as contour pixel. Then, for each angle shot is calculated the distance between the contour pixel and the centre of the bitmap. These distances are annotated for each sequential section. Then, the discontinuity between the sections is diminished by the calculation of the average of the

contour pixels with the same angle in the adjacent pictures. The new correct distance of any contour pixel will be

$$L_n = \frac{l_n + l_{n+1}}{2}. \quad (3)$$

" $n$ " is the progressive number of the section and " $l$ " is the distance between the contour pixel and the centre of

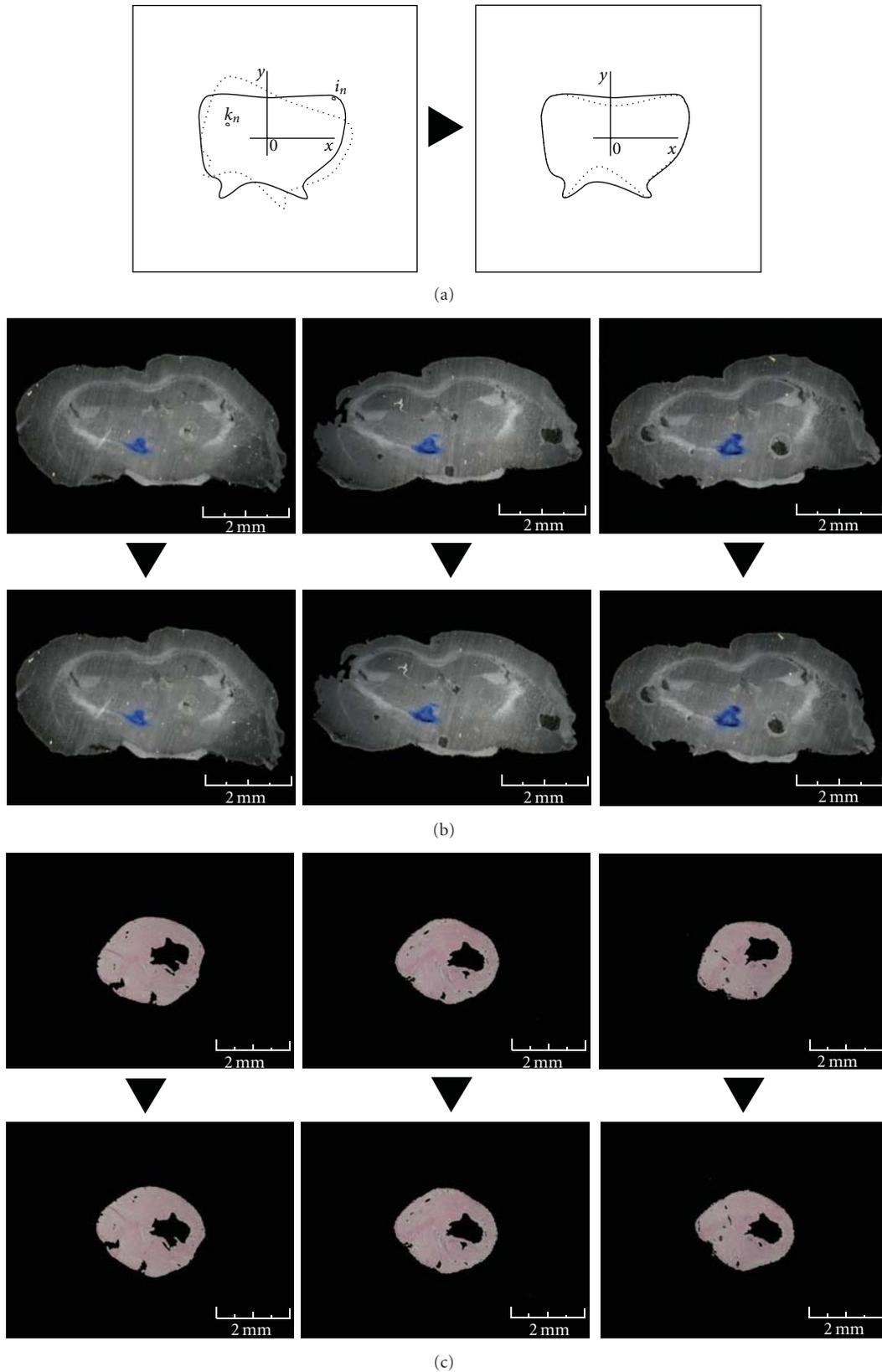


FIGURE 3: Woven image orientation. (a) Outline of the image orienting routine (see routine 2). The images of the woven area were oriented regarding the adjacent slides by the routine that rotates the woven area of the bitmap 360°, then it chooses the angle where the differences between the adjacent images are the smallest, and hence, the images are the most similar.  $i$  is the intensity of the pixel number of the first image and  $k$  is the intensity of the pixel  $n$  of the previous image. The image chosen is the one with the smallest value of  $d$ . (b) Example of the image-orienting procedure: In the panels we can see two of the images obtained from the brain serial sections that are rotated with respect to the adjacent image. (c) The same procedure has been used for the heart section images.

the bitmap for the angle “ $\alpha$ ”. With the distance of the contour pixels, the internal area pixels of the section are also adjusted by a proportional linear correction (Figure 4).

Absence of image contour could be recognized using image from mouse brain atlas atlas that display tissue sections identical to our.

*2.6.5. To Reconstruct 3D Model and Calculate Tissue Volume.* The model has been assembled in a file containing a three-dimensional matrix. For the correct dimensioning of the model, we need the relative dimensions “ $x$ ” and “ $y$ ” of a pixel and its depth “ $z$ ”. To this aim, we have calculated the depth of the pixels covering the distances between the sections by carrying out the right proportion between the real dimensions of the tissue section and the pixel dimensions of the tissue area in the picture as follows:

$$X_{\text{rel}} = Y_{\text{rel}} = \frac{L_w}{n_x} = \frac{H_w}{n_y}, \quad (4)$$

“ $X_{\text{rel}}$ ” and “ $Y_{\text{rel}}$ ” are the relative dimensions (height, width) represented by one pixel, “ $L_w$ ” and “ $H_w$ ” are the effective dimensions of a part of the tissue, and “ $n_x$ ” and “ $n_y$ ” are the number of pixels that represent these parts of tissue.

$$Z_{\text{rel}} = X_{\text{rel}} \cdot d_s, \quad (5)$$

“ $Z_{\text{rel}}$ ” is the real depth represented by one pixel and “ $d_s$ ” is the effective distance between two sequential sections.

The measure of volume has been carried out by calculating each not black voxel (i.e., every voxel representing a part of tissue).

To visualize and navigate the 3D models, we loaded the files containing data in software allowing the volume visualization of 3D medical images. The software used is the demo version of VolView 2.0 produced by Kitware.

We did not perform data interpolation. In our experimental condition, the isinterpolation 3D volume doesn’t increase the precision since the error order of the interpolation of the volume is much lower than the order error occurring during the reconstruction of the model from dissected slice sections.

*2.6.6. Isolation of a Molecular Signal into a 3D Model.* In order to view only the pixels that represent the  $\beta$ -Galactosidase (X-Gal = blue signal), we selected the areas where the blue component of the intensity RGB of the pixels are greater than the other components for a value defined by the customer.

### 3. Results and Discussion

We developed a computational procedure for the generation of a 3D model starting from *postmortem* tissues/organs serial slices in which it is possible to analyze and highlight target molecular signals as well as anatomical areas of interest.

*3.1. Routines Development.* The total numbers of serial slices processed were 113 for the brain and 100 for each heart,

taken from coronal brain serial sections of treated mice and heart serial sections of wild-type mice. Before building the 3D model, we have optimized procedures consisting of four steps.

(i) The first manipulation consisted of soil and glare deletion. The intensity of the pixels outside of the woven area in the slides was selected, and these pixels were transformed into black colour. All sections with strong background noise were excluded from the analysis. However, salt/pepper noise may be eliminated through the selection of colour levels adequately (Figure 1).

(ii) In the second step, images of each serial slices were centred with respect to the total area of the bitmap. The routine calculates the barycentre of the section area in the image and moves it to the centre of the bitmap (Figure 2). This step generates a compacted model that allows the alignment of the molecular signal (e.g., X-Gal marks within the brain model) and the accurate reconstruction of the tissues/organs (e.g., measure of the heart volume).

(iii) In the third step, each image from tissue slices was oriented in comparison with the adjacent image by a routine that rotates the section area of the bitmap  $360^\circ$ . For each fraction of rotation, the application calculates and memorizes the summary of the difference between the intensity of each pixel in the image with each pixel in the adjacent image. Finally, the routine chooses the image in which the tissue area is oriented with the best angle for the alignment between the tissues in the adjacent images, that is, the angle with the smallest summary memorized (Figure 3). These elaborations were necessary in order to eliminate the inaccuracy generated during the phase of fixing brain or heart slice sections.

To evaluate the alignment/rotation error, we performed comparative analyses, elaborating brain sections acquired by MRI. The images were downloaded by “The centre of translational neuroimaging” web site, and the model downloaded was “3D MRI Digital Atlas database of an adult C57BL/6J mouse brain” (<http://www.bnl.gov/CTN/mouse/>). The tissues in the images scanned by MRI were moved and rotated randomly, then, using our application, we have processed these pictures in order to match them with originals.

Comparative analysis indicated that the inaccuracy was just limited to the vertical curvature of the posterior portion of brain that unaffected the correct reconstruction of the model, even if small error occurs, those errors could at least generate tiny modification of the volume shape but did not interfere with the analysis.

Thus, our application allowed maintaining the structure of the brain with the attack of the spine.

(iv) The last step was the correction of the deformation caused by the technical slices preparation. Based on the anatomical structure of tissues, the sections deformations were reduced by the comparison of each section with the adjacent sections. Notably, our method allows a correct analysis of data even tissue/organ are not symmetric. As reported above, modification of the volume (such as shape distortion) did not interfere with the analysis.

In respect to the active contour model [23, 24], the algorithm that we used extracted the contour of each section

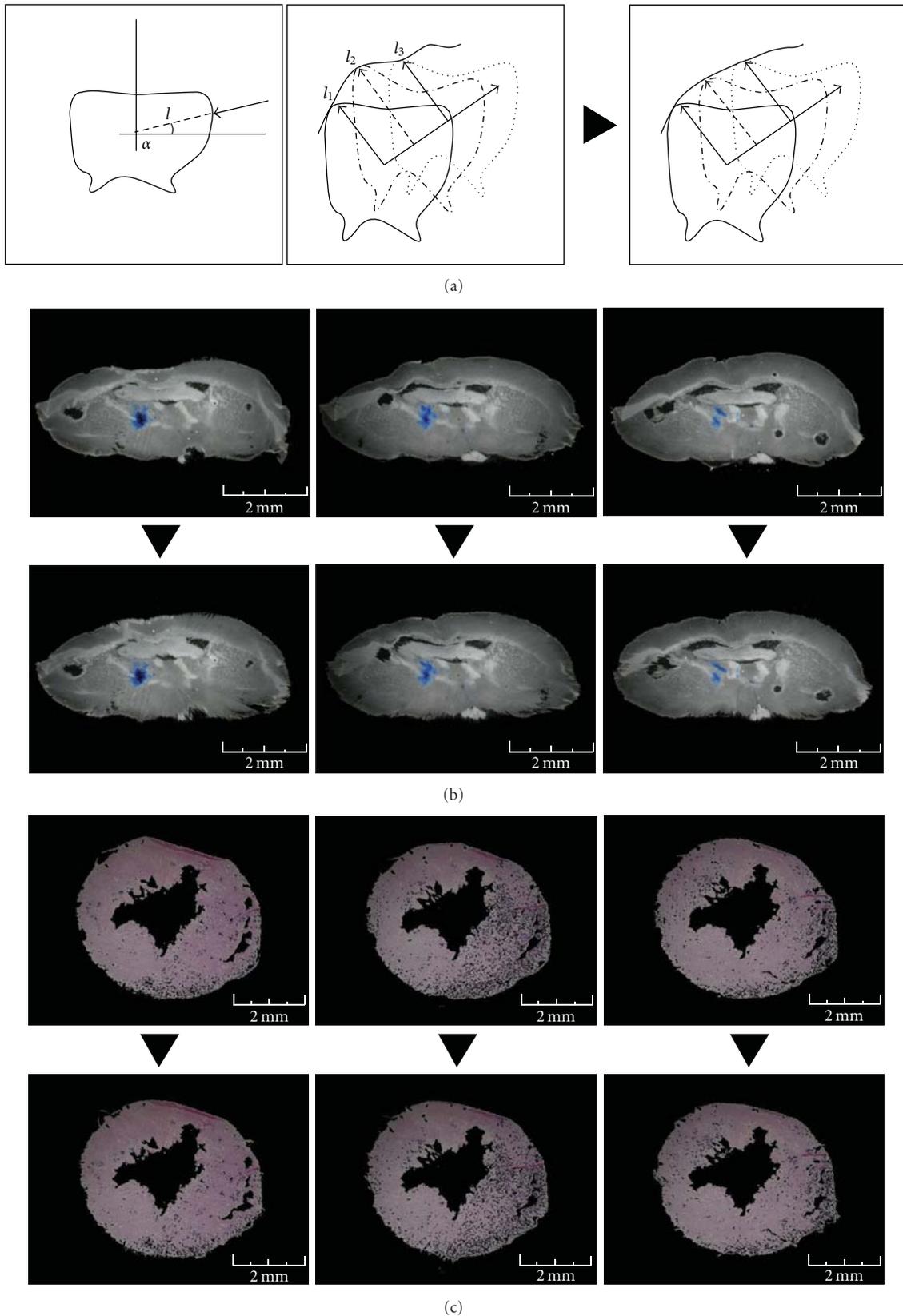


FIGURE 4: Reduction of the tissue deformation. (a) Outline of the tissue deformation routine (see routine 3). The discontinuity between the sections is diminished by the calculation of the medium of the contours.  $l_n$  is the distance between the contour pixel and the centre of the bitmap of the  $n$  serial section for the angle  $\alpha$ . With the distance of the contour pixels, the internal area pixels of the section are also adjusted by a proportional linear correction. (b) The images represent three original sequential sections of the brain. In the images below, the deformations have been reduced in order to carry out sequential images which are more homogenous. (c) The images represent three original sequential sections of the heart.

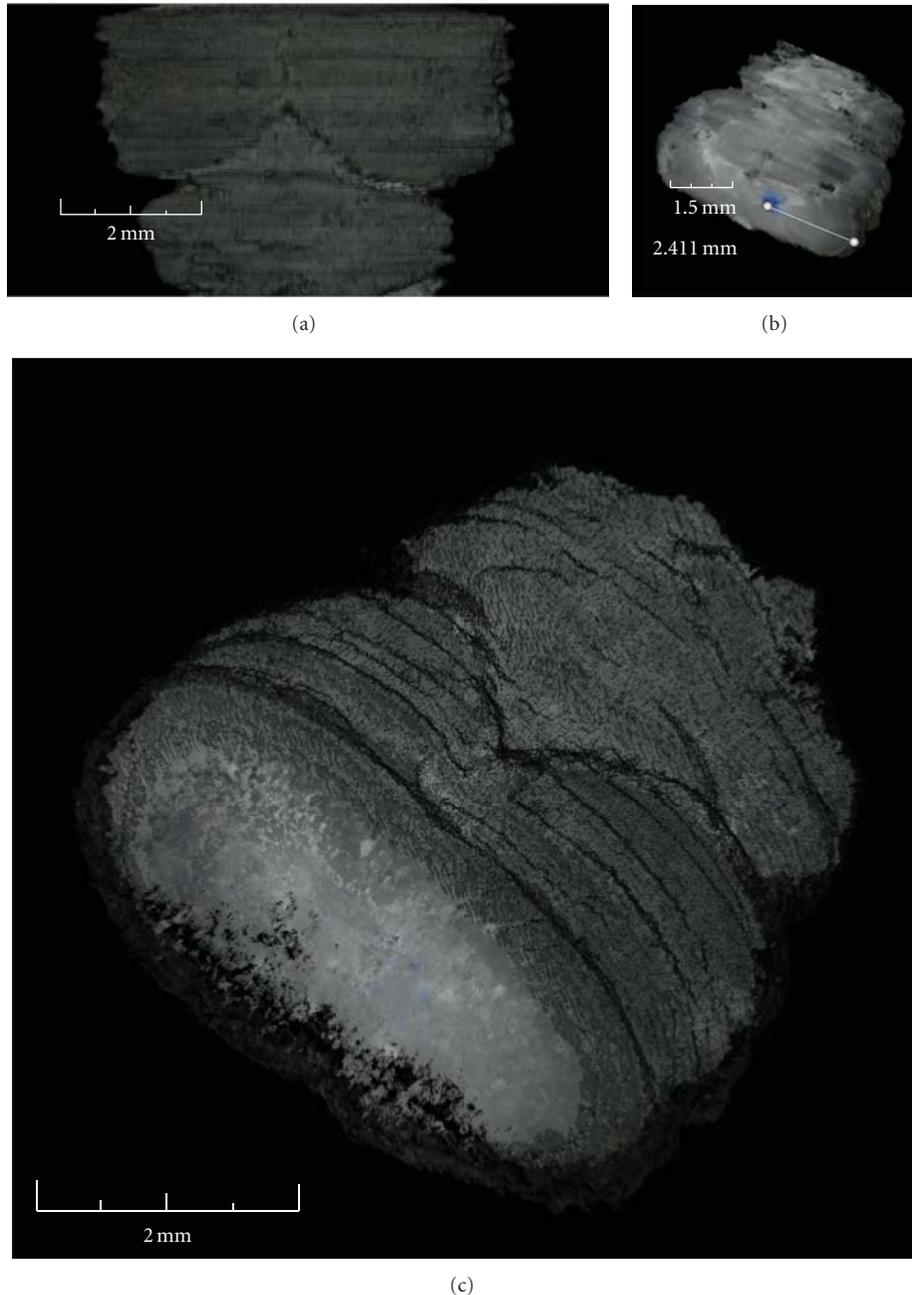


FIGURE 5: 3D views of the brain model. (a) Superior view of the model. (b) Isometric view of the horizontal section of the brain model. In this image we can see the internal structure of the mouse brain and the distance from the external of the brain to the injection point. (c) Isometric view of the full model.

and memorized for each degree the distance between the centre and the first not null-pixel by the control from the outside towards the centre of the bitmap. This procedure was replicated for each image. In this way, the surface of model, composed by the contours of the serial images, is memorized and can be smoothed by calculating the average of the contours (Figure 4). Moreover, for the correct reconstruction of the models, the distances of the sections compared to the width of the pixel were calculated.

*3.2. Assembly and Model Navigation.* The 3D model can be visualized and highly manipulated by a volume visualization software. It can be magnified, moved, and turned on three axes. The zones of greater interest can be isolated, measured, and analyzed at molecular level (Figure 5).

Here, we report two different examples of this computational procedure. The first consists in the generation of a 3D brain model using C57/BL6 mice injected with HSV-T0Z herpes simplex viral vector. The second consists in

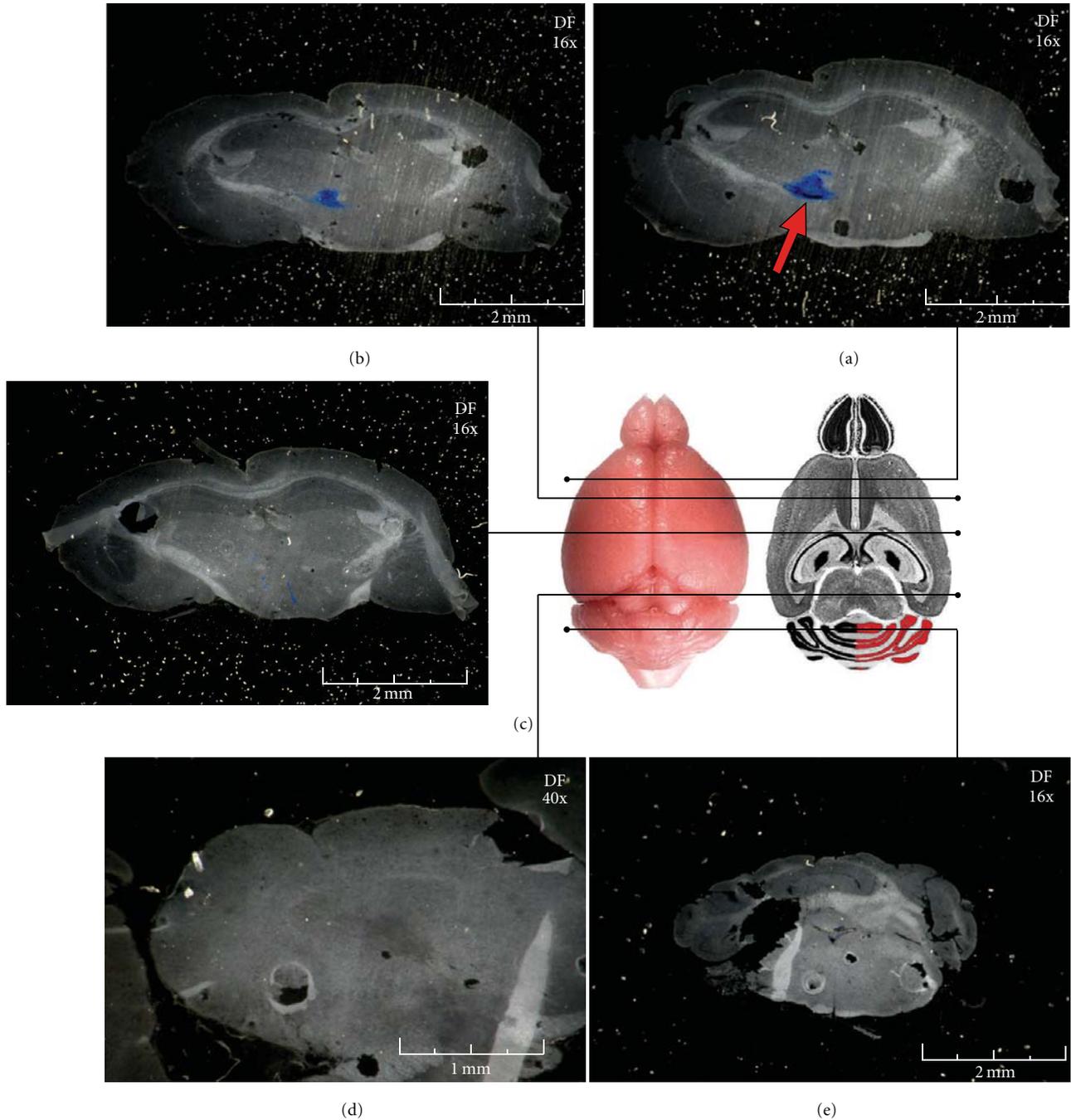


FIGURE 6: Representative sections of HSV-T0Z distribution in the mouse central nervous system. Serial brain sections were produced dissecting animals in coronal orientations (a–e). Here, we show a part of the representative coronal sections. Sections were stained with the X-Gal substrate (blue signal) as described in method paragraph. In the dark field (DF) images are indicated the magnification and measurement bars. Point of injection is shown (red arrow) into image (a) on the left hemisphere close to bregma line. In (e) there is a representative section of cerebellum.

the generation a 3D model of mouse heart. We have used this model to measure the heart volume value.

3.3. *Generation of 3D Model of Mouse Brain after Gene Transfer Approach.* We used HSV-T0Z, a nonreplicating

herpes simplex viral vector reporting the  $\beta$ -Galactosidase gene [22, 25]. HSV-1 has the ability to infect a wide variety of cell types in the nonreplicating phase, for example, neurons, as well as the intrinsic capacity to be transported in a retrograde manner to motor and/or sensory neuronal

cell bodies following peripheral inoculation [22, 25, 26]. We injected the vector into the internal capsule of the left brain hemisphere. We designed an experimental plan composed of two groups of five-month-old mice. Group 1, C57/BL6 + HSV-T0Z,  $5 \times 10^6$  total PFU; group 2, untreated C57/BL6 mice. The animals were sacrificed after 72 h and 1 month, and the brains were sectioned on serial slices by a cryostat. The slices were analysed by a NIKON-Eclipse-TE2000 microscope equipped with an Olympus F-View camera. The coronal serial slices images were used to compose the 3D virtual brain model as above described (see video S1 and video S2 in Supplementary Material available online at doi:10.1155/2011/236854).

We analysed the viral vector distribution by monitoring the X-gal staining in coronal, transversal, and sagittal high-throughput brain serial sections. New results confirmed our previous work with a wide viral vector spreading in both injected and uninjected hemispheres (Figure 6). Our computational method allowed a more accurate step-forward analysis of the distribution of those stained proteins.

First, we isolated the molecular signal by the selection and the amplification of its levels of tonality (Figure 7). The model allowed the detection of the  $\beta$ -Galactosidase stained (the tonality of X-Gal staining is blue and always greater than the general tonality of the brain section which is grey) and the identification of the signal position in all the brain areas. The 3D model allows to appreciate the high distribution of the HSV-T0Z viral vector in brain area far away from the site of injection, in both treated and untreated hemisphere and cerebellum. Notably, the model allowed following the viral vector within the brain after administration over time (Figure 7; video S1). Further, in this model, it is possible to merge the signal with the brain anatomical structure allowing the investigation of putative mechanisms involving the transgenic distribution (video S1, video S2).

**3.4. Generation of 3D Model of Mouse Heart and Measurement of Tissue Volume.** To be useful, a 3D model of mouse heart has to recreate the cardiac tissue in physiologic and pathologic state or after therapeutic treatment (i.e., cell transplantation). The current echocardiography is not always comprehensive. Here, we showed that report an example of our method allowed the generation of a 3D mouse heart model (Figure 8) and permitted the measure of its volume. This parameter is particularly relevant for evaluating morphological changes during aging and also for monitoring the clinical progression of cardiac disease in murine animal models [27–29].

The measure of heart tissue volume has been carried out by the calculation of every voxel representing tissue after an accurate evaluation of the virtual voxel volume with respect of real tissue dimensions. We found that the total heart volumes were  $65.5 (\pm 0.3) \text{ mm}^3$  and  $84 (\pm 0.4) \text{ mm}^3$ , the total ventricle chamber volumes were  $5.2 \pm 0.1 \text{ mm}^3$  and  $7.3 \pm 0.2 \text{ mm}^3$ , and the total ventricle volumes were  $58.3 \pm 0.2 \text{ mm}^3$  and  $79.8 \pm 0.4 \text{ mm}^3$ , in female and male C57/BL6 mice, respectively. The data were consistent in terms of

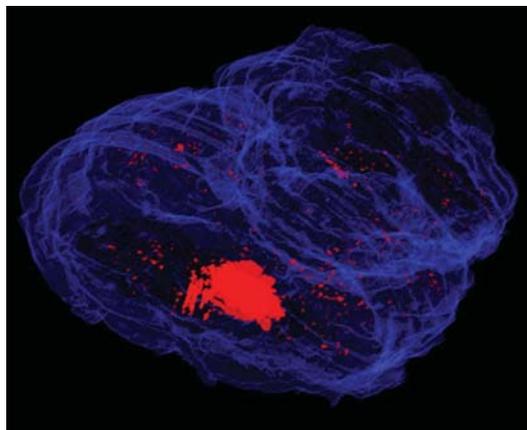


FIGURE 7: Isolation of the X-Gal staining. An isometric view with the distribution of the X-Gal staining (in red) is shown, highlighting a magnification of the signal. The model has been rendered with a demo version of VolView 2.0 produced by Kitware, USA.

values, and the difference between male and female values was statistically significant.

**3.5. Conclusion.** We have developed a new computational procedure able to overcome the limitations of visual analysis emerging by the examination of a molecular signal within images of serial tissues/organs sections by using the conventional techniques (Computerized Tomography, Positron Emission Tomography, Magnetic Resonance Imaging, 3D ultrasound, X-ray, and 3D IR imaging). Despite the high-quality performance in evaluating samples for analytical medicine, these techniques fall on the analysis at molecular level of therapeutics compound (e.g., generic drugs and recombinant proteins). In our case, the method that we developed represents the simplest way to follow the three-dimensional distribution of gene product in an *in vivo* gene transfer approach or to measure volume of an organ such as the heart. The only limitation of this method could be in the state of the original tissue slices. In fact, much damaged brain slices must be discarded. However, due to the availability of adjacent sections to the disrupted slices, the relative loss of data in the final model is minimized.

The 3D position of these molecules as well as their distribution from the site of injection within the tissue/organ architecture represents essential parameters for developing efficacious clinical approaches and, also, for therapeutic drugs design. The main characteristic of our method is in the possibility of having a 3D-imaging apparatus where other classical instruments are not applicable. This procedure is suitable for every condition of imaging analysis postmortem and without particular, sophisticated, or expensive instruments.

However, on the basis of the easy application of the procedure and mostly for the high-quality information that we obtained, we consider our method to be a simple and valid 3D-modeling instrument for studying the expression and localization of stained cells/proteins/genes/RNA/drugs

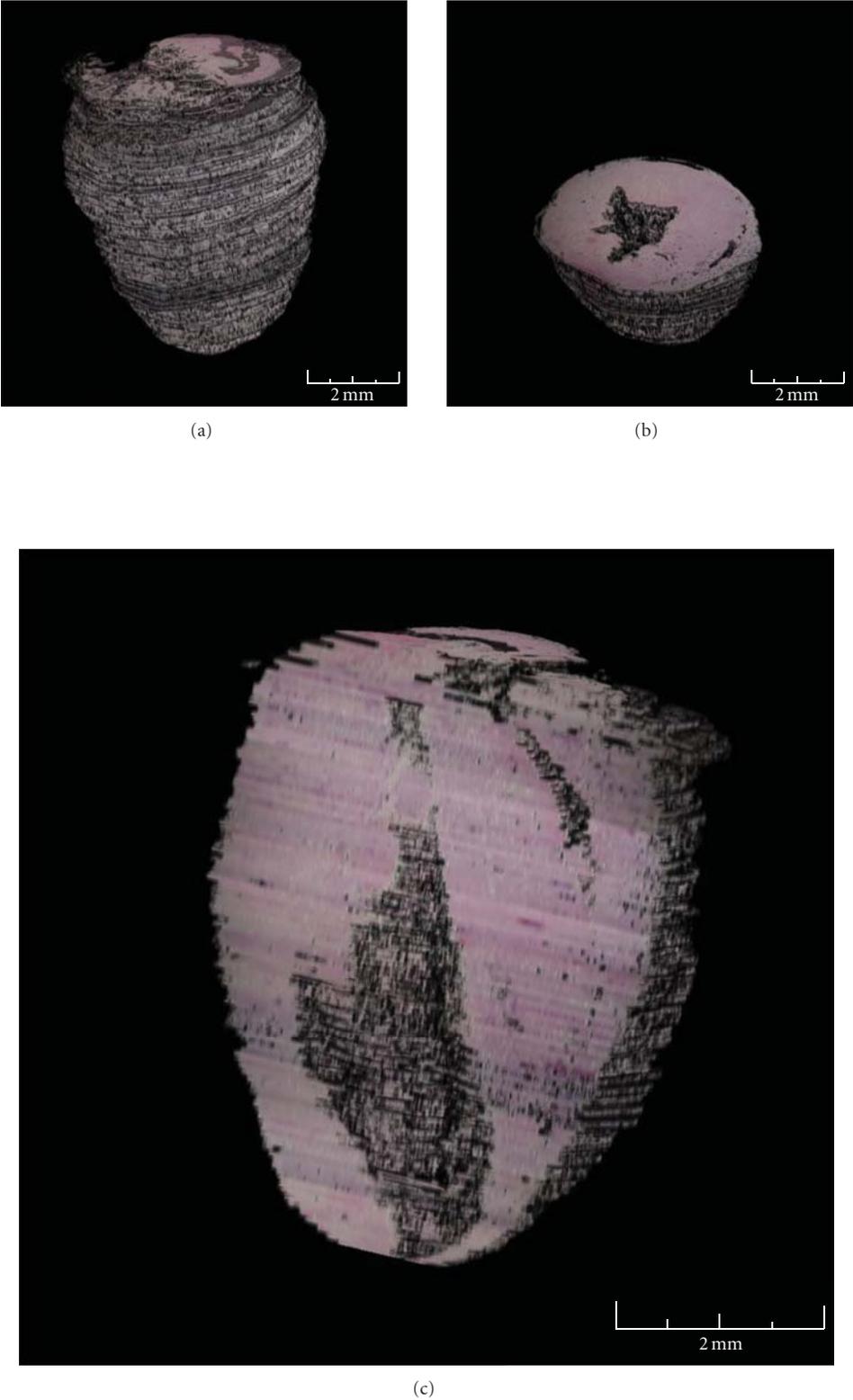


FIGURE 8: 3D views of the heart model. (a) Isometric view of the whole heart model. (b) Isometric view of the horizontal section of the heart model. (c) Isometric view of the vertical section of the heart model. In this image, we can see the ventricles of the mouse heart and the thickness of the heart walls.

as well as the tissue architecture, even in the presence of biomaterials for regenerative medicine.

## Conflict of Interests

The authors declare that there is no conflict of interests.

## Acknowledgments

This study was supported by the Italian *Ministero della Salute* [Grant no. RF-UMB-2006-339457 to Antonio Orlacchio. (REG.17O) and Aldo Orlacchio], the *Comitato Telethon Fondazione Onlus*, Italy (Grant no. GGP10121 to Antonio Orlacchio), the Italian *Ministero dell'Istruzione, dell'Università e della Ricerca* (PRIN 2008—Grant nos. 020903052 to Antonio Orlacchio and 20084XRSBS.001 to Aldo Orlacchio), the *Università di Roma "Tor Vergata"*, Italy (FAA 2008—Grant no. 020903023 to Antonio Orlacchio), the *Fondazione Cassa di Risparmio di Perugia* (Grant no. 2009.020.0050 to Aldo Orlacchio), and the *INBB* (to Aldo Orlacchio). We thank Hilary Giles (MA) for language advice and proof reading.

## References

- [1] S. Martino, F. D'Angelo, I. Armentano et al., "Hydrogenated amorphous carbon nanopatterned film designs drive human bone marrow mesenchymal stem cell cytoskeleton architecture," *Tissue Engineering. Part A*, vol. 15, no. 10, pp. 3139–3149, 2009.
- [2] S. Martino, I. di Girolamo, A. Orlacchio, A. Datti, and A. Orlacchio, "MicroRNA implications across neurodevelopment and neuropathology," *Journal of Biomedicine and Biotechnology*, vol. 2009, Article ID 654346, 13 pages, 2009.
- [3] S. Martino, I. di Girolamo, C. Cavazzin et al., "Neural precursor cell cultures from GM2 gangliosidosis animal models recapitulate the biochemical and molecular hallmarks of the brain pathology," *Journal of Neurochemistry*, vol. 109, no. 1, pp. 135–147, 2009.
- [4] M. Quattrocchi, M. Cassano, S. Crippa, I. Perini, and M. Sampaoli, "Cell therapy strategies and improvements for muscular dystrophy," *Cell Death and Differentiation*, vol. 17, no. 8, pp. 1222–1229, 2010.
- [5] A. Lattanzi, M. Neri, C. Maderna et al., "Widespread enzymatic correction of CNS tissues by a single intracerebral injection of therapeutic lentiviral vector in leukodystrophy mouse models," *Human Molecular Genetics*, vol. 19, no. 11, pp. 2208–2227, 2010.
- [6] B. Gentner, I. Visigalli, H. Hiramatsu et al., "Identification of hematopoietic stem cell-specific miRNAs enables gene therapy of globoid cell leukodystrophy," *Science Translational Medicine*, vol. 2, no. 58, p. 58ra84, 2010.
- [7] F. D'Angelo, I. Armentano, S. Mattioli et al., "Micropatterned hydrogenated amorphous carbon guides mesenchymal stem cells towards neuronal differentiation," *European Cells & Materials*, vol. 20, pp. 231–244, 2010.
- [8] S. J. Ferguson, J. T. Bryant, and K. Ito, "Three-dimensional computational reconstruction of mixed anatomical tissues following histological preparation," *Medical Engineering and Physics*, vol. 21, no. 2, pp. 111–117, 1999.
- [9] J. S. S. Wu and J. H. Chen, "Clarification of the mechanical behaviour of spinal motion segments through a three-dimensional poroelastic mixed finite element model," *Medical Engineering and Physics*, vol. 18, no. 3, pp. 215–224, 1996.
- [10] S. Kobayashi, M. Tateno, K. Utsumi et al., "Quantitative analysis of brain perfusion SPECT in Alzheimer's disease using a fully automated regional cerebral blood flow quantification software, 3DSRT," *Journal of the Neurological Sciences*, vol. 264, no. 1-2, pp. 27–33, 2008.
- [11] T. Auer, A. Schwarcz, T. Doczi, K. D. Merboldt, and J. A. Frahm, "A novel group analysis for functional MRI of the human brain based on a two-threshold correlation (TTC) method," *Journal of Neuroscience Methods*, vol. 167, no. 2, pp. 335–339, 2008.
- [12] D. Liu and D. J. Smith, "Voxelation and gene expression tomography for the acquisition of 3-D gene expression maps in the brain," *Methods*, vol. 31, no. 4, pp. 317–325, 2003.
- [13] T. E. Yankeelov, L. R. Arlinghaus, X. Li, and J. C. Gore, "The role of magnetic resonance imaging biomarkers in clinical trials of treatment response in cancer," *Seminars in Oncology*, vol. 38, no. 1, pp. 16–25, 2011.
- [14] J. Kuharić, N. Kovacic, P. Marusic, A. Marusic, and V. Petrovecki, "Positive identification by a skull with multiple epigenetic traits and abnormal structure of the neurocranium, viscerocranium, and the skeleton," *Journal of Forensic Sciences*, vol. 56, no. 3, pp. 788–793, 2011.
- [15] W. Yang, X. Xu, L. Duan et al., "High-throughput measurement of rice tillers using a conveyor equipped with x-ray computed tomography," *Review of Scientific Instruments*, vol. 82, no. 2, article 025102, 7 pages, 2011.
- [16] T. Stürzel, M. Bieberle, E. Laurien et al., "Experimental facility for two- and three-dimensional ultrafast electron beam x-ray computed tomography," *Review of Scientific Instruments*, vol. 82, no. 2, article 023702, 7 pages, 2011.
- [17] P. Ranjanomennahary, S. S. Ghalila, D. Malouche et al., "Comparison of radiograph-based texture analysis and bone mineral density with three-dimensional microarchitecture of trabecular bone," *Medical Physics*, vol. 38, no. 1, pp. 420–428, 2011.
- [18] T. Frosch, K. L. Chan, H. C. Wong, J. T. Cabral, and S. G. Kazarian, "Nondestructive three-dimensional analysis of layered polymer structures with chemical imaging," *Langmuir*, vol. 26, no. 24, pp. 19027–19032, 2010.
- [19] K. Janssens, J. Dik, M. Cotte, and J. Susini, "Photon-based techniques for nondestructive subsurface analysis of painted cultural heritage artifacts," *Accounts of Chemical Research*, vol. 43, no. 6, pp. 814–825, 2010.
- [20] X. Chen, Y. Lin, C. Wang, G. Shen, S. Zhang, and X. Wang, "A surgical navigation system for oral and maxillofacial surgery and its application in the treatment of old zygomatic fractures," *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 7, no. 1, pp. 42–50, 2011.
- [21] S. J. Taka and S. Srinivasan, "NIRViz: 3D visualization software for multimodality optical imaging using Visualization Toolkit (VTK) and Insight Segmentation Toolkit (ITK)," *Journal of Digital Imaging*. In press.
- [22] S. Martino, P. Marconi, B. Tancini et al., "A direct gene transfer strategy via brain internal capsule reverses the biochemical defect in Tay—Sachs disease," *Human Molecular Genetics*, vol. 14, no. 15, pp. 2113–2123, 2005.
- [23] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

- [24] J. A. Sethian, *Level Set Methods and Fast Marching methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, Cambridge University Press, Cambridge, UK, 1999.
- [25] E. A. Burton, S. Huang, W. F. Goins, and J. C. Glorioso, "Use of the herpes simplex viral genome to construct gene therapy vectors," *Methods in Molecular Medicine*, vol. 76, pp. 1–31, 2003.
- [26] P. Marconi, R. Argnani, E. Berto, A. L. Epstein, and R. Manservigi, "HSV as a vector in vaccine development and gene therapy," *Human Vaccines*, vol. 4, no. 2, pp. 91–105, 2008.
- [27] M. Cassano, S. Biressi, A. Finan et al., "Magic-factor 1, a partial agonist of Met, induces muscle hypertrophy by protecting myogenic progenitors from apoptosis," *PLoS One*, vol. 3, no. 9, article e3223, 2008.
- [28] B. G. Gálvez, D. Covarello, R. Tolorenzi et al., "Human cardiac mesoangioblasts isolated from hypertrophic cardiomyopathies are greatly reduced in proliferation and differentiation potency," *Cardiovascular Research*, vol. 83, no. 4, pp. 707–716, 2009.
- [29] M. Quattrocchi, G. Palazzolo, I. Agnolin et al., "Synthetic sulfonyl-hydrazone-1 positively regulates cardiomyogenic microRNA expression and cardiomyocyte differentiation of induced pluripotent stem cells," *Journal of Cellular Biochemistry*. In press.