MULTIRESOLUTION IMAGE MODELING AND BAYESIAN RECONSTRUCTION ALGORITHMS WITH APPLICATIONS TO EMISSION TOMOGRAPHY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Thomas Frese

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2001

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisor, Prof. Charles A. Bouman, for his exceptional guidance and support going far beyond what I could ever have imagined. His tireless involvement and high standards have taught me to analyze problems with much greater precision. His strong interest in my personal and professional development has resulted in countless opportunities and significant exposure to the academic community. Prof. Bouman, thank you for going the ten extra miles and making this such a great experience.

My wife, Charu, has been an integral part of this accomplishment right from the beginning, when she submitted my application to the Ph.D. program while I was still in Germany. I cannot express how much her love and support have meant to me during this challenging time. Charu, what would I do without you?

I would like to thank my parents and grandparents for providing me with the foundation that made it all possible. They have given me the interest, ability, and strength required for this endeavor. I would also like to thank Charu's parents for their substantial support and for being so wonderful.

There are many friends and colleagues who have shaped this experience through their personal and professional involvement. A list of them would have to include Ufuk Agar, Brian Atkins, Jau-Yuen Chen, Hui Cheng, Jim Grice and Christopher Taylor, however, there are many others who are not listed here. Dhiraj Kacker has been an important friend and colleague whom I would like to thank for numerous arguments and discussions.

I would like to thank Prof. Jan P. Allebach, Prof. Peter C. Doerschuk and Prof. Zygmunt Pizlo for their guidance and for being members of my advisory committee. I would also like to thank Dr. Bernice E. Rogowitz for many interesting discussions and for providing opportunities at IBM Research. Finally, I gratefully acknowledge Dr. Ned C. Rouze for his help and his invaluable comments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Frese, Thomas. Ph.D., Purdue University, May, 2001. Multiresolution Image Modeling and Bayesian Reconstruction Algorithms With Applications to Emission Tomography. Major Professor: Charles A. Bouman.

Presented in this thesis are multiresolution image models and Bayesian algorithms for statistical image reconstruction. The proposed methods are generally applicable to inverse problems in image processing, however, the focus of this work is the application to tomographic imaging.

Chapter two presents a wavelet graph image model designed to perform space-adaptive regularization. The model is based on a novel hierarchical dependency structure in the wavelet tree. An efficient multiresolution Bayesian reconstruction algorithm is proposed that allows for enforcement of space-domain constraints such as positivity despite using a wavelet prior model. Chapter three presents a comparison of Bayesian algorithms for tomographic reconstruction with emphasis on accurately modeling the tomography scanner. Included in this work is a new approach to obtaining an empirical model of the tomography scanner's system kernel. Chapter four presents a Bayesian multiresolution framework for discrete-valued image reconstruction in transmission and emission tomography. The approach includes a new method for efficient and accurate estimation of the discrete levels in the reconstruction image. Chapter five addresses the problem of image modeling for content-based database search and presents a subject study on human perception of the similarity of images.

1. INTRODUCTION

My research over the last four years was largely concerned with statistical image models and their application to Bayesian tomographic reconstruction. In general, I have found statistical modeling and multiresolution techniques to provide an extremely flexible framework with significant potential for improving image quality. An important lesson learned was, that the most significant gains in image quality could be realized by working with a specific physical data acquisition system and modeling its imperfections and hardware characteristics. However, I also found that once an accurate model of the physical system was used, large additional performance improvements were possible by improving the statistical assumptions such as the prior model used for Bayesian reconstruction.

Chapter 2 of this thesis presents a nonlinear wavelet prior model designed to perform space-adaptive regularization. The basic idea of this hierarchical model is to exploit inter-scale dependencies of wavelet coefficients in order to locally adapt the regularization to the image characteristics. This allows for strong regularization of uniform image regions, while textures and sharp edges are preserved. Critical contributions in this work are a new wavelet dependency structure, called the wavelet graph model, and a multiresolution optimization algorithm for wavelet priors that can enforce positivity in the space-domain. Relating to the work in this chapter, Appendix A.6 presents a literature review on existing multiresolution image models.

Chapter 3 presents a quantitative comparison of Bayesian and non-Bayesian reconstruction algorithms for tomography. The comparison is based on data acquired from a specific positron emission tomography system, the IndyPET scanner. In addition to comparing different reconstruction algorithms, this work emphasizes the accurate modeling of the scanning system. In contrast to previous approaches that largely rely on analytical modeling of detector properties, I have developed an empirical model of the scanner's system function based on scans of line sources. Reconstruction quality is compared for the empirical system function and a simple, analytical model.

Chapter 4 presents a Bayesian multiresolution framework for discrete-valued tomographic reconstruction. Importantly, the approach includes a new method for estimating the discrete levels in the data. The multiresolution algorithm presented in this work is shown to be faster and more robust as compared to a fixed resolution version.

Chapter 5 presents research on a very different aspect of image modeling, human image similarity perception. Modeling of human similarity perception is required for developing algorithms for content-based image database search. Presented here is a human subject study that is aimed at identifying the most salient dimensions of human image similarity discrimination.

2. ADAPTIVE WAVELET GRAPH MODEL FOR BAYESIAN TOMOGRAPHIC RECONSTRUCTION

Abstract

We introduce an adaptive wavelet graph image model applicable to Bayesian tomographic reconstruction and other problems with non-local data models. The proposed model captures coarse-to-fine scale dependencies in the wavelet tree by modeling the conditional distribution of wavelet coefficients given overlapping windows of scaling coefficients containing coarse scale information. This results in a graph dependency structure which is more general than a quadtree, enabling the model to produce smooth estimates even for simple wavelet bases such as the Haar basis. The interscale dependencies of the wavelet graph model are specified using a spatially non-homogeneous Gaussian distribution with parameters at each scale and location. The parameters of this distribution are selected adaptively using nonlinear classification of coarse scale data. The nonlinear adaptation mechanism is based on a set of training images. In conjunction with the wavelet graph model, we present a computationally efficient multiresolution image reconstruction algorithm. This algorithm is based on iterative Bayesian space-domain optimization using scale recursive updates of the wavelet graph prior model. In comparison to performing the optimization over the wavelet coefficients, the space-domain formulation facilitates enforcement of pixel positivity constraints. Results indicate that the proposed framework can improve reconstruction quality over fixed resolution Bayesian methods.

2.1 Introduction

A major challenge for Bayesian image reconstruction methods is the design of image models that accurately account for edges as well as uniform and textured regions in images, yet result in tractable estimation algorithms. In comparison to Markov random field (MRF) models, multiresolution methods can improve image modeling and increase computational efficiency. However, little work has been done on applying the multiresolution framework to Bayesian tomographic reconstruction and other problems with non-local data models. In addition, a well known problem with many multiresolution models is that the resulting estimates exhibit blockiness which is usually a result of the quadtree structure of the model. In this work, we address these issues by introducing an adaptive wavelet graph prior model. The inter-scale dependencies of this model are not limited to a quadtree structure, resulting in smooth estimates even for simple wavelet bases such as the Haar basis. In conjunction with this prior model, we propose a fast, iterative multiresolution reconstruction algorithm that can incorporate space domain constraints such as positivity, and thus, is applicable to Bayesian tomographic reconstruction.

Markov random field prior models [1, 2, 3] have enjoyed considerable success in Bayesian image reconstruction[4] and restoration[1]. However, MRF approaches are typically limited to modeling very local interactions in images. Furthermore, accounting for both edges and uniform image regions is difficult within the MRF framework. To avoid this problem, several MRF models have been proposed that provide good edge preservation without explicitly modeling edges[5, 6, 7, 8, 9, 10, 11].

In comparison to MRF priors, multiresolution models can improve modeling and offer fast estimation algorithms[12, 13, 14, 15, 16, 17]. Multiresolution models better account for long range interactions and are more practical with respect to treating edges, smooth and textured regions differently. Furthermore, multiresolution models are suitable for very efficient and robust estimation algorithms.

In recent years, multiresolution techniques have been developed which use linear system models on trees[18, 12, 13, 14, 19, 20, 21, 22, 23]. Nonlinear extensions of these models have been applied to image restoration with both Gaussian and Poisson noise[24, 16, 17, 25, 26]. Other methods have been developed for image segmentation[27, 28, 29, 30].

Most of the existing work on multiresolution models has focused on applications where the observations are spatially localized. Typically, the observations are assumed to be conditionally independent given the local state of the model[18, 12, 20, 21, 16, 17]. For this class of problems, the application of multiresolution models defined on quadtrees is very appealing because it leads to non-iterative, scale-recursive estimation and realization algorithms.

Little work, however, has been done on applying multiresolution Bayesian estimation to problems with non-local observations as encountered in tomographic reconstruction. In the case of tomographic reconstruction, application of multiresolution models is further complicated by the need to enforce pixel positivity constraints. This poses a problem for wavelet based multiresolution approaches, since it is difficult to enforce pixel positivity when the reconstruction optimization is performed in the wavelet domain.

As a result, few attempts have been made to use multiresolution models for iterative Bayesian tomographic reconstruction. In [31], Wu, Herman, and Browne used a wavelet domain prior model for MAP reconstruction of tomographic data. The prior used in this method is a space-variant simultaneous autoregressive (SAR) model whose coefficients are extracted using anisotropic diffusion. Nowak and Kolaczyk [25, 26] have proposed a tomographic reconstruction technique using a wavelet prior model. Their approach uses the expectation maximization (EM) algorithm to decouple the estimation problem from the linear projection transformation of the observations [6]. The quadtree structure of their prior model allows for closed-form EM updates, however, since the overall EM technique remains iterative, the computational benefit of using a quadtree model is less clear. For the Poisson case, their approach is limited to the Haar wavelet basis. Saquib, Bouman, and Sauer proposed a prior model for tomography that used a multiresolution pyramid representation[32].

However, a disadvantage of this pyramid representation is that different scales contain redundant information. This makes formulation of a consistent Bayesian model difficult.

A general problem for multiresolution models formulated on quadtrees is blockiness of the resulting estimates. In quadtree models, nodes that are spatially adjacent can be far apart in the tree so that their correlation is poorly modeled. This results in blocky estimates as observed in [20, 33, 34, 28, 35, 36, 21]. A popular fix is to average multiple estimates obtained for different spatial alignments of the tree or wavelet basis[37, 33, 36]. More elegant approaches have used trees with nodes corresponding to overlapping portions of the image domain[35] or have performed state augmentation to model the dependencies of general wavelet bases from within a quadtree structure[23]. These approaches have in common that their data representation is highly overcomplete. This overcomplete representation can make accurate modeling of sampled data difficult.

A more direct way to avoid blockiness is to use a dependency structure that is more general than the quadtree. For image segmentation, Bouman and Shapiro[28] have used a pyramidal graph where each node depends on a fixed size window at the next coarser scale. Kato, Berthod and Zerubia [27, 38] proposed a fully 3-D MRF where each node's neighborhood consists of adjacent nodes at the same scale and its quadtree parent. A disadvantage of violating the tree constraint is that Bayesian reconstruction must be performed iteratively as compared to the recursive algorithms available for quadtree models. For applications such as tomographic reconstruction this is not a limitation since the forward model requires iterative optimization in any case.

In this work, we develop a stochastic multiresolution framework for Bayesian image reconstruction involving data models with spatially non-local measurements. We propose a wavelet graph prior model in combination with a computationally efficient multiresolution reconstruction algorithm applicable to iterative tomographic reconstruction.

The basic concept of the proposed wavelet graph model is to exploit dependencies of wavelet coefficients across scales. We capture these dependencies by modeling the wavelet coefficients at each scale and location as a function of a window of scaling coefficients at the same scale. This structure has several important implications: First, by conditioning the wavelet coefficients on overlapping windows of scaling coefficients, the model dependencies are not limited to a quadtree structure. This reduces blockiness, enabling the model to produce smooth estimates even for simple wavelet bases such as the Haar basis. Secondly, the structure is such that the optimal wavelet graph model for a stationary process is homogeneous at each scale, resulting in a substantial reduction in the number of model parameters. Finally, the model is causal in scale, not overcomplete, and each wavelet coefficient is a function of only a few scaling coefficients. The model is not suitable for non-iterative scale-recursive optimization; however, due to the causality in scale and the spatially localized dependencies, the model allows for very efficient iterative optimization using scale recursive updates.

The wavelet graph model is specified through the conditional distributions of the wavelet coefficients given the window of scaling coefficients. We model these distributions using a spatially non-homogeneous Gaussian model with image-adaptive parameters. As compared to a fully nonlinear model, the Gaussian model is suitable for global optimization in a Bayesian framework while the adaptation allows the model to account for non-linear dependencies. The parameter adaptation is based on nonlinear classification of coarse scale data. The classifiers and class parameters used for the adaptation are obtained using training data. The training procedure allows the overall model to incorporate characteristics of typical reconstructions as prior information.

In conjunction with the wavelet graph model, we propose a fast multiresolution reconstruction algorithm applicable to Bayesian tomographic reconstruction. This reconstruction algorithm computes a sequence of Bayesian MAP estimates. Each estimate in the sequence is computed with fixed adaptation of the wavelet graph

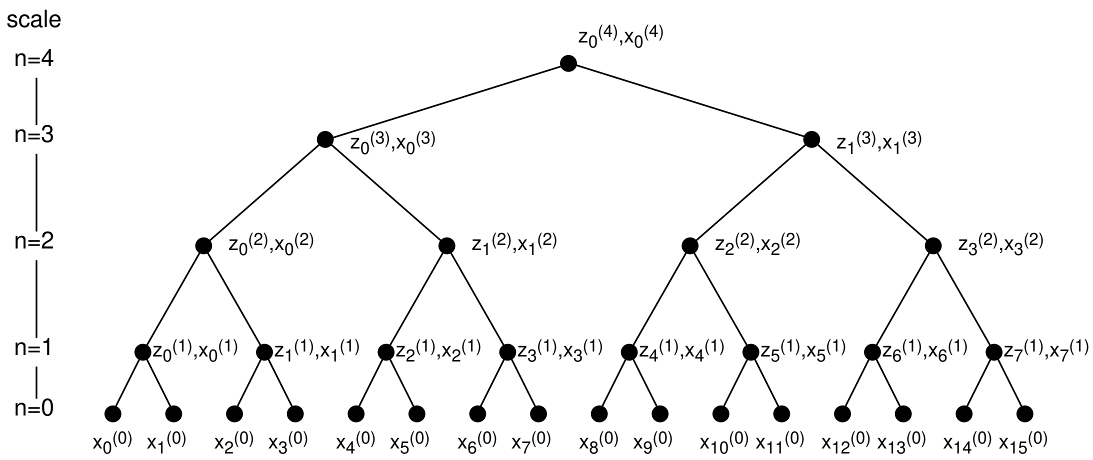


Fig. 2.1. Wavelet decomposition in 1-D for $L = 4$ resolution levels. The wavelet transform of the original image $x^{(0)} = x$ is given by $z = (z^{(1)}, \cdots, z^{(L)}, x^{(L)})$.

model. After computing each estimate, the parameters of the wavelet graph model are re-adapted. The MAP estimates are computed in the space-domain using scale-recursive updates of the multiresolution model. The space domain formulation of the optimization is essential for application of our model to tomographic reconstruction. Since the optimization is performed over the scaling coefficients, positivity constraints are easily enforced.

The rest of this chapter is organized as follows: Section 2.2 introduces the image model and develops the overall structure of the multiresolution reconstruction algorithm. Following in Section 2.2.3, we discuss in detail the space-domain optimization algorithm to compute the MAP estimates. Section 2.2.4 describes our implementation of the nonlinear classifiers used for the adaptation of the multiresolution prior. In Section 2.3, we apply the proposed model to tomographic reconstruction. Experimental results for simulated tomography data are presented in Section 2.4.

2.2 Image Model and Multiresolution Reconstruction Algorithm

2.2.1 Wavelet Graph Model

In order to define the notation for the wavelet decomposition, consider the one-dimensional case. Let x be the N-dimensional vector of the image pixel values in raster order and let x_s denote the pixel value at location s. We now consider the class of wavelet decompositions that can be computed using the recursions

$$\begin{aligned} x_s^{(n+1)} &= \sum_k h_{k-2s} x_k^{(n)} && (2.1) \\ z_s^{(n+1)} &= \sum_k g_{k-2s} x_k^{(n)} && (2.2) \end{aligned}$$

where h is the analysis lowpass and g is the analysis highpass filter. We assume both h and g to have compact support. In this notation, $x_s^{(n)}$ denotes the scaling coefficient and $z_s^{(n)}$ the wavelet or detail coefficient at scale n and location s. We denote the finest scale as $n = 0$ such that $x^{(0)} = x$. An illustration of this decomposition is shown in Fig. 2.1.

In the following, we will assume an L resolution wavelet decomposition. The non-overcomplete wavelet transform of x is then specified by the wavelet coefficients $z^{(1)} \ldots z^{(L)}$ and the coarsest scale scaling coefficients $x^{(L)}$. Using matrix notation, the L resolution wavelet transform of x can be written as

$$z \triangleq \begin{bmatrix} z^{(1)} \\ \vdots \\ z^{(L)} \\ x^{(L)} \end{bmatrix} = Wx \qquad (2.3)$$

where W denotes the wavelet transform matrix, and z includes the coarsest scaling coefficients $x^{(L)}$.

The basic concept of our image model is to exploit the dependencies of the wavelet coefficients across scales. To formulate the approach, we write the distribution $\log p_z(z)$ in terms of the conditional distribution at each scale given the information at all coarser scales.

$$\log p_z(z) = \log p_{x^{(L)}}(x^{(L)}) + \sum_{n=1}^{L} \log p_{z^{(n)}|z^{(n+1)},\ldots,z^{(L)},x^{(L)}}(z^{(n)}|z^{(n+1)},\ldots,z^{(L)},x^{(L)}) \tag{2.4}$$

Since the scaling coefficients $x^{(n)}$ contain exactly the same information as $z^{(n+1)},\ldots,z^{(L)},x^{(L)}$, we may rewrite (2.4) as

$$\log p_z(z) = \log p_{x^{(L)}}(x^{(L)}) + \sum_{n=1}^{L} \log p_{z^{(n)}|x^{(n)}}(z^{(n)}|x^{(n)}) \tag{2.5}$$

where $x^{(n)}$ is a function of $z^{(n+1)},\ldots,z^{(L)},x^{(L)}$. The distribution assumption for $x^{(L)}$ is typically not important due to the high signal-to-noise ratio at the coarsest resolution. Therefore, we assume that $x^{(L)}$ is uniformly distributed. Thus, we can write

$$\log p_z(z) = \sum_{n=1}^{L} \log p_{z^{(n)}|x^{(n)}}(z^{(n)}|x^{(n)}) + \text{const} . \tag{2.6}$$

To obtain a practical model, we assume the wavelet coefficients at different locations to be conditionally independent given the scaling coefficients at the same scale. Furthermore, we assume the wavelet coefficients at each location only to depend on a small window of scaling coefficients. Let s denote a spatial location at a given scale n such that $z_s^{(n)}$ is the vector of the wavelet coefficients at location s. For the 2-D case, $z_s^{(n)}$ has three components corresponding to the high-low, low-high, and high-high coefficients of a separable wavelet decomposition. We then define ∂s as a window with compact support centered at position s with circular boundary conditions. Then $x_{\partial s}^{(n)}$ is the set of scaling coefficients within the window ∂s at resolution n. Further, let $S^{(n)}$ denote the set of all locations s of the wavelet decomposition at scale n. Using this notation, our assumptions are that the $z_s^{(n)}$ are conditionally independent and that $z_s^{(n)}$ depends only on $x_{\partial s}^{(n)}$. Applying these assumptions to (2.6) results in the model

$$\boxed{\log p_z(z) = \sum_{n=1}^{L} \sum_{s\in S^{(n)}} \log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(z_s^{(n)}|x_{\partial s}^{(n)}) + \text{const} .} \tag{2.7}$$

We will call any model of the form (2.7) a wavelet graph model. Figure 2.2 illustrates

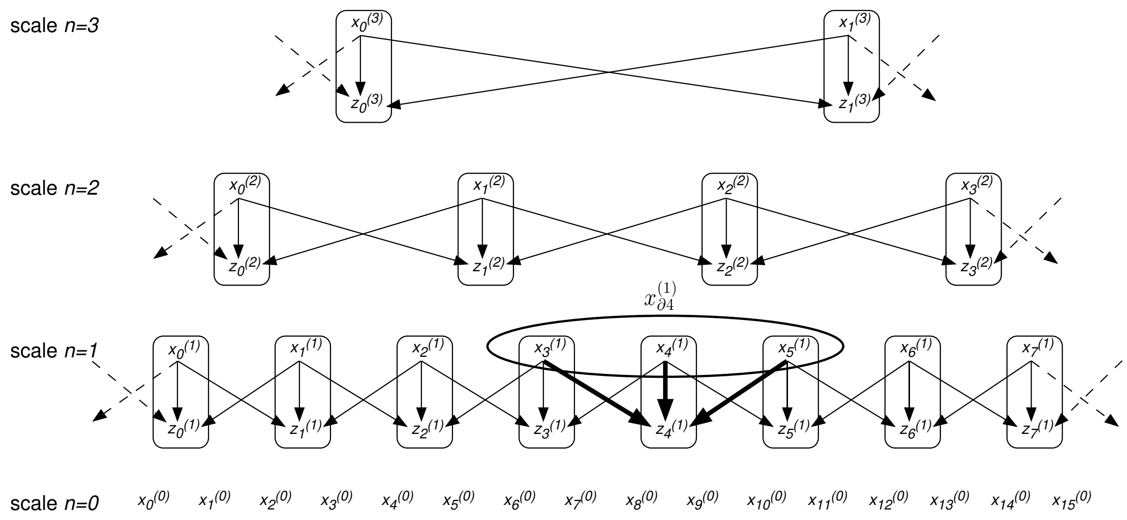


Fig. 2.2. Spatial dependencies for the 1-D case. Using a prediction window $\partial s = \{s-1, s, s+1\}$, each wavelet coefficient $z_s^{(n)}$ depends on the three scaling coefficients $\{x_{s-1}^{(n)}, x_s^{(n)}, x_{s+1}^{(n)}\}$. Notice that the scaling coefficients $x^{(n)}$ contain all the information at coarser scales $l > n$.

the spatial dependencies of the model (2.7) for the case of a one dimensional signal and a three point window $\partial s = \{s-1, s, s+1\}$. In this case, the conditional distribution of $z_s^{(n)}$ depends only on the three scaling coefficients $x_{\partial s}^{(n)} = \{x_{s-1}^{(n)}, x_s^{(n)}, x_{s+1}^{(n)}\}$.

An important advantage of the structure (2.7) is that the optimal wavelet graph model for a stationary process is homogeneous. By homogeneous, we mean that $\log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(z_s^{(n)}|x_{\partial s}^{(n)}) = f_n(z_s^{(n)}, x_{\partial s}^{(n)})$ for some functions f_n that do not depend on s. The following theorem, proven in Appendix A.7, makes the statement above precise:

Theorem 1 *Let $\{X_s^{(0)}\}_{s=0}^{2^N-1}$ be a 1-D discrete-time random process that is circularly stationary, i.e. $X_s^{(0)}$ and $V_s = X_{(s-k)\mathrm{mod}2^N}^{(0)}$ have the same distribution. Let $X_s^{(n)}$ and $Z_s^{(n)}$ for $1 \leq n \leq L$ be the wavelet decomposition of $X_s^{(0)}$ as specified by (2.1) and (2.2) using circular boundary conditions and let $\log p_z(z)$ be a wavelet graph model of the form (2.7) with*

$$\log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(z_s^{(n)}|x_{\partial s}^{(n)}) = f_n(z_s^{(n)}, x_{\partial s}^{(n)}, \nu_s^{(n)})$$

where the $\nu_s^{(n)}$ are parameters of the model. Assume there exists a unique minimizer *of the Kullback-Leibler distance*

$$\nu^* = \arg\max_{\nu} E[\log p_z(Z)] \ . \tag{2.8}$$

Then, $(\nu_s^{(n)})^*$ *is not a function of* s*, implying a homogeneous wavelet graph model* $\log p_z(z)$*.*

The property stated in this theorem greatly simplifies parameter estimation since we only need to estimate a single set of parameters at each scale. In general, this would not be the case if we conditioned the inter-scale dependencies on wavelet coefficients instead of the window of scaling coefficients. This homogeneity property is very important since it dramatically reduces the number of free parameters in the model, thereby allowing practical model estimation from sampled data. In section 2.2.4, we will use this property to justify the design of a single nonlinear classifier for each scale.

We first consider the case of a spatially non-homogeneous Gaussian model. In this case, the conditional distributions $\log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(z_s^{(n)}|x_{\partial s}^{(n)})$ must be of the form

$$\log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(z_s^{(n)}|x_{\partial s}^{(n)}) = -\frac{1}{2}\|z_s^{(n)} - A_s^{(n)}x_{\partial s}^{(n)} - \sigma_n b_s^{(n)}\|^2_{\sigma_n^{-2}B_s^{(n)}} + \text{const} \tag{2.9}$$

where $A_s^{(n)}$ is a matrix, $b_s^{(n)}$ is a column vector, $B_s^{(n)}$ is a positive definite matrix, σ_n is a scaling constant, and $\|\cdot\|_B$ denotes the norm such that $\|\varepsilon\|_B^2 = \varepsilon^t B\varepsilon$. We note that for this model, the conditional mean of $z_s^{(n)}$ is an affine function of $x_{\partial s}^{(n)}$ given by

$$\mu_s^{(n)} \triangleq A_s^{(n)}x_{\partial s}^{(n)} + \sigma_n b_s^{(n)} \ , \tag{2.10}$$

and when $\sigma_n^2 = 1$, the conditional covariance of $z_s^{(n)}$ is given by $R_s^{(n)} \triangleq (B_s^{(n)})^{-1}$. Consequently, the model is parameterized by $A_s^{(n)}, b_s^{(n)}, B_s^{(n)}$ and a global scaling constant σ_n for each scale n. The scaling constants σ_n will play an important role later by allowing us to use the same adaptation mechanism for different amplitude scalings of the image x. Assuming the wavelet transform W is orthonormal, then $\det W = 1$, and we may express $\log p_x(x)$ as

$$\log p_x(x) \ = \ \log p_z(Wx) \tag{2.11}$$

$$= -\frac{1}{2}\sum_{n=1}^{L}\sum_{s\in S^{(n)}} \|z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - \sigma_n b_s^{(n)}\|^2_{\sigma_n^{-2} B_s^{(n)}} + \text{const} \quad (2.12)$$

where $z_s^{(n)} = (Wx)_s^{(n)}$. The model (2.12) is used as the prior distribution for the Bayesian reconstruction of $x = x^{(0)}$.

In order to formulate a multiresolution reconstruction algorithm, we also want to directly calculate coarse scale reconstructions using the prior model only for coefficients at scales coarser than the reconstruction scale. Thus, we define a coarse scale prior model for the direct reconstruction of the scaling coefficients $x^{(l)}$ at scale l as

$$\begin{aligned}\log p_{x^{(l)}}(x^{(l)}) &= \log p_{z^{(l+1)},\ldots,z^{(L)},x^{(L)}}(z^{(l+1)},\ldots,z^{(L)},x^{(L)}) \quad (2.13)\\ &= -\frac{1}{2}\sum_{n=l+1}^{L}\sum_{s\in S^{(n)}} \|z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - \sigma_n b_s^{(n)}\|^2_{\sigma_n^{-2} B_s^{(n)}} + \text{const} \quad (2.14)\end{aligned}$$

Given noisy measurements y and a physical data model $\log p_{y|x}(y|x)$, we obtain the data model $\log p_{y|x^{(l)}}(y|x^{(l)})$ at scale l as

$$\log p_{y|x^{(l)}}(y|x^{(l)}) = \log p_{y|x}(y|\mathcal{I}_l^0 x^{(l)}) \quad (2.15)$$

where $\mathcal{I}_l^n$ denotes the interpolation from scale l to scale n. The interpolation $\mathcal{I}_l^0 x^{(l)}$ is obtained as the wavelet reconstruction of $x^{(l)}$ assuming that $z^{(n)} = 0$ for $1 \le n \le l$.

Based on (2.14) and (2.15), the Bayesian MAP estimate $\hat{x}^{(l)}$ of the scaling coefficients $x^{(l)}$ at scale l is the solution to the optimization problem

$$\hat{x}^{(l)} = \arg\max_{x^{(l)}\ge 0}\{\log p_{y|x^{(l)}}(y|x^{(l)}) + \log p_{x^{(l)}}(x^{(l)})\}\ . \quad (2.16)$$

We have included the positivity constraint $x^{(l)} \ge 0$ since it is important for tomographic reconstruction. Note that for the special case $l = 0$, expression (2.16) is the standard MAP equation for the image x.

2.2.2 Spatially Adaptive Multiscale Reconstruction

We use a multiresolution algorithm to perform the image reconstruction and to adaptively select the parameters of the linear model. As illustrated in Fig. 2.3, the basic concept of the multiresolution algorithm is to compute a sequence of Bayesian

MAP estimates from coarse to fine scale. The algorithm starts with the reconstruction of the scaling coefficients $x^{(L)}$ at the coarsest scale L and then successively performs the reconstructions at the finer scales $l = L-1, \ldots, 0$. At each step in this sequence, the current reconstruction is used to initialize the model parameters at the next finer scale and to re-adapt the parameters at the coarser scales. Let $\theta_s^{(l)}$ denote the vector of model parameters at location s and scale l

$$\theta_s^{(l)} = [A_s^{(l)}, b_s^{(l)}, B_s^{(l)}] \ . \tag{2.17}$$

After computing the MAP reconstruction $\hat{x}^{(l)}$ at scale l, we update all the parameters $\theta_s^{(n)}$ for $n \geq l$. These new parameters are then used to reconstruct $x^{(l-1)}$ at the next finer scale.

We update the parameters $\theta_s^{(n)}$ using a nonlinear classification method derived from recent work in image interpolation[39]. More specifically, we update $\theta_s^{(n)}$ by applying the nonlinear operator $T^{(n)}[\cdot]$ to the window of scaling coefficients $\hat{x}_{\partial s}^{(n)}$.

$$\theta_s^{(n)} \leftarrow T^{(n)} \left[\frac{\hat{x}_{\partial s}^{(n)}}{\sigma_n} \right] \tag{2.18}$$

Note that based on Theorem 1, we can use a single $T^{(n)}[\cdot]$ for all locations s at scale n. The normalization by σ_n in (2.18) is included to account for possible scaling of the image $x^{(n)}$. To obtain a robust estimate, we compute the 10% trimmed mean[40] of $\hat{x}^{(n)}$ over the approximate support of the active image region.

$$\sigma_n \leftarrow AVG_{trim10\%}[\hat{x}^{(n)}] \tag{2.19}$$

The nonlinear operators $T^{(n)}[\cdot]$ are obtained during a training phase. The structure of the $T^{(n)}[\cdot]$ and the training procedure are explained in section 2.2.4.

The coarse-to-fine scale multiresolution reconstruction algorithm is summarized in Fig. 2.4. Note that the final fine scale reconstruction produced by this algorithm is not a conventional MAP estimate. Rather, it is a MAP estimate with respect to a data dependent prior. This prior is the spatially non-homogeneous linear model at all scales with the parameters obtained in the last adaptation step.

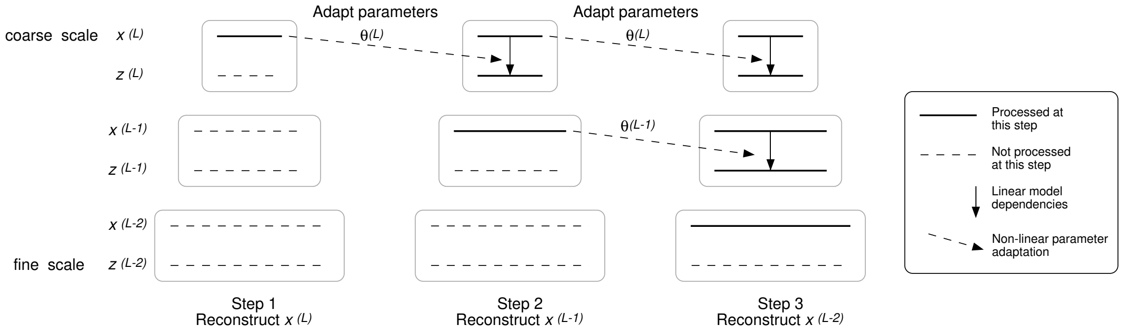


Fig. 2.3. Illustration of the multiresolution reconstruction algorithm. Starting at the coarsest scale $l = L$, we calculate a coarse-to-fine scale sequence of MAP estimates $x^{(l)}$. Each estimate $x^{(l)}$ takes into account the contributions of the adapted prior model at *all* coarser scales $n > l$. After computing $x^{(l)}$, we initialize the parameters $\theta_s^{(l)}$ and re-adapt $\theta_s^{(n)}$ for $n > l$.

1. Initialize $\hat{x}^{(n)} = 0$ and $\hat{z}^{(n)} = 0$ for all scales n.
2. Compute maximum likelihood (ML) reconstruction $\hat{x}^{(L)}$.
3. For $l = L - 1$ downto 0 {
   (a) Adapt Gaussian model: For $l < n \leq L$, $\forall s$, assign $\theta_s^{(n)}$ and σ_n using (2.18), (2.19).
   (b) Initialize $x^{(l)} \leftarrow \mathcal{I}_{l+1}^{l} \hat{x}^{(l+1)}$.
   (c) Calculate MAP reconstruction $\hat{x}^{(l)}$ for Gaussian wavelet graph prior model, using (2.16).
   (d) Recompute $\hat{x}^{(n)}, \hat{z}^{(n)}$ for $l \leq n \leq L$ as wavelet decomposition of $\hat{x}^{(l)}$.

   }

Fig. 2.4. Summary of multiresolution reconstruction algorithm.

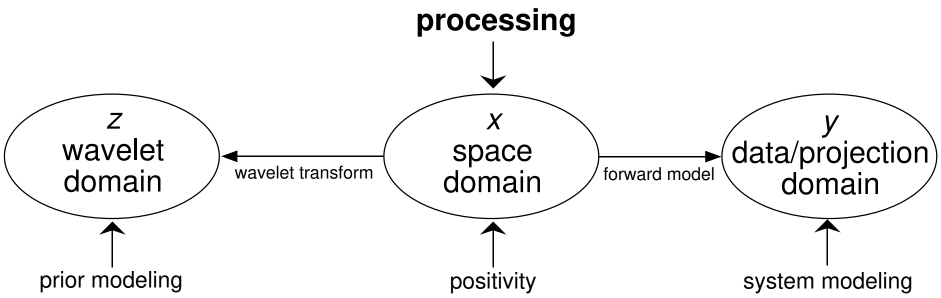


Fig. 2.5. The image is modeled in the wavelet domain but the MAP optimization is performed in the space domain. This allows the positivity constraint to be easily enforced.

2.2.3 MAP Optimization for Gaussian Wavelet Graph Model

In this section, we describe a computationally efficient implementation of the MAP optimization in step 3.c of Fig. 2.4. The MAP optimization is performed for the Gaussian wavelet graph prior model with fixed model parameters $\theta_s^{(l)} = [A_s^{(l)}, b_s^{(l)}, B_s^{(l)}]$ and σ_n. As discussed in the previous section, these parameters are obtained using nonlinear classification of coarse scale data. The parameter selection mechanism is described section 2.2.4.

The positivity constraint, $x^{(l)} \geq 0$, is an essential component of the MAP optimization equation (2.16). However, enforcement of positivity can be very difficult in the wavelet domain, particularly for general wavelet transforms. Figure 2.5 illustrates our approach for solving this problem. The optimization is performed in the space domain, while the prior model is formulated in the wavelet domain and the system model is formulated in the projection domain. This makes enforcement of positivity simple. Another advantage of space domain optimization is that it simplifies the forward model in tomography. This is because the transformation from z to y is generally less sparse than the transformation from x to y.

Below, we derive a general MAP optimization algorithm in the space domain. We then show how the algorithm can be implemented in a computationally efficient manner by incorporating our model assumptions of conditional independence and limited spatial support. Since (2.16) has the same structure for any l, we will simplify the presentation by assuming reconstruction at the finest scale $l = 0$. Thus, we will omit the superscript (l) and write $x = x^{(0)}$. The coarse scale reconstructions $\hat{x}^{(l)}$ for $l > 0$ can be obtained by using the same optimization algorithm and treating scale l as the finest scale $n = 0$ as discussed at the end of this section.

To derive the optimization algorithm, define matrices $U^{(n)}$ as the subsets of the inverse wavelet transform W^{-1} such that

$$x^{(n)} = U^{(n)} z = U^{(n)} \begin{bmatrix} z^{(1)} \\ \vdots \\ z^{(L)} \\ x^{(L)} \end{bmatrix} . \tag{2.20}$$

Further, let $U^{(n)}_{\partial s}$ denote the rows of $U^{(n)}$ such that $x^{(n)}_{\partial s} = U^{(n)}_{\partial s} z$. Using this notation, we can rewrite (2.10) as

$$\mu^{(n)}_s = A^{(n)}_s U^{(n)}_{\partial s} z + \sigma_n b^{(n)}_s . \tag{2.21}$$

To simplify the notation, let us define $\tilde{A}^{(n)}_s = A^{(n)}_s U^{(n)}_{\partial s}$, $\tilde{b}^{(n)}_s = \sigma_n b^{(n)}_s$, and $\tilde{B}^{(n)}_s = \sigma_n^{-2} B^{(n)}_s$. Furthermore, let $\tilde{A}$, $\tilde{b}$ and $\tilde{B}$ denote the parameters for all locations s and all scales n. We can then re-write the model (2.12) as

$$\begin{aligned} \log p_x(x) &= -\frac{1}{2} \| z - \tilde{A} z - \tilde{b} \|^2_{\tilde{B}} + \text{const} && (2.22) \\ &= -\frac{1}{2} \| (I - \tilde{A}) W x - \tilde{b} \|^2_{\tilde{B}} + \text{const} && (2.23) \end{aligned}$$

where $x = W^{-1} z$.

To optimize $\log p_x(x)$ with respect to a single pixel value x_i, we need the first and second derivatives with respect to x_i. Let us define ε as the prediction error $z - \mu$ of the linear model in the wavelet domain

$$\varepsilon = (I - \tilde{A}) W x - \tilde{b} . \tag{2.24}$$

Based on (2.23), we can then write the gradient $\nabla_x \log p_x(x)$ as

$$\begin{aligned} \nabla_x \log p_x(x) &= x^t W^t (I - \tilde{A})^t \tilde{B} (I - \tilde{A}) W - \tilde{b}^t \tilde{B} (I - \tilde{A}) W && (2.25) \\ &= \varepsilon^t \tilde{B} (I - \tilde{A}) W . && (2.26) \end{aligned}$$

If we now let e_i denote the unitary vector in direction x_i, we can write the first derivative α_1 as

$$\begin{aligned} \alpha_1 &= \frac{\partial}{\partial x_i} \log p_x(x) && (2.27)\\ &= (\nabla_x \log p_x(x))\ e_i && (2.28)\\ &= \varepsilon^t \tilde{B}(I-\tilde{A})W_{*i} && (2.29) \end{aligned}$$

where W_{*i} denotes the i^{th} column of W. Similarly, for the second derivative α_2 we obtain

$$\begin{aligned} \alpha_2 &= \frac{\partial^2}{\partial x_i^2} \log p_x(x) && (2.30)\\ &= W_{*i}^t (I-\tilde{A})^t \tilde{B}(I-\tilde{A})W_{*i}\ . && (2.31) \end{aligned}$$

Notice that $(I-\tilde{A})W_{*i}$ is the derivative of the prediction errors in the wavelet domain with respect to x_i. Let us define $d\varepsilon = (I-\tilde{A})W_{*i}$, then

$$\begin{aligned} \alpha_1 &= \varepsilon^t \tilde{B} d\varepsilon && (2.32)\\ \alpha_2 &= (d\varepsilon)^t \tilde{B} d\varepsilon\ . && (2.33) \end{aligned}$$

The prior $\log p_x$ can now be written as a function of the perturbation γ of pixel x_i

$$\log p_x(x+\gamma e_i) = \alpha_1\gamma + \frac{1}{2}\alpha_2\gamma^2 + \text{const}\ . \qquad (2.34)$$

This expression can easily be optimized with respect to γ as part of a standard iterative coordinate descent (ICD) [41, 42] algorithm in x. Figure 2.6 summarizes the basic steps of the ICD optimization algorithm for MAP reconstruction using the Gaussian wavelet graph prior model. The prediction errors ε are kept as a state vector. The set $S = S^{(0)}$ denotes the set of image pixels at scale $n = 0$.

In the following, we show how the algorithm can be implemented in a computationally efficient manner by incorporating our model assumptions of conditional independence and limited spatial support. To illustrate the implementation, we augment the tree structure of the linear model as shown in Fig. 2.7 for the 1-D case. In this representation, each tree node $\mathcal{N}_s^{(n)}$ contains the wavelet coefficients $z_s^{(n)}$, the scaling

1. initialize $\varepsilon = (I - \tilde{A})Wx - \tilde{b}$
2. until convergence do

   for $i \in S$ {

   (a) $d\varepsilon = (I - \tilde{A})W_{*i}$
   (b) $\alpha_1 = \varepsilon^t \tilde{B} d\varepsilon$, $\alpha_2 = (d\varepsilon)^t \tilde{B} d\varepsilon$
   (c) $\Delta = \arg\max_\gamma \{\log p_{y|x}(y|x + \gamma e_i) + \alpha_1 \gamma + 0.5\alpha_2 \gamma^2\}$
   (d) $x_i \leftarrow x_i + \Delta$
   (e) $\varepsilon \leftarrow \varepsilon + d\varepsilon\Delta$

   }

Fig. 2.6. General formulation of the iterative coordinate descent (ICD) optimization algorithm for space-domain MAP reconstruction using the Gaussian wavelet graph prior model. The prediction errors ε are kept as a state vector. The set $S = S^{(0)}$ denotes the set of image pixels at scale $n = 0$.

coefficient $x_s^{(n)}$ as well as the current prediction error $\varepsilon_s^{(n)} = z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - \tilde{b}_s^{(n)}$. In addition, $\mathcal{N}_s^{(n)}$ contains the temporary variables $dz_s^{(n)}$ and $dx_s^{(n)}$. These variables are the derivatives of $z_s^{(n)}$ and $x_s^{(n)}$ with respect to the pixel value x_i that is currently being updated in the space domain. To compute $dz_s^{(n)}$ and $dx_s^{(n)}$, let $W^{(n)}$ denote the subset of the wavelet transform such that $z^{(n)} = W^{(n)} x$. Further, let $D^{(n)}$ denote the decimation operation used to obtain the scaling coefficients $x^{(n)}$ from the original image x as

$$x^{(n)} = D^{(n)} x^{(0)} = U^{(n)} W x^{(0)} . \tag{2.35}$$

For the update of pixel value x_i, we compute $dz_s^{(n)}$ and $dx_s^{(n)}$ as

$$dz_s^{(n)} = W_{si}^{(n)} \tag{2.36}$$

$$dx_s^{(n)} = D_{si}^{(n)} . \tag{2.37}$$

The notation in (2.36) is for the 1-D case, where both $dz_s^{(n)}$ and $dx_s^{(n)}$ are scalars.

For the 2-D case, $dz_s^{(n)}$ is the three-component vector containing the derivatives corresponding to the high-low, low-high and high-high components of $z_s^{(n)}$. Define $dx_{\partial s}^{(n)}$ as the vector with components $dx_k^{(n)}$ for $k \in \partial s$. Using this notation, we can write $d\varepsilon_s^{(n)}$ as

$$d\varepsilon_s^{(n)} = dz_s^{(n)} - A_s^{(n)} dx_{\partial s}^{(n)} \ . \tag{2.38}$$

In order to compute $d\varepsilon$ efficiently, we want to consider only the locations (s, n) for which $d\varepsilon_s^{(n)}$ is nonzero. Let us define the sets $S_i^{(n)}$ as

$$S_i^{(n)} = \{s : dz_s^{(n)} \neq 0 \text{ or } dx_s^{(n)} \neq 0\} \ . \tag{2.39}$$

Notice, that the sets $S_i^{(n)}, 1 \leq n \leq L$, are only a function of the wavelet transform W_{*i}. The change $dx_{\partial s}^{(n)}$ is nonzero only at locations whose prediction window includes nodes in $S_i^{(n)}$. We define the set of these locations as $\tilde{S}_i^{(n)}$

$$\begin{aligned} \tilde{S}_i^{(n)} &= \{k : s \in \partial k \text{ for some } s \in S_i^{(n)}\} && (2.40) \\ &= \bigcup_{s \in S_i^{(n)}} \partial s \ . && (2.41) \end{aligned}$$

The last equation is a result of the symmetry of the prediction window. Notice, that since the prediction window ∂s includes s, we have $S_i^{(n)} \subset \tilde{S}_i^{(n)}$ such that both $dz_s^{(n)} = 0$ and $dx_{\partial s}^{(n)} = 0$ for $s \notin \tilde{S}_i^{(n)}$ and consequently $d\varepsilon_s^{(n)} = 0$ for $s \notin \tilde{S}_i^{(n)}$. Figure 2.7 illustrates the definitions of $S_i^{(n)}$ and $\tilde{S}_i^{(n)}$ for the specific example of a 1-D Haar wavelet decomposition. We can now compute the nonzero components of $d\varepsilon$ as

$$d\varepsilon_s^{(n)} = dz_s^{(n)} - A_s^{(n)} dx_{\partial s}^{(n)} \qquad s \in \tilde{S}_i^{(n)} \ . \tag{2.42}$$

The first and second derivatives α_1 and α_2 are then given by

$$\begin{aligned} \alpha_1 &= \sum_{n=1}^{L} \sum_{s \in \tilde{S}_i^{(n)}} \varepsilon_s^{(n)} \tilde{B}_s^{(n)} d\varepsilon_s^{(n)} && (2.43) \\ \alpha_2 &= \sum_{n=1}^{L} \sum_{s \in \tilde{S}_i^{(n)}} d\varepsilon_s^{(n)} \tilde{B}_s^{(n)} d\varepsilon_s^{(n)} \ . && (2.44) \end{aligned}$$

The derivatives α_1 and α_2 are used to perform the MAP optimization with respect to x_i in steps 2.c and 2.d of Fig. 2.6. After updating x_i, the state variables $\varepsilon_s^{(n)}$ are

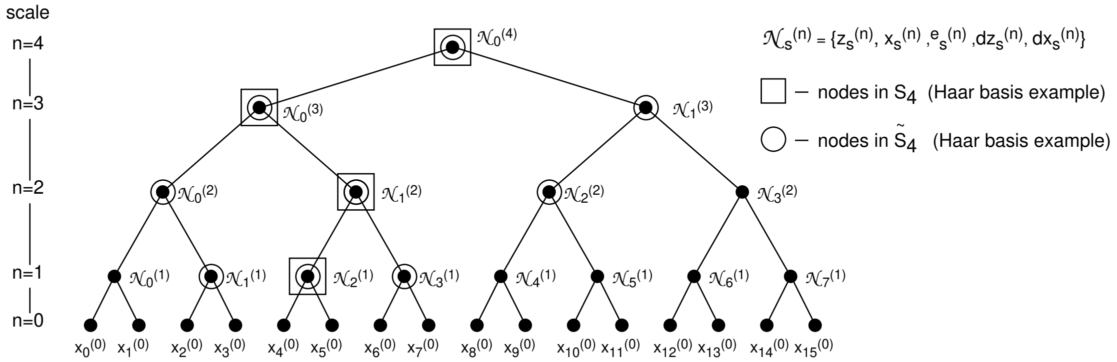


Fig. 2.7. Illustration of ICD update computation for pixel $x_4^{(0)}$ using a Haar wavelet basis and a 3-point window ∂s. To update $x_4^{(0)}$, only the nodes in the set $\tilde{S}_4^{(n)}$, shown in circles, must be considered. Each node $\mathcal{N}_s^{(n)}$ for $n > 0$ contains the variables $z_s^{(n)}, x_s^{(n)}, \varepsilon_s^{(n)}$ as well as the temporary variables $dz_s^{(n)}$ and $dx_s^{(n)}$ required for the computation of the ICD update.

updated as

$$\varepsilon_s^{(n)} \leftarrow \varepsilon_s^{(n)} + d\varepsilon_s^{(n)}\Delta \qquad s \in \tilde{S}_i^{(n)} \tag{2.45}$$

for $1 \leq n < L$. This completes the efficient implementation of the iterative coordinate descent optimization. Note that $dx_s^{(n)}, dz_s^{(n)}, S_i^{(n)}, \tilde{S}_i^{(n)}, d\varepsilon_s^{(n)}$ and α_2 are not data dependent but are only a function of i. If desired, these variables can be precomputed and stored for all i. Figure 2.8 summarizes the optimization algorithm.

While the development above assumes optimization at the finest scale $l = 0$, the same algorithm can be used for the coarse scale reconstructions $\hat{x}^{(l)}$ for $l > 0$. To optimize (2.16) for $l > 0$, we treat scale l as if it were the finest scale $n = 0$. Thus, instead of iterating over the pixel locations i at fine scale, we now index over $i \in S^{(l)}$. The state variables $z_s^{(n)}, x_s^{(n)}, \varepsilon_s^{(n)}$ as well as the directions $dz_s^{(n)}, dx_s^{(n)}, d\varepsilon_s^{(n)}$ are computed using the wavelet transform of $x^{(l)}$ and are computed for $n > l$ only.

The computational complexity associated with the optimization of the linear image model is on the order of $N^{(l)}(\log N^{(l)})^2$ multiplications for one full update of $\hat{x}^{(l)}$, where $N^{(l)}$ is the number of image pixels at the reconstruction scale l. For the special case of a Haar wavelet basis, this complexity reduces to $N^{(l)} \log N^{(l)}$. The

1. initialize $\varepsilon = (I - \tilde{A})Wx - \tilde{b}$
2. until convergence do

   for $i \in S$ {

   (a) (*) compute $dz_s^{(n)} = W_{si}^{(n)},\ dx_s^{(n)} = D_{si}^{(n)}, \forall s, n > 0$
   (b) (*) assign $S_i^{(n)} = \{s : dz_s^{(n)} \neq 0 \vee dx_s^{(n)} \neq 0\}$
   (c) (*) assign $\tilde{S}_i^{(n)} = \cup_{s \in S_i^{(n)}} \partial s$
   (d) (*) $d\varepsilon_s^{(n)} = dz_s^{(n)} - A_s^{(n)} dx_{\partial s}^{(n)}$ for $s \in \tilde{S}_i^{(n)}$
   (e) $\alpha_1 = \sum_{n=1}^{L} \sum_{s \in \tilde{S}_i^{(n)}} \varepsilon_s^{(n)} (\sigma_n^{-2} B_s^{(n)}) d\varepsilon_s^{(n)}$
   (f) (*) $\alpha_2 = \sum_{n=1}^{L} \sum_{s \in \tilde{S}_i^{(n)}} d\varepsilon_s^{(n)} (\sigma_n^{-2} B_s^{(n)}) d\varepsilon_s^{(n)}$
   (g) $\Delta = \arg\max_\gamma \left\{ \log p_{y|x}(y|x + \gamma e_i) + \alpha_1 \gamma + 0.5 \alpha_2 \gamma^2 \right\}$
   (h) $x_i \leftarrow x_i + \Delta$
   (i) $\varepsilon_s^{(n)} \leftarrow \varepsilon_s^{(n)} + d\varepsilon_s^{(n)} \Delta$ for $s \in \tilde{S}_i^{(n)}$

   }
3. recompute $\hat{x}^{(n)}, \hat{z}^{(n)}$ for $1 \leq n \leq L$ as wavelet decomposition of $\hat{x}$

Fig. 2.8. Detailed algorithm for efficient ICD updates using wavelet graph model. The operations marked with (*) can be pre-computed. Excluding the forward model term of step 2.f, the computational complexity is order $N(\log N)^2$ for one full update of $\hat{x}^{(0)}$.

details can be found in Appendix A.8.

2.2.4 Nonlinear Classifiers for Parameter Selection

The nonlinear operators $T^{(n)}[\cdot]$ used for the parameter selection (2.18) are obtained using a method derived from recent work in image interpolation[39]. The selection of $\theta_s^{(n)}$ is performed by first classifying the input vector $x_{\partial s}^{(n)}/\sigma_n$ into a class $k_{n,s}$ and then selecting $\theta_s^{(n)}$ as a parameter vector associated with this class. To simplify the notation, we denote the parameter vector associated with class $k_{n,s}$ as $\theta_{k_{n,s}}^{(n)} = [A_{k_{n,s}}^{(n)}, b_{k_{n,s}}^{(n)}, B_{k_{n,s}}^{(n)}]$. Let $\mathcal{K}^{(n)}[\cdot]$ denote the classifier at scale n. The classification and parameter assignment can then be written as

$$k_{n,s} \leftarrow \mathcal{K}^{(n)}\left[\frac{\hat{x}^{(n)}_{\partial s}}{\sigma_n}\right] \tag{2.46}$$

$$\theta^{(n)}_s \leftarrow \theta^{(n)}_{k_{n,s}} . \tag{2.47}$$

The classifiers $\mathcal{K}^{(n)}[\cdot]$ and parameters $\theta^{(n)}_{k_{n,s}}$ are obtained during a training phase. For our implementation, we use a tree-based agglomerative clustering method which is described in detail in Appendix A.9. To summarize the approach, we initially partition the space of training samples $\{x^{(n)}_{\partial s}/\sigma_n, z^{(n)}_s/\sigma_n\}$ by performing a vector quantization (VQ) on $\{x^{(n)}_{\partial s}/\sigma_n\}$. For each cluster k, we then calculate the filters $[A^{(n)}_k, b^{(n)}_k]$ as the minimum mean square error linear predictors for $z^{(n)}_s$. Starting with this initial partitioning, we then form a cluster tree by merging pairs of clusters in a greedy fashion. At any given stage, we combine the two clusters whose merging results in the smallest increase in prediction error on the training set. Thus we form a binary tree where each node is associated with its optimal linear prediction filter for the conditional mean. In order to not overfit the classification model, we perform optimal tree pruning[43, 44] using a second data set for cross-validation. The pruning is performed based on the prediction error for the conditional mean only.

The matrices $B^{(n)}_k$ are computed as

$$B^{(n)}_k = \frac{1}{w^2} R^{-1}_k \tag{2.48}$$

where R_k is the conditional sample covariance for class k (see Appendix), and w is a regularization parameter.

We have found it to be advantageous to constrain the classifiers $\mathcal{K}^{(n)}[\cdot]$ and the linear model predictors $A^{(n)}_s x^{(n)}_{\partial s}$ to only depend on $x^{(n)}_{\partial s} - x^{(n)}_s$, that is the difference between $x^{(n)}_{\partial s}$ and its center scaling coefficient. This constraint makes the model invariant to additive shifts in the gray value of the image and therefore improves robustness of training on smaller training sets. The details are listed in Appendix A.9.

2.3 Tomographic Reconstruction

In this section, we discuss the application of the proposed model to Bayesian tomographic reconstruction. In the following, we present the statistical data model $\log p_{y|x}(y|x)$ for both emission and transmission tomography[42]. In both cases, we use the exact Poisson counting statistics. Let y denote the vector of photon counts for all M projections at different angles and displacements. Furthermore, let us define P as the tomographic projection matrix so that P_{i*} denotes the vector formed by its i^{th} row. For transmission tomography, the log-likelihood $\log p_{y|x}(y|x)$ may then be written as

$$\log p_{y|x}(y|x) = \sum_{i=1}^{M} \left(-y_T e^{-P_{i*}x} + y_i(\log y_T - P_{i*}x) - \log(y_i!) \right) \quad \text{(transmission)} \tag{2.49}$$

where y_T denotes the dosage. For emission tomography, the log-likelihood is given by

$$\log p_{y|x}(y|x) = \sum_{i=1}^{M} (- P_{i*}x \ + y_i \log(P_{i*}x) - \log(y_i!)) \qquad \text{(emission)} \ . \tag{2.50}$$

Both (2.49) and (2.50) have the common form

$$\log p_{y|x}(y|x) = -\sum_{i=1}^{M} f_i(P_{i*}x) \tag{2.51}$$

where the $f_i(\cdot)$ are convex and differentiable.

Based on the fine scale model above, we can compute the coarse scale data models $\log p_{y|x^{(l)}}(y|x^{(l)})$ for $l > 0$. Combining (2.15) and (2.51), we obtain

$$\begin{aligned} \log p_{y|x^{(l)}}(y|x^{(l)}) &= -\sum_{i=1}^{M} f_i(P_{i*}\mathcal{I}_l^0 x^{(l)}) & (2.52) \\ &= -\sum_{i=1}^{M} f_i(P_{i*}^{(l)} x^{(l)}) & (2.53) \end{aligned}$$

where $P_{i*}^{(l)} = P_{i*}\mathcal{I}_l^0$. Thus, the coarse scale data models are equivalent to a standard model of the form (2.51) with projection matrix $P^{(l)}$. The columns of $P^{(l)}$ can be computed as linear combinations of the columns of P.

In the following, we discuss the MAP optimization for the tomographic data model. We write the equations for the emission case only, however, all methods analogously apply to the transmission case. Since the form of $\log p_{y|x^{(l)}}(y|x^{(l)})$ is the same for any l, we simplify the notation by omitting the superscripts (l) and assuming reconstruction at the finest scale $l = 0$. The solutions for $l > 0$ are obtained by replacing x with $x^{(l)}$ and P with $P^{(l)}$. To implement the optimization of section 2.2.3 we need to solve

$$\Delta = \arg\max_{\gamma} \left\{ \log p_{y|x}(y|x + \gamma e_i) + \alpha_1 \gamma + \frac{1}{2}\alpha_2 \gamma^2 \right\} . \tag{2.54}$$

The basic concept of the ICD algorithm[41, 42] is to solve (2.54) using a Newton-Raphson strategy. Importantly, the algorithm exploits the sparse nature of the projection matrix P by maintaining a state vector $\tilde{p} = Px$ of the current forward projection of x. Given $\tilde{p}$, we can write the first and second derivatives ψ_1 and ψ_2 of $\log p_{y|x}(y|x)$ with respect to the pixel value x_i as

$$\psi_1 = \frac{\partial}{\partial x_i} \log p_{y|x}(y|x) = - \sum_{\{k: P_{ki} > 0\}} P_{ki} \left(1 - \frac{y_k}{\tilde{p}_k} \right) \tag{2.55}$$

$$\psi_2 = \frac{\partial^2}{\partial x_i^2} \log p_{y|x}(y|x) = - \sum_{\{k: P_{ki} > 0\}} y_k \left(\frac{P_{ki}}{\tilde{p}_k} \right)^2 . \tag{2.56}$$

Based on the notation above, the second order approximation to $\log p_{y|x}(y|x + \gamma e_i)$ with respect to γ is

$$\log p_{y|x}(y|x + \gamma e_i) \approx \log p_{y|x}(y|x) + \psi_1 \gamma + \frac{1}{2}\psi_2 \gamma^2 . \tag{2.57}$$

Thus, the Newton-Raphson update of (2.54) is given by

$$\Delta = -\frac{\psi_1 + \alpha_1}{\psi_2 + \alpha_2} . \tag{2.58}$$

The state vector $\tilde{p}$ can be updated efficiently using $\tilde{p}_k{}' = P_{ki}\Delta + \tilde{p}_k$ for $\{k : P_{ki} \neq 0\}$.

2.4 Experimental Results

In this section, we compare the tomographic reconstruction performance of the proposed framework to that of two fixed-resolution Bayesian methods using Markov random field prior models and to convolution backprojection (CBP). Our results are based on two sets of simulated emission tomography data. The first data set is based on a discrete-valued bar phantom and is used to compare the algorithms quantitatively in terms of reconstruction bias and noise variance. The second data set is a more realistic case and is used to compare the reconstruction methods in terms of visual characteristics, mean square error and speed of computation.

We implemented the wavelet graph model (WGM) and multiresolution reconstruction algorithm using an orthonormal 2-D Haar wavelet decomposition. The number of resolution levels was set to $L = 5$. For a reconstruction resolution of 256 by 256 pixels, this resulted in an array of 16 by 16 scaling coefficients at the coarsest scale. We compared reconstruction results for two different sets of training images used to obtain the nonlinear classifiers and class parameters as described in section 2.2.4. For each case, the training and pruning samples were obtained by computing the wavelet transform of the training images. The number of reconstruction iterations was a function of the scale n such that $K_0 * (\sqrt{2})^n$ iterations were performed at scale n.

In order to demonstrate how the wavelet graph structure by itself can reduce the blockiness commonly encountered with Haar wavelet models, we also implemented a spatially homogeneous linear version of the wavelet graph model. The linear model uses no adaptation and only a single parameter vector $\theta_s^{(n)} \neq f(s)$ for all locations s at scale n. This can be implemented simply by setting the number of classes for the nonlinear classifiers in (2.46) to one.

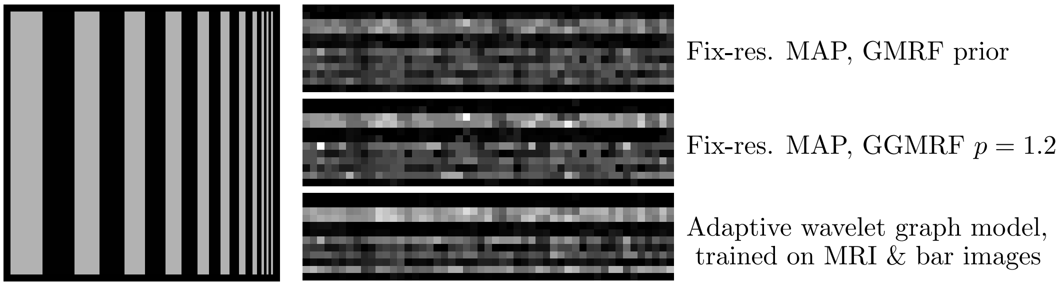
The two fixed resolution MAP reconstruction algorithms were based on a Gaussian Markov random field (GMRF) and a generalized Gaussian Markov random field (GGMRF) prior model respectively. The GGMRF[10] is an edge-preserving, spatially homogeneous MRF that uses a non-quadratic penalty term. For the results shown here, the generalized Gaussian parameter was set to $p = 1.2$. The fixed resolution MAP reconstructions were computed using iterative coordinate descent optimization. The convolution backprojection (CBP) algorithm was implemented for both a ramp filter and a generalized Hamming reconstruction filter with frequency response $H(\omega) = H_{id}(\omega)\,(0.5 + 0.5\cos(\pi\omega/\omega_c))$ for $|\omega| \leq \omega_c$ where $H_{id}(\omega)$ denotes the ideal reconstruction ramp filter.

2.4.1 Bar Phantom Results

The first data set was generated using the bar phantom image shown in Fig. 2.9(a). The binary image of size 115 by 115 pixels has pixel values of 1 for the bars and 0.02 for the background. The projection data was generated by forward projecting a 256 by 256 pixel image with the bar image in the center and a zero background. The projection data was calculated at 128 evenly spaced angles each with 256 displacements. The projection beam was assumed to be an ideal line. The data samples were formed by Poisson random variables with the appropriate mean. The average number of counts per projection was 83.

For this data set, we used two different sets of training images for the nonlinear prediction model for the prior parameters. To approximate the characteristics of typical medical images, the first training set consisted of 40 MRI images, each of size 256 by 256 pixels. In order to show how the proposed method can be adapted in cases where *a priori* knowledge about special characteristics of the phantoms is available, the second training set contained 3000 amplitude and rotational variations of a bar phantom in addition to the 40 MRI images. The phantom shown in Fig. 2.9(a) was not included in the training. Because of the discrete nature of the second training set, we increased the value of δ in (A.16) of the Appendix to $\delta = 0.3$ to obtain reliable covariance parameters R_k. The size of the window ∂s was set to 3 by 3 coefficients.

In order to quantitatively compare the different algorithms, we calculated reconstruction bias and variance statistics for the reconstructed image $\hat{x}_{ij}$. The bias and variance measures use the fact that the original phantom is constant along columns. We first calculated $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ as the mean and variance of the j^{th} column in the



(a) Bar phantom (b) Magnified reconstruction of high frequency region (rotated)

Fig. 2.9. Bar phantom and magnified high frequency region of sample reconstructions.

reconstruction $\hat{x}_{ij}$. Bias and variance were then computed as

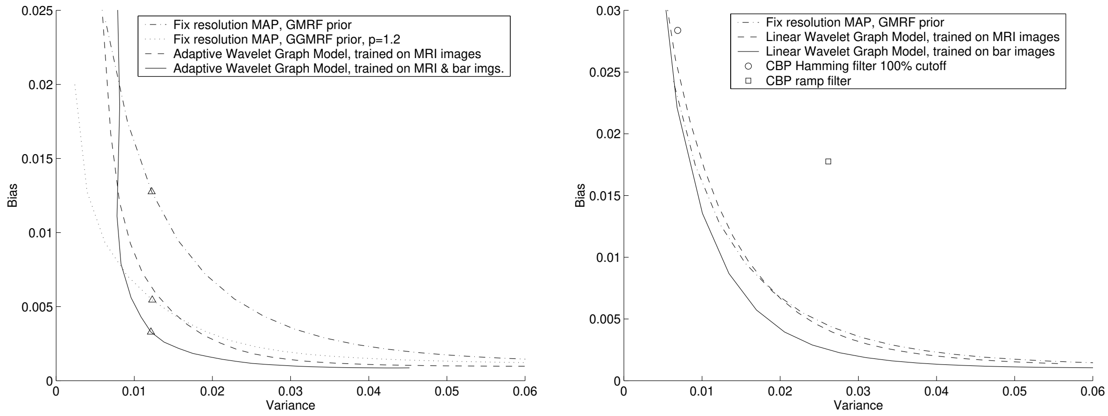
$$\text{bias} = \frac{1}{N} \sum_{j=1}^{N} (\hat{\mu}_j - \mu_j)^2 \tag{2.59}$$

$$\text{var} = \frac{1}{N} \sum_{j=1}^{N} \hat{\sigma}_j^2 \tag{2.60}$$

where μ_j is the value of the j^{th} column in the original phantom and $N = 115$ is the number of columns.

The bias and variance statistics were computed for each reconstruction technique as a function of the regularization parameter. For the wavelet graph model, we varied the parameter w in (2.48) in the range $[0.1, 10]$. For the MRF priors, we varied the scale parameter σ in the range $[0.1, 1.4]$ and for the CBP reconstruction we used a ramp filter as well as a Hamming filter with cutoff frequency $\omega_c = 1$. The proposed algorithm was initialized with a constant image and performed a fixed number of $25 * (\sqrt{2})^n$ reconstruction iterations at each scale n. The fixed resolution ICD MAP algorithms were initialized with a CBP result and then performed 40 reconstruction iterations.

Figure 2.10(a) shows a comparison of the results for the adaptive wavelet graph model and the two fixed resolution Bayesian methods. Each plot corresponds to the bias/variance curve of a single reconstruction method as a function of the regulariza-



(a) Adaptive wavelet graph model (b) Linear models

Fig. 2.10. Bias-variance reconstruction performance of each algorithm as a function of regularization parameter. The triangles in (a) correspond to the bias-variance values for the sample reconstructions in Fig. 2.9(b).

tion parameter. The upper left hand corner of the graphs corresponds to strong regularization (high bias, low variance) whereas the lower right hand corner corresponds to weak regularization (low bias, high variance). The results indicate that the adaptive wavelet graph model trained on the set of MRI images performed significantly better than the GMRF method and performed comparably to the GGMRF based method. While the GGMRF performed best in the low variance region, its residual bias in the high variance region was slightly higher. Specifically, the GGMRF prior model has a tendency to introduce bias resulting in a loss of detail even when very little regularization is applied. When trained on the combination of bar phantoms and MRI images, the adaptive wavelet graph model outperformed both fixed resolution Bayesian methods. Figure 2.9(b) shows the magnified high frequency region of sample reconstructions corresponding to the bias/variance points marked by triangles in Fig. 2.10(a). The adaptive wavelet graph model reconstruction better resolves the high frequency bars than the two fixed resolution Bayesian reconstructions.

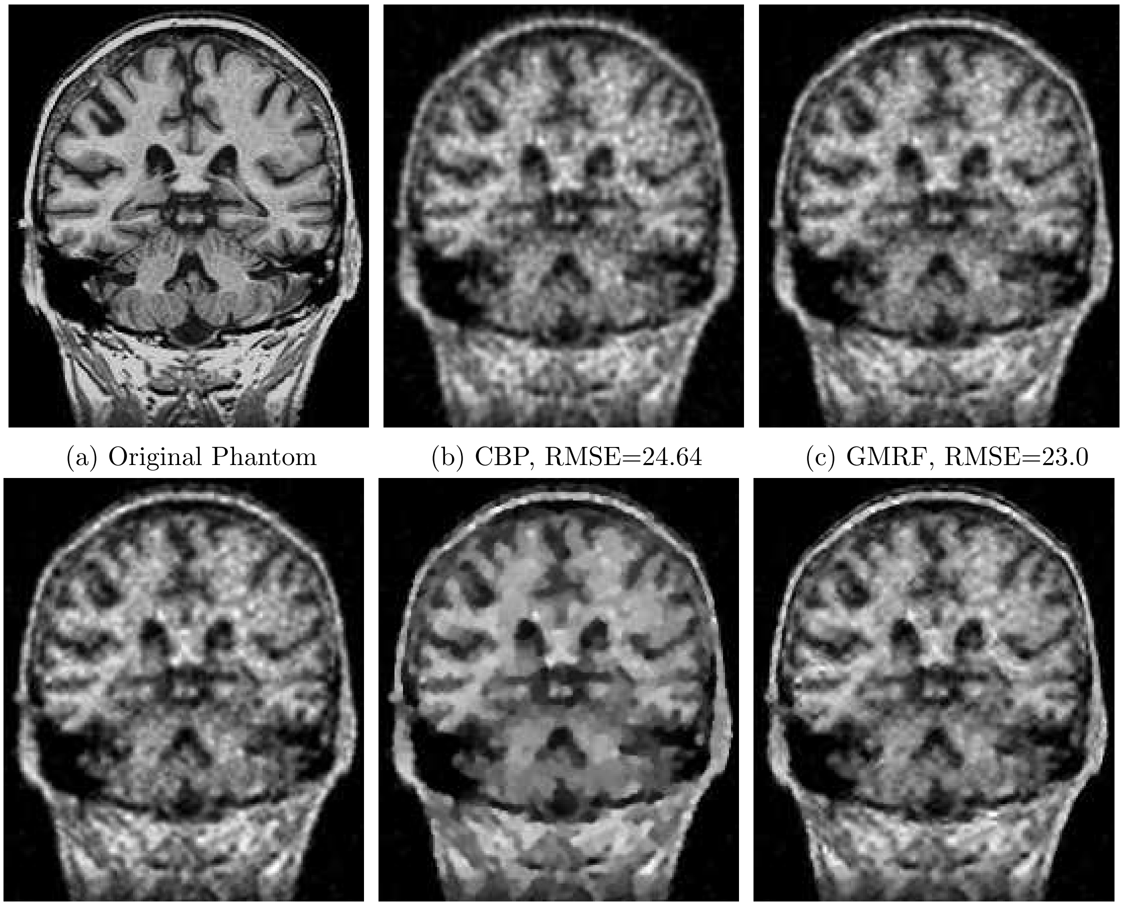
Figure 2.10(b) shows the bias/variance curves for the linear, non-adaptive, wavelet graph model in comparison to those for the fixed resolution GMRF technique and for CBP. Trained on the set of MRI images, the linear wavelet graph model performed comparably to the GMRF method. When trained on the combination of bar phantoms and MRI images, the linear wavelet graph model achieved lower bias at equal variance as compared to the GMRF result.

2.4.2 Medical Image Phantom Results

For our second data set, we used used a magnetic resonance imaging (MRI) reconstruction image as our phantom data. The original phantom is shown in Fig. 2.11(a). The resolution of the original image was 256 by 256 pixels with a mean pixel value of 0.32. The projection data was calculated at 128 evenly spaced angles each with 256 parallel projections assuming a field of view of 20cm square. To better illustrate the deblurring potential of the algorithm, we assumed a projection beam of width 2.34mm, corresponding to three times the projection spacing. The beam profile was assumed to be triangular. The data samples were formed by Poisson random variables with the appropriate means. The average number of counts per projection was 235.

For this data set, the adaptive and linear wavelet graph models were trained using the same set of 40 MRI images as for the first data set. The phantom in Fig. 2.11(a) was not included in the training set. The size of the window ∂s was set to 5 by 5 coefficients.

To compare the results from different reconstruction algorithms, the regularization parameter for each algorithm was adjusted to minimize the mean square error of the reconstruction. This resulted in a cutoff frequency of $\omega_c = 0.63$ for the CBP using a generalized Hamming filter, $\sigma = 0.2$ for the GMRF prior, $\sigma = 0.24$ for the GGMRF prior, $w = 1.075$ for the linear wavelet graph model and $w = 1.0$ for the adaptive wavelet graph model. All iterative algorithms were run for a large number of iterations to insure sufficient convergence. The proposed multiresolution algorithm was initialized with a constant image and performed $148 * (\sqrt{2})^n$ iterations at each scale n. The fixed resolution Bayesian methods were initialized with the CBP reconstruction and performed 500 iterations.



(a) Original Phantom (b) CBP, RMSE=24.64 (c) GMRF, RMSE=23.0

(d) Lin. WGM, RMSE=23.59 (e) GGMRF, RMSE=22.21 (f) Ad. WGM, RMSE=22.60

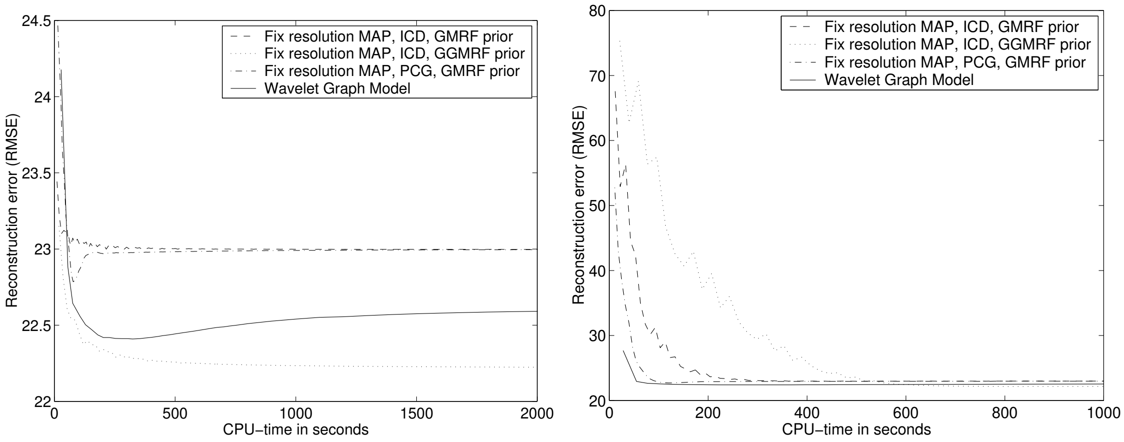
Fig. 2.11. Original phantom (a) and reconstructions: (b) convolution backprojection, (c) fixed resolution MAP with GMRF prior, (d) proposed algorithm using linear wavelet graph model, (e) fixed resolution MAP with GGMRF prior, (f) proposed method using adaptive wavelet graph model.

The CBP reconstruction in Fig. 2.11(b) is noisy in the uniform image regions and contains blurry edges. The GMRF MAP reconstruction in Fig. 2.11(c) is less noisy than the CBP; however, the quadratic regularization function of the Gaussian MRF results in blurred edges. The reconstruction (d) obtained using the linear wavelet graph model is visually roughly comparable to the GMRF result (c) but contains slight blocking artifacts and has higher RMSE. However, considering that this is a spatially homogeneous linear model using a Haar wavelet prior without shift-averaging, the reconstruction is surprisingly smooth. This is a result of the overlapping dependency structure of the wavelet graph structure. Fig. 2.11(e) shows the fixed resolution GGMRF MAP reconstruction. The result has sharper edges as compared to the GMRF case (c) and achieves the lowest mean square error of all four methods. Visually, however, the reconstruction (e) is of poor quality since it suffers from considerable loss of detail. The reconstruction using the adaptive wavelet graph model in Fig. 2.11(f) is superior to that of the other four methods. In comparison to the GMRF case (c), the reconstruction (f) contains sharper edges while the noise in the uniform regions is better suppressed. The mean square error is lower than for the GMRF case (c) but higher than for the GGMRF result (e). In comparison to (e), however, the proposed method preserves more detail.

2.4.3 Computational Efficiency Comparison

We compared the reconstruction error convergence of the proposed algorithm to iterative coordinate descent (ICD) and preconditioned gradient descent (PCG) fixed resolution Bayesian reconstruction techniques. The PCG algorithm used a simple bent line search and was implemented as described in [45] except that it did not use a factored system matrix. The convergence measurements for all algorithms were taken for reconstructions of our second data set as in Fig. 2.11. The ICD algorithm was used with the GMRF and GGMRF prior models. The PCG algorithm was used only with the GMRF prior since the Newton-Raphson line search of our implementation was not applicable to the GGMRF prior.

Figure 2.12(a) and (b) show the reconstruction error convergence as a function



(a) CBP initialization

(b) Constant initialization

Fig. 2.12. Error convergence of the proposed algorithm compared to fixed resolution MAP reconstructions with GMRF and GGMRF prior models.

of CPU-time for the different algorithms. The reconstruction error was calculated as the root mean square error between the reconstruction and the ground truth. For the ICD and PCG algorithms, the graphs were obtained by measuring the CPU-time after each iteration. For the proposed algorithm, the graphs were obtained by measuring the CPU-times for full reconstructions using $K_0(\sqrt{2})^n$ iterations at scale n and varying K_0. The CPU-times were measured on a 700 MHz Pentium III CPU. They do not include the time needed for pre-processing such as the computation of the tomographic system matrices that can be performed off-line. Figure 2.12(a) shows the error convergence for the case where all algorithms were initialized with the CBP reconstruction of Fig. 2.11(b). For this initialization, ICD and PCG converge very quickly, particularly for the GMRF prior. The proposed algorithm has slighly slower convergence. Figure 2.12(b) shows the convergence results when all reconstruction algorithms were initialized to a constant image whose forward projection matched the total number of counts of the projection data. For this initialization, the fixed resolution ICD algorithm has slow convergence due to the slow low-frequency convergence of the ICD algorithm [41]. The proposed algorithm converges fastest of all the methods in this comparison. Thus, the multiresolution technique can provide a computational advantage in cases where an initialization with the correct low-frequency behavior is not available. Specifically, the multiresolution approach allows us to use an ICD optimization technique without requiring a CBP initialization. This is an advantage for systems with non-circular geometry[46] or limited angle problems where CBP initializations are not easily obtained.

2.5 Conclusions

We propose a wavelet graph prior model in conjunction with a multiresolution Bayesian reconstruction algorithm applicable to tomographic reconstruction. The wavelet graph prior model has a dependency structure that is more general than a quadtree. This enables the model to produce smooth estimates even for a Haar wavelet basis. Furthermore, the wavelet graph structure is such that the optimal model for a stationary process is homogeneous, resulting in a substantial reduction in the number of model parameters. The multiresolution reconstruction algorithm uses the wavelet graph prior model but performs a sequence of MAP optimizations in the space-domain. The space domain formulation allows us to efficiently enforce the pixel positivity constraint and to preserve the sparseness of the tomographic projection operator. Our experimental results indicate that the proposed framework can improve reconstruction quality over commonly used fixed resolution Bayesian methods.

3. QUANTITATIVE COMPARISON OF FBP, EM, AND BAYESIAN RECONSTRUCTION ALGORITHMS, INCLUDING THE IMPACT OF ACCURATE SYSTEM MODELING, FOR THE INDYPET SCANNER

Abstract

We quantitatively compare filtered backprojection (FBP), expectation maximization (EM) and Bayesian reconstruction algorithms for the IndyPET animal tomography scanner. The IndyPET scanner is a small to intermediate field of view scanner with non-circular detector geometry. In contrast to previous approaches that largely rely on Monte Carlo simulations, we obtained an empirical system kernel for the IndyPET scanner directly based on scans of line source phantoms. The empirical system kernel is incorporated into the forward model of EM and the Bayesian reconstruction algorithms to achieve resolution recovery.

Reconstruction quality is compared for FBP, EM and Bayesian techniques with three different prior models. In addition, we compare reconstruction quality for the empirical system kernel and an analytically calculated, triangular kernel. Reconstruction quality is evaluated for data acquired on the IndyPET scanner using a Hoffman 3-D brain phantom and a bar phantom. Furthermore, we present an MSE analysis for a simulated phantom containing a lesion on an approximately uniform background. The results indicate, that using an accurate system kernel significantly improves the bias-variance envelope of the iterative techniques over FBP. We found that when an inaccurate kernel was used, EM and Bayesian MAP performed similarly. However, when an accurate system kernel was used for both algorithms, Bayesian techniques achieved significantly higher reconstruction quality than EM.

3.1 Introduction

Iterative methods for positron emission tomography (PET) reconstruction can theoretically improve reconstruction quality by accurately modeling the data acquisition process and by regularizing the ill-posed inverse problem. While iterative methods have been studied extensively in the literature, clinical PET remains largely dominated by filtered backprojection (FBP) with some growing interest in the ordered subsets expectation maximization (OSEM) algorithm[47].

An important advantage of iterative algorithms that is often not exploited in practical implementations is the ability to incorporate an accurate model of the tomography scanner into the reconstruction algorithm. Specifically, iterative algorithms can achieve resolution recovery by accounting for sinogram blurring due to detector crystal penetration, inter-crystal scatter, depth of interaction and other effects[48, 49, 50]. Early investigations for the expectation maximization (EM) reconstruction algorithm[51] have found that a more accurate system kernel can reduce bias in region-of-interest (ROI) quantitation[52, 48]. However, the more accurate kernel increased the variance of the ROI estimates obtained from EM reconstructions, even for simulated projection data that had been generated with the same system kernel. This effect is likely due to the fact that a more accurate system kernel results in a more ill-posed reconstruction problem[52], which is consistent with the observation of ringing artifacts[53]. The increased ill-posedness adversely affects the quality of EM reconstructions, particularly, since EM does not use any explicit regularization.

Bayesian reconstruction techniques[6, 7, 41, 54, 55, 56, 57, 42] on the other hand, explicitly address the ill-posedness by regularizing the reconstruction through use of a prior model such as a Markov random field (MRF)[5, 6, 7, 8, 9, 10, 11]. Consequently, we would expect Bayesian methods to provide better reconstruction quality as compared to EM, especially when a more accurate system kernel is used. Qi, *et al.*, [53, 45] have successfully applied 3D Bayesian MAP reconstruction combined with an advanced response model[50] to data from the microPET[58] and the Siemens/CTI

HR+[59] scanners. The results demonstrate that this approach can increase ROI quantitation accuracy and FWHM resolution at lower variance as compared to FBP and OSEM[45, 60].

Our objective for this study is to provide a quantitative comparison of FBP, EM, and Bayesian reconstruction algorithms and to investigate the impact of accurately modeling the scanner for iterative reconstruction. We quantitatively evaluate reconstruction quality for simulated data and for phantom data acquired on the IndyPET tomography scanner[46]. The IndyPET scanner has been developed as a high resolution dedicated research scanner for use in small animal, intermediate sized animal and small field of view human (brain, breast) imaging applications. As shown in Fig. 3.1, the distinguishing feature of this scanner is a non-circular detector geometry consisting of two, approximately planar detector banks mounted on a rotatable gantry.

To evaluate the impact of accurately modeling the scanner for iterative reconstruction, we have obtained an empirical system kernel for the IndyPET scanner based on scans of line source phantoms. This approach is different from previous studies which have typically relied on analytical modeling of detector properties[61, 62] or on Monte Carlo simulations based on statistical detector models[63, 64, 65, 50]. For the IndyPET scanner, obtaining an empirical system kernel is greatly facilitated by its non-circular detector geometry. Specifically, the detector geometry enables us to measure the full range of incident photon angles using only sources positioned near the center of the field of view. In contrast, a circular detector geometry would require more measurements since the incident photon angle is a strong function of radial source position, and for a source in the center of the field of view, all photon incidences are approximately normal to the detector surfaces. As shown in Fig. 3.1(c), our results indicate that, even for a centered source and normal photon incidence, our empirical system kernel is very different from the triangular shape expected for ideal detectors. In fact, treating the detectors as a black box may be simpler and potentially more accurate than analytical modeling. One approximation in our current model is that we do not account for the depth dependence of the system kernel, however, the approach can be extended to take this effect into account.

We compare reconstruction quality for FBP, EM and Bayesian techniques. For the Bayesian techniques, we compare maximum *a posteriori* (MAP) reconstructions using two different priors, a Gaussian Markov random field (GMRF) and a generalized Gaussian Markov Random field (GGMRF)[10] model. In addition, we evaluate a recently proposed Bayesian algorithm that employs a nonlinear wavelet prior model to perform space-adaptive regularization[66, 67]. For all Bayesian methods, the MAP optimization was implemented using iterative coordinate descent (ICD) optimization[10, 42]. For EM and Bayesian GMRF, we compare reconstructions for the empirical system kernel and an analytically calculated triangular system kernel. In addition, we compare Bayesian GMRF reconstructions for parametric system kernels of different widths.

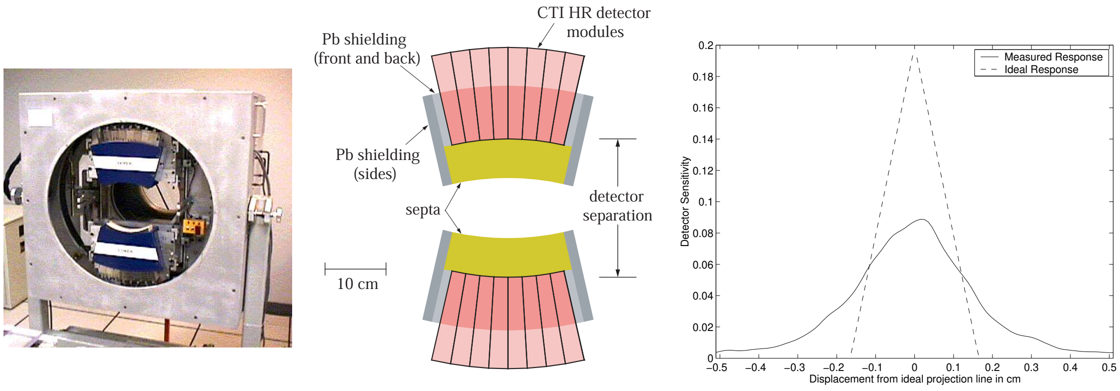
Three data sets are used for this study. The first data set was acquired on the IndyPET scanner using a bar phantom. This phantom is similar to an X-ray resolution target and generates square-wave emission profiles with different periodicities. Reconstructions of this phantom are analyzed in terms of spatial bias and variance statistics. Bias and variance are calculated by fitting Fourier series to the reconstructed wave forms and comparing the fitted coefficients to those of an ideal square wave. Different combinations of reconstruction algorithms and system kernels are compared in terms of their bias-variance envelopes which are obtained by varying a smoothing parameter for each reconstruction method. In addition, we use the bar phantom data to compare the convergence speed of EM and Bayesian GMRF MAP. The second data set was collected on the IndyPET scanner using a Hoffman 3-D brain phantom. Reconstruction images generated by different combinations of reconstruction algorithms, system kernels and smoothing parameters are compared in terms of visual quality. The third data set consists of simulated projection data corresponding to a circular lesion that was added to an approximately uniform background. Twenty realizations of noisy projection data were generated using the empirical system kernel for the IndyPET scanner. Reconstructions for different algorithms and smoothing parameters are analyzed quantitatively in terms of the expected mean square error (MSE) for the lesion and the background regions. While mean square error is not a direct indicator of clinical performance, it provides an objective measure of the quantitative accuracy of the reconstruction. In comparison, we have found the clinically more relevant contrast recovery ratio (CRC) to be a strong function of the procedure and regions used for its computation, an observation that is consistent with earlier reports[48]. For this study, we limit our analysis to MSE to avoid the additional complication of choosing a task-specific computation method for the CRC.

One shortcoming of Bayesian MAP techniques is that the reconstructions have spatially varying resolution[68]. Specifically, high activity regions are regularized more strongly, which can be suboptimal for lesion data as in our simulated data set. Modified algorithms have been proposed that result in constant resolution[68, 69] or constant contrast-to-noise ratio[70, 71] reconstructions. For our simulated data set, we evaluate the impact of the modification proposed by Fessler and Rogers[68].

3.2 IndyPET Scanner

The IndyPET[46] positron emission tomography (PET) scanner has been developed as a high resolution, high sensitivity dedicated research scanner for use in small animal, intermediate sized animal, and small field of view human (brain, breast) imaging applications. As illustrated in Fig. 3.1, the distinguishing feature of the scanner is the use of two, approximately planar detector banks with adjustable separation which are mounted on a rotatable gantry. In comparison to using a full ring of detectors, the scanner's geometry results in reduced parallax and approximately uniform resolution throughout the field of view. The scanner uses removable septa which can be placed in front of the detectors. The use of septa greatly reduces the influence of scatter. Septa were included for all measurements reported in this study.

The two detector banks of the scanner consist of 8 CTI-HR[72] BGO detector modules each. The detector crystal in each of the 16 detector modules is cut into



(a) Scanner (b) Detector geometry (c) Detector kernel

Fig. 3.1. The IndyPET tomography scanner (a) employs two approximately planar detector banks mounted on a rotatable gantry (b). The plot (c) shows an empirical system kernel corresponding to the response of a single detector pair as a function of source offset from the ideal projection line. This kernel was obtained for a projection with normal photon incidence and a source centered between the detectors. The empirical result is considerably different from the analytically calculated, triangular kernel for ideal detectors.

a 7 (transaxial) by 8 (axial) array of crystal segments with a transaxial spacing of 3.3 mm and an axial spacing of 6.3 mm. The crystal depth is 30 mm.

The scanner's bucket controllers and coincidence electronics are built by CTI. Coincidence events are sorted into sinogram bins corresponding to angle and displacement using the CTI ACS-II real-time sorter. For the data sets used in this study, the ACS-II sorter was configured to perform automatic randoms subtraction, thus our projection data were precorrected for randoms. The bucket electronics report the singles count rates for each bucket with and without deadtime correction. For the results reported here, the ratios of the reported corrected and uncorrected singles rates were used to obtain deadtime correction factors for the reconstructions.

Figure 3.2(a) shows the resolution of the IndyPET scanner as a function of source offset from the center for different detector separations. The resolution was measured as the FWHM of line source reconstructions obtained using FBP with a ramp filter.

The resolution throughout the field of view is within 25% of the resolution at the center.

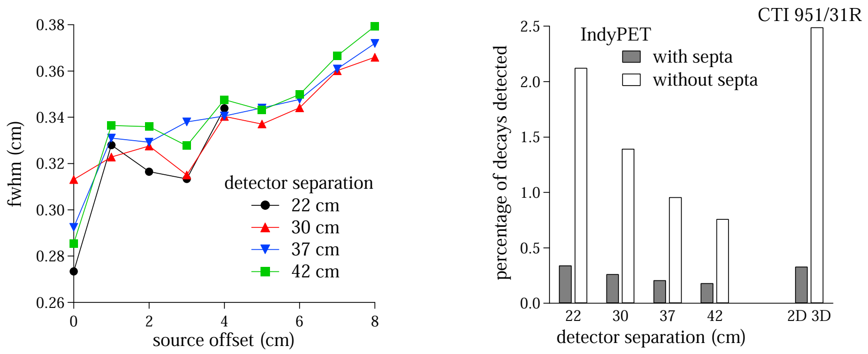
In addition to the approximately uniform resolution, a second advantage of the IndyPET scanner's geometry is the variable detector separation which can be reduced to increase sensitivity for smaller subjects. This results in added flexibility since the geometry can be adapted to perform well in a large variety of imaging tasks from scanning of mice to human heads. However, the scanner's overall sensitivity is lower than that of a circular scanner with the same radius. Figure 3.2(b) shows the sensitivity of the IndyPET scanner with and without septa in comparison to that of a CTI 951/31 scanner. The sensitivity was obtained by calculating the percentage of detected decays for a 1ml sphere source in the center of the field of view. Because the detectors do not form a full ring, data must be collected at different gantry angles, resulting in increased imaging time. However, the geometry is more economical with respect to the number of detectors required as compared to a circular scanner.

3.2.1 IndyPET Data Acquisition

Data collection on the IndyPET scanner is performed in a 'step and shoot' mode where multiple frames of data are collected with equal durations at uniformly spaced gantry angles. Because the detectors do not form a complete ring, the data collected for each frame form a diamond shaped segment from a complete sinogram as illustrated in Fig. 3.3. In cases where the total acquisition time is such that radioactive decay becomes significant, data can be collected using multiple gantry sweeps of 180 degrees with shorter acquisition duration per frame. This procedure avoids the problem of imaging different angular positions at significantly different activity levels.

3.2.2 Non-uniform Sinogram Spacing

The geometry of the IndyPET scanner results in projection measurements that are non-uniformly spaced in angle and displacement. Iterative reconstruction algorithms such as EM and Bayesian techniques can account for the non-uniform spacing in their forward projection operator. FBP, however, requires a sinogram with a full set of evenly spaced angles and uniform displacements. To obtain projection mea-



(a) Resolution

(b) Sensitivity

Fig. 3.2. Resolution and sensitivity of the IndyPET scanner. The plots in (a) show the resolution of the IndyPET scanner as a function of source offset from the center for different detector separations. The resolution was measured as the FWHM of line source reconstructions using FBP with a ramp filter. The sinogram for the resolution measurement was generated as described in section 3.2.2, with the exception of using only the 12 central rows of the sinogram segments for each frame. The barplot in (b) shows the sensitivity of the scanner for different detector separations. The sensitivity was obtained by calculating the percentage of detected decays for a 1ml sphere source in the center of the field of view.

surements that can be used to generate an approximate sinogram suitable for FBP, we adjust the detector bank separation so that the 180 degree range for the gantry rotation is an integral multiple of the angular size of a single detector segment. The gantry rotation steps can then be selected as integer multiples of the detector width, which results in approximately uniformly spaced projection angles. Based on these considerations, all projection data for this study were acquired using a detector separation of 37.815 cm and 45 frames with gantry rotation steps of 4 degrees between frames. To generate a sinogram for FBP reconstruction, the diamond shaped data sets for all frames are first individually corrected for relative detector sensitivity, deadtime and decay. A full sinogram is then obtained by summing the corrected diamond shaped segments after they have been shifted by the appropriate number of rows to account for the gantry rotation. The entries of the summed sinogram are individually normalized by the number of frames that contributed to the respective

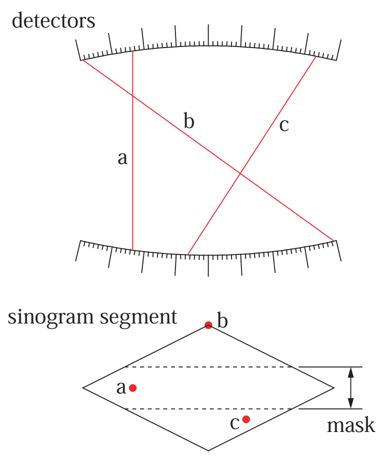


Fig. 3.3. Illustration of diamond shaped sinogram segment for a single gantry position. The points a,b,c illustrate the positions of sample lines of response relative to the detectors and the sinogram. For FBP reconstructions, an approximate complete sinogram is generated using a mask to limit the angular range of each frame. This limits the degree of non-uniform angular projection spacing due to the curvature of the detector banks.

position in the complete sinogram. The resulting sinogram is approximate because the center of curvature of the detector banks does not coincide with the center of ganty rotation, and thus, even when rotating by integer multiples of detector segments, the projection angles are non-uniformly spaced. To limit the effect of this approximation, we only use use the central angular portion of the sinogram segment for each frame as illustrated in Fig. 3.3. For the FBP results reported in this study, only the central 32 (out of 56) rows of each sinogram segment were used. Because of the diamond shape, the 32 central rows contain 81% of the total number of elements in each sinogram segment.

We emphasize, that the generation of a complete sinogram and the masking of sinogram segments were only used for FBP reconstruction. All iterative reconstructions reported in this study used the unmodified, original projection data where each entry in each frame sinogram segment was parametrized as a separate projection associated with values for the projection angle and the displacement that were calculated taking the curvature of the detector banks into account. Furthermore, for iterative reconstruction, the corrections for detector sensitivity, deadtime and decay were not applied to the projection data, but rather, were incorporated into the forward projection operator of the reconstruction algorithm.

3.2.3 Modeling the System Kernel

The basic concept of EM and iterative Bayesian reconstruction algorithms is to forward project a trial image at each iteration and to compare the calculated projections to the measured ones. Importantly, the forward operator can precisely model the scanner's geometry and can incorporate an accurate system kernel that accounts for sinogram blurring due to the width of the detector segments and effects such as crystal penetration and inter-crystal scatter. Incorporating the sinogram blurring into the forward model of the reconstruction algorithm effectively deblurs the reconstruction image and thus allows for resolution recovery.

For this study, we obtained an empirical model of the system kernel for the IndyPET scanner based on scans of line sources. This approach is different from previous studies which have typically relied on statistical detector modeling[61, 62] and Monte Carlo simulations[63, 64, 65, 50]. We compare the reconstruction quality of iterative reconstruction algorithms for the empirical kernel and the triangular kernel shown in Fig. 3.1(c). The triangular kernel is the analytical result for a source centered between ideal detectors and is obtained by calculating the solid angle spanned by the detectors seen from the source as a function of source displacement from the ideal projection line. In addition, we perform a limited experiment to evaluate reconstruction quality for a set of parametric kernels. Note, that the triangular and the parametric kernels are 1-D functions of source displacement and do not incorporate any dependencies on projection angle θ. In contrast, the empirical kernel is a 2-D kernel that is a function of both the projection angle and the displacement of the source from the ideal projection line.

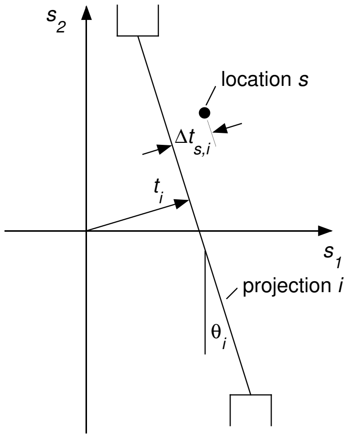


Fig. 3.4. We parameterize the detector kernel $h_i(s)$ by the displacement difference $\Delta t_{s,i}$ between projection i and location s.

Parameterization of System Kernel

First, we introduce our notation and parameterization of the system kernel which is incorporated into the tomographic projection matrix. Let x be the N-dimensional vector of emission rates of the pixels in raster order. Further, let y denote the vector of photon counts for the M projections for all projection angles and displacements over all frames. We then define the $M \times N$ tomographic projection matrix P such that P_{ij} is proportional to the probability that an emission from pixel j is registered by the i^{th} detector pair. Given this notation, the expectation $E[y]$ is given by the forward projection Px,

$$E[y] = Px \ . \tag{3.1}$$

For our model, the entries P_{ij} are computed based on the scanner's geometry and its system kernel. Corrections for effects such as attenuation, detector normalization, decay and deadtime are discussed in section 3.3. The system kernel specifies the response of the detector pairs as a function of source position. Let $s = (s_1, s_2)$ denote the Cartesian coordinates in the field of view of the scanner and define $h_i(s)$

as the kernel of detector pair i. Then P_{ij} can be computed as the 2-D integral

$$P_{ij} = \int_{s \in S_j} h_i(s)\, ds \tag{3.2}$$

where S_j is the set of coordinates s in pixel j. In the following, we assume that pixels in the reconstruction image are considerably smaller than the detectors so that $h_i(s)$ is approximately constant over the support of a single pixel. Then

$$P_{ij} \approx h_i(s_{c_j}) A_j \tag{3.3}$$

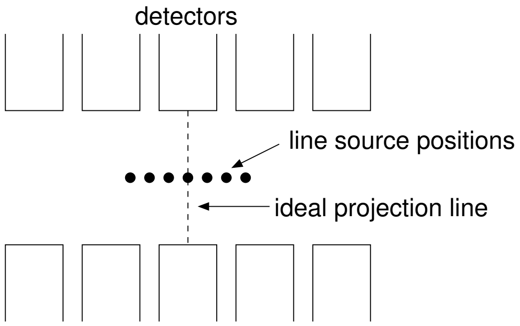
where A_j is the area and s_{c_j} is the center coordinate of pixel j. Note, that for larger reconstruction pixels, this assumption can be enforced by computing P for a sub-sampled pixel grid and summing the contributions for each pixel[53].

Based on the tomographic coordinate system of projection angles and displacements, we model the kernel $h_i(s)$ of a single detector pair as i follows. As illustrated in Fig. 3.4, define the ideal projection i as the line connecting the surface centers of the two detectors of pair i. Let θ_i be the projection angle and define t_i as the displacement of projection i from the coordinate origin. For an arbitrary location $s = (s_1, s_2)$, we can calculate the displacement difference $\Delta t_{s,i}$ between location s and projection i as $\Delta t_{s,i} = s_1 \cos(\theta_i) + s_2 \sin(\theta_i) - t_i$. We then model the detector kernel $h_i(s)$ as

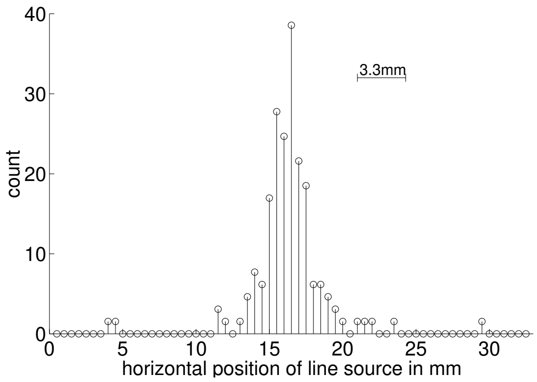
$$\begin{aligned} h_i(s) &= \tilde{h}_i(\Delta t_{s,i}) && (3.4) \\ &= \tilde{h}_i(s_1 \cos(\theta_i) + s_2 \sin(\theta_i) - t_i) && (3.5) \end{aligned}$$

where $\tilde{h}_i$ is a 1-D function of $\Delta t_{s,i}$. The parameterization above ignores the depth dependence of detector kernel, i.e., the coordinate parallel to the projection line. This dependence could be added as an additional parameter.

Considering effects such as detector penetration, we expect the differences between the $\tilde{h}_i(\Delta t_{s,i})$ for different detector pairs i to be predominantly due to the incident photon angles for the different projections. For the IndyPET scanner, it is a good approximation to assume the incident photon angle to be equal to the



(a) Scans at different line source positions



(b) Detector pair count vs. source pos.

Fig. 3.5. Scans of line source phantoms were used to obtain an empirical model for the scanner's system kernel. We note that the FWHM of the detector response in (b) is approximately equal to the detector crystal spacing. The scale-bar indicates the crystal spacing of 3.3 mm.

projection angle since the detector banks have low curvature. Based on these considerations, we model the detector kernels h_i as a function of an overall system kernel h that is parameterized by projection angle

$$h_i(s) = \tilde{h}_i(\Delta t_{s,i}) = h(\theta_i, \Delta t_{s,i}) \ . \tag{3.6}$$

The parameterization above does not account for the boundaries between different detector blocks, an effect that has typically been ignored[50, 53, 45].

The simplicity of the parameterization in (3.6) is a direct consequence of the geometry of the IndyPET scanner. In contrast, the parameterization for a circular detector geometry would require an additional parameter since the detector orientation relative to the projection lines is a function of the displacement of the projection from the center of the field of view.

Empirical Model Based on Line Source Scans

We obtained an empirical system kernel $h(\theta_i, \Delta t_{s,i})$ for the IndyPET scanner using scans of an axially oriented line source phantom, resulting in an approximate point source in the transaxial 2-D reconstruction plane. The line source was mounted on a translation stage positioned near the center of the field of view and data sets were collected as the source was moved horizontally in steps of 0.5 mm, see Fig. 3.5(a). Figure 3.5(b) shows an example of the signal measured for a single detector pair as a function of the horizontal position of the line source.

Based on the collected line source measurements, we have developed the following modeling procedure to obtain an empirical system kernel. Define s_k as the k^{th} horizontal position of the line source on the translation stage and let $y_{k,i}$ denote the measured projection count for detector pair i at source position k. To eliminate differences due to the different overall detector gains c_i, the count data are normalized to $\tilde{y}_{k,i} = y_{k,i}/c_i$ where the c_i are obtained from a normalization scan. For each $\tilde{y}_{k,i}$ we calculate the projection angle θ_i and the displacement difference $\Delta t_{s_k,i}$ between the ideal projection i and source position s_k. Based on this parameterization, the $\tilde{y}_{k,i}$ are accumulated onto a 180×200 point meshgrid with dimensions $-\theta_{max} \leq \theta \leq \theta_{max}$ and -10 mm $\leq \Delta t \leq 10$ mm, where for our detector separation of 37.815 cm, the maximum projection angle is $\theta_{max} = 24.87$ degrees. The accumulation is performed using nearest neighbor interpolation since due to the geometry of the IndyPET, the projection data are non-uniformly spaced in angle and displacement. After accumulating the counts on the grid, the grid values are normalized by the number of data points that have been assigned to each grid location. Grid points that have not been assigned any data points are computed using bilinear interpolation. To reduce the effect of measurement noise, the gridded data are smoothed using a 2-D Gaussian lowpass kernel with a standard deviation of 2 samples, corresponding to 0.2 mm in the displacement coordinate and to 0.5 degrees in the angular direction. The filtered, gridded data are our model for the system kernel $h(\theta, \Delta t)$ as shown in Fig. 3.6. This kernel was computed using $K = 65$ different line source positions.

The result in Fig. 3.6 does not indicate a significant dependence on projection angle. This effect is due to the limited range of projection angles acquired with the IndyPET scanner. Furthermore, our measurements become noisier towards larger angles since, due to the geometry, fewer data points were available in this range.

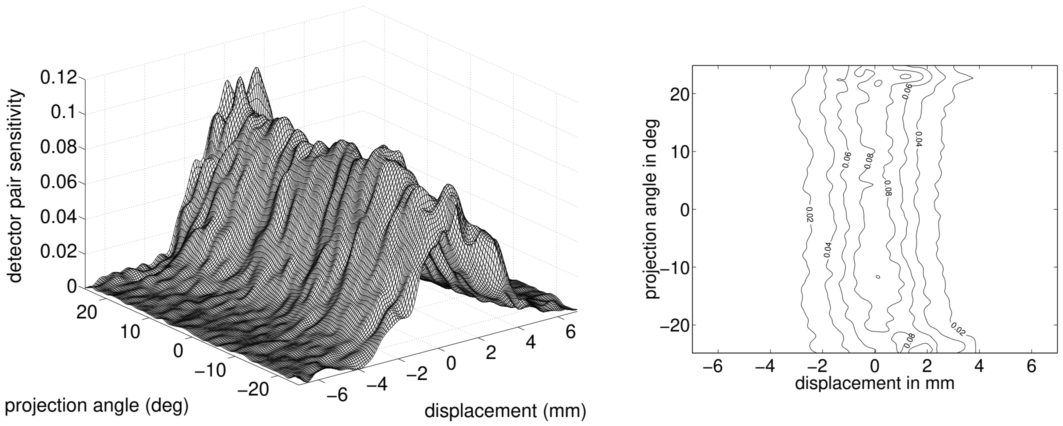


Fig. 3.6. Empirical system kernel $h(\theta, \Delta t)$ for the IndyPET scanner. The kernel is parameterized by projection angle θ and the displacement Δt of the source from the ideal projection line.

Note, that for iterative reconstruction, we truncate the system kernel to the range $|\Delta t| < 5$ mm.

A practical challenge in performing the modeling procedure above is to estimate precisely the physical positions s_k of the line source. While the horizontal increments are accurately known by virtue of the translation stage, it is difficult to obtain an accurate estimate for the initial position s_0. Performing the modeling procedure with a constant offset in the estimated s_k implies that reconstructions of projection data for a single, zero gantry position will be shifted by the same offset. However, when the gantry is rotated, so is the shift in each single-frame reconstruction, such that the reconstructed positions for a point source will move on a circle. For reconstructions of complete data sets consisting of multiple frames, this results in blurring. Since our objective in modeling the system is to achieve resolution recovery, it is critical that we estimate the s_k with an uncertainty much smaller than the width of a detector crystal segment.

Interestingly, the circular blurring effect described above can be used to obtain an accurate estimate of s_0. To summarize the approach, we first obtain an approximate initial estimate of s_0 directly from the sinogram. Based on this estimate, we perform the modeling procedure to obtain an approximate kernel $h(\theta, \Delta t)$. This kernel is then used to reconstruct projection data of a point source taken at different gantry angles. By reconstructing the data set frame-by-frame, we can estimate the reconstructed position of the point source for each gantry angle. This allows us to estimate the phase and radius of the circular blurring, which in turn give an improved estimate of s_0. Note, that this procedure calibrates the offset relative to the center of rotation of the gantry, independent of errors in the detector mounting on the gantry. The details of the estimation and claibration procedure are listed in Appendix A.10.

3.3 Reconstruction Algorithms

In this study, we compare reconstructions obtained using FBP, EM and Bayesian techniques with three different prior models. For each algorithm, we evaluate reconstruction quality as a function of one smoothing parameter; filter cutoff frequency for FBP, the number of iterations for EM, and the smoothing parameter of the prior for the Bayesian methods.

3.3.1 Statistical Model for Iterative Reconstruction

For EM and the Bayesian methods, we use the exact Poisson counting statistics to model the data acquisition. Let y denote the vector of photon counts for all M projections. For the data collected on the IndyPET scanner, the number of projections for a single reconstruction plane is $M = 45 \times 3136$, corresponding to 45 acquired frames, each associated with a diamond shaped sinogram segment of 3136 entries. As mentioned in section 3.2, the counts y are pre-corrected for random coincidences. Let P be the tomographic projection matrix so that P_{i*} denotes the vector formed by its i^{th} row. The shifted Poisson log-likelihood[73] may then be written as

$$\log p_{y|x}(y|x) = \sum_{i=1}^{M} \Big(-(P_{i*}x + s + 2r) + (y_i + 2r)\log(P_{i*}x + s + 2r) - \log((y_i + 2r)!) \Big) \tag{3.7}$$

where s and r denote contributions due to scatter and random coincidences respectively. Note that in general, scatter and randoms are a function of the projection index i; however, for this study, we assumed identical scatter and randoms values for all projections. The scatter value s was obtained by examining sinogram activity outside the support of the imaged object. The randoms value was selected experimentally such that $y_i + 2r$ was mostly positive. For all data sets, r was less than 2 counts per projection. Importantly, for a given set of projection data, the same values of s and r were used for all reconstruction algorithms to ensure a fair comparison.

The system matrix P was calculated from the system kernel $h(\theta, \Delta t)$ using (3.6) and (3.3), i.e., by approximating the kernel as constant over the support of a single pixel. If the entries P_{ij} are computed on-the-fly for each forward projection, the P_{ij} for non-zero gantry angles are obtained simply by using the difference between the projection angle θ_i and the gantry angle to index into the system kernel h. However, in our implementation, we speed up computation by storing P. To reduce the storage space required, we only store the matrix for a zero gantry position. In this case, the entries P_{ij} for non-zero gantry angles are obtained by rotating the coordinates of pixel j by the negative gantry angle and mapping the rotated coordinates into the pixel grid for the zero gantry position. Corrections for attenuation, deadtime, and decay were performed by multiplying the entries P_{ij} by the appropriate factors. Since different frames require different corrections, these multiplications were performed on the fly for each forward projection. The attenuation coefficients were calculated analytically by forward projecting the known or estimated support of the imaged object using a constant attenuation coefficient.

3.3.2 Bayesian Reconstruction Algorithms

Bayesian reconstruction methods regularize the tomographic reconstruction by using a prior probability distribution $p(x)$ that incorporates assumptions about typical images such as local smoothness. The image reconstruction is computed as the approximate maximum *a posteriori* (MAP) estimate

$$\hat{x}_{MAP} = \arg\max_x (\log p(y|x) + \log p(x)) \tag{3.8}$$

where $\log p(y|x)$ is the Poisson likelihood function (3.7).

For this study, we compare three Bayesian techniques using a Gaussian Markov random field prior (GMRF), a generalized Gaussian Markov random field (GGMRF) prior and an adaptive wavelet graph model (WGM) prior respectively. The commonly used GMRF prior can be written as

$$\log p(x) = -\frac{1}{2\sigma^2} \sum_{\{k,j\} \in \mathcal{N}} b_{k-j} (x_k - x_j)^2 + \text{constant} \tag{3.9}$$

where σ is a smoothing parameter and $\mathcal{N}$ is the set of all neighboring pixel pairs in the image x.

A disadvantage of the GMRF prior model is that the square potential function tends to oversmooth edges. To address this problem, several 'edge-preserving' potential functions have been proposed[5, 6, 7, 8, 9, 10, 11]. One such model is the GGMRF[10] which can be written as

$$\log p(x) = -\frac{1}{p\sigma^p} \sum_{\{k,j\} \in \mathcal{N}} b_{k-j} (x_k - x_j)^p + \text{constant} \tag{3.10}$$

where p is called a shape parameter. For the results presented in this study, we chose a GGMRF shape parameter of $p = 1.5$ which should produce sharper edges than a GMRF but result in reconstructions that appear less segmented as compared to using a smaller p-value or a Huber[74] prior model.

The adaptive wavelet graph model (WGM)[66, 67] is a nonlinear prior model formulated in the wavelet domain. The WGM exploits dependencies between wavelet coefficients at different resolutions to adapt the degree of regularization to the local image statistics. Specifically, the coarse scale wavelet coefficients are used to estimate the regularization parameters at the finer scales. This procedure allows for strong regularization of uniform image regions to reduce noise while edges and textures are preserved. Since the prior model for this method is data dependent, it is not a

MAP reconstruction technique. However, for simplicity, below we will use the term MAP optimization when referring to this algorithm. The WGM requires a set of training data to estimate the parameters of the wavelet prior model. For the results in this study, the WGM was trained on the same set of MRI reconstructions as in references[66, 67].

For all three prior models, the MAP optimization is performed using the iterative coordinate descent (ICD) algorithm[10, 42]. The ICD algorithm sequentially optimizes the MAP equation (3.8) with respect to single pixel values x_j. In comparison to gradient methods[75, 54, 56, 76], the ICD algorithm allows for simple enforcement of pixel positivity, resulting in more robust convergence. However, a disadvantage of ICD as compared to gradient methods is that the computational complexity cannot be reduced by using a factored matrix model as in [56, 50, 45].

Compared to previous implementations, the ICD algorithm in this study uses a simple multiresolution initialization strategy instead of FBP initialization. The multiresolution initialization is performed by calculating a coarse-to-fine scale sequence of MAP reconstructions, starting at a coarsest resolution of 16 by 16 pixels with a zero initial condition. The reconstruction at each resolution is interpolated and then used to initialize the reconstruction at the next finer scale. An advantage of this strategy is that the computationally inexpensive coarse scale reconstructions provide a good initialization for the fine scale reconstructions. This initialization typically reduces the number of computationally expensive iterations at the finest scale. In addition, the multiresolution initialization eliminates the need for an FBP reconstruction, which on the IndyPET scanner requires the preliminary generation of a full sinogram as described in section 3.2.2. One approximation made in the multiresolution initialization is that the assumption of constant system kernel per pixel (3.3) does not hold at coarse scales. However, this approximation does not affect the final fine scale reconstructions since the coarse scale solutions are only used for purposes of initialization. The choice of MRF smoothing parameters at coarse scales is not critical. For all results presented in this study, we have used $\sigma_n = 2^{-n}\sigma_0$

where σ_n denotes the smoothing parameter at scale n and $n = 0$ denotes the finest scale. For all reconstructions shown here, 25 iterations were performed at each scale to ensure sufficient convergence.

One concern with Bayesian reconstruction algorithms for emission tomography is that the reconstructions have spatially varying resolution[68] which can result in over-regularization of high count regions. This effect is undesirable, particularly, for ROI studies with high dynamic range. To address this issue, several modified algorithms have been proposed[68, 69, 70, 71]. In this study, we evaluate the impact of the modifications proposed by Fessler and Rogers[68] applied to the GMRF and GGMRF priors for a simulated data set containing a lesion on an approximately uniform background.

3.3.3 Expectation Maximization Reconstruction Algorithm

The EM algorithm used in this study is the basic algorithm[51] extended to include corrections for scatter and random coincidences. The EM update equation used was

$$x_j^{n+1} = \frac{x_j^n}{\sum_{i=1}^M P_{ij}} \sum_{i=1}^{M} P_{ij} \frac{y_i + 2r}{2r + s + \sum_{k=1}^N P_{ik} x_k^n} \tag{3.11}$$

where x^n is the reconstruction estimate after the n^{th} iteration. The smoothness of the reconstructions was controlled by the number of iterations performed. For all reconstructions, the EM algorithm was initialized with a constant image whose forward projection matched the total number of counts in the measured projection data. To ensure a fair comparison, our implementation of EM uses the same programming code as the Bayesian ICD-MAP implementation, with the only modification being the appropriate update equation and elimination of the prior model. Thus, exactly the same system matrix computation and corrections are used for both algorithms.

3.3.4 Filtered Backprojection

For filtered backprojection reconstruction, the frame sinogram segments were first precorrected for deadtime, exponential decay, overall detector sensitivity, scatter and randoms. An approximate complete sinogram was then generated as described in section 3.2.2. The smoothness of the FBP reconstructions was controlled by varying the parameter α of a reconstruction filter with frequency response

$$H(f) = H_{ramp}(f)\left(0.5 + 0.5\cos\left(\frac{\pi f}{\alpha f_N}\right)\right) \quad \text{for} \quad |f| \leq \min(\alpha f_N, f_N) \tag{3.12}$$

and zero elsewhere, where f_N is the Nyquist frequency. Note, that for $\alpha = 1$, we obtain a Hanning filter, and for $\alpha > 1$, the Hanning response smoothly transitions to a ramp filter. For the results presented in this paper, α was varied in the range from 0.3 to 8.0 and, in addition, a ramp filter was used to obtain reconstructions with minimal smoothing.

3.4 Experimental Results

Phantom data acquired on the IndyPET scanner as well as simulated data were used to evaluate reconstruction quality for different reconstruction algorithms and for different system kernels. Three data sets were used, a bar phantom, a Hoffman 3-D brain phantom, and a simulated data set containing a circular lesion on an approximately uniform background. Quantitative analyses of reconstruction quality for FBP, EM and the Bayesian techniques were performed for the bar phantom and the simulated data set. For the bar phantom, we also compare reconstruction quality for the empirical and the triangular system kernels; and we evaluate reconstruction quality for a set of parametric system kernels with different widths. Note, that whenever we refer to the empirical kernel, we mean the 2-D kernel shown in Fig. 3.6 including the dependence on projection angle. Finally, we compare the computational efficiency of EM and Bayesian GMRF MAP and evaluate different initialization strategies for the ICD-MAP algorithm.

3.4.1 Bar Phantom

As illustrated in Fig.3.7, the bar phantom consists of four arrays of parallel Plexiglas bars with thicknesses of 1.55, 2.31, 2.90, and 4.32 mm separated by gaps of equal widths. The space between the bars was filled with FDG, resulting in square wave emission profiles with different periods for each quadrant. The projection data were acquired on the IndyPET scanner. Fourteen different data sets were collected

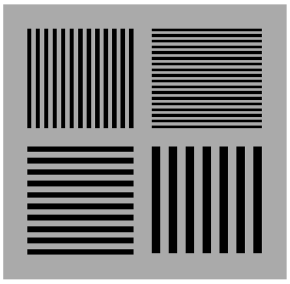
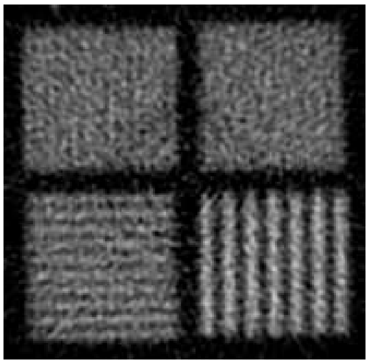


Fig. 3.7. Illustration of bar phantom. The bar phantom consists of four arrays of parallel bars with thicknesses of 1.55, 2.31, 2.90, and 4.32 mm separated by gaps of equal widths.

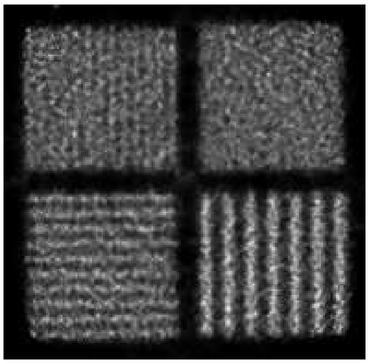
sequentially, each corresponding to a full 180 degree gantry sweep using 45 frames. The acquisition time per frame was identical for all 14 data sets. By summing different subsets of these data sets frame by frame, we generated 4 combined data sets with a total of 10.4M, 4.4M, 2M, and 586,000 counts in the reconstructed plane.

FBP, EM, and Bayesian reconstructions of the four combined data sets were computed using a wide range of smoothing parameters for each algorithm. The reconstruction resolution was 256 by 256 pixels for a field of view of 18 cm, resulting in a pixel size of 0.7 mm square. For EM and Bayesian MAP with a GMRF prior, we compare reconstructions for both the empirical system kernel shown in Fig. 3.6 and the triangular kernel shown in Fig. 3.1(c). For Bayesian GGMRF MAP and the Bayesian wavelet graph model (WGM), only the empirical system kernel was used. In a separate experiment described in section 3.4.1, we compare Bayesian GMRF reconstructions for a set of parametric system kernels of different widths.

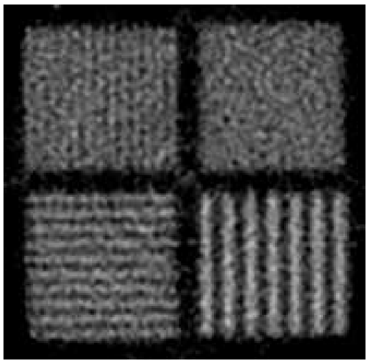
Figures 3.8 and 3.9 show sample reconstructions for the data set with 10.4M counts. The images were cropped to 192×196 pixels, corresponding to a physical size of 13.5×13.78 cm. The smoothing parameters for the reconstructions shown in Fig. 3.8 were chosen to minimize the expected spatial reconstruction error (bias^2+variance, see section 3.4.1) for the 4.32 mm bars in quadrant 4. Similarly, Fig. 3.9 shows reconstructions with minimum error for the 2.90 mm bars in quadrant 3. The images of the EM reconstructions are slightly clipped due to the high



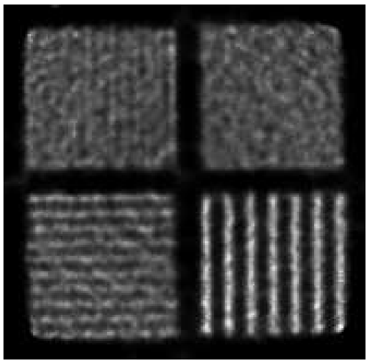
FBP ramp filter



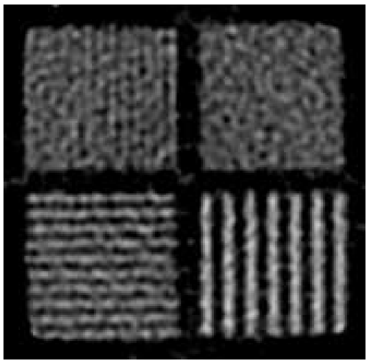
EM, 198 iterations
triangular kernel



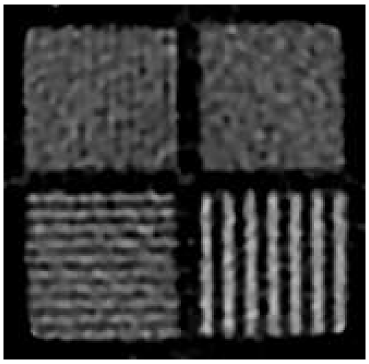
MAP GMRF, $\sigma = 5.5$
triangular kernel



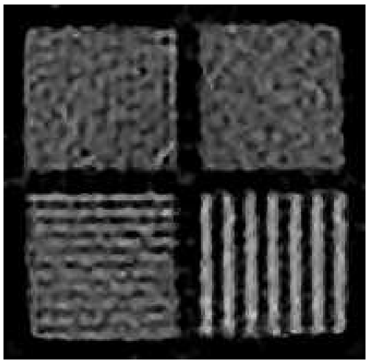
EM, 348 iterations
empirical kernel



MAP GMRF, $\sigma = 6.9$
empirical kernel

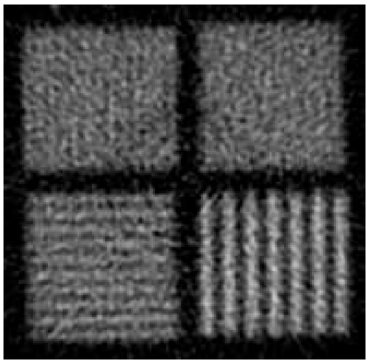


MAP GGMRF, $\sigma = 6.9$
empirical kernel

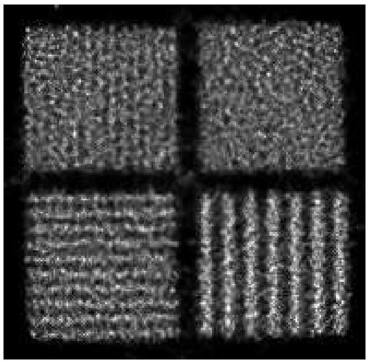


WGM, $\omega = 2.5$
empirical kernel

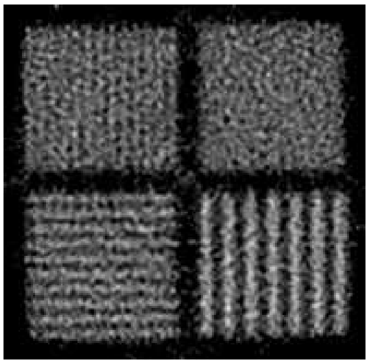
Fig. 3.8. Bar phantom reconstructions for 10.4M total counts. The images correspond to a physical size of 13.5×13.78cm at a resolution of 0.70mm/pixel. The smoothing parameter for each method was selected to minimize the expected error for the bar profile in quadrant 4 (4.32 mm bars).



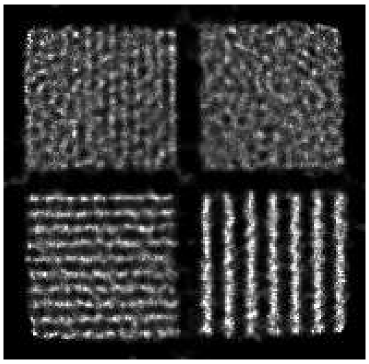
FBP ramp filter



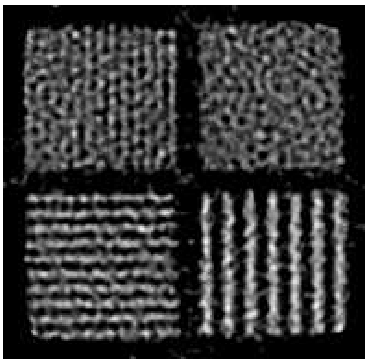
EM, 342 iterations
triangular kernel



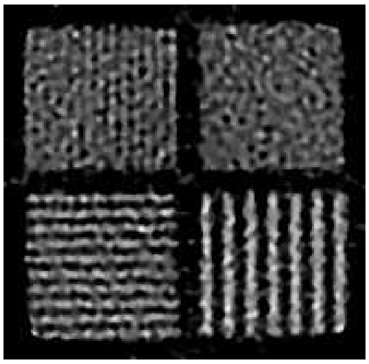
MAP GMRF, $\sigma = 8.7$
triangular kernel



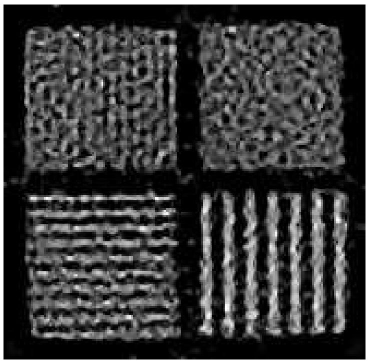
EM, 1068 iterations
empirical kernel



MAP GMRF, $\sigma = 13.0$
empirical kernel



MAP GGMRF $\sigma = 15.0$
empirical kernel



WGM, $\omega = 6.0$
empirical kernel

Fig. 3.9. Bar phantom reconstructions for 10.4M total counts with smoothing parameters selected to minimize the expected error for the bar profiles in quadrant 3 (2.90mm bars).

amplitude of the noise peaks in the reconstructions.

Based on visual inspection, the empirical system kernel significantly improves the quality of the iterative reconstruction techniques. Also, when the empirical system kernel is used, the iterative techniques provide improved resolution over FBP. For the Bayesian reconstructions, the non-Gaussian GGMRF and WGM prior models provide sharper edges as compared to the GMRF prior. However, the GMRF prior appears to have a wider frequency response, allowing for good reconstruction quality of all quadrants using a single regularization parameter. Note, that none of the reconstruction methods can resolve the horizontal 1.55 mm bars in quadrant 1, thus our quantitative analysis below is limited to quadrants 2-4.

Bar Phantom Bias and Variance Analysis

The bar phantom reconstructions were analyzed by comparing the reconstructed wave forms to the expected square wave profile. Bias and variance measures were computed separately for each quadrant by fitting harmonic series to the reconstructed bar profiles. Let $x_Q(r,n)$ denote the reconstruction values in a 50 by 50 pixel window centered over quadrant Q. Let n denote the coordinate running perpendicular to the bars (e.g. horizontal for quadrants 2 and 4) and let r be the coordinate parallel to the bars. The 1-D average reconstruction bar profile $\bar{x}_Q(n)$ is obtained as

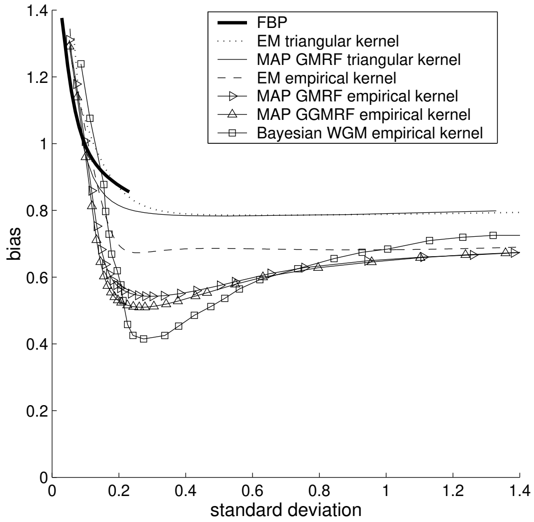
$$\bar{x}_Q(n) = \frac{1}{R}\sum_{r=1}^{R} x_Q(n,r) \tag{3.13}$$

where $R = 50$ is the size of the window and $1 \leq n \leq R$. We then determine the coefficients a_k and b_k of a harmonic series $\tilde{x}_Q(n)$

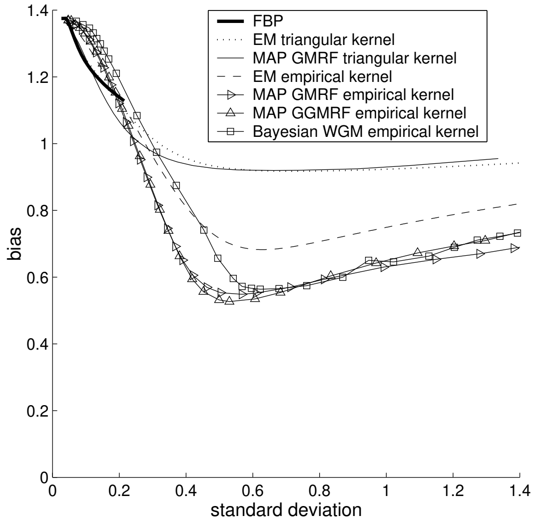
$$\tilde{x}_Q(n) = b_0 + \sum_{k=1}^{K}(a_k \sin(2\pi k f_Q n) + b_k \cos(2\pi k f_Q n)) \tag{3.14}$$

by using a least squares optimization to fit $\tilde{x}_Q(n)$ to $\bar{x}_Q(n)$. The frequency f_Q is a fixed frequency appropriate for each quadrant. For the results presented in this study, the order of the fitted series was $K = 7$.

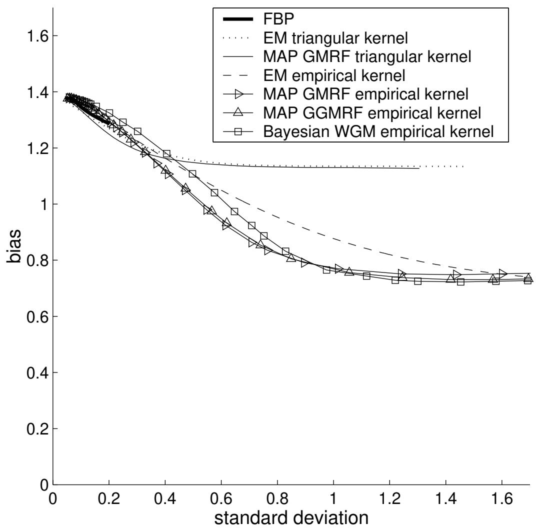
The reconstruction bias is calculated by comparing the fit $\tilde{x}_Q(n)$ to an ideal square wave. Let α_k and β_k be the Fourier coefficients of an ideal square wave with unit



(a) quadrant 4, 4.32 mm bars



(b) quadrant 3, 2.90 mm bars



(c) quadrant 2, 2.31 mm bars

Fig. 3.10. Bias vs. standard deviation for the bar phantom reconstructions for 10.4M total counts. Bias and standard deviation were calculated using the Fourier analysis described in section 3.4.1. The curve for each reconstruction algorithm was obtained by varying the smoothing parameter.

amplitude, a DC value of one, and a phase equal to the phase of the fundamental of $\tilde{x}_Q(n)$. We then calculate the squared reconstruction bias for quadrant Q as

$$\text{bias}_Q^2 = \min_c \left\{ \sum_{k=1}^{K} (\frac{a_k}{c} - \alpha_k)^2 + \sum_{k=0}^{K} (\frac{b_k}{c} - \beta_k)^2 \right\} . \tag{3.15}$$

The minimization over c accounts for the fact that the absolute scaling of the reconstruction algorithms is not calibrated. Thus, we scale the reconstructed signal to the ideal square wave such that bias is minimized. The minimization in (3.15) is performed analytically. The bias is affected by the choice of the scatter and randoms parameters r and s in (3.7) used for the reconstruction. Specifically, a smaller value for s results in increased bias due to an increased DC offset in the reconstructed wave profile. Thus, we emphasize that, for a given set of projection data, we use the same scatter and randoms values for all reconstruction methods to ensure a fair comparison.

While the bias characterizes the systematic deviation of the reconstruction from the true signal, the variance characterizes the random variation of the reconstruction around its average. The variance for quadrant Q is calculated as the combined variance over all rows r around the fit $\tilde{x}_Q(n)$

$$\text{std}_Q^2 = \frac{1}{c^2(R^2-1)} \sum_{r=1}^{R} \sum_{n=1}^{R} (x_Q(n,r) - \tilde{x}_Q(n))^2 \tag{3.16}$$

where c is the scaling constant determined in (3.15). Note, that the variance defined here is the spatial variance characterizing the noisiness of a single reconstruction as compared to the variance of the reconstruction estimator.

Figure 3.10 shows reconstruction bias vs. standard deviation for the 10.4M count case separately for the three quadrants. The curves were obtained by varying the smoothing parameter for each reconstruction method. Note, that the points closest to the coordinate origin in Fig. 3.10(a) and (b) correspond to the reconstruction images shown in Fig. 3.8 and Fig. 3.9 respectively.

The bias vs. standard deviation plots indicate, that using the empirical system kernel as compared to the triangular kernel significantly lowers the bias at equal

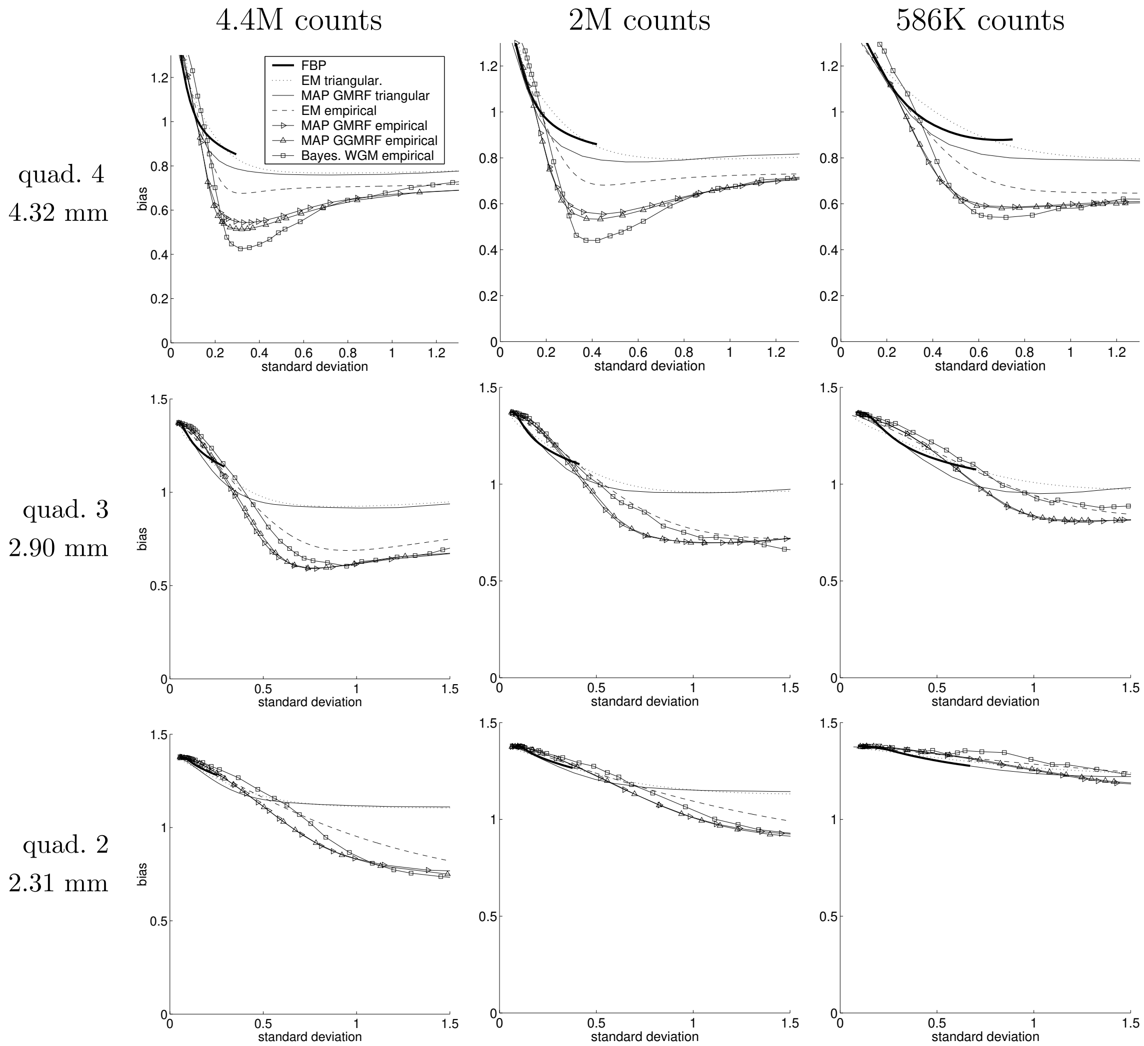


Fig. 3.11. Reconstruction bias vs. standard deviation for the different quadrants of the bar phantom for different count levels.

variance for EM and Bayesian GMRF MAP. Interestingly, when the triangular kernel is used, EM and GMRF MAP perform similarly. However, when the empirical kernel is used, GMRF MAP achieves significantly lower bias at equal variance as compared to EM. For the empirical kernel, the GGMRF prior performs slightly better than the GMRF model. The WGM model performs well in quadrant 4 but not as well in quadrants 2 and 3. The FBP reconstructions have low variance and high bias even for a ramp filter since no resolution recovery is performed. FBP only partly resolves the bars quadrant 3 and does not resolve quadrant 2. For quadrant 4, the FBP bias vs. standard deviation is similar to the low variance range of the curves for the iterative algorithms using a triangular system kernel. However, when the empirical system kernel is used, the iterative techniques achieve significantly lower bias at equal variance as compared to FBP.

Figure 3.11 shows the bias vs. standard deviation curves for the bar phantom data sets with 4.4M, 2M and 586K total counts. The results are very consistent with those for the 10.4M count data set. Due to the lower signal to noise ratio at lower count levels, the advantage of using the empirical system kernel decreases slightly. This is expected, as the potential for resolution recovery is best at high signal to noise ratios. Note, that the reconstructions for 586K counts does not resolve the bars in quadrant 2.

In general, we would expect the reconstruction bias to decrease as the standard deviation increases, i.e. as less smoothing is performed. However, most of the curves in Figs. 3.10 and 3.11 indicate that after reaching a minimum, the bias increases for larger values of the standard deviation. We attribute the increase to two factors. First, our reconstructions have slight circular artifacts resulting from inaccuracies in the normalization scan. Thus, the reconstruction profile deviates systematically from an ideal square wave as less smoothing is performed. Secondly, the bias is calculated using a limited window of 50 by 50 pixels for each quadrant which may result in noisy estimates of the higher order Fourier coefficients.

The reduction in bias achieved by the empirical system kernel is mainly due to

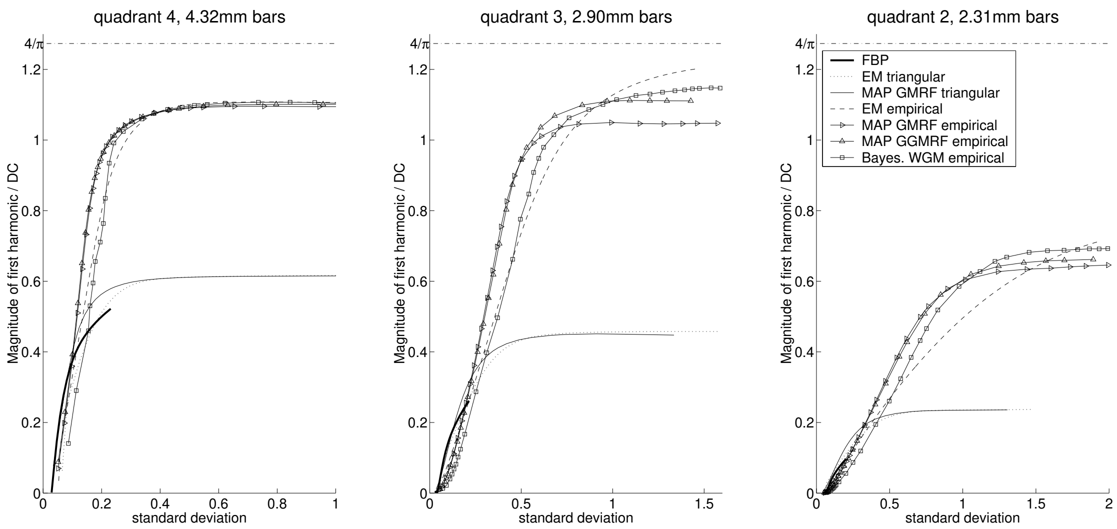


Fig. 3.12. Normalized magnitude of the fundamental Fourier component of the reconstructed bar profiles (10.4M counts). The normalized magnitude was calculated from the Fourier coefficients a_k, b_k of the harmonic fit as $\sqrt{a_1 + b_1}/b_0$. The ideal value for a square wave of unit amplitude and DC of one is $4/\pi$. The empirical system kernel significantly improves the amplitude of the reconstructed bar profiles around the DC level.

a much larger amplitude of the reconstructed bar profiles. Figure 3.12 shows the normalized magnitude $\sqrt{(a_1^2 + b_1^2)}/b_0$ of the fundamental frequency component of the reconstructed wave form as a function of reconstruction standard deviation. The ideal value for a square wave of unit amplitude and DC value of $b_0 = 1$ would be $4/\pi \approx 1.27$. The curves in Fig. 3.12 demonstrate that using the empirical system kernel results in a much larger variation of the reconstructed profile around the DC value as compared to using the triangular kernel or FBP.

Figure 3.13 shows the reconstructed bar profiles in quadrant 4 for the reconstruction images shown in Fig. 3.8. The plotted profiles are the averages $\bar{x}_4(n)$ as defined in (3.13) normalized by their DC value. For EM and GMRF MAP, the dashed lines correspond to the reconstructions using the triangular system kernel. Again, the iterative reconstructions using the empirical system kernel achieve much larger amplitudes in comparison to DC. EM using the empirical kernel produces a slightly

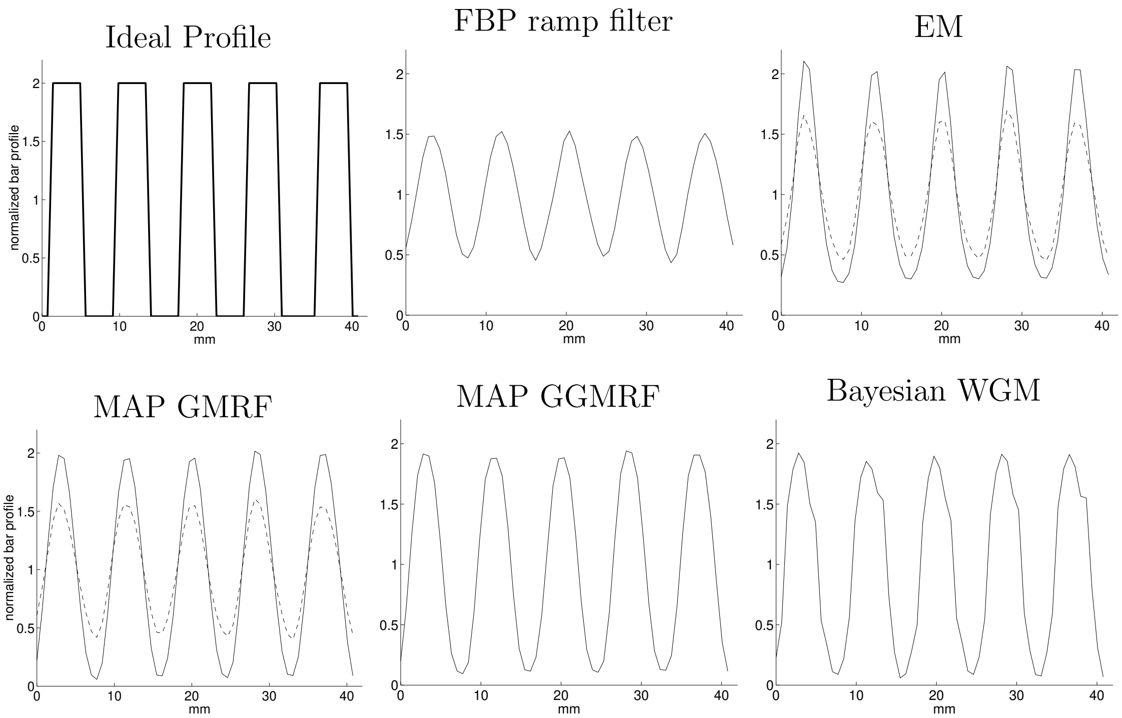


Fig. 3.13. Bar profile reconstruction averages $\bar{x}_4(n)$ for quadrant 4 (4.32 mm bars). The profiles were calculated from the reconstruction images shown in Fig. 3.8, i.e. for smoothing parameters minimizing the error in quadrant 4. The dashed lines for EM and MAP GMRF correspond to reconstructions using the triangular system kernel.

asymmetric profile with peaks overshooting the ideal profile. The GMRF MAP profile is symmetric but has very smooth transitions resulting in blurred edges. The GGMRF MAP method produces slightly sharper transitions. The WGM prior results in much sharper transitions but has some artifacts in the very high amplitude region.

Comparison of Parametric System Kernels of Different Widths

To evaluate the impact of deviations in the system kernel used for iterative reconstruction, we compare Bayesian GMRF reconstructions of the bar phantom for a set of parametric kernels of different widths. The parametric kernels were obtained by convolving the triangular kernel of Fig. 3.1(c) with Gaussians of different standard deviations β. To obtain a baseline 1-D kernel that is similar to the 2-D empirical

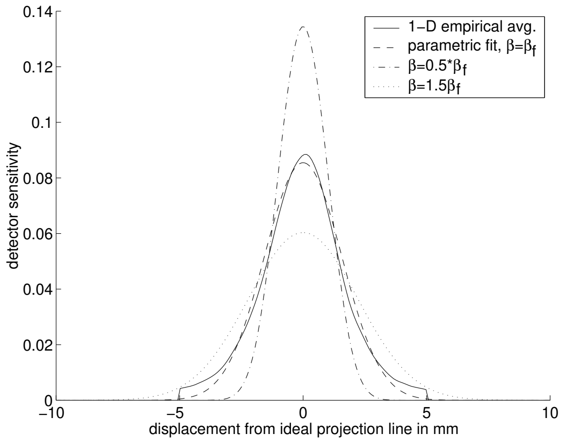
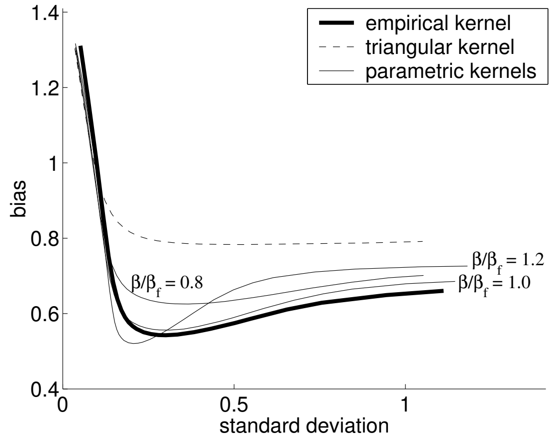


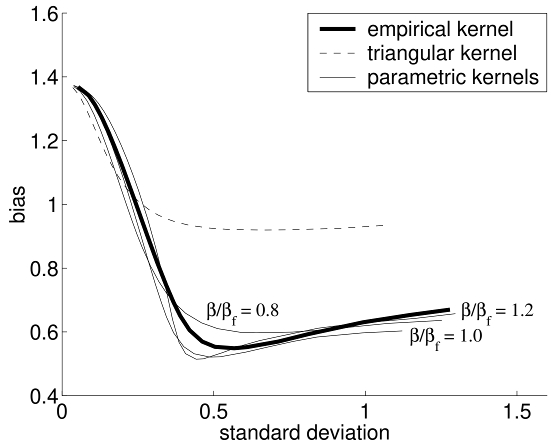
Fig. 3.14. Parametric kernels for sensitivity analysis. The empirical 2-D system kernel $h(\theta, \Delta t)$ was averaged in the angular coordinate θ to obtain a 1-D average kernel (solid line). A parametric fit (dashed line) to the averaged empirical kernel was obtained by convolving the ideal triangular kernel with a Gaussian. Parametric kernels of different widths are then generated by varying the standard deviation of the convolving Gaussian relative to the value obtained for the fit. The 1-D kernels are normalized for equal area.

system kernel, we first computed an average of the empirical kernel in Fig. 3.6 over the 80% central portion of the angular range. The β value for the baseline parametric kernel was then estimated to obtain a least-squares fit to the averaged empirical kernel. For the reconstructions, we consider parametric kernels with parameters from $\beta = 0$ (triangular kernel) to $\beta = 2\beta_f$, where β_f is the fitted value. Figure 3.14 shows examples for $\beta = 0.5\,\beta_f$ and $\beta = 1.5\,\beta_f$. All parametric kernels were normalized such that the area under the curve was equal to that of the 1-D averaged empirical kernel.

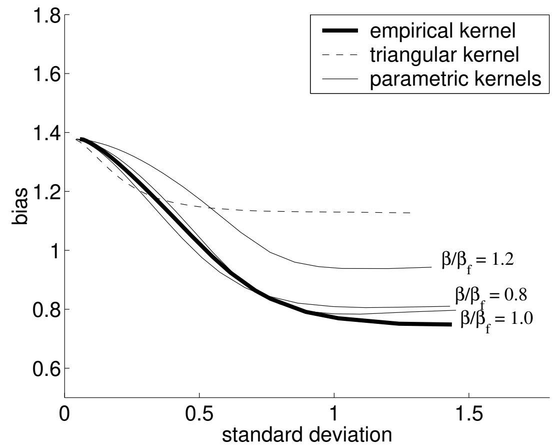
Figure 3.15 compares the bias vs. standard deviation curves of reconstructions using parametric kernels of different widths and for the triangular and the original 2-D empirical kernel shown in Fig. 3.6. We observed that some improvements in reconstruction quality can be achieved by increasing the width of the parametric kernel as compared to the triangular kernel. However, the result for $\beta = \beta_f$ deviates noticeably from that for the original 2-D empirical system kernel. This indicates, that



Quadrant 4



Quadrant 3



Quadrant 2

Fig. 3.15. Impact of deviations in the system kernel used for iterative reconstruction. Shown are the bias vs. standard deviation plots for Bayesian GMRF reconstructions of the bar phantom (10.4M counts). The β/β_f parameters indicate the width of the parametric kernels where β_f corresponds to fitting the averaged empirical kernel.

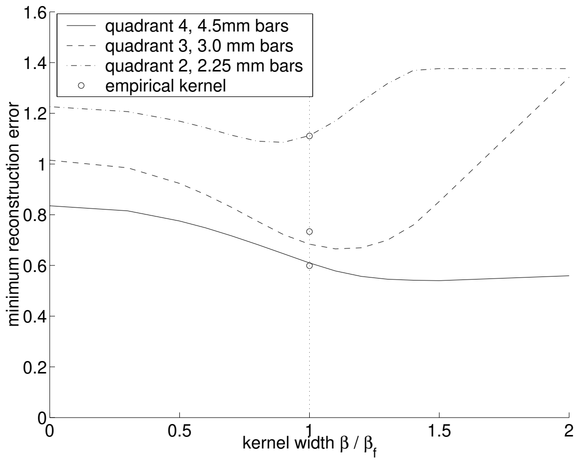


Fig. 3.16. Minimum reconstruction error as a function of the width parameter β of the parametric system kernels. The plotted value for each width is obtained as the minimum $\sqrt{(\text{bias}^2 + \text{variance})}$ for quadrant 4 taken over a range of smoothing parameters for the GMRF prior.

small deviations in the system kernel can significantly affect reconstruction quality. For very high values of the reconstruction standard deviation, i.e. as the influence of the Bayesian prior model is minimal, the empirical system kernel results in the lowest bias for quadrants 2 and 4. This effect may indicate, that the empirical kernel more accurately describes the scanner as compared to the parametric models.

Figure 3.16 shows minimum reconstruction error as a function of the width of the parametric system kernels. The minimum expected reconstruction error is calculated as the minimum value of $\sqrt{(\text{bias}^2 + \text{variance})}$ over all smoothing parameters for fixed β. The results indicate, that the fitted width $\beta = \beta_f$ provides a good compromise for all quadrants. Using a wider system kernel increases the error in quadrant 2 but improves the result for quadrants 3 and 4. However, as can be seen from the curve for $\beta = 1.2\beta_f$ in Fig.3.15(a), a wider kernel improves the result only for a very small range of standard deviations. This effect may be due to interactions with the GMRF prior, i.e. over-sharpening the reconstruction through the forward model may counteract the blurring of the GMRF prior.

3.4.2 Hoffman 3-D Brain Phantom

The second data set used in this study was acquired on the IndyPET scanner using a Hoffman 3-D brain phantom. The data set was reconstructed at a resolution of 256 by 256 pixels with a field of view of 20 cm, resulting in a pixel size of 0.78 mm square. The reconstructed plane contained a total of 2M counts. Reconstructions were performed using FBP, EM and the Bayesian techniques. For EM and Bayesian GMRF MAP, we compare the empirical and the triangular system kernels. For the other Bayesian techniques, we used the empirical system kernel only.

Figure 3.17 shows reconstructions of the Hoffman 3-D brain phantom. Note, that due to the planar structure of the phantom, the reconstructions correspond to overlapping physical planes in locations where the shapes of the adjacent planes in the phantom differ. The individual reconstruction images were cropped to 177×148 pixels, corresponding to a physical size of 13.8×11.5 cm. Each row in Fig. 3.17 corresponds to a single reconstruction algorithm with three hand-picked smoothing parameters. The first three rows correspond to FBP and to algorithms using the triangular system kernel. The bottom two rows show EM and GMRF MAP reconstructions using the 2-D empirical system kernel. In general, using the empirical kernel improves the quality of the iterative reconstructions. The EM reconstructions deteriorate with increasing numbers of iterations before some of the detail is converged.

Figure 3.18 compares EM and Bayesian reconstructions using the empirical system kernel. We observe that EM does not reproduce the interior contours as well as the Bayesian techniques. As compared to the GMRF, the GGMRF prior produces sharper contours with few artifacts. The WGM produces a much sharper image, however, the result has significant artifacts.

3.4.3 Simulated Phantom

A simulated data set was generated by adding a constant circular lesion of 4 mm diameter and 5 times background activity to an approximately uniform background as shown in Fig. 3.19. The background image was an FBP reconstruction of an FDG

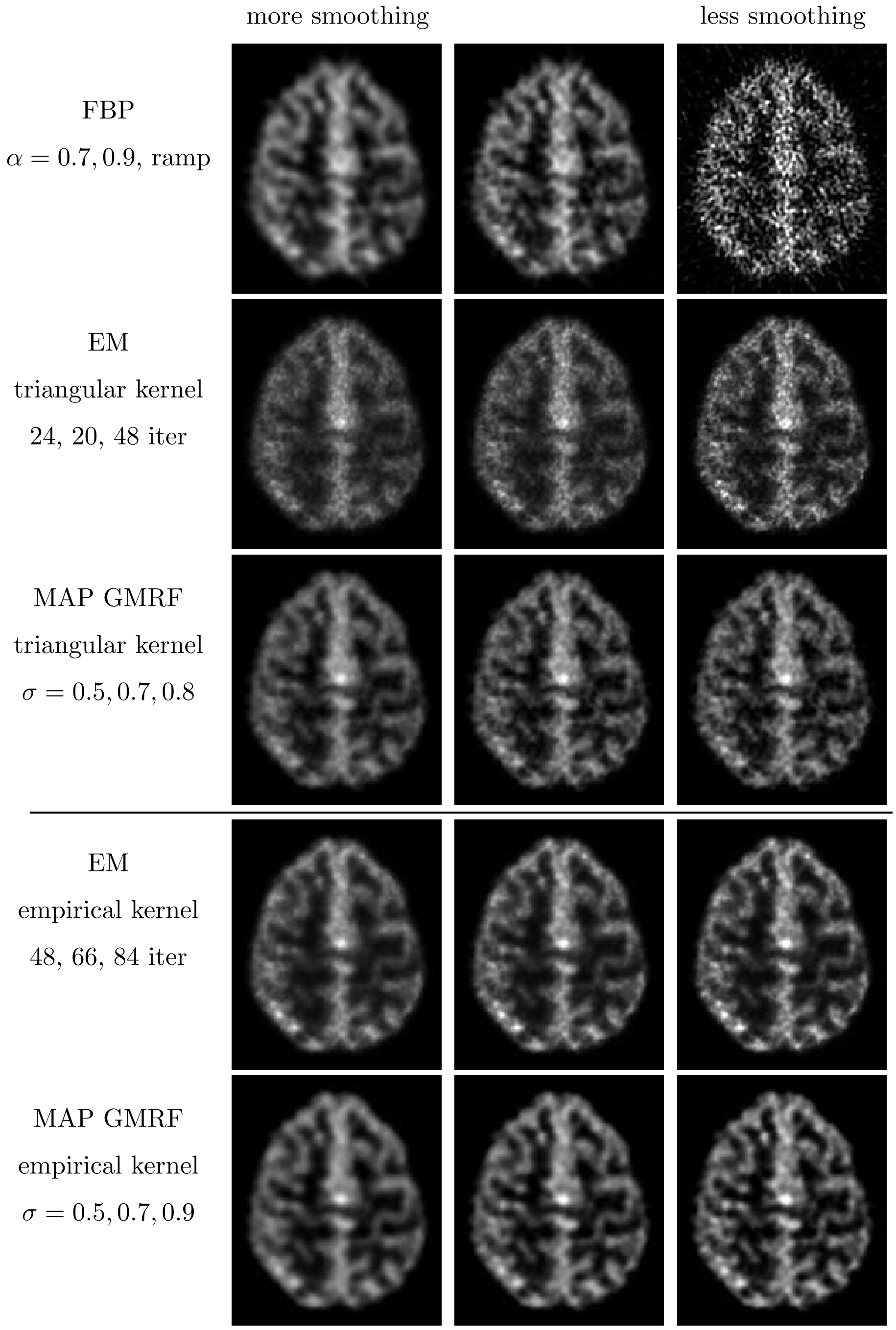


Fig. 3.17. Reconstructions of Hoffman 3-D brain phantom. Each row shows reconstructions for a single algorithm using three hand-picked regularization parameters. The first three rows correspond to FBP and iterative methods using the triangular system kernel. The last two rows correspond to iterative methods using the empirical system kernel.

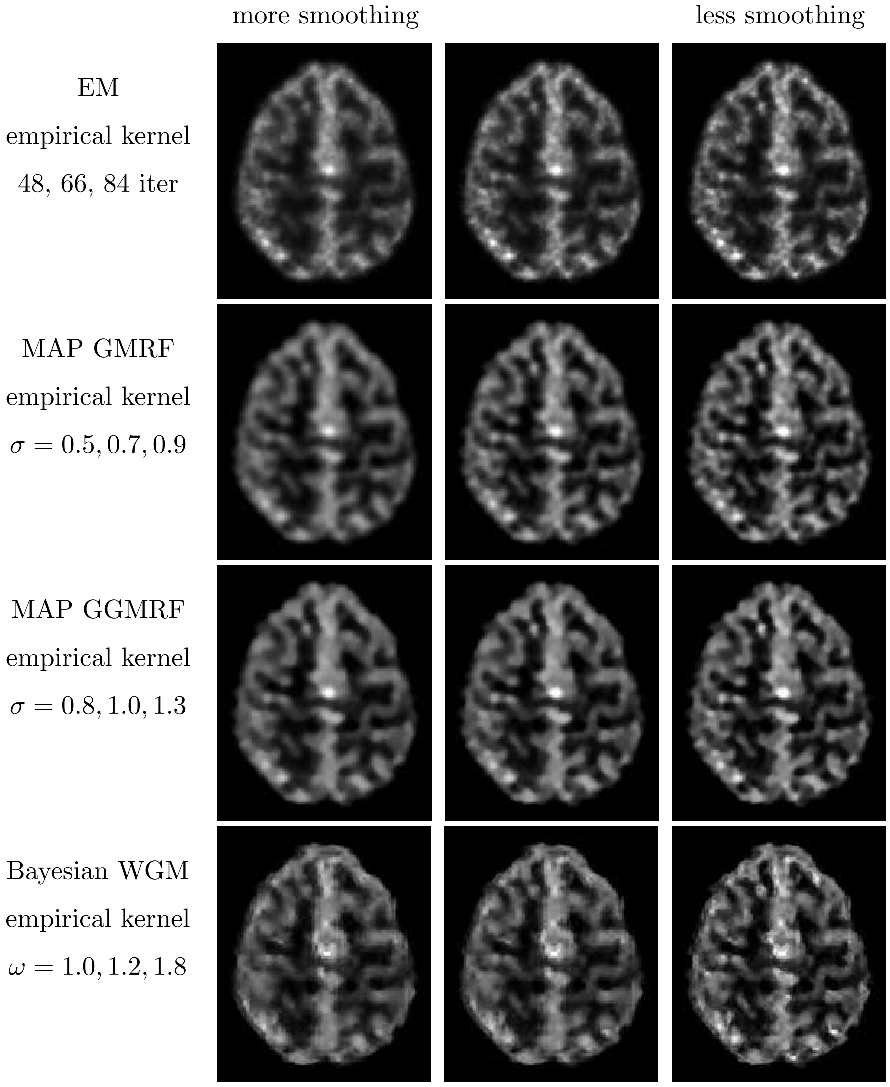


Fig. 3.18. Comparsion of EM and Bayesian reconstructions using different prior models. In all cases, the empirical system kernel was used.

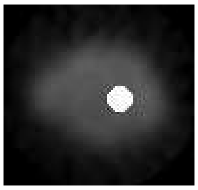


Fig. 3.19. Simulated phantom. A circular lesion of five times background activity was added to an approximately uniform background. The image was cropped to a size of 101×106 pixels, corresponding to a simulated physical size of 3.16×3.31 cm.

mouse study on the IndyPET scanner. Simulated projection data were generated by forward projecting the image in Fig. 3.19 using the empirical system kernel for the IndyPET scanner. The data samples were formed by Poisson random variables with the appropriate means. Dead-time, decay, and relative detector sensitivity were not modeled. Forward projection and reconstruction were performed at a resolution of 256 by 256 pixels assuming a field of view of 8 cm. This results in a pixel size of 0.31 mm square which is more than 10 times the detector resolution of the scanner. The total number of counts in the reconstruction plane was 130,000. Twenty separate noisy realizations of projection data were generated.

Reconstructions of the simulated data set were performed using FBP, EM and the Bayesian techniques. The empirical system kernel was used for EM and the Bayesian techniques. Even though the use of the same system kernel for forward projection and reconstruction will bias the results towards the iterative reconstruction techniques[77], we can still obtain fair comparisons between different iterative techniques and, in addition, obtain an approximate idea about their performance relative to FBP.

The reconstructions were evaluated quantitatively in terms of mean square error estimator performance for each reconstruction technique as a function of smoothing parameter. We do not use the clinically more relevant contrast recovery ratio (CRC) since we have found the CRC values and the associated rank order of performance for the different algorithms to be a strong function of the exact procedure and the regions chosen for the CRC calculation. Thus, for clinical evaluation, we consider it necessary to precisely specify a task with an associated ROI estimator or to perform human observer studies[78]. For this study, we limit ourselves to MSE analysis to decouple the reconstruction problem from the ROI estimation problem. Let x be the ground truth image and let $\hat{x}$ be a single reconstruction. Define L and B as the known sets of lesion pixels and background pixels. The set B was obtained by thresholding the ground truth image, eroding the resulting mask and excluding the lesion plus a 1 mm thick annulus around it. Further, let μ_L and μ_B be the true mean activity values in the lesion and background. We then calculate the relative MSE errors e_L^2 and e_B^2 for the lesion and background regions as

$$e_L^2 = \frac{1}{\mu_L^2 N_L} \sum_{j \in L} (\hat{x}_j - x_j)^2 \tag{3.17}$$

$$e_B^2 = \frac{1}{\mu_B^2 N_B} \sum_{j \in B} (\hat{x}_j - x_j)^2 \tag{3.18}$$

where N_L and N_B are the number of pixels in the lesion and background regions L and B respectively. We note that e_L^2 and e_B^2 are normalized relative to the true signal amplitudes since one is typically more interested in the ratio of lesion activity to background activity as compared to the absolute values. The errors e_L^2 and e_B^2 are averaged over the 20 reconstructions of the different noise realizations to give estimates $\hat{E}[e_L^2]$ and $\hat{E}[e_B^2]$ for the expected relative errors for each reconstruction method as a function of regularization parameter.

Figure 3.20(a) shows a plot of the expected reconstruction error for the lesion $\sqrt{\hat{E}[e_L^2]}$ vs. the expected reconstruction error for the background $\sqrt{\hat{E}[e_B^2]}$ for different algorithms. The curve for each algorithm was obtained by varying the smoothing parameter. The results indicate, that the Bayesian methods can achieve lower error in the lesion region than EM. The GGMRF prior model performs significantly better than the GMRF prior for this problem. The curve for the GGMRF prior is closest to the coordinate origin, signifying lowest combined error.

An interesting characteristic of the Bayesian methods is that the minimum error in the lesion region and the minimum error in the background region are achieved at

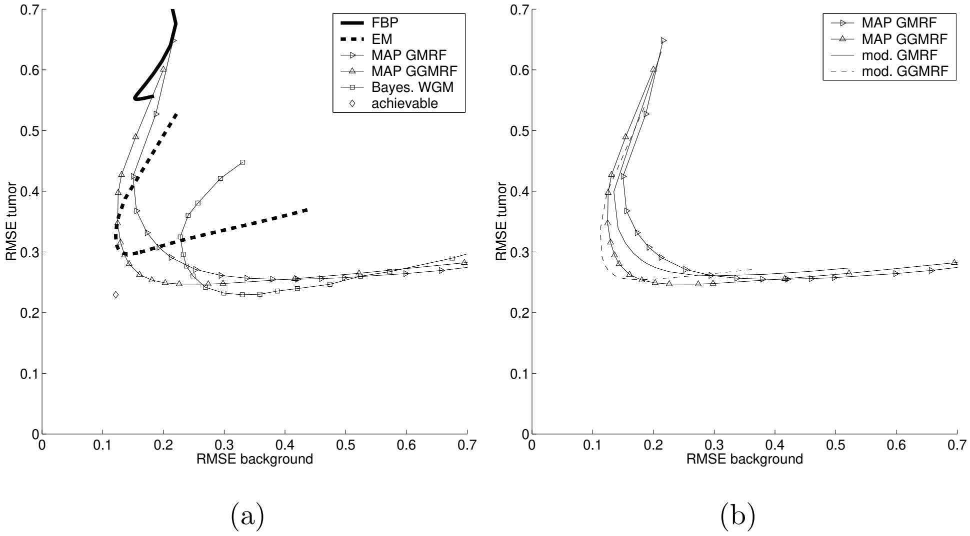


Fig. 3.20. Simulated phantom. Expected error for the known lesion region as a function of the expected error in the background region. The curves were obtained by varying the smoothing parameter for each method. The errors are averages over reconstructions of 20 noisy realizations of the projection data. Shown in (a) is a comparison for the different iterative methods and FBP. Shown in (b) is the effect of using the modification of the MAP prior proposed in [68].

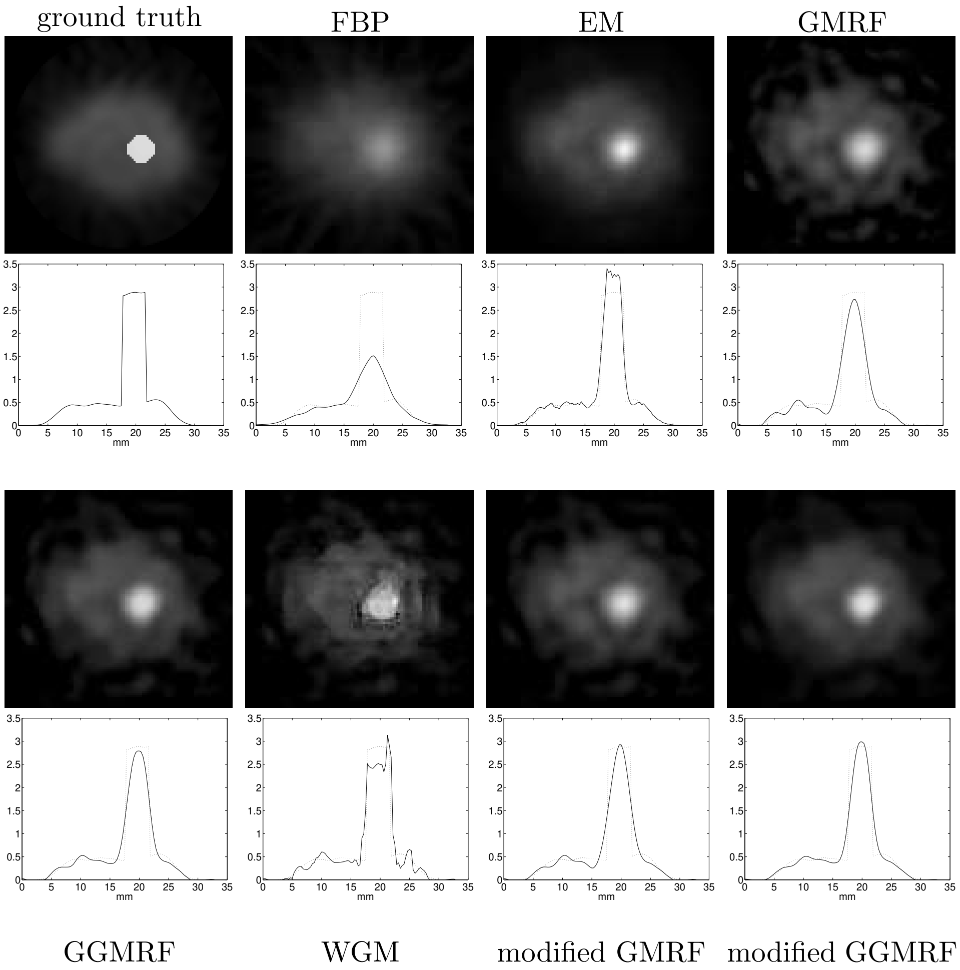
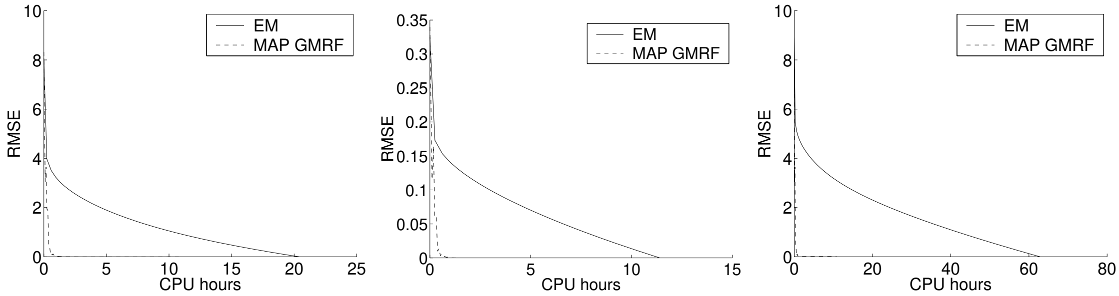


Fig. 3.21. Reconstruction images and profiles for the simulated tumor phantom. The images correspond to a physical area of 3.16×3.31cm at a pixel resolution of 0.31mm. The smoothing parameter for each method was selected to minimize the sum of the relative errors for the tumor and the background region. The 1-D profiles correspond to a single row of pixels intersecting the center of the lesion.

different points on the curve, i.e. for two different values of the smoothing parameter. In other words, there is no single smoothing parameter that optimizes algorithm performance for both lesion and background. This behavior may be a result of the activity dependency of the regularization of Bayesian methods[68]. Note, that as indicated by the single diamond in Fig. 3.20(a), a lower bound for achievable performance for this problem is given by the combination of the lowest value of lesion error over all methods and lowest value of background error over all methods. Specifically, since we assume known size and location of the lesion, we can trivially achieve this bound by using one reconstruction method for the background and another one for the lesion. The distance of the bound to the measured curves indicates, that there is significant room for further improvement.

Figure 3.20(b) shows the reconstruction error for Bayesian GMRF and GGMRF reconstructions using the modification proposed by Fessler and Rogers[68] to achieve approximately space-invariant regularization. This modification is effective in reducing the combined lesion and background errors since the curves move closer to the coordinate origin. For the GGMRF prior, the modification results in a much smaller compromise between choosing the optimal smoothing parameter for background or lesion. However, the minimum error in the lesion region is slightly higher for the modified priors.

Figure 3.21 shows reconstruction images as well as 1-D profiles through the reconstructed lesion for the different algorithms. The images were cropped to a size of 101×106 pixels corresponding to a simulated physical size of 3.16×3.31 cm. The reconstruction resolution is 0.31 mm per pixel which is ten times higher than the 3.3 mm nominal detector resolution of the IndyPET scanner. The smoothing parameters for the different algorithms correspond to the points closest to the origin in Fig. 3.20, minimizing the sum $\hat{E}[e_L^2] + \hat{E}[e_B^2]$ of the relative errors for the lesion and background regions. We observed that the iterative algorithms resolve the lesion substantially better than FBP. In comparison to EM, the Bayesian GMRF and GGMRF reconstructions more accurately quantify the lesion. The WGM prior model



(a) 10.4M counts, smoothing optimized for quadrant 4

(b) 586K counts, smoothing optimized for quadrant 4

(c) 10.4M counts, smoothing optimized for quadrant 3

Fig. 3.22. Computational efficiency of EM and Bayesian ICD-MAP with a GMRF prior and multiresolution initialization. The curves show reconstruction root mean square error relative to the converged image for MAP and relative to the reconstruction after the desired number of iterations for EM. The smoothing parameters for (a) and (c) correspond to the reconstructions in Fig. 3.8 and Fig. 3.9 respectively.

segments the lesion surprisingly accurately, however, it slightly underestimates lesion activity and has some artifacts. Overall, the modified GGMRF model appears to achieve the most accurate result. Note, that the reconstructions have ringing artifacts in the background which is consistent with observations in earlier studies[53]. The ringing is due to the extremely high reconstruction resolution, which in combination with the wide detector kernel results in a highly ill-posed reconstruction problem.

3.4.4 Comparison of Computational Efficiency

Convergence speed was compared for EM and Bayesian MAP using a GMRF prior for reconstructions of the bar phantom data sets. Figure 3.22 shows root mean square error (RMSE) of the reconstruction images as a function of CPU-time. For EM, the RMSE values were calculated in reference to the final reconstruction after the desired number of iterations. For MAP, the RMSE was calculated in reference to the converged reconstruction image. To account for the multiresolution initialization, the RMSE curves for MAP were obtained by separately running the complete multiresolution strategy for different numbers of iterations, using the same number of iterations at each scale. The CPU-times were measured on a 700 MHz Pentium III CPU and do not include the time needed for pre-processing such as the computation of the tomographic system matrices that can be performed off-line. The CPU-time required for a single iteration of EM was approximately the same as for a single MAP iteration at the finest reconstruction scale.

The results in Fig. 3.22 demonstrate that ICD-MAP converges dramatically faster than EM. In addition, the convergence speed for EM is a strong function of the smoothing parameter and the signal to noise ratio of the projection data whereas the convergence speed of MAP is not affected. Based on the large difference between the algorithms, it is questionable if an ordered subsets acceleration[47] of EM can result in convergence speeds that are comparable to MAP.

We note that due to the large dimensionality of the problem, even the MAP reconstructions take about 40 minutes for the reconstruction of a single 2-D plane. Specifically, the acquisition of 45 frames on the IndyPET scanner results in 45×3136 separate projection measurements. Furthermore, the sparsity of the projection matrix is reduced by using the relatively wide empirical system kernel. Storage of the complete projection matrix for all 45 frames for a reconstruction resolution of 256 by 256 pixels would require on the order of 5GB, however, by storing the matrix for a single gantry position only, this is reduced to 120MB.

Figure 3.23 compares the convergence speed of ICD-MAP for different initializations of the algorithm. The graphs are based on reconstructions of the bar phantom data set with 10.4M counts using a GMRF prior. We compare the multiresolution strategy to initialization by a constant and by an FBP ramp reconstruction image. Both the constant and the FBP image were scaled such that their forward projection matched the total number of counts in the measured sinogram. The graph for the multiresolution initialization was obtained as in Fig. 3.22 and accounts for the total CPU-time needed for the reconstructions at all resolutions. The results indicate, that the multiresolution initialization significantly improves the convergence speed of the algorithm. Based on the known convergence properties of ICD[41], we do not

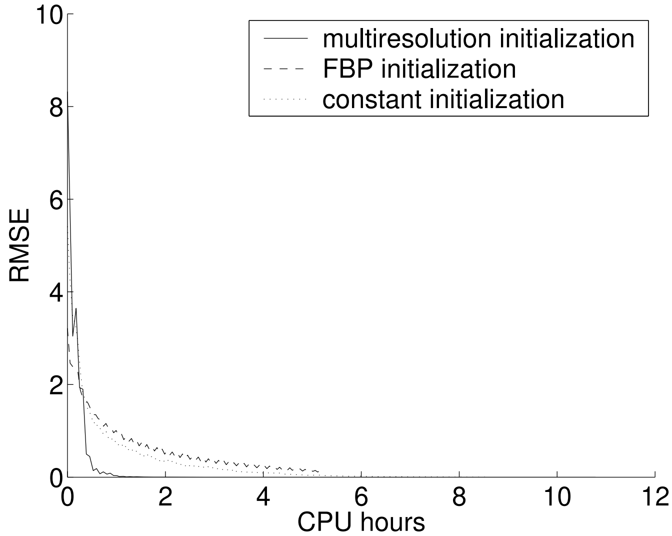


Fig. 3.23. Convergence speed of ICD-MAP for different initializations. The graph for the multiresolution initialization includes the CPU-time for the reconstructions at all resolutions. The graphs are based on the reconstruction of the 10.4M count bar phantom data using a GMRF prior with smoothing parameters selected to minimize error in quadrant 4.

expect a constant initialization to perform well. However, it is surprising, that the convergence for FBP initialization is actually worse than that for constant initialization. This effect may be due to local or global scaling differences of the FBP reconstruction in comparison to the MAP estimate. While the convergence for FBP initialization can potentially be improved through more careful scaling, the multiresolution initialization is uncomplicated, very effective and does not require separate implementation of FBP.

3.5 Conclusions

We have presented a quantitative comparison of FBP, EM and Bayesian reconstruction algorithms in combination with different system kernels for the IndyPET animal tomography scanner. The results demonstrate that accurately modeling the tomography scanner considerably improves the reconstruction quality of the iterative techniques.

When an inaccurate system kernel is used, EM and Bayesian MAP with a GMRF prior produce comparable reconstruction quality. For this case, the quality of the iterative methods is not clearly superior to FBP as evidenced by the similar bias vs. standard deviation envelopes for the bar phantom reconstruction.

However, when an accurate system kernel is used, the iterative methods produce much higher reconstruction quality as compared to FBP. Furthermore, when an accurate system kernel is used, the Bayesian methods perform considerably better than EM. Specifically, Bayesian MAP techniques using GMRF and GGMRF prior models achieve lower bias at equal variance as well as visually superior reconstructions as compared to EM.

For the Bayesian methods, we have compared Gaussian and generalized Gaussian Markov random field prior models as well as a recently proposed wavelet graph prior model. Our results indicate, that the GGMRF can improve quantitative accuracy and can produce sharper edges over the GMRF. The wavelet graph model produced sharper edges than the MRF priors but the results had artifacts. For a simulated phantom containing a lesion on an approximately uniform background, we have found the modifications proposed by Fessler and Rogers[68] to improve the performance of the Bayesian GMRF and GGMRF techniques.

In terms of computational efficiency, Bayesian MAP using an ICD optimization algorithm with multiresolution initialization was much faster than EM.

4. MULTISCALE BAYESIAN METHODS FOR DISCRETE TOMOGRAPHY

Abstract

Statistical methods of discrete tomographic reconstruction pose new problems both in stochastic modeling to define an optimal reconstruction, and in optimization to find that reconstruction. Multiscale models have succeeded in improving representation of structure of varying scale in imagery, a chronic problem for common Markov random fields. This chapter shows that associated multiscale methods of optimization also avoid local minima of the log *a posteriori* probability better than single-resolution techniques. These methods are applied here to both segmentation/reconstruction of the unknown cross-sections, and estimation of unknown parameters represented by the discrete levels.

4.1 Introduction

The reconstruction of images from projections is important in a variety of problems including tasks in medical imaging and non-destructive testing. Perhaps, the reconstruction technique most frequently used in commercial applications is convolution backprojection (CBP) [79]. While CBP works well for reconstruction problems with a complete set of projections having high signal-to-noise ratio (SNR), special cases benefit from alternative algorithms which can better model the imaging geometry and measurement process. These cases arise, for example, in low dosage medical imaging [80], non-destructive testing of materials with widely varying densities [81] and applications with limited angle projections [82] or hollow projections [83]. In such cases, statistical and discrete-valued methods can substantially improve the reconstruction quality by incorporating important prior information about both the imaging system and the object being imaged. Discrete reconstruction methods are based on the assumption that the object being imaged is composed of a discrete set of materials each with uniform properties. Therefore, an ideal reconstruction should only contain pixel or voxel values from a corresponding set of discrete levels. In this case, the problem of reconstruction reduces to one of determining the specific levels present in a reconstruction and then classifying each pixel to one of these discrete levels. Discrete reconstruction methods impose a very strong constraint on the reconstruction process, and therefore can substantially improve reconstruction quality.

Early methods for discrete-valued reconstruction focused on reconstructions of binary arrays from only the horizontal and vertical projections [84]. The deterministic projections were treated as a system of linear equations. Attention was particularly paid to the ambiguity of reconstructions which was formulated in the context of switching components [84, 85]. Algorithms for unambiguous reconstruction were developed by assuming object constraints such as connectedness in 2-D [86] or convexity in 3-D [87]. In addition, these concepts were extended to four or more projection angles including the analysis of the ambiguity problem [85, 88]. However, all of these techniques assume deterministic projection measurements and do not perform optimally under high noise conditions.

A second approach to discrete-valued reconstruction detects parameterized objects directly in the projection domain. This strategy is applicable when the objective is to detect specific objects or regions such as tumors in medical imaging or material defects in non-destructive testing. Rossi and Willsky [89] introduced this approach by performing maximum likelihood (ML) estimation of the location of a single object in the imaging plane. This concept was later extended to a three-dimensional parameterization supporting multiple objects per plane [90]. Here, constrained objects in 3-D are formed as a combination of basic cylinders whose parameters are estimated as part of the reconstruction. A review of object parameterization methods as well as a new algorithm for the approximate reconstruction of compact objects modeled by polyhedral shapes was presented by by A. Mohammad-Djafari and C. Soussen in [91]. Parameterized object reconstruction methods are specifically designed for low SNR conditions. However, they rely on *a priori* knowledge about shape characteristics of the objects in the cross-section. These methods are therefore not applicable in cases where such information is unavailable or the objects in the cross-section cannot easily be parameterized.

In this work, we focus on discrete-valued reconstruction from noisy projections using statistical methods. Statistical methods model the random nature of the physical data collection process, then seek the solution that best matches the probabilistic behavior of the data. Consequently, statistical methods can improve performance considerably in cases of low SNR. Statistical approaches also easily incorporate special geometries such as limited or missing angle projection measurements. Common statistical techniques incorporate implicit information about desired characteristics of the reconstruction without explicit modeling of objects in the cross-section.

A statistical method that is well suited for tomographic reconstruction is Bayesian maximum *a posteriori* (MAP) estimation. Bayesian methods in general have been shown to improve performance in many emission and transmission tomography problems [92, 6, 7, 41] as well as in image restoration tasks [93, 1]. Bayesian MAP estimation reconstructs the image as a tradeoff between matching the projection data and regularizing the solution by a prior probability distribution. The regularization imposed by the prior reflects assumed characteristics of feasible reconstructions. Due to this regularization, the MAP estimation problem is well-posed and avoids the high noise sensitivity frequently encountered in maximum likelihood (ML) estimation.

Priors for Bayesian reconstruction methods are often chosen to impose smoothness constraints on the reconstruction to eliminate high frequency noise. A prior model that has generally proven to be useful in the tomographic setting is the Markov random field (MRF) image model [92, 94, 6, 7, 10]. The chapter by M. T. Chan, G. T. Herman and E. Levitan in [91] presents a new MRF for modeling image prior distributions and describes methods for estimating the MRF's parameters. These parameter estimation techniques allow the model to be adapted to the specific characteristics of an ensemble of images. Importantly, the Bayesian estimate is computed directly from the convolution backprojection reconstruction, rather than from the original projection data. This approach has the advantage of reducing computation.

In contrast, we use a simple discrete MRF model [95, 2, 96], and instead focus on the computational difficulties resulting from direct Bayesian reconstruction from the tomographic data. In order to solve this optimization problem, we employ multiscale algorithms to both reduce computation and improve convergence to the global minimum. In addition, we introduce a method for estimating the densities of the discrete regions as part of the reconstruction process. This is important because precise knowledge of these discrete densities is required for accurate Bayesian reconstruction.

The MAP reconstruction itself can be formulated as an optimization problem which can be solved using a number of different techniques. The expectation-maximization (EM) algorithm, suitable for ML reconstruction [51], has been adopted for MAP estimation with Gaussian priors [97, 98, 99, 100]. Extensions of these models to more general MRF priors were proposed in [6, 101, 7, 57]. However, application of the EM algorithm for MAP estimation is difficult and usually suffers from slow convergence.

Instead of using EM techniques, we focus on the direct optimization of the MAP equation. We adopt a pixel-wise update method known as iterative coordinate descent (ICD) [41, 42] which maximizes the MAP criterion by iteratively updating each pixel of the image. The discrete version of ICD used here essentially implements the iterated conditional modes (ICM) technique introduced by Besag [2]. However, while ICM was designed for image restoration tasks, the ICD algorithm is specifically designed for the tomographic reconstruction problem resulting in dramatically improved computational efficiency.

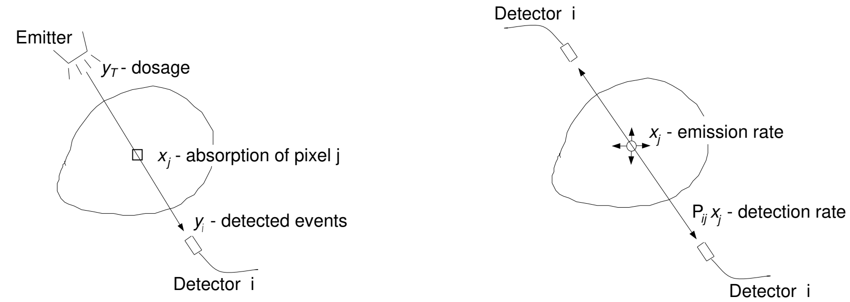
In addition to solving the optimization problem, the discrete-valued reconstruction requires knowledge of the density or emission rate levels in the cross-section. In practice, exact information about these discrete levels is often unavailable. In such cases it is desirable to estimate the densities or emission rates as part of the reconstruction. In Section 4.5, we discuss our method to estimate these discrete classes concurrently with the reconstruction process. We show how the class estimation can be formulated as a continuous-valued tomographic reconstruction problem with the number of points equal to the number of classes.

Finally, we extend our reconstruction method to a multiresolution algorithm. Multiresolution techniques achieve performance improvements in a variety of imaging problems [20, 102] including image segmentation [96, 28] and continuous-valued tomographic reconstruction [32]. Multiresolution algorithms reconstruct the image at different resolutions, typically progressing from coarse to fine scale. The coarse scale solutions serve as initialization or prior information for reconstructions at finer scales. Due to the improved initialization and the higher SNR at coarse scale, multiresolution algorithms are typically more robust with respect to local minima. In addition, local pixel interactions at coarse scale are equivalent to large scale interactions at fine scale. This combined with the low computational complexity at coarse scale makes multiscale algorithms very efficient.

The multiscale algorithm presented here is a straightforward extension of our fixed scale algorithm. The reconstruction is performed in a coarse-to-fine fashion by initializing each resolution level with the interpolated reconstruction of the next coarser level. The reconstructions at each level are computed using the fixed resolution method. Our experimental results demonstrate that this multiscale algorithm is less prone to being trapped in local minima and in many cases, computationally more efficient than the fixed resolution version.

4.2 Stochastic data models for tomography

In this section, we will develop the statistical framework for MAP reconstruction in computed tomography. Our framework is applicable to both transmission and emission measurements and supports general imaging geometries. The models presented here are based on the exact Poisson statistics of the photon measurements.



(a) Transmission setup (b) PET emission setup

Fig. 4.1. Physical setup for transmission and positron emission tomography (PET). In transmission tomography (a), photons are induced into the cross-section. After attenuation by the absorptive material, detectors measure the remaining photon rates. In positron emission tomography (b), the cross-section contains a radioisotope and is surrounded by a ring of detectors. If two detectors register photons at the same time, a pixel emission is assumed to have occurred on the projection line between them.

Computationally more efficient but approximate models can be obtained by using a Taylor expansion of the likelihood function [41, 42].

In transmission tomography, the objective is to measure photon attenuation for different projections through a cross-section of absorptive material. An illustration of the physical setup is shown in Fig. 4.1(a). The cross-section is surrounded by a ring of photon emitters and detectors. The emitters induce a calibrated photon rate y_T directed along certain angular and parallel projections. After attenuation by the absorptive material, the photon rates are measured by the detector opposite the respective emitter. The photon rates measured by the detectors are not direct measurements of attenuation. Instead they are noisy photon counts which can be modeled as Poisson-distributed random variables.

In order to write the probability density for the Poisson measurements, define X as the N-dimensional vector of attenuation densities of the pixels in raster order. Let Y denote the vector of photon counts for all M projections at different angles and parallel offsets. Furthermore, let P_{ij} correspond to the length of intersection between the j^{th} pixel and the i^{th} projection. Then P is the matrix of elements P_{ij} and P_{i*}

denotes the vector formed by its i^{th} row. Given these assumptions, the photon count Y_i, corresponding to projection i, is Poisson distributed with mean $y_T \exp(-P_{i*}x)$. The distribution of the Y_i may then be written as

$$\mathcal{P}(Y = y|x) = \prod_{i=1}^{M} \frac{\exp(-y_T e^{-P_{i*}x})(y_T e^{-P_{i*}x})^{y_i}}{y_i!}. \tag{4.1}$$

We use upper case letters for random variables and lower case letters for particular realizations. Taking the logarithm of (4.1), we obtain the log-likelihood

$$\begin{aligned}\text{(transmission)} \quad L(y|x) = \log \mathcal{P}(Y = y|x) = \\ \sum_{i=1}^{M} \left(-y_T e^{-P_{i*}x} + y_i(\log y_T - P_{i*}x) - \log(y_i!)\right).\end{aligned} \tag{4.2}$$

In emission tomography, no dosage is induced into the cross-section. Instead, the image plane contains some photon emitting material. A physical setup for the specific example of positron emission tomography (PET) is shown in Fig. 4.1(b). In this case, the cross-section contains a radioactive isotope. Recombination of positrons in the radioisotope results in emission of gamma rays in two opposite directions. These gamma rays are detected by a ring of detectors around the cross-section. If two detectors register photons at the same time, this is counted as an emission on the projection line between them. In the following, we will develop the statistical model for the general emission case.

The objective in emission tomography is to reconstruct the emission rates of all pixels in the image plane. Again, the photon detections can be modeled as Poisson distributed random variables. In order to emphasize the similarity to the transmission problem, we will use the same notation, but interpret x as the vector of emission rates for all N pixels and Y as the observed photon counts. We define P_{ij} as the probability that an emission from pixel j is registered by the i^{th} detector pair. The photon counts Y are then Poisson distributed with parameter $P_{i*}x$ which yields the distribution

$$\mathcal{P}(Y = y|x) = \prod_{i=1}^{M} \frac{\exp(-P_{i*}x)(P_{i*}x)^{y_i}}{y_i!}. \tag{4.3}$$

The log-likelihood is therefore given by

$$\text{(emission)} \qquad L(y|x) = \log \mathcal{P}(Y = y|x) = \sum_{i=1}^{M} \left(-P_{i*}x \; + y_i \log(P_{i*}x) - \log(y_i!) \right). \tag{4.4}$$

The log-likelihood functions for both the transmission and the emission case have the form

$$\log \mathcal{P}(Y = y|x) = -\sum_{i=1}^{M} f_i(P_{i*}x) \tag{4.5}$$

where $f_i(\cdot)$ are convex and differentiable functions. This common form will lead to similar methods of solving these two problems. In the following, we will write all equations for the emission case; however, all methods apply analogously to the transmission case.

For the emission problem, maximum likelihood (ML) estimation of x from y yields the optimization problem

$$\hat{x}_{ML} = \arg\min_{x} \sum_{i=1}^{M} \left(P_{i*}x \; - y_i \log(P_{i*}x) \right) \; . \tag{4.6}$$

For low signal-to-noise-ratio medical imaging problems, the ML estimate has well documented shortcomings [103, 104, 105]. Noise and sampling limitations can produce high frequency noise in the ML reconstruction that is not present in the original cross-section. It is therefore desirable to regularize tomographic inversion by some means. Maximum *a posteriori* probability (MAP) estimation addresses this problem by treating the original image as a random field, X, with prior distribution, $p(x)$. Again, we use a lower case x to denote a particular realization of the random vector X. The prior distribution regularizes the optimization problem so that a unique solution always exists [106]. The logarithm of the *a posteriori* distribution of X given Y may be computed using Bayes' formula.

$$\begin{aligned} L_p(x|y) &\stackrel{\triangle}{=} \log \mathcal{P}(X = x|Y = y) \\ &= L(y|x) + \log p(x) - \log \mathcal{P}(Y = y) \end{aligned} \tag{4.7}$$

The maximum *a posteriori* (MAP) estimate is then the value of $\hat{x}$ which maximizes the *a posteriori* density given the observations y

$$\begin{aligned} \hat{x} &= \arg\max_x L_p(x|y) \\ &= \arg\max_x \{L(y|x) + \log p(x)\} \\ &= \arg\max_x \{L(y,x)\} . \end{aligned} \tag{4.8}$$

The last equation indicates that the MAP estimate also maximizes the log of the joint distribution, $L(y,x) = \log \mathcal{P}(X = x, Y = y)$.

The MAP estimate has been shown to substantially improve performance in many image reconstruction and estimation problems. While computation of the exact MAP estimate is computationally intractable for the discrete problem, approximate solutions can be obtained with reasonable complexity as outlined in the next section. We will treat only the MAP estimation problem, since the ML estimate is the special case of a constant prior distribution.

4.3 Markov random field prior models

While the likelihood term $L(y|x)$ in the MAP equation (4.8) is determined by the physics of the data collection process, the prior distribution is selected by the experimenter to model desired characteristics of typical reconstructions. Most commonly, the prior models are chosen to reflect the high correlation of adjacent pixels. A model that has proven particularly useful is the Markov random field (MRF) [107]. Similar to a Markov chain in one dimension, the 2D MRF limits pixel interactions to a local neighborhood of pixels. This localization allows for efficient optimization of the MAP equation.

In order to write the equations for the MRF prior, we first need to define the concept of a neighborhood. If i denotes a single pixel location, we will denote its neighborhood by ∂i. This neighborhood can consist of any set of pixels $\{k : k \neq i\}$ which satisfies the symmetry property that $i \in \partial k \Rightarrow k \in \partial i$. Given this definition, a MRF is a random field which has the property

$$p(x_i|x_j, j \neq i) = p(x_i|x_{\partial i}). \tag{4.9}$$

In other words, the prior conditional probability of a pixel value depends only on a local neighborhood of pixels. Under some weak technical conditions, a random field is a MRF if and only if it has a probability distribution corresponding to a Gibbs distribution [107, 108]. This result, which is known as the Hammersley-Clifford theorem, may be used to express the likelihood function $\log p(x)$. While the theory of MRF's is quite extensive [109, 110, 111], we will restrict ourselves to a simple model based on at most an 8-point neighborhood.

Since we are interested in discrete-valued tomographic reconstruction, we assume that each pixel takes on a value from a set $\mathcal{E}$ of K discrete emission rates. We then apply a discrete MRF prior model that is frequently used in segmentation problems [95, 2, 96]. The model encourages neighboring locations to have the same states or, in our case, emission rates. To define the model, we must first define two simple functions, $t_1(x)$ and $t_2(x)$. $t_1(x)$ is the number of horizontally and vertically neighboring pixel pairs with different emission rates in x, and $t_2(x)$ is the number of diagonally neighboring pixel pairs with different emission rates in x. The discrete density function for $x \in \mathcal{E}^N$ is then assumed to be of the form

$$\log p(x) = -(\beta_1 t_1(x) + \beta_2 t_2(x)) + \log(Z) \tag{4.10}$$

where Z is an unknown constant called the partition function. The regularization parameters β_1 and β_2 weight the influence of the prior in comparison to the likelihood term. Larger values of β_1 and β_2 assign higher cost to local pixel differences which will result in a smoother reconstruction. Based on the geometry of the 8-point neighborhood, β_2 is often chosen as $\beta_2 = \beta_1/\sqrt{2}$. In the following, we will often write β for β_1 and assume $\beta_2 = \beta_1/\sqrt{2}$.

Substituting the prior (4.10) into the MAP and likelihood equations (4.8) and (4.4) for the emission case, we obtain the optimization criterion

$$\begin{aligned} \hat{x} = \arg\max_{x \in \mathcal{E}^N} \Big\{ \sum_{i=1}^{M} & (-P_{i*}x \ + y_i \log(P_{i*}x)) \\ & -(\beta_1 t_1(x) + \beta_2 t_2(x)) \Big\}. \end{aligned} \tag{4.11}$$

4.4 Optimization techniques

In order to compute the MAP reconstruction, we must perform the optimization of (4.11). Gradient methods such as steepest descent or conjugate gradient optimization are not directly applicable, since the discrete prior (4.10) is non-differentiable.

A method that is well suited for the MAP optimization is a discrete version of iterative coordinate descent (ICD) [41, 42]. The ICD method sequentially updates each pixel of the image. With each update, the current pixel is chosen to maximize the posterior probability (4.11). Therefore, the discrete ICD algorithm essentially implements the well known ICM optimization introduced by Besag [2]. However, while ICM was originally developed for image restoration tasks, the ICD implementation is specifically designed for the tomographic reconstruction problem. The ICD algorithm takes advantage of the sparse structure of the forward projection matrix P to dramatically speed-up the optimization. Furthermore, ICD initializes the optimization with the convolution backprojection instead of the ML initialization used by ICM. The ML estimate is not a good initialization for tomographic reconstruction problems and, since the pixel likelihoods are not independent, the ML estimate is computationally expensive to compute. The convolution backprojection, in comparison, is inexpensive to compute and captures most of the low-spatial frequency behavior of the reconstruction. This makes the CBP a suitable initialization, especially since coordinate-wise update methods have slow convergence for low spatial frequencies and fast convergence for high-spatial frequencies. In the following, we will show how the ICD can be used to efficiently compute the MAP estimate.

Let $v_1(z, x_{\partial j})$ be the number of horizontal and vertical neighbors of x_j which do not have emission rate z, and $v_2(z, x_{\partial j})$ be the number of diagonal neighbors of x_j which do not have emission rate z. Then, the maximization of the MAP equation with respect to pixel x_j can be written as

$$\begin{aligned} x_j^{n+1} = \arg\min_z \{ & -L(y|X_j = z, X_k = x_k^n, k \neq j) \\ & + (\beta_1 v_1(z, x_{\partial j}) + \beta_2 v_2(z, x_{\partial j}))\}. \end{aligned} \tag{4.12}$$

In our notation, x^n is the image containing all previous pixel updates. Thus, the reconstruction x^{n+1} differs from x^n only at pixel j. A full update of the reconstruction requires N applications of (4.12).

Computation of the log-likelihood $L(y|z, x^n)$ using (4.4) for each pixel update would still lead to prohibitive computational complexity. This can be avoided by using only the change in the log-likelihood

$$\Delta L(z) = \sum_{i \in \mathcal{I}_j} \Big(-P_{ij} z + y_i \log(P_{i*} x^n + P_{ij}(z - x_j^n)) - y_i \log(P_{i*} x^n) \Big) \tag{4.13}$$

where I_j is the set of projections i which intersect pixel x_j, i.e. $\mathcal{I}_j = \{i : P_{ij} \neq 0 \ , \ 1 \leq i \leq M\}$. Leaving out the terms which are constant with respect to z, the update equation for x_j can then be written as

$$\begin{aligned} x_j^{n+1} = \arg\min_z \Big\{ & \sum_{i \in \mathcal{I}_j} (P_{ij} z - y_i \log(P_{i*} x^n + P_{ij}(z - x_j^n))) \\ & + (\beta_1 v_1(z, x_{\partial j}) + \beta_2 v_2(z, x_{\partial j})) \Big\}. \end{aligned} \tag{4.14}$$

Assuming a reasonably small set $\mathcal{E}$ of K fixed emission rates, the minimization can be carried out by trying all $z \in \mathcal{E}$ and selecting the one which minimizes (4.14). We store the M-dimensional state vector $S = Px$ between iterations. After a pixel x_j is updated, the components of S can be efficiently updated using

$$S_i^{n+1} = S_i^n + P_{ij}(x_j^{n+1} - x_j^n). \tag{4.15}$$

This update is necessary only for the components $i \in \mathcal{I}_j$ since for all other projections $P_{ij} = 0$.

In order to assess the computational complexity of the reconstruction, we first define M_0 as the average number of projections passing through a single pixel

$$M_0 = \frac{1}{N} \sum_{j=1}^{N} |\mathcal{I}_j| \ . \tag{4.16}$$

The computational cost for a pixel update is then on the order of KM_0 operations. The complexity of a full-update of the reconstruction is therefore NKM_0. This is quite reasonable, considering that due to the sparsity of P, M_0 is typically small compared to M, i.e. $M_0 \ll M$. In cases where K is large, it might be desirable to reduce computation by using a global second order approximation to the likelihood functions (4.2) and (4.4) as described in [41, 42]. Using these techniques, computation for a single pixel update can be reduced to order $K+M_0$, resulting in $N(K+M_0)$ complexity for a full reconstruction update.

4.5 Estimation of discrete levels

So far, we have assumed the set $\mathcal{E}$ of discrete emission rates or densities to be known. In practice, however, the exact emission rates corresponding to different regions in the cross-section may not be known. Even if a good initial guess is available, the accuracy of the emission rates is critical for the reconstruction. For illustration, assume that the emission rate of a particular region is over-estimated by some amount. For projections y_i which pass through this region, the forward projected reconstruction $P_{i*}x$ will be larger than the measured photon count, i.e. $P_{i*}x > y_i$. To compensate for this mismatch, the reconstruction algorithm may misclassify large numbers of pixels. Therefore, it is desirable to estimate the discrete emission rates as part of the reconstruction algorithm.

In this section, we show how the emission rates can be estimated concurrently with the reconstruction. We implement ML estimation of the emission rates by iteratively updating entire regions of pixels with equal emission rates [112]. We will show that this estimation is equivalent to a continuous-valued tomographic ML reconstruction problem with K pixels. The updates of the emission rates will be performed between each full ICD update of the reconstruction.

Let $\theta_1 \ldots \theta_K$ denote the discrete emission rates so that $\mathcal{E} = \{\theta_1, \cdots, \theta_K\}$. Changing a single emission rate θ_k is equivalent to changing all pixels in the reconstruction that are classified to have emission rate θ_k. If we define a region as the collection of all pixels with the same emission rate, we obtain K different regions in the reconstruction. Analogously to the projection matrix P for individual pixels, we can now define a projection matrix Q for the regions. Given the region geometries, we can compute the entry Q_{ik} as the probability that an emission from the k^{th} region is registered by the i^{th} detector.

Assuming knowledge of Q, the likelihood for the emission rates can be computed analogously to the pixel likelihoods in (4.2) and (4.4). The resulting optimization problem is equivalent to a continuous-valued reconstruction with K pixels and projection matrix Q.

In practice, direct computation of Q from the geometry of the regions would be computationally involved and difficult to update. Instead, we can obtain an expression for the entries of Q by adding the contributions of all pixels in a region. We can rewrite the i'th forward projection $P_{i*}x$ as follows

$$\begin{aligned} P_{i*}x &= \sum_{j=1}^{N} P_{ij}x_j \\ &= \sum_{k=1}^{K}\left(\theta_k \sum_{\{j:x_j=\theta_k\}} P_{ij}\right) \\ &= \sum_{k=1}^{K}\theta_k Q_{ik} \quad = \quad Q_{i*}\theta \end{aligned} \tag{4.17}$$

where

$$Q_{ik} = \sum_{\{j:x_j=\theta_k\}} P_{ij}. \tag{4.18}$$

For the emission case, this yields the likelihood function

$$\log \mathcal{P}(Y=y|\theta) = \sum_{i=1}^{M}\left(-Q_{i*}\theta + y_i \log(Q_{i*}\theta) - \log(y_i!)\right). \tag{4.19}$$

This log-likelihood function is clearly of the same form as (4.4), except that the discrete-valued N-component vector x is replaced by the continuous-valued K-component vector θ. Also, the new projection matrix Q is of size $M \times K$ instead of $M \times N$. Thus, maximum likelihood estimation of θ is equivalent to a continuous-valued tomographic ML reconstruction with K pixels.

Since θ is continuous-valued, the optimization of (4.19) is different from the discrete case. In general, all methods proposed for continuous-valued tomographic reconstruction can be applied. Again, we will use ICD optimization since it is easily implemented with constraints such as positivity of the emission rates. The ICD update equation of the θ_k is analogously to (4.14) given by

$$\theta_k^{n+1} = \arg\min_{v \geq 0} \left\{ \sum_{i \in \tilde{I}_k} (Q_{ik} v - y_i \log(Q_{i*}\theta^n + Q_{ik}(v - \theta_k^n))) \right\} \tag{4.20}$$

where $v \geq 0$ enforces the non-negativity of the emission rates and $\tilde{I}_k$ is defined as $\tilde{I}_k = \{i : Q_{ik} \neq 0 \ , \ 1 \leq i \leq M\}$. Since the cost function in (4.20) is well approximated by a quadratic, the optimization can be efficiently implemented using Newton's method.

Let ϕ_1 and ϕ_2 be the first and second derivatives of the log-likelihood function evaluated at the current emission rate θ_k. The Newton update for minimization of (4.20) is then given by

$$\theta'_k = \max\left\{ \theta_k - \frac{\phi_1}{\phi_2} \ , \ 0 \right\} \tag{4.21}$$

where the derivatives ϕ_1 and ϕ_2 are computed as

$$\phi_1 = \sum_{i \in \tilde{I}_k} Q_{ik} \left(1 - \frac{y_i}{Q_{i*}\theta} \right) \tag{4.22}$$

$$\phi_2 = \sum_{i \in \tilde{I}_k} y_i \left(\frac{Q_{ik}}{Q_{i*}\theta} \right)^2 . \tag{4.23}$$

The Newton updates (4.21)-(4.23) are repeatedly applied until $|\phi_1| < \epsilon$. For our experimentation, we have found an accuracy of $\epsilon = 0.001$ to be sufficient. For efficient computation, we store and update the same state vector $S = Px = Q\theta$ as in the discrete MAP reconstruction (4.15). If θ_k^n is updated to θ_k^{n+1}, S can be updated as

$$S_i^{n+1} = S_i^n + Q_{ik}(\theta_k^{n+1} - \theta_k^n) \tag{4.24}$$

for all $i \in \tilde{I}_k$.

In order to assess computational complexity, we define H as the average number of Newton-iterations per class update (4.20). Furthermore, we use the number of projections M as a bound for the size of the sets $\tilde{I}_k$. Computation of a single emission rate θ_k then requires on the order of MH multiplies and divides. Typically, only a few Newton-iterations are necessary to obtain sufficient accuracy ($H < 2$). The computational complexity of updating the ML estimate of θ is then KMH. Note that KMH is typically small in comparison to a complete update of the discrete reconstruction x of order NKM_0.

The iterations for the estimation of θ can be performed between full reconstruction updates. Each time a pixel changes during the reconstruction, the new Q matrix can be obtained as follows: If $x_j^n = \theta_k$ and $x_j^{n+1} = \theta_l$, then

$$\begin{aligned} Q'_{ik} &= Q_{ik} - P_{ij} \\ Q'_{il} &= Q_{il} + P_{ij} \end{aligned} \tag{4.25}$$

for all projections $i \in I_j$. This recursion results in a computationally efficient algorithm since it avoids recomputing Q using (4.18) after each reconstruction update.

In order to apply the estimation of the emission rates as described above, it is necessary to obtain initial values for the estimates of θ_k. In practice, initial values for the θ_k can be extracted from the convolution backprojection reconstruction. One possibility is to extract the initial θ_k manually by taking the average value of approximately uniform regions in the CBP. This ensures that the estimated emission rates correspond to the regions of interest in the reconstruction and minimizes chances of the estimation getting trapped in local minima.

If, on the other hand, a fully unsupervised algorithm is desired, clustering techniques can be applied to the CBP reconstruction to estimate the initial emission rates θ_k. To do this, we used a clustering method based on Gaussian mixture models and the EM algorithm [113]. This method used the Rissanen criterion to estimate both the number of clusters K as well as the mean emission rate of each cluster. These estimates were then used to initialize the estimation of the emission rates which resulted in a fully unsupervised reconstruction algorithm.

4.6 Multiscale approaches

We now extend the previous results to a multiresolution framework. Multiresolution algorithms reconstruct the cross-section at different resolutions, typically starting at coarse resolution and progressing to the desired finest resolution. Fig.

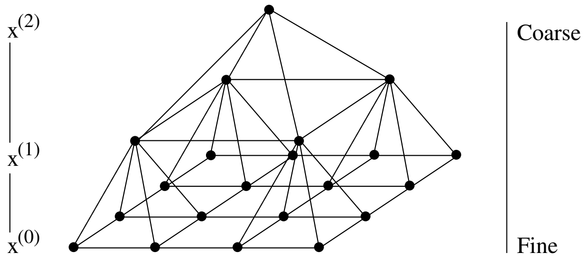


Fig. 4.2. Illustration of multiresolution structure. Shown are pixels in three different resolution levels. We assume a quadtree structure in which each coarse scale pixel corresponds to four pixels at the next finer scale.

4.2 illustrates the multiresolution structure used in our algorithm. We assume a quadtree model in which each coarse scale pixel corresponds to four pixels at the next finer scale. Each resolution level is half the size of the next finer level in each direction and therefore contains only $1/4^{th}$ the number of pixels. The small number of pixels at coarse scales implies lower computational complexity at these levels.

Compared to the fixed scale reconstruction, the multiscale approach has several significant advantages. While at first, reconstructing the cross-section at several resolutions might seem like additional overhead, the multiscale algorithm has substantially faster convergence behavior. The fixed scale ICD reconstruction algorithm updates one pixel at a time using a prior that only depends on a small pixel neighborhood. As a result, propagation of information per iteration is limited which results in slow convergence for low spatial frequencies. The multiscale version of the algorithm improves this by first reconstructing the image at coarse resolutions where local interactions are equivalent to large scale propagation at fine resolutions. The coarse scale reconstructions then serve as an initialization for the finer reconstructions. Since the coarse reconstructions already contain the large scale behavior of the solution, substantially fewer iterations are necessary at the finer scales. This, combined with the fact that the coarse scale reconstructions are of low complexity, makes the multiscale algorithm very efficient. In addition to increased efficiency, the multiresolution algorithm is more robust with respect to local minima in the optimization. This increased robustness holds for both the reconstruction and the estimation of the emission rates.

4.7 Multiscale MRF

The multiscale MRF model is a straightforward generalization of the fixed resolution model in section 4.3. For the multiresolution case, we essentially use the fixed scale algorithm for each resolution level and use the result to initialize the next finer level [96].

Let $x^{(n)}$ denote the reconstruction at resolution n, where $n = 0$ is the finest resolution and $n = L-1$ is the coarsest resolution. In order to calculate the log-likelihood function for level n, we simply compute a new projection matrix $P^{(n)}$ which incorporates the larger pixel size at level n. The matrix $P^{(n)}$ is of dimension $M \times 4^{-n}N$. The log-likelihood for the emission case is then given by

$$L^{(n)}(y|x^{(n)}) = \sum_{i=1}^{M} \left(-P_{i*}^{(n)} x^{(n)} \ + y_i \log(P_{i*}^{(n)} x^{(n)}) - \log(y_i!) \right) . \tag{4.26}$$

This yields the MAP equation

$$\hat{x}^{(n)} = \arg\min_{x^{(n)}} \left\{ -L^{(n)}(y|x^{(n)}) + \beta_1^{(n)} t_1(x^{(n)}) + \beta_2^{(n)} t_2(x^{(n)}) \right\} . \tag{4.27}$$

The remaining question is how to choose the coarse resolution parameters $\beta_1^{(n)}$ and $\beta_2^{(n)}$. An intuitive approach is to choose these parameters so that the cost functions for any two adjacent resolutions are equal when the finer reconstruction $x^{(n-1)}$ equals the coarser reconstruction $x^{(n)}$ [96]. This assumes that the finer reconstruction $x^{(n)}$ is constant on blocks of 2 by 2 pixels.

Let I denote the operator which interpolates by a factor of two using pixel-replication. The equality of adjacent levels can then be written as $x^{(n-1)} = Ix^{(n)}$. We now observe that a horizontal or vertical pixel difference in $x^{(n)}$ results in two horizontal or vertical plus two diagonal differences in the pixel-replicated $Ix^{(n)}$, and one diagonal pixel difference in $x^{(n)}$ yields one diagonal pixel difference in $Ix^{(n)}$. Therefore, $x^{(n-1)} = Ix^{(n)}$ implies

$$t_1^{(n-1)} = 2\, t_1^{(n)}$$

$$t_2^{(n-1)} = 2\,t_1^{(n)} \;+\; t_2^{(n)}\;. \tag{4.28}$$

Consequently, the fine and coarse resolution cost functions will be equal if

$$\begin{aligned} \beta_1^{(n)} &= 2(\beta_1^{(n-1)} \;+\; \beta_2^{(n-1)}) \\ \beta_2^{(n)} &= \beta_2^{(n-1)} \end{aligned} \tag{4.29}$$

for all resolutions n. By using these parameters, minimization of (4.27) corresponds to the minimization of the original MAP equation (4.11) under the constraint that the solution be constant on the appropriately sized blocks.

Coarse resolution minimization using the parameters given by (4.29) will effectively minimize (4.11) if the correct segmentation is approximately block constant. However, this recursion for the parameters has an undesirable property. It implies that the MRF models for coarser resolution segmentations should have progressively higher spatial correlation, or alternatively, finer resolution segmentations should have lower correlation. This, of course, runs counter to normal assumptions of spatial coherence in images, and will tend to cause insufficient spatial correlation at finer resolutions or excessive correlation at coarse resolutions. A more reasonable approach is to assume that the spatial correlation is independent of the resolution since this avoids the problem of excessive correlation at coarse resolutions. Also, this assumption is appropriate when prior information is unavailable about the likely scale of regions in the image. Therefore, in all experimentation, we will fix the parameters of the MRF as a function of scale

$$\begin{aligned} \beta_1^{(n)} &= \beta_1 \\ \beta_2^{(n)} &= \beta_2. \end{aligned} \tag{4.30}$$

The L level Multiresolution MAP reconstruction algorithm may then be summarized as follows:

1. Compute CBP and estimate initial emission rates θ.

2. Classify CBP pixels into discrete emission rates θ_k using thresholding. Decimate the result $(L-1)$-times to initialize $x^{(L-1)}$. Set $n = L-1$.

Table 4.1
Computational complexity for reconstruction and emission rate updates. While the complexity of the reconstruction is a function of the resolution n, the cost for the estimation of θ is constant. For the overall reconstruction, the time spent on estimation of θ is typically less than 10% of total execution time.

	Reconstruction Update	Emission Rate Update
Fixed Scale	NKM_0	KMH
Multiscale	$4^{-n}NKM_0^{(n)} \approx 2^{-n}NKM_0$	KMH

N	Number of pixels in the reconstruction
K	Number of discrete emission rates θ_k
M	Number of projections
$M_0^{(n)}$	Average number of projections intersecting a pixel at scale n, $M_0 = M_0^{(0)}$
H	Average number of Newton iterations for update of single θ_k

3. Compute reconstruction $x^{(n)}$ using the following method:

   (a) Update $x^{(n)}$ using one full pass of discrete ICD algorithm.

   (b) If no pixel change occurs, goto 4.

   (c) Perform six full updates of discrete levels θ, goto (a).

4. If $n = 0$ stop.

5. Initialize $x^{(n-1)}$ with pixel-replicated $x^{(n)}$.

6. Set $n = n - 1$, goto 3.

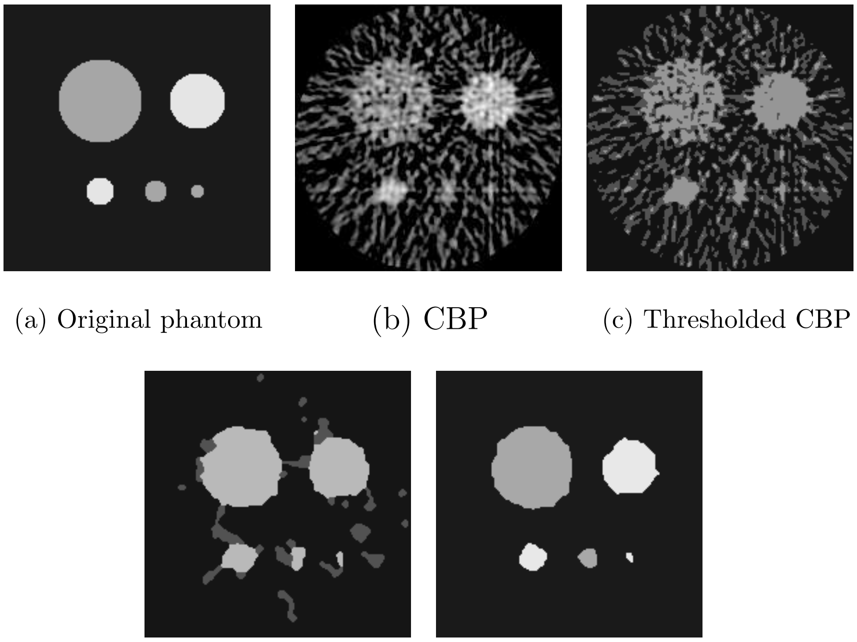
The parameters β_1 and β_2 can be chosen manually to achieve the amount of regularization desired.

4.8 Computational complexity

Table 4.1 compares the computational complexity for one full update of the reconstruction to one update of the emission rates θ. The complexity of the reconstruction update NKM_0 depends on the number of pixels N and therefore on the resolution of the reconstruction. In the multiresolution framework, each level contains $1/4^{th}$ the number of pixels of the next finer level. The reconstruction complexity at scale n is therefore $4^{-n}NKM_0^{(n)}$, where $M_0^{(n)}$ replaces M_0. Due to the larger pixel size at coarse resolution, more projections intersect each coarse-scale pixel and $M_0^{(n)}$ increases with n. In order to compare the complexities for different scales, we can approximate $M_0^{(n)}$ as follows: As the size of a pixel doubles in each direction, we assume that the number of parallel projections intersecting the pixel at each angle doubles. The cost of a reconstruction update at scale n is therefore approximately $2^{-n}NKM_0$. Assuming that the multiscale reconstruction performs the same number of iterations at each scale as the fixed resolution algorithm, the multiscale overhead is bounded by a factor of 2. In most real applications, however, the multiscale algorithm performs considerably fewer computationally expensive iterations at fine scale than the fixed scale method.

The complexity of the emission rate update is not a function of resolution. To obtain a bound on the computational cost, we assume that all M projections intersect each region. Since the number of Newton iterations per class update H is usually small, the upper bound KMH for the emission rate update is small in comparison to the cost NKM_0 for a reconstruction update at finest or fixed scale. At coarse resolutions, however, the complexity for the emission rate update may become comparable to the cost for a reconstruction update. Again, the advantage of the multiscale method is that the emission rates often converge after the reconstruction of the coarser scales. Therefore, fewer iterations for reconstruction and emission rate estimation are necessary at finer scale.

In practice, the total cost of the emission rate estimation is usually small compared to that of the reconstruction updates. Performing six updates of θ between full reconstruction updates, we find that the cost of the emission rate updates is typically less than 10% of the total execution time. In addition, the multiscale algorithm is typically faster than the fixed resolution method.



(a) Original phantom (b) CBP (c) Thresholded CBP

(d) Fix-Resolution MAP $\beta = 1.0$ (e) Multiresolution MAP $\beta = 1.0$

Fig. 4.3. Results for synthetic cross-section. Shown in (a) is the original cross-section. The continuous-valued CBP (b) contains considerable noise which is still present in the thresholded version (c), using the thresholds determined by unsupervised clustering. The fixed resolution algorithm (d), gets trapped in a local minimum resulting in class estimates close to the initialization. The multiresolution algorithm (e) estimates the classes correctly and achieves higher reconstruction performance.

4.9 Results

Reconstructions using the fixed and multiscale algorithms on synthetic data are shown in Fig. 4.3. Fig. 4.3(a) shows the original cross-section of size 192 by 192 pixels where each pixel is of both width and height 3.13mm. The cross-section contains pixels with three different emission rates as shown in Table 4.2. The projection data was calculated at 16 evenly spaced angles each with 192 parallel projections. The

Table 4.2
Original and estimated emission rates for synthetic cross-section. While the fixed scale algorithm gets trapped near the clustering initialization, the multiscale method estimates the emission rates quite accurately. All units are in mm^{-1}.

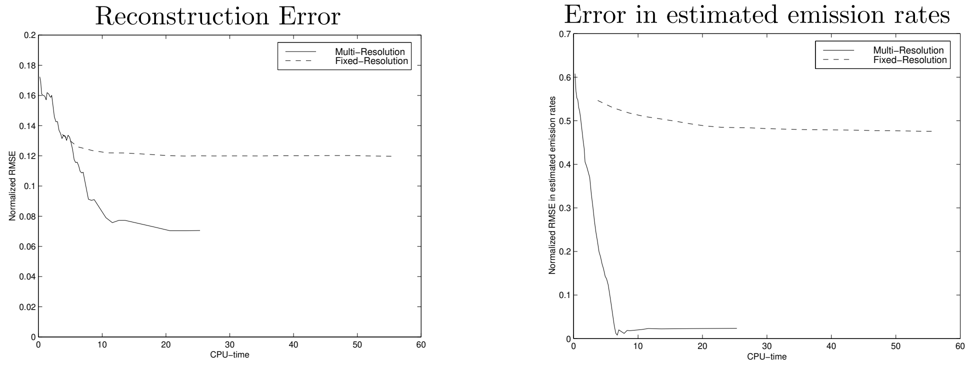
	Em.-Rate θ_1	Em.-Rate θ_2	Em.-Rate θ_3
Original Phantom	0.001	0.05	0.1
CBP Clustering	0.0005	0.0108	0.04
Fixed-Res. MAP	0.0007	0.0105	0.0632
Multi-Res. MAP	0.0010	0.0512	0.1028

Table 4.3
CPU-time in seconds for fixed- and multiresolution algorithms. Both algorithms were run until convergence and terminated when no pixel change occurred in a discrete reconstruction update.

	Multi-Resolution	Fixed-Resolution
Phantom I	27	59
Phantom II	349	404

projection beam was assumed to be infinitely thin. The data samples where formed by Poisson random variables with the appropriate means.

Fig. 4.3(b) shows the convolution backprojection (CBP) reconstruction using a generalized Hamming filter weighted by a Gaussian envelope. The CBP reconstruction was used to obtain initial values for the emission rates θ. The unsupervised clustering routine using a Gaussian mixture model applied to the CBP reconstruction identified three clusters with mean emission rates $\theta = [0.005, 0.0108, 0.04]$. The clustering result consists of two classes with very low emission rates corresponding to background pixels and only one class with higher emission rate corresponding to the discs in the foreground. A first discrete-valued reconstruction can be obtained

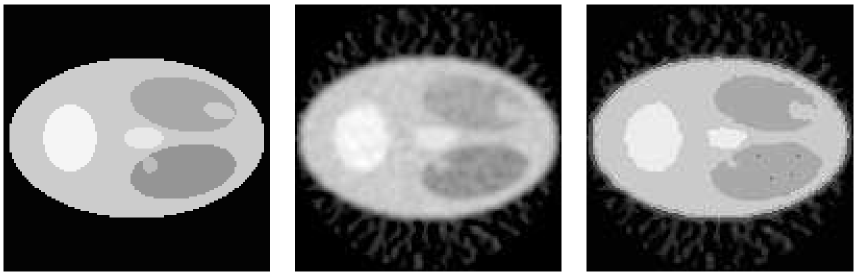


(a) (b)

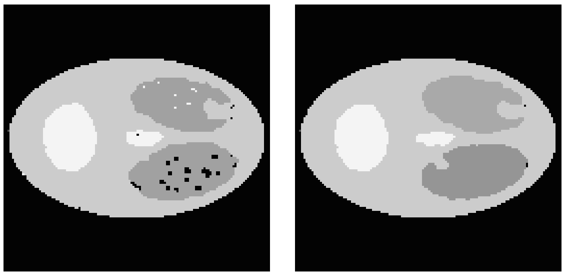
Fig. 4.4. Comparison of convergence for the fixed- and multiresolution MAP reconstructions of Fig. 4.3(d)-(e). Shown in (a) is the normalized reconstruction error as a function of CPU-time. The multiscale algorithm converges considerably faster than the fixed resolution method. This is partly due to the fast convergence of the emission rate estimates in the multiscale case (b). The multiresolution algorithm achieves lower final error in estimation and reconstruction.

by thresholding the CBP reconstruction using the midpoints between the emission rates determined by the clustering routine. The resulting thresholded CBP is shown in Fig. 4.3(c). In addition to the errors in class estimates, the result contains noise and aliasing effects. Fig. 4.3(d) shows the fixed resolution MAP reconstruction using $\beta = 1.0$ where we assume that $\beta_1 = \beta$ and $\beta_2 = \beta/\sqrt{2}$. The fixed resolution reconstruction was initialized to the thresholded CBP reconstruction and the class estimates were initialized to the clustering result. While the reconstruction is less noisy than the thresholded CBP, the estimation of emission rates is trapped in a local minimum close to the initial values from the clustering result. This results in the classification of all 5 discs into the same class of emission rates.

The reconstruction result using the multiscale algorithm with $L = 5$ resolution levels and $\beta = 1.0$ is shown in Fig. 4.3(e). The algorithm was initialized as in the fixed scale case. The estimated emission rates using the multiscale technique are very close to the true values as shown in Table 4.2. This results in correct classification of the 4 larger discs in the cross-section. Only the smallest disc is misclassified to a smaller area but higher emission rate than in the original phantom. This is not surprising, considering the high level of noise and the small size of the disc which practically eliminates it from coarser resolution levels. For $\beta = 1.0$, there is essentially no high frequency noise in the reconstruction. Overall, the quality of the multiscale reconstruction is superior to the fixed scale MAP reconstruction and the thresholded CBP. The multiscale method is particularly robust with respect to the estimation of the emission rates.



(a) Original phantom (b) CBP (c) Thresholded CBP



(d) Fix-Resolution MAP $\beta = 1.0$ (e) Multiresolution MAP $\beta = 1.0$

Fig. 4.5. Results for second synthetic cross-section. The clustering only identifies 2 out of 7 classes correctly which is reflected in the thresholded CBP (c). The fixed resolution algorithm (d) estimates 3 of 7 classes correctly, and fails to divide the two gray ovals on right into two distinct classes, which results in compensation artifacts. The multiresolution algorithm (e) performs better, estimating 5 out of seven classes correctly. Neither algorithm divides the bright left and center spot into two distinct classes. Both also miss the emission rate of the small patch within the lower right oval.

In addition to the superior reconstruction quality, the multiresolution method is faster than the fixed scale algorithm. Table 4.3 shows the execution times for both the fixed- and multiresolution reconstructions (Phantom I). Both algorithms were run until convergence and terminated when no pixel change occurred in a discrete reconstruction update. While the multiscale method terminated after 25 seconds, the fixed scale method needed 56 seconds to converge. Fig. 4.4 compares the error convergence of the two algorithms. Fig. 4.4(a) shows the reconstruction error as a function of CPU-time. The reconstruction error is calculated as normalized root mean square error, i.e.

$$E = \sqrt{\frac{\sum_{j=1}^{N}(\hat{x}_j - x_j)^2}{\sum_{j=1}^{N} x_j^2}} \tag{4.31}$$

where x denotes the true cross-section and $\hat{x}$ denotes the current estimate of the cross-section. In the multiscale case, $\hat{x}$ is computed by interpolating $x^{(n)}$ using pixel replication. The multiscale algorithm converges considerably faster and achieves lower final reconstruction error than the fixed resolution method. We also observed that the multiscale method achieves larger posterior likelihood, confirming that the fixed scale algorithm gets trapped in a local minimum.

The difference in convergence speed and final reconstruction error between the fixed and multiscale algorithm is reflected in the convergence behavior of the emission rate estimates. Fig. 4.4(b) shows the convergence of the normalized root mean square error of the emission rate estimates. For the multiscale algorithm, the θ are essentially converged after only a few coarse scale iterations. This reduces the number of computationally expensive iterations at finer scales, thereby accelerating overall convergence substantially.

Results for a second synthetic cross-section are shown in Fig. 4.5. The original cross-section in Fig. 4.5(a) contains 7 discrete levels of emission rates. The size of the phantom is 128 by 128 pixels where each pixel is of both width and height 1.56mm. The projection data was calculated at 128 evenly spaced angles, each with 128 parallel projections. Again, the data samples were obtained as Poisson random variables with the appropriate mean. Fig. 4.5(b) shows the convolution backprojection which is blurred and contains considerable amounts of noise. The clustering routine was used to obtain initial values for the emission rates. Due to the

Table 4.4
Original and estimated emission rates for second synthetic cross-section. Bold faced numbers represent emission rates that were estimated within a reasonable tolerance of their true values. Both, the fixed and multiscale algorithm were initialized to the emission rates determined by the clustering algorithm. All units are in mm^{-1}.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
Orig. Phantom	0.001	1.2	1.6	2.0	2.4	3.2	3.6
CBP Clustering	0.0005	0.028	0.094	0.307	**1.606**	**2.359**	3.335
Fixed-Res. MAP	**0.001**	0.028	0.094	0.338	1.445	**2.403**	**3.574**
Multi-Res. MAP	**0.001**	0.144	0.336	**1.211**	**1.602**	**2.404**	**3.578**

high noise in the CBP reconstruction the number of clusters was not estimated but manually set to seven. As shown in Table 4.4, the clustering only identified 2 out of the 7 classes within a reasonable tolerance of their true values. The thresholded CBP shown in Fig. 4.5(c) shows that the two gray ovals on the right are erroneously classified to have the same emission rate. Similarly, the bright patch on the left is set to the same emission rate as the center spot. In addition, the small patch within the gray oval on the lower right is misclassified. Fig. 4.5(d) shows the reconstruction using the fixed resolution algorithm. As before, the emission rates were initialized to the cluster means and the reconstruction was initialized to the thresholded CBP. As indicated in Table 4.4, the fixed resolution algorithm terminates with class estimates close to the initial cluster values. While the background level of 0.001 is now correctly estimated, the emission rate for the two gray ovals with original rates 1.2 and 1.6 has moved between the two values to $\theta_5 = 1.4458$. Notice that the overestimation of the lower right oval yields a mismatch between observed projection counts y and the forward projected reconstruction Px. Since the pixel values x within the oval are too large, we obtain $P_{i*}x > y_i$ for many projections intersecting this region. To compensate for this, the reconstruction algorithm inserts a pattern of black spots into the region which lowers the average projection count. Analogously, the gray

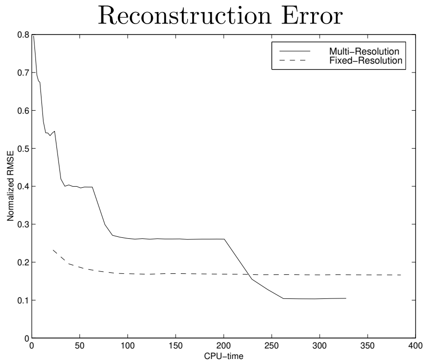


Fig. 4.6. Error convergence for fixed- and multiresolution MAP reconstructions of the second phantom. While the multiscale error at coarser scales is comparably high, convergence at each resolution is rapid. Again, the multiresolution algorithm terminates earlier and achieves lower final reconstruction error than the fixed-resolution method.

oval on the upper right contains white spots to compensate underestimation of the emission rate 1.6 by 1.4458. Effects such as these result from the algorithm's being trapped in local minima of the MAP cost function. This is particularly critical for fixed resolution reconstructions that include the estimation of emission rates.

The multiscale algorithm is less prone to being trapped in local minima. For this phantom, 5 out of 7 emission rates are estimated correctly. The reconstruction as shown in Fig. 4.5(e) contains little noise and classifies the two gray ovals correctly. However, the left and center bright regions with emission rates 3.6 and 3.2 are still both classified as having a single emission rate of 3.5783. This can be improved by initializing the emission rates closer to their true values. In general, by varying the initial estimate for the emission rates, it is often possible for the fixed scale algorithm to obtain reconstructions comparable to the multiscale reconstruction. However, the fixed resolution algorithm is less robust with respect to the emission rate estimation and close initialization does not guarantee a comparable reconstruction.

In almost all cases, with and without estimation of emission rates, the multiresolution algorithm is faster than the fixed resolution method. As shown in Table 4.3 (Phantom II), the multiscale algorithm terminates after 349 seconds compared

Table 4.5
Percentage of CPU-time used for estimation of emission rates. In all cases, the complexity for estimating θ is smaller than 10% of the total execution time. Shown in brackets are the CPU-times for estimating θ over the total execution times in seconds.

	Multi-Resolution	Fixed-Resolution
Phantom I	6.48% (1.75/27s)	1.64% (0.97/59s)
Phantom II	8.70% (30.37/349s)	2.77% (11.21/404s)

to 404 seconds for the fixed resolution method. Fig. 4.6 compares the error convergence for the fixed and multiscale reconstructions of the second phantom. The multiscale error at coarse scales is comparably high since the coarse scale reconstructions cannot account for the phantom's fine structure. However, the multiscale algorithm converges rapidly at each scale, resulting in lower total execution time. Again, the multiresolution algorithm achieves lower final reconstruction error.

Finally, we examine the computational complexity of the estimation of emission rates in relation to the overall complexity of the algorithms. Table 4.5 shows the percentage of CPU-time used for the estimation of θ. For all reconstructions, six full updates of θ were performed between full reconstruction updates. In all cases, the emission rate updates make up less than 10% of the total CPU-time. The percentages are smaller for the fixed resolution algorithm than for the multiresolution algorithm. Since the fixed resolution algorithm performs more computationally expensive iterations at fine scale, the relative cost for reconstruction updates is higher than for the multiscale method. Added over all resolutions, however, the multiscale algorithm performs more iterations than the fixed scale method. Since an iteration at any scale includes a fixed-cost update of θ, the multiscale method spends more absolute time on estimating the emission rates. This may be reduced by introducing a convergence criteria for the emission rate updates instead of running a fixed number of iterations between reconstruction updates.

In conclusion, the results indicate that the multiresolution algorithm can achieve reconstruction results superior to the fixed scale method. It is particularly robust with respect to the initialization of emission rates. Furthermore, the multiscale method is typically faster than the fixed resolution algorithm. For both algorithms, the computational cost for estimating the emission rates is small in comparison to the reconstruction complexity.

4.10 Conclusion

In this work, we have described a fixed- and multiscale method for discrete-valued Bayesian reconstruction. The multiscale MRF reconstruction algorithm is a straightforward extension of the fixed scale model. Interaction between resolution levels is obtained by initialization of each reconstruction with the previous coarser reconstruction. The algorithm includes an efficient method for estimating the discrete emission rates. The quality of the multiresolution reconstructions is significantly better than thresholded CBP reconstructions. In comparison to a fixed scale MAP reconstruction, the multiresolution method is less prone to local minima and converges faster.

5. PERCEPTUAL IMAGE SIMILARITY EXPERIMENTS

Abstract

In this chapter, we study how human observers judge image similarity. To do so, we have conducted two psychophysical scaling experiments and have compared the results to two algorithmic image similarity metrics. For these experiments, we selected a set of 97 digitized photographic images which represent a range of semantic categories, viewing distances, and colors. We then used the two perceptual and the two algorithmic methods to measure the similarity of each image to every other image in the data set, producing four similarity matrices. These matrices were analyzed using multidimensional scaling techniques to gain insight into the dimensions human observers use for judging image similarity, and how these dimensions differ from the results of algorithmic methods. This chapter also describes and validates a new technique for collecting similarity judgments which can provide meaningful results with a factor of four fewer judgments, as compared with the paired comparisons method.

5.1 Introduction

Advances in digital cameras, large accessible data storage, internet repositories, and image applications have fueled the recent development of methods for searching, retrieving, and navigating through a set of images [114, 115, 116, 117, 118, 119]. In a typical search task, the user selects an image and asks the computer to retrieve images which are similar. The computer compares the features of the selected image with the characteristics of the other images in the set and returns the most similar images for inspection. Typically, this is done by computing, for each image, a vector containing the values of a number of attributes and computing the distance between image feature vectors. The best image matches are typically displayed to the user as an array of "postage stamp" sized images, in descending order of this computed distance. These methods produce several well-documented artifacts. The methods, for one, do not include any knowledge about the intrinsic organization of the images. For example, if a query image is equidistant between two clusters, the Euclidean metric will return the images in metric order, interdigitating images from the two clusters. The success of these metrics, furthermore, depends largely on the power of the identified features. Often, these features are thought to contribute to human judgments of image similarity (e.g. color, texture, and shape), but with few exceptions[120, 119], the characteristics of human similarity perception have not been included in the selection of these features. The third artifact of these methods is that the images are displayed in metric order, but not in an order which is conducive for navigation. Similar images are in order by row, but even if the ordering is perfect, similar images are not near each other in the array. Recently, the idea of representing images in a multidimensional, browsable space has been proposed[121, 122]. These navigation spaces, however, are not perceptual, and therefore do not give the user intuitive cues for finding images which are "redder", "darker", "out-of-doors", "more natural", or "containing more people".

The purpose of this perceptual image similarity project is to examine the issue of image similarity from the perspective of the human observer. In this work, we present two psychophysical scaling experiments in which we measure the similarity of 97 carefully selected digital photographic images. In these experiments, each image is compared with every other image, and the observer judges their perceived similarity. The users are free to use whatever criteria they choose for making this decision. We compare these perceptual judgments to similarity ratings produced by two algorithmic methods, one based on similarities in the color histograms of the images, the other based on more sophisticated perceptually-relevant features [119]. We evaluate these perceptual and algorithmic similarity ratings using multidimensional scaling techniques and explore the results of two- and three-dimensional MDS solutions.

The next step in this research is to model perceptual image similarity judgments in terms of calculable image features. This would allow us to extend our findings to the development of better image search and retrieval methods, and to the development of more intuitive navigation spaces.

5.2 Experimental Design

In these experiments, human observers judged the similarity of 97 carefully-selected JPEG images. Two psychophysical scaling methods were used to measure the perceived similarity of each image with every other image in the set. In the "Table Scaling" experiment, observers organized printed thumbnail images on a tabletop so that similar images would be near each other and dissimilar images would be far apart. This design allowed the observer to see all the images at once. In the "Computer Scaling" Experiment, the images were presented on a computer display. On each trial, a reference image was compared with eight randomly-chosen images, and the observer selected the image which was most similar. Both experimental procedures produced a similarity matrix which served as the input to multidimensional scaling algorithms. We also calculated image similarity matrices for these 97 images using the algorithmic similarity metrics. The multidimensional scaling results for these algorithmic matrices were compared with the psychophysical results.

5.2.1 Selecting the Stimuli

Ninety-seven JPEG images were selected from the PhotoDisc collection of 5000 photographic images. This set was chosen because they were reputed to be of good photographic quality, and had been used by other researchers in this field[116, 119]. These images were selected according to three explicit criteria. First, we wanted to make sure that we included a wide range of topics. To achieve this goal, we consulted books designed to teach photography and matched our selections to these focus areas. These were: animals, people, indoor scenes, nature, buildings, textures, and man-made. The photographic textbooks also focused on the distance of the object from the viewer, and so, for each category, we explicitly selected images which were wide-angle, normal (e.g., 50 mm for 35 mm film), or close-up. In order to insure that we had selected a set of images which covered a broad range of colors and light levels, we measured each image in calibrated CIELab space, and iterated on the image samples until we achieved a balanced distribution in all three dimensions of this color space. The images were in both landscape and portrait modes.

5.2.2 Observers

Fifteen volunteer observers were recruited from the T.J. Watson Research Center, Hawthorne Laboratory. They were 12 men and 3 women, ranging in age from approximately 23 to 45. Observers received a lunch voucher for each session.

5.2.3 Experiment 1: Table Scaling

For this experiment, we printed the 97 JPEG images on a color printer at 300 dots per inch. The size of the prints was 3 cm by 2 cm, on a 3.5 cm by 3.5 cm white background.

The images were placed randomly on a large round table, and the observers' task was to arrange them so that the physical distances between them were inversely proportional to their perceived similarity. That is, the more similar the images appeared, the closer they were to be placed next to each other. Since the table surface was, by definition, a 2-D surface, this experiment forced the observers to project the multidimensional relationships between the images down to a two-dimensional space.

Nine subjects served in this experiment. Each session took 30-45 minutes, and the observers found this to be a "fun, puzzle-like" task. For each observer's solution, we measured the physical distance between each pair of images and created a similarity matrix. We created a pooled matrix by accumulating these distances across subjects.

5.2.4 Second Experiment: Computer Scaling

In the second experiment, fifteen observers compared each of the 97 thumbnail images with every other image. In the traditional multidimensional scaling paradigm, these measurements would be made using a "paired comparisons" design. In paired comparison, all stimulus pairs are compared, and for each pair, the observer assigns a number proportional (or inversely proportional) to perceived similarity. We modified this procedure for two reasons. First, with $n = 97$ stimuli, the paired comparisons procedure would require $n(n-1)/2$ comparisons, which, in this case, would be 4656 trials. Second, we were seeking a procedure which would not depend on subjective magnitude judgments, since these can be prone to bias. We therefore developed a variation on the traditional paradigm. In this variation, each image was compared with each other, taken eight at a time, thereby reducing the number of trials by a factor of four. Thus, on each trial, the observer viewed a randomly-selected reference stimulus and eight test stimuli, selected randomly from the set of 97, and judged which of the eight appeared most similar.

The experiment was conducted on a color-calibrated display monitor, using the Netscape browser to present the stimulus images. The display measured 32 cm by 24 cm, and subtended roughly 40 by 30 degrees of visual angle when viewed at a distance of approximately 47 cm. The image bitmaps used in this experiment measured 123 by 83 pixels. Viewed on the display monitor, the size of each image was approximately 4.7 cm by 3.1 cm, and subtended approximately 6 by 4 degrees of visual angle. Figure 5.1 shows one trial of this experiment, as displayed to the observer on the computer display. The reference stimulus was presented along the left edge of the display, accompanied by two rows of four test stimuli running horizontally along the display. On each trial, the observer evaluated the eight test stimuli, judged their similarity to the reference stimulus, then used a mouse to "click" on the most similar test image. This response launched the next trial of the experiment. The experiment was self-paced, and observers could take breaks whenever they chose. The 1164 trials and 20 practice trials typically took three 1-hour sessions to complete.

In addition to these objective judgments, we also collected verbal protocols from the observers, asking them to free-associate about the reasons they had for making their selections. The purpose of this technique was to develop a better intuition for what observers thought they were doing, and to gain insight into candidate

Click on the image on the right which is most similar to the one on the left.

Fig. 5.1. One trial of the computer scaling experiment. The reference stimulus on the left is compared to the eight test stimuli on the right. The task was to select the one test stimulus that is most similar to the reference.

dimensions of image similarity.

Validation of the Experimental Design

In our Computer Scaling paradigm, each image appears as a reference stimulus twelve times, and on each of these trials, is paired with a randomly-selected set of eight test stimuli. Each observer selects just one "most-similar" image on each of these trials, and so, over the course of the experiment, data are only entered into 12 cells in each 97-cell row of the similarity matrix. To create the overall similarity matrix, results of the individual observers were accumulated. In particular, if a subject selected test image k as a match to reference image r, then the rk-th entry of the similarity matrix was incremented by one. With fifteen observers, there are only 15 x 12 votes in each row of the similarity matrix, or 180 distributed over the 97 images. This produces a rather sparse similarity matrix.

In order to verify that these sparse measurements can lead to a meaningful dataset, we performed a Monte Carlo simulation. The simulation assumes a "true" similarity matrix S whose entries s_{ij} are between zero (least similar) and one (most similar). Preliminary experiments showed that most images are perceived as being very dissimilar to each other. Therefore, similarity matrices tend to be sparse and contain mostly small entries. For the simulation, the similarities were assumed to be approximately exponentially distributed[123] with mean $\lambda^{-1} = 0.5$. Furthermore, the matrix was assumed to be 30% sparse which we consider a conservative estimate. This leads to the probability distribution

$$p_{s_{ij}}(s) = 0.3\delta(s) + 0.7\frac{\lambda e^{-\lambda s}}{1 - e^{-\lambda}}(\mathrm{u}(s) - \mathrm{u}(s-1)) \qquad \text{for} \quad i \neq j \tag{5.1}$$

where δ is the Dirac-delta and u denotes the unit-step function. The similarity matrices used for the simulation were obtained by sampling from this distribution and symmetrizing the result. Notice that the symmetry assumption is a consequence of our intent to embed the estimated similarity matrix into a metric space using multidimensional scaling.

In order to model the subject's choice behavior, we assume a choice process based on Gaussian confusion probabilities. For illustration, assume that the subject is presented with one reference image r and two test images t_1 and t_2. Let s_{rt_1} and s_{rt_2} denote the similarities of the reference image to test images one and two respectively. Then, assuming without loss of generality that $s_{rt_1} < s_{rt_2}$, the probability of selecting the test images is assumed to be

$$\begin{aligned} P(t_1) &= \frac{1}{2}e^{-\frac{(s_{rt_1}-s_{rt_2})^2}{2\sigma^2}} && (5.2)\\ P(t_2) &= 1 - \frac{1}{2}e^{-\frac{(s_{rt_1}-s_{rt_2})^2}{2\sigma^2}}. && (5.3) \end{aligned}$$

For our subject experiment where eight test images are displayed at each trial, this generalizes as follows: Assume we wish to calculate the probability for choosing test image t_j, $1 \leq j \leq 8$. We first divide the 7 remaining test images t_i into two sets $T_j^{(1)}$ and $T_j^{(2)}$. The first set $T_j^{(1)}$ contains those test images that are more similar to the reference image than t_j, i.e. $T_j^{(1)} = \{t_i : s_{rt_i} > s_{rt_j}\}$. The second set $T_j^{(2)}$ contains the test images that are less or equally similar to the reference image than t_j, i.e. $T_j^{(2)} = \{t_i : s_{rt_i} \leq s_{rt_j}, i \neq j\}$. Generalizing the two-test case from above, we obtain

$$P(t_j) = \frac{\prod_{t_i \in T_j^{(1)}} \frac{1}{2} e^{\frac{-(s_{rt_i} - s_{rt_j})^2}{2\sigma^2}} \prod_{t_i \in T_j^{(2)}} (1 - \frac{1}{2} e^{\frac{-(s_{rt_i} - s_{rt_j})^2}{2\sigma^2}})}{\sum_{k=1}^{8} \left(\prod_{t_i \in T_k^{(1)}} \frac{1}{2} e^{\frac{-(s_{rt_i} - s_{rt_k})^2}{2\sigma^2}} \prod_{t_i \in T_k^{(2)}} (1 - \frac{1}{2} e^{\frac{-(s_{rt_i} - s_{rt_k})^2}{2\sigma^2}}) \right)} \tag{5.4}$$

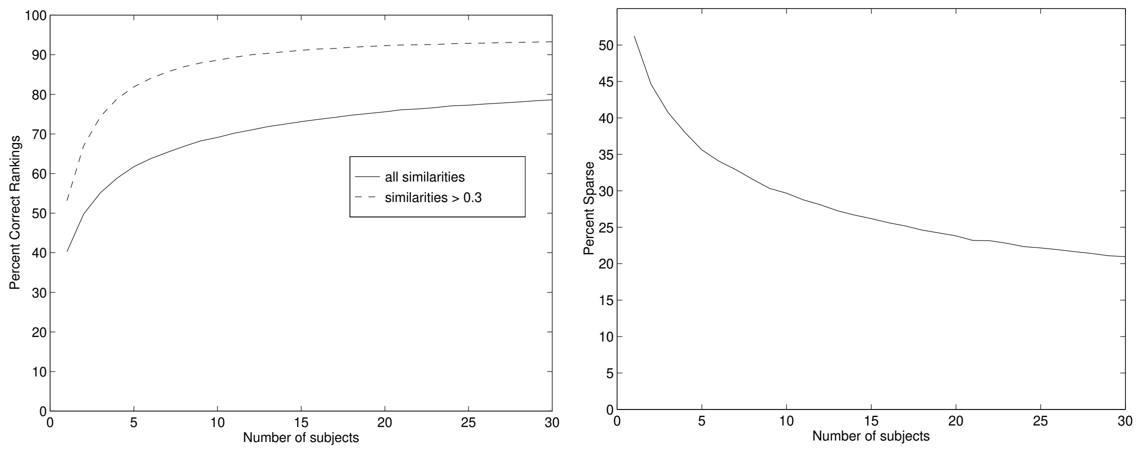
where the denominator normalizes the sum of the probabilities $P(t_j)$ to one. The standard deviation used was $\sigma = 0.1$.

Using the assumptions above, we performed the Monte Carlo simulation by generating experimental trials and selecting image matches according to (5.4). The estimate $\hat{S}$ was then formed as follows: Starting with a matrix of zeros, the entry $\hat{s}_{ij}$ was incremented by one if the test image j was selected as a match to reference i. The matrices were added across the hypothetical subjects, symmetrized and divided by the number of subjects.

Notice, that the estimates $\hat{s}_{ij}$ are biased, i.e. they will not converge to the true s_{ij} as the number of subjects increases. However, the magnitude *rankings* of the $\hat{s}_{ij}$ converge to those of s_{ij}. Since we evaluate the data with respect to its value for the multidimensional scaling procedure we are mostly interested in correct rankings as compared to the correct numerical values.

In order to evaluate the simulation results, we calculated the sparsity of the estimate as well as the percentage of correct ranking relationships of matrix entries. The percentage of correct ranking relationships was computed by examining all pairs {(i,j),(k,l)} of matrix entries and comparing whether their ranking relationship was the same in S and $\hat{S}$.

The results are shown in Fig. 5.2. Figure 5.2(a) shows the percentage of correct similarity rankings as a function of the number of subjects. The solid curve shows correct rankings evaluated over all matrix entries. For 15 subjects, the percentage of correct rankings is 73 %. The dashed curve evaluates correct rankings only for entries s_{ij} greater than 0.3. In particular, these entries were paired with all other entries and ranking relationships were compared as in the general case. The percentage correct in this case is higher than in the general case. For 15 subjects, we obtain



(a) (b)

Fig. 5.2. Results of the Monte Carlo simulation. Shown in (a) is the percentage of correct similarity rankings as a function of the number of subjects. The solid line evaluates correct rankings for all matrix entries, whereas the dashed line shows percentage correct for entries greater 0.3 only. Shown in (b) is the sparsity of the estimate.

91.1% correct rankings. The interpretation is that we measure the larger similarities more exactly than the small similarities. This is consistent with our design using sparse measurements and selecting only the highest similarity match at each trial. Figure 5.2(b) shows the sparsity of the estimate. Notice that sparsity here is defined as the percentage of zero estimates for non-zero entries in the true S-matrix. Ideally, the sparsity should be zero. However, considering, that most of the true matrix entries are assumed to be small, we consider a sparsity of 26.2% for 15 subjects satisfactory. Notice that these results are fairly robust with respect to small changes in the simulation assumptions and parameters. Concluding, the results suggest that using our experimental design, data from 15 subjects should be sufficient to obtain a meaningful estimate of the similarity matrix for multidimensional scaling evaluation.

5.2.5 Computing Similarity Matrices for Image Processing Algorithms

In the above experiments, we use perceptual judgments to measure the distance between the 97 images. For comparison, we also used algorithmic image similarity metrics to characterize these distances. The first metric is based only on global color-histograms of the images. In particular, this metric computes image dissimilarity as the L_1-norm of color-histograms in CIELab color-space. The number of bins used for the histograms was 8/16/16 for the L, a^* and b^* channels respectively. While this is a very simple algorithm, algorithms like these are often an important component of more sophisticated measures[116, 118].

The second metric[119] is a multichannel model derived from models for image quality assessment. This metric is based on a multiresolution framework of color, contrast and orientation-selective attributes. Image dissimilarity is calculated as a weighted combination of attribute differences. The relative weights of the attributes are based on a training set of perceptual judgments.

5.3 Results

5.3.1 Qualitative Findings

Figure 5.3 shows the accumulated results for one image for all fifteen observers in the computer scaling experiment. The number of "votes" is indicated below each of the ten most frequently selected images. Although individual observers used different verbal descriptors of their behavior when making these judgments, their selections were very similar. In this example, an image of three snowdrop flowers is judged to be most similar to other images showing a countable number of colorful flowers, followed by other foliage and flower scenes, outdoor scenes with lots of green foliage and color, and later by an image of foliage with animal figures.

Figure 5.4 shows a similar result for a very different image. Here, the picture of the young girl with a baseball glove is matched most frequently to other images including children, then to a solitary portrait, followed by group images which include either a child or outdoor sports. Qualitatively, thus, there seemed to be a high agreement across subjects in their rating of image similarity.

Although these matches were very consistent across observers, they were not always symmetric. For example, a nature scene might commonly be matched to nature scenes including children. However, when these nature scenes containing children are shown as reference, they will most frequently be matched to images of other children and not to nature scenes without children. This is an important point;

the MDS analysis does not capture such non-symmetric behavior, since the images are embedded in a metric space. In future work we will explore methods to analyze these aspects of the data.

5.3.2 Multidimensional Scaling

Multidimensional scaling is a psychometric procedure originally designed to estimate the perceived distances between stimuli which vary along a large number of dimensions, where the goal is to uncover the dimensions along which these judgments are made[124]. This concept has been used, for example, to discover the perceptual dimensions of perceptual relationships between notes in a musical scale[125] and the dimensions of color sensation[126]. It has been used in the image analysis community to identify the perceptual dimensions of textures[127], and in the image retrieval community to measure the similarity of images based on their color histograms[121, 122] as well as to explore the behavior of a texture perception algorithm[128].

The goal of the multidimensional scaling procedure is to place objects that are specified only by their distances into a lower dimensional space. In particular, the input to the MDS algorithm is a distance matrix of the pairwise object distances. The procedure then places the objects into a metric space while preserving the distances as much as possible. Notice that there is a large variety of MDS algorithms which differ in their assumptions and optimization strategies. For an excellent overview, the reader is referred to [129].

The multidimensional scaling procedure employed here is metric least-squares scaling. We first convert the similarity matrix S from the subject experiment to a dissimilarity matrix D as

$$d_{ij} = 1 - s_{ij}. \tag{5.5}$$

Now assume that the images i and j are placed at position x_i and x_j in the metric space. We then define $\tilde{d}_{ij}$ as the Euclidean distance $\tilde{d}_{ij} = \| x_i - x_j \|_2$. The optimization or stress function θ used for the MDS algorithm is then

$$\theta = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(\tilde{d}_{ij} - d_{ij})^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}^2} \tag{5.6}$$

where N is the number of images in the set. The minimization of θ is performed using a conjugate gradient optimization.

We used this multidimensional scaling algorithm to reduce the dimensionality of the four similarity matrices we described above, derived from, 1) the color histogram metric, 2) the perceptually- based image similarity metric, 3) the Perceptual Table Scaling data, and 4) the Perceptual Computer Scaling experiment. In order to compare the MDS results for the different methods, we allowed for translations and rotations to minimize the differences in a least square sense.

5.3.3 The multidimensional scaling results

We performed a multidimensional scaling analysis in two and three dimensions for each of the similarity matrices derived from the different methods.

MDS in 2 dimensions

Figure 5.5 shows the two-dimensional result obtained when using the color histogram algorithm to compute distances between images. Not surprisingly, the two-dimensional solution reflects the influence of color. We have added some color words to the figure to simplify interpretation. Although the images do seem to segment into color regions, these regions are rather diffuse. For example, images with a lot of green span from the landscape view of foliage at the bottom left through the landscape view of agricultural terraces in the middle of the image, and beyond. Looking at this projection, we clearly see that equal distances in the space do not correspond to equal differences in perceived color. For example, the foliage image in the bottom left and the terraces image are very similar in overall color but are quite far apart in this space. Conversely, the terraces image is right next to a sailboat image whose overall color is quite different. This may be due to the high stress (0.27) of the MDS solution. It may also be due to the fact that overall perceived color is

Reference Test images in order of decreasing number of votes

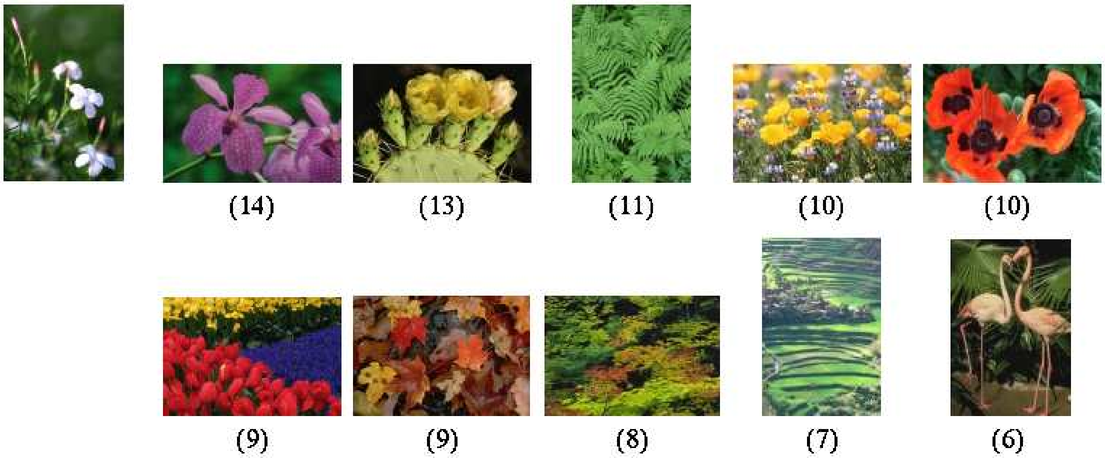


Fig. 5.3. Computer Scaling matching results for snowdrop-flowers image. The two rows on the right show the test images that were most frequently matched to the reference on the left. Indicated below each image is the number of votes this match received across the 15 subjects. In this case, the reference was most frequently matched to other images showing a small number of flowers, followed by other foliage and flower scenes.

Reference Test images in order of decreasing number of votes

Fig. 5.4. Computer Scaling matching results for "girl-baseball-glove" image. This image is most frequently matched to images of other children, followed by a portrait and group images containing either children or sports scenes.

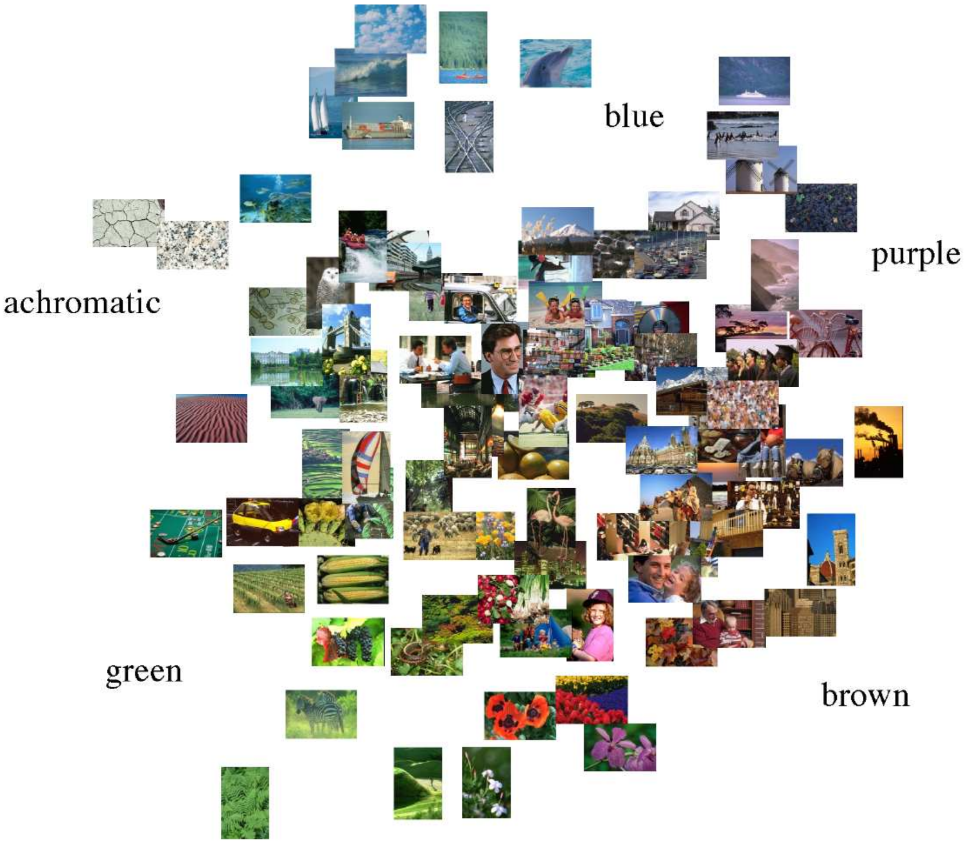


Fig. 5.5. Multidimensional Scaling result using the color histogram distance metric

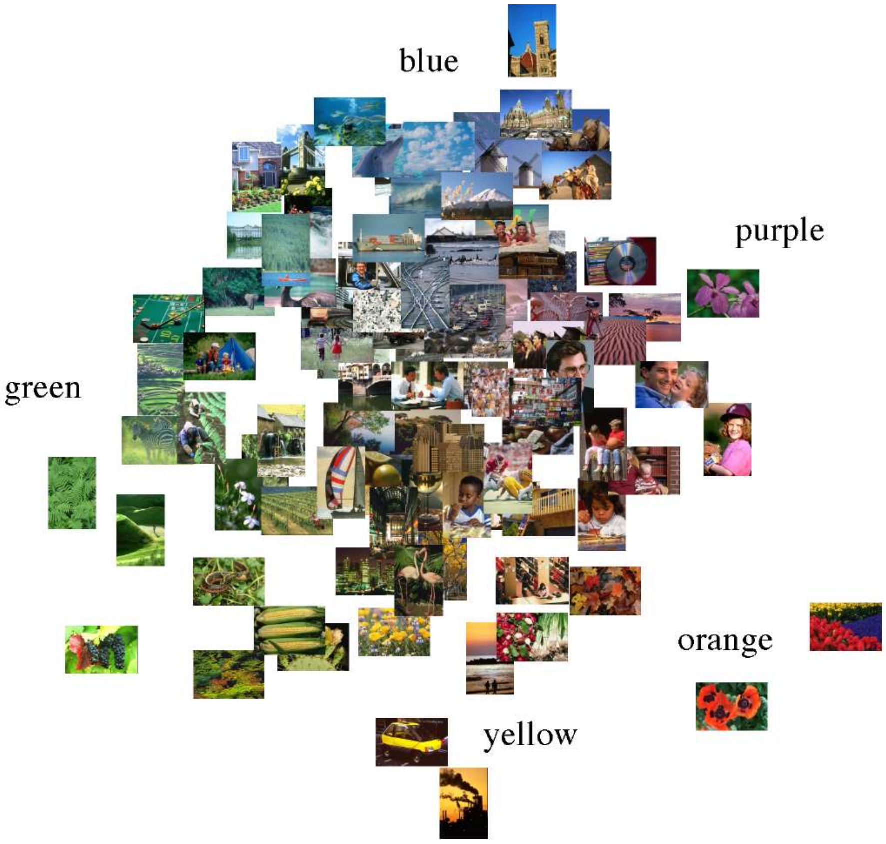


Fig. 5.6. Multidimensional Scaling result for the perceptually optimized distance metric.

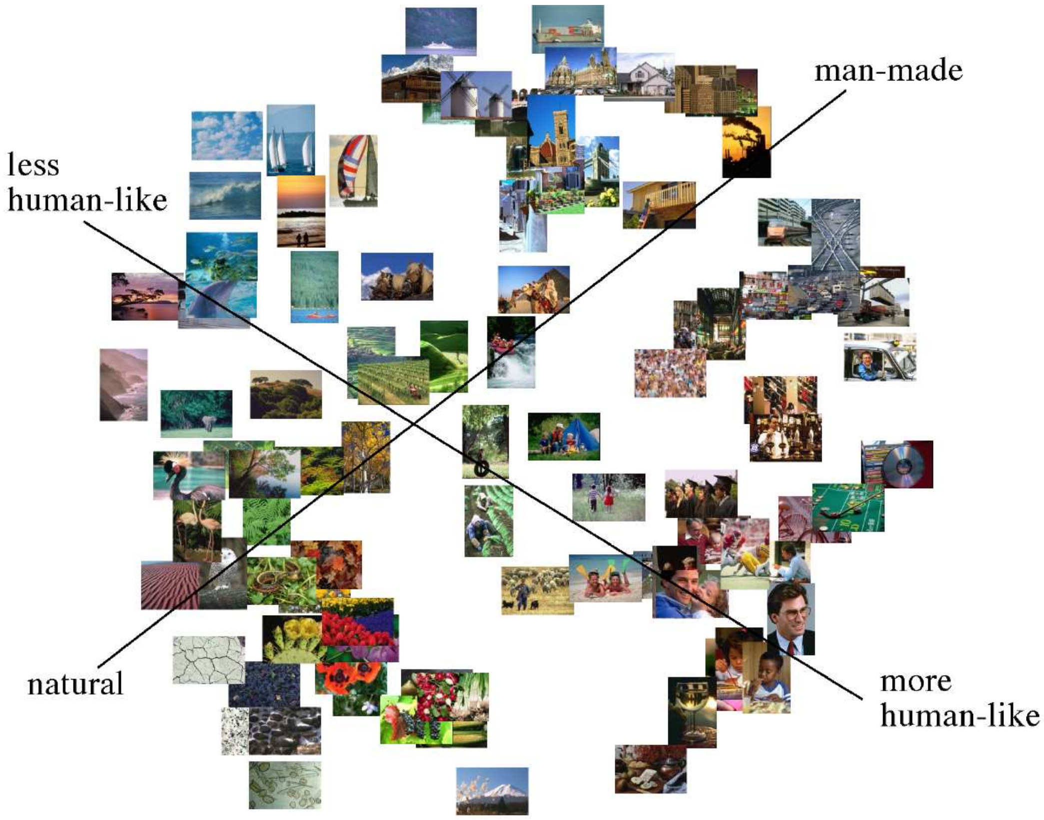


Fig. 5.7. Multidimensional Scaling result using the image distances from the table scaling experiment

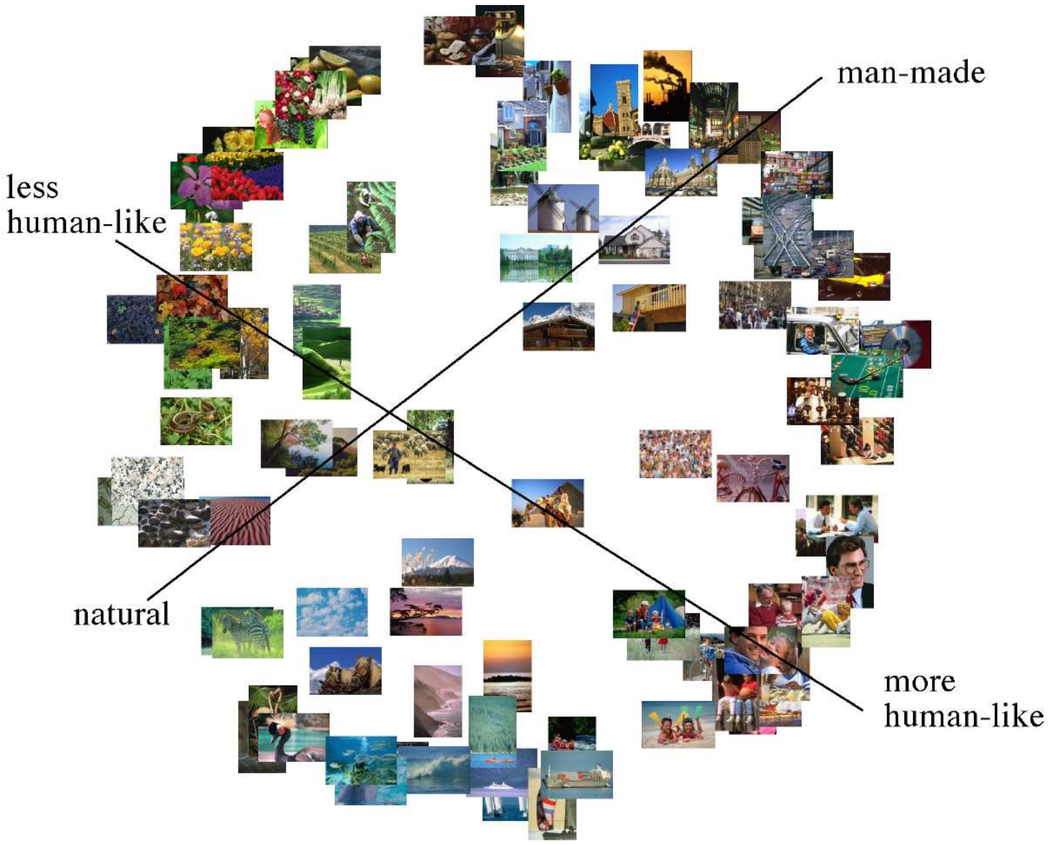


Fig. 5.8. Multidimensional Scaling result using the image distances from the computer scaling experiment

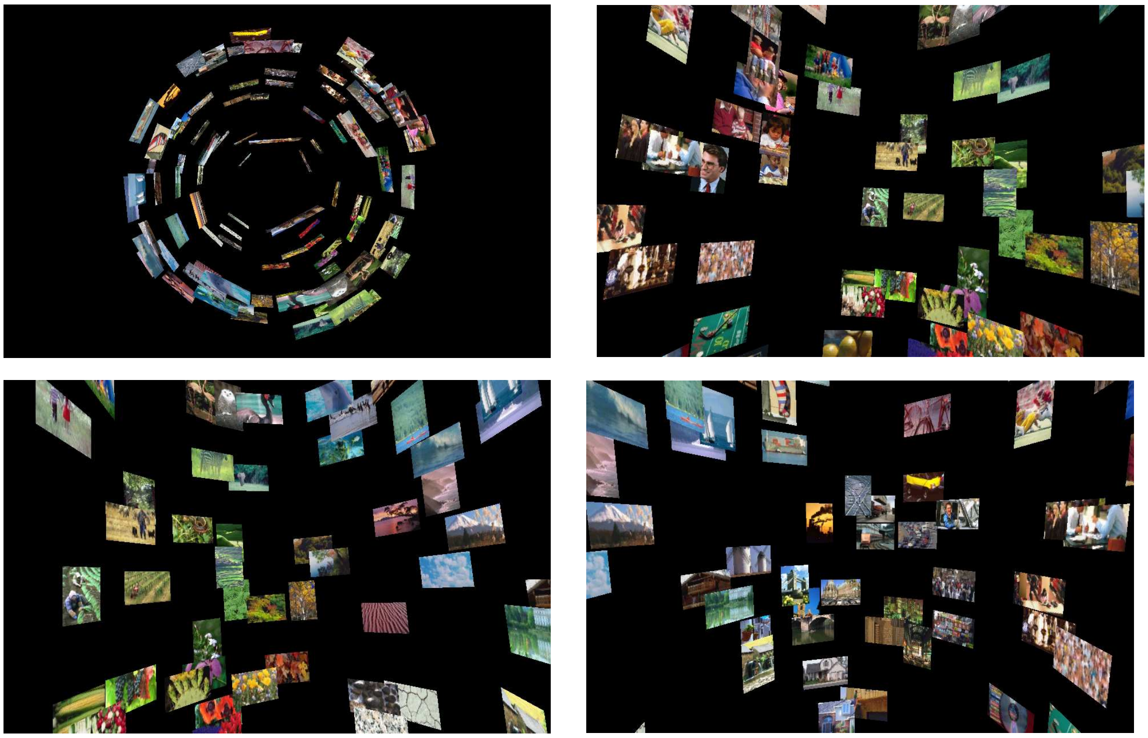


Fig. 5.9. Multidimensional scaling results for 3 dimensions as seen in the VRML browser. The top left shows a top view of the 3-D space. The three other panels (top right; bottom left, bottom right) show three sequential views as the viewer looks out from the center of the space.

not well-described by this simple algorithm. We believe, for example, that the color of the main object in the image contributes more to the overall impression of the color than would be expected given its pixel contribution. That is, we would expect close-up images of people to be close together perceptually because the color of the skin is more important to the overall impression of image color than the color of the background or the color of the clothing. To test results of this type, we are planning to compare them with semantic color descriptors for these 97 images, such as the "overall color", the "color of the main object", "the background color", etc.

One very interesting observation from these data is that overall color alone does capture information about the semantic meaning of these images. For example, a large number of the outdoor, natural scenes appear near to each other in this projection, suggesting that color alone carries semantic information. Since it is very difficult to develop algorithms which capture the semantic information in images, understanding the correlations between color and semantics could be a useful tool in developing semantic image descriptors which rely on calculated image attributes.

Figure 5.6 shows the results of the 2D MDS for the perceptually-based image processing metric. This algorithm contains terms which capture color attributes, and the influence of color is apparent in this projection. With this algorithm, the achromatic and brown images are at the center of the projection, and the spectral colors rotate around this hub. This algorithm is designed to capture additional features in the image. Low spatial-frequency color images are grouped together which can be seen especially within color regions in this projection. For example, the images with large blue expanses with brown subjects in the foreground are together in the top right region, including buildings, windmills, and horses. The portraits of humans, with large expanses of skin color against a darker background, are also organized together. Images with high-spatial-frequency luminance modulations are also near each other. Looking toward the center of the projection, for example, there is a cluster of traffic, crowd, city street and texture images. The perceptually weighted attributes, low-spatial-frequency color and high-spatial-frequency luminance, do seem to produce a more perceptually-plausible organization of these images than did the color histogram alone, capturing some compositional attributes in the images.

Figure 5.7 shows the two-dimensional multidimensional scaling result of the perceptual table scaling experiment. The popularity of color-histogram-based image similarity metrics is validated by these perceptual results. Overall color does seem to play a significant role in the perceptual organization of these images. The color organization, however, is much less pronounced, suggesting that other factors are playing significant roles in these judgments. There appear to be groupings of blue, green, and brown/orange images, and perhaps achromatic images. The attributes described in the result for the perceptually-weighted algorithm also seem to be operating. Within color areas, there appear to be clusters of low-spatial-frequency color images and clusters of high-spatial-frequency luminance images.

Another organization, however, appears to emerge from this projection. In multidimensional scaling, the two-dimensional algorithm maps the stimuli onto a two-dimensional plane, and analysts typically try to interpret these dimensions in conceptual terms. Following this tradition, we have superimposed two candidate axes onto this image. One axis is what we call the natural vs. man-made axis, running from the rock and flower images in the bottom left through the nature scenes, to nature scenes with man-made objects, to man-made objects with nature, to the buildings and shipping docks in the upper right-hand side of the projection. The other axis is the human vs. non-human a axis, running from the less human-like images of sunsets and clouds in the top left through scenes with animals, to scenes with small images of humans in various settings to images featuring large full-face portraits of humans in the bottom right side of the projection. These candidate dimensions also seem to describe the images as we move around the obviously circular projection. Starting with natural images of seascapes, in the upper left, we move to increasingly man-made, less natural images, which do not contain humans, then move through the man-made objects, which likewise do not include humans, through to images which increasingly include humans, on to the end of the "more human axis". Continuing our path, we have humans in nature, then natural objects without humans.

Figure 5.8 shows the 2D result for the Perceptual Computer Scaling experiment. Again, this result supports the idea that overall color is important in image similarity judgments, and also provides support for the low-spatial-frequency color and high-spatial-frequency luminance hypothesis. There appear to be clusters within this organization which clearly reflect these features.

Perhaps more striking is the similarity in organization of these results to those of the table-sorting experiment. In this image we have superimposed onto the 2D projection the same qualitative axes as in the previous figure. Although there are differences in the fine structure of how the images are organized, the overall structure produced by the two psychophysical experiments is extremely similar. In the computer scaling experiment, there is a progression across the circular array from natural to man-made, and, in what appears to be an orthogonal dimension, an equally compelling progression from less human-like to more human-like. Around the circle, we go from nature scenes which do not include humans or man-made objects, to nature scenes which include some man-made objects, such as buildings, through to images of man-made objects without nature or humans, to images of man-made objects with humans, but not nature. From the end of the human axis, we progress to images with humans and natural scenes, and back to images with purely natural scenes without humans or man-made objects. Interestingly, in both psychophysical projections, animals are organized more closely with natural scenes than they are with humans, perhaps because in these images, all the animal images were taken in nature.

MDS in 3 dimensions

One of the most interesting features of the two dimensional projections of the psychophysical results is the circular structure. Using the Diamond visualization tool[130], we were able to look more closely at this three-dimensional MDS structure. We found that the 3-D solution is best described as a sphere of images, where the sphere has a thickness of not more than 10 percent of the distance across the diameter. This means that the images in the central region of the 2-D projections are really images which would belong the shell of the 3D solution. To examine this further, we created a three-dimensional geometry and pasted the center point of each image onto its coordinate position in this 3D space. In order to view these images, we positioned them so that each was normal to a vertical axis in this space. For example, all the images on the top of the sphere were effectively rotated 90 degrees.

We created a VRML "world" which allowed us to view the images in this 3-D geometry from a number of different perspectives. Perhaps the most interesting view of these data was to zoom into the center of this rotating set of images, and watch the images move slowly past. Figure 5.9 shows a matrix of four views generated from the VRML world. The top left panel contains a view from the top, showing concentric rings of images forming a shell around the center of the space. The remaining panels contain three sequential views around the interior of the 3D shape, looking outward. The shapes of the images appear distorted because we have introduced a fisheye transformation in order to see more images. The view on the top right shows a progression from images with human figures, with children organized together, moving through to outdoor scenes with humans as small, subordinate features, to green outdoor nature scenes with animals, to outdoor, nature scenes without animals or people. A grouping of nature scenes with flowers and vegetables is at the bottom right-hand corner of the panel, and includes the snow-drop image discussed in Figure 5.3. It is interesting to notice that the ordering of similarity across observers in Figure 5.3 is well-matched to the distances in this projection.

The panel at the lower left shows another snapshot of this 3D space, the next view in the sequence. This view overlaps the view just discussed, beginning with images of humans in natural scenes. The top of the view shows nature scenes with animals on green backgrounds followed by animal images on blue ocean backgrounds (fish and birds), water scenes without people, then sea scenes with man-made objects. The right-hand side of the panel is filled with man-made objects: ships, bridges, trains and train tracks. The panel at the bottom right picks up where the previous one left off and shows the progression from ocean scenes with boats, to outdoor scenes with buildings, to man-made scenes with humans, back to where we began.

The overall impression generated by this 3D view is that the results generated by human observers are very systematic, with sensible progressions, following semantic, color, and structural characteristics.

MDS Stress

As expected, very low dimensional spaces cannot represent the full complexity of perceptual similarity judgments. This is reflected in the stress values shown in Table 5.1. Comparing the two psychophysical techniques, we notice that the stress in the table scaling method is lower than for the computer scaling method. This suggests that, when constrained to a 2-D projection, subjects are more likely to agree on the

Table 5.1
Stress values for MDS results as a function of dimensionality.

Dimensions MDS space	Computer scaling	Table Scaling	Color-histogram algorithm	Perceptually optimized algorithm
2	0.33	0.25	0.27	0.26
3	0.23	0.16	0.17	0.19
4	0.18	0.10	0.13	0.14

two most fundamental dimensions, thereby reducing the dimensionality of the data across subjects.

The similarity between the MDS solutions for the Table Scaling and Computer Scaling experiments, and their intuitive appeal, suggests that, despite this high stress, they do capture some of the most important dimensions of image similarity.

5.4 Discussion

Algorithmic image similarity metrics commonly make use of attributes which are thought to influence human judgments of image similarity. In this work, we have explicitly studied human visual similarity judgments in an effort to develop similarity metrics better matched to human judgments. Using psychophysical scaling techniques, we have compared the results of several methods for measuring the similarity of images. The first is a simple color histogram algorithm, the second is a metric with parameters weighted by psychophysical judgments, and the remaining two are the results of psychophysical scaling experiments which explicitly measure human judgments of image similarity.

5.4.1 Evaluating the psychophysical results

Although the two scaling results were designed to test the visual similarity of our 97 test images, the experimental conditions were quite dissimilar. In Table Sorting, all images were in view simultaneously whereas in Computer Scaling, the observer selected the one image among a set of eight which looked most similar. In

Table Scaling, the images were printed; in Computer Scaling, they were presented on a computer display. In Table Scaling, the experiment constrained the solution to the 2D tabletop, whereas the Computer Scaling experiment did not. In Table Scaling, the similarity matrix was fully populated, since a physical distance was computed between each pair of images once the solution was obtained. In Computer Scaling, the matrix was sparse. Given these differences, it is quite remarkable that the two techniques produced such similar 2D solutions when their respective similarity matrices were scaled using multidimensional scaling. This similarity suggests that the two techniques tapped the same perceptual processes, independent of differences in experimental methodology.

5.4.2 Using these Results

The systematic nature of these data make them an excellent basis for future research in image similarity. It is clear that human observers use many dimensions in their evaluations of image similarity, including color, high-frequency luminance information, low-frequency color information, and perhaps most important, semantic information. That the two- and three-dimensional solutions to the scaling judgments produce smooth, orderly transitions from image to image suggests that these cues may be smoothly combined. This suggests that it may be possible to develop descriptors which describe these smooth transitions as a function of measurable features of the images.

In the analysis of the 2D MDS projections for the psychophysical scaling experiments, we suggested the possibility of dimensions running from more- to less-human-like, and from natural to man-made. Returning to these images, it is clear that other organizations are also possible. For example, it is easy to see that the images are grouped according to semantic category: buildings, animals, boats, children, and man-made objects are organized near each other. This suggests that it might also be useful to use cluster analysis techniques.

5.4.3 Modeling

The psychological image similarity spaces we have uncovered using the multidimensional scaling techniques appear to be visually intuitive, and if this were the entire set of existing images in the universe, we might be inclined to offer our 3D VRML world as a navigation tool. Certainly images which appear similar to each other appear near each other, and it is easy to understand how to navigate within this space. However, this is not the universe of images, and in order to make these results useful, it must be possible to compute where new images should go. This means that we need to model image positions in terms of images features which can be computed automatically. To do so, our first task will be to try to model the positions of the images in terms of measurable image characteristics, then test this model with a new set of images.

Another use for these data is to test the performance of various image similarity metrics. We have seen in this study that, at least qualitatively, the color histogram and the perceptually optimized metric capture some aspects of the human judgment data, but do not adequately model the human similarity space. At minimum, an error measure could be devised comparing the results of the proposed metric with the perceptual results. As more is learned about the visual mechanisms underlying these results, however, it is hoped that a more theoretical, quantitative model could be developed which could describe the deficiencies of image similarity metrics in terms of visual processes.

5.4.4 Image Semantics

It is quite clear that these results suggest semantic categories. It is easy to see semantic clusters in the data, of people, outdoor scenes, seashore scenes, etc. These semantic categories, however, appear to correlate with image descriptors. For example, images with indoor scenes tend to be brownish, have low light levels, and many straight edges. One idea for getting a handle on these semantic categories is to explore how much of the information contained in these semantic categories can be accounted for in terms of calculable image processing descriptors and image statistics.

5.4.5 Why are the MDS results so circular/spherical?

The circular structure of the computer-scaling result could reflect the sparsity of the similarity matrix. In particular, the distance matrix contains a considerable number of entries equal to one. The optimal geometric solution to preserve these distances in two dimensions is a circular structure. This explanation, however, does not explain the circular structure of the table-scaling result, since that matrix was fully populated. The circular structure might disappear if a non-metric MDS algorithm is used instead of the metric algorithm discussed above.

5.5 Conclusions

We have conducted two psychophysical scaling experiments aimed at uncovering the dimensions human observers use in rating the similarity of photographic images, and have compared the results with two algorithmic image similarity methods. Although these experiments were conducted with different media, different tasks, and different methodologies, they produced very similar multidimensional scaling results. First, the overall color of the images was an important factor in judging similarity, and the dimensions "Human vs. non-human" and "natural vs. man-made" were very salient. These low-dimensional solutions did not capture all the richness in these multidimensional judgments, as reflected in the overall stress of the multidimensional analysis, but they did provide systematic structures with relatively smooth transitions and intuitive organizations. These features encourage us to use these results as a basis for developing perceptually-based image similarity metrics and intuitive navigation environments.

LIST OF REFERENCES

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, November 1984.

[2] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 259–302, 1986.

[3] R. Chellappa and A. Jain, eds., *Markov Random Fields: Theory and Applications*. Boston: Academic Press, Inc., 1993.

[4] E. Levitan and G. Herman, "A maximum a *posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-6, pp. 185–192, September 1987.

[5] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, Massachusetts: MIT Press, 1987.

[6] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, pp. 194–202, June 1989.

[7] P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. on Medical Imaging*, vol. 9, pp. 84–93, March 1990.

[8] R. Stevenson and E. Delp, "Fitting curves with discontinuities," *Proc. of the first international workshop on robust computer vision*, pp. 127–136, October 1-3 1990.

[9] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 367–383, March 1992.

[10] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, pp. 296–310, July 1993.

[11] S. Brette, J. Idier, and A. Mohammad-Djafari, "Scale invariant Markov models for Bayesian inversion of linear inverse problems," in *Maximum Entropy and Bayesian Methods* (J. Skilling and S. Sibisi, eds.), pp. 199–212, Netherlands: Kluwer Academic Publishers, 1996.

[12] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. on Information Theory*, vol. 38, pp. 766–784, March 1992.

[13] M. Basseville, A. Benveniste, and A. Willsky, "Multiscale autoregressive processes, part I: Schur-Levinson parametrizations," *IEEE Trans. on Signal Processing*, vol. 40, pp. 1915–1934, August 1992.

[14] M. Basseville, A. Benveniste, and A. Willsky, "Multiscale autoregressive processes, part II: Lattice structures for whitening and modeling," *IEEE Trans. on Signal Processing*, vol. 40, pp. 1935–1954, August 1992.

[15] K. C. Chou, S. Golden, and A. S. Willsky, "Multiresolution stochastic models, data fusion and wavelet transforms," *Signal Processing*, vol. 34, pp. 257–282, December 1993.

[16] R. D. Nowak, "Multiscale hidden Markov models for Bayesian image analysis," Tech. Rep. 004-98, Michigan State University, Ann Arbor, MI, 1998.

[17] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. on Signal Processing*, vol. 46, pp. 886–902, April 1998.

[18] K. Chou, A. Willsky, A. Benveniste, and M. Basseville, "Recursive and iterative estimation algorithms for multi-resolution stochastic processes," in *$28^t h$ Conference on Decision and Control*, vol. 2, (Tampa, Florida), pp. 1184–1189, December 13-15 1989.

[19] B. Claus and G. Chartier, "Multiscale signal processing: Isotropic random fields on homogeneous trees," *IEEE Trans. on Circuits and Systems:Analog and Digital Signal Proc.*, vol. 41, pp. 506–517, August 1994.

[20] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. on Image Processing*, vol. 3, pp. 41–64, January 1994.

[21] P. W. Fieguth, W. C. Karl, and A. S. Willsky, "Efficient multiresolution counterparts to variational methods for surface reconstruction," *Computer Vision and Image Understanding*, vol. 70, pp. 157–176, May 1998.

[22] A. B. Frakt and A. S. Willsky, "Efficient multiscale stochastic realization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, (Seattle, WA), pp. 2249–2252, May 12-15 1998.

[23] K. Daoudi, A. B. Frakt, and A. S. Willsky, "Multiscale autoregressive models and wavelets," *IEEE Trans. on Information Theory*, vol. 45, pp. 828–845, April 1999.

[24] H.-C. Yang and R. Wilson, "Adaptive image restoration using a multiresolution Hopfield neural network," in *Fifth International Conference on Image Processing and its Applications, (IEE Conference Publication No.410)*, (Edinburgh, UK), pp. 198–202, July 4-6 1995.

[25] R. Nowak and E. D. Kolaczyk, "A multiscale MAP estimation method for Poisson inverse problems," in *32nd Asilomar Conference on Signals, Systems & Computers*, vol. 2, (Pacific Grove, CA), pp. 1682–1686, November 1-4 1998.

[26] R. D. Nowak and E. D. Kolaczyk, "A statistical multiscale framework for Poisson inverse problems," *IEEE Trans. on Information Theory*, vol. 46, pp. 1811–1825, August 2000. Special Issue on Information-Theoretic Imaging.

[27] Z. Kato, M. Berthod, and J. Zerubia, "Multiscale Markov random field models for parallel image classification," in *International Conference on Computer Vision*, (Berlin, Germany), pp. 253–237, May 1993.

[28] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. on Image Processing*, vol. 3, pp. 162–177, March 1994.

[29] M. L. Comer and E. J. Delp, "Segmentation of textured images using a multiresolution Gaussian autoregressive model," *IEEE Trans. on Image Processing*, vol. 8, pp. 408–420, March 1999.

[30] J.-M. Laferte, P. Perez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. on Image Processing*, vol. 9, pp. 390–404, March 2000.

[31] Z. Wu, G. T. Herman, and J. A. Browne, "Edge preserving reconstruction using adaptive smoothing in wavelet domain," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 3, (San Francisco, California), pp. 1917–1921, October 31 - November 6 1993.

[32] S. S. Saquib, C. A. Bouman, and K. Sauer, "A non-homogeneous MRF model for multiresolution Bayesian estimation," in *IEEE International Conference on Image Processing*, vol. 2, (Lausanne Switzerland), pp. 445–448, September 16-19 1996.

[33] P. W. Fieguth, W. C. Karl, W. S. Willsky, and C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 33, pp. 280–292, March 1995.

[34] J.-M. Laferte, F. Heitz, P. Perez, and E. Fabre, "Hierarchical statistical models for the fusion of multiresolution image data," in *International Conference on Computer Vision*, (Cambridge, MA), pp. 908–913, June 20-23 1995.

[35] W. W. Irving, P. W. Fieguth, and A. S. Willsky, "An overlapping tree approach to multiscale stochastic modeling and estimation," *IEEE Trans. on Image Processing*, vol. 6, pp. 1517–1529, June 1997.

[36] R. D. Nowak, "Shift invariant wavelet-based statistical models and 1/f processes," in *IEEE DSP Workshop*, 1998.

[37] R. Coifman and D. Donoho, "Translation invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*, pp. 125–150, New York: Springer Verlag, 1995.

[38] Z. Kato, M. Berthod, and J. Zerubia, "A hierarchical Markov random field model and multitemperature annealing for parallel image classification," *Graphical Models and Image Processing*, vol. 58, pp. 18–37, January 1996.

[39] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Tree-based resolution synthesis," in *Proceedings of the Image Processing, Image Quality, Image Capture Systems Conference (PICS '99)*, (Savannah, GA), pp. 405–410, April 25-28 1999.

[40] E. L. Lehmann, *Theory of Point Estimation.* New York: John Wiley & Sons, 1983.

[41] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, pp. 534–548, February 1993.

[42] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, pp. 480–492, March 1996.

[43] S. B. Gelfand, C. S. Ravishankar, and E. J. Delp, "An iterative growing and pruning algorithm for classification tree design," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 163–174, February 1991.

[44] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees.* Belmont, CA: Wadsworth International Group, 1984.

[45] J. Qi, R. M. Leahy, C. Hsu, T. H. Farquhar, and S. R. Cherry, "Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+," *IEEE Trans. on Nuclear Science*, vol. 45, pp. 1096–1103, June 1998.

[46] N. C. Rouze, W. Winkle, and G. D. Hutchins, "IndyPET - a high resolution, high sensitivity dedicated research scanner," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 3, (Seattle, WA), pp. 1460–1464, October 24-30 1999.

[47] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 601–609, December 1994.

[48] R. E. Carson, Y. Yan, B. Chodkowski, T. K. Yap, and M. E. Daube-Witherspoon, "Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 526–537, September 1994.

[49] C. A. Johnson, Y. Yan, R. E. Carson, R. L. Martino, and M. E. Daube-Witherspoon, "A system for the 3D reconstruction of retracted-septa PET data using the EM algorithm," *IEEE Trans. on Nuclear Science*, vol. 42, pp. 1223–1227, August 1995.

[50] E. *Ü*. Mumcuoğlu, R. M. Leahy, S. R. Cherry, and E. Hoffman, "Accurate geometric and physical response modelling for statistical image reconstruction in high resolution PET," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 3, (Anaheim, CA), pp. 1569–1573, November 2-9 1996.

[51] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-1, pp. 113–122, October 1982.

[52] J.-S. Liow and S. C. Strother, "Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions," *IEEE Trans. on Medical Imaging*, vol. 10, pp. 563–571, December 1991.

[53] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner," *Phys. Med. Biol.*, vol. 43, pp. 1001–1013, April 1998.

[54] L. Kaufmann, "Maximum likelihood, least squares, and penelized least squares for PET," *IEEE Trans. on Medical Imaging*, vol. 12, pp. 200–214, June 1993.

[55] J. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithms," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. 42, pp. 2664–2677, October 1994.

[56] E. *Ü*. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 687–701, December 1994.

[57] A. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. on Medical Imaging*, vol. 14, pp. 132–137, March 1995.

[58] S. R. Cherry, Y. Shao, S. Siegel, R. W. Silverman, K. Meadors, J. Young, W. F. Jones, D. Newport, C. Moyers, E. *Ü*. Mumcuoğlu, M. Andreaco, M. Paulus, D. Binkley, R. Nutt, and M. E. Phelps, "MicroPET: a high resolution PET scanner for imaging small animals," *IEEE Trans. on Nuclear Science*, vol. 44, pp. 1161–1166, June 1997.

[59] G. Brix, J. Zaers, L.-E. Adam, M. E. Bellemann, H. Ostertag, H. Trojan, U. Haberkorn, J. Doll, F. Oberdorfer, and W. J. Lorenz, "Performance evaluation of a whole-body PET scanner using the NEMA protocol," *Journal of Nuclear Medicine*, vol. 38, pp. 1614–1623, October 1997.

[60] A. Chatziioannou, J. Qi, A. Moore, A. Annala, K. Nguyen, R. M. Leahy, and S. R. Cherry, "Comparison of 3D maximum a posteriori and filtered backprojection algorithms for high resolution animal imaging with microPET," *IEEE Trans. on Medical Imaging*, vol. 19, pp. 507–512, May 2000.

[61] T. J. Holmes and D. C. Ficke, "Analysis of positron-emission tomography scintiallation-detectors with wedge faces and inter-crystal septa," *IEEE Trans. on Nuclear Science*, vol. 32, pp. 826–830, February 1985.

[62] Z. Liang, "Detector response restoration in image reconstruction of high resolution positron emission tomography," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 314–321, June 1994.

[63] N. A. Keller and L. R. Lupton, "PET detector ring aperture function calculations using monte carlo techniques," *IEEE Trans. on Nuclear Science*, vol. 30, pp. 676–680, February 1983.

[64] R. Lecomte, D. Schmitt, and G. Lamoureux, "Geometric study of a high resolution PET detection system using small detectors," *IEEE Trans. on Nuclear Science*, vol. 31, pp. 556–561, February 1984.

[65] R. H. Huesman, E. M. Salmeron, and J. R. Baker, “Compensation for crystal penetration in high resolution positron tomography,” *IEEE Trans. on Nuclear Science*, vol. 36, pp. 1100–1107, February 1989.

[66] T. Frese, C. A. Bouman, N. C. Rouze, G. D. Hutchins, and K. Sauer, “Bayesian multiresolution algorithm for pet reconstruction,” in *IEEE International Conference on Image Processing*, (Vancouver, BC, Canada), pp. 613–616, September 10-13 2000.

[67] T. Frese, C. A. Bouman, and K. Sauer, “Adaptive wavelet graph model for Bayesian tomographic reconstruction,” *Submitted to IEEE Trans. on Image Processing*.

[68] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs,” *IEEE Trans. on Image Processing*, vol. 5, pp. 1346–1358, September 1996.

[69] J. W. Stayman and J. A. Fessler, “Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction,” *IEEE Trans. on Medical Imaging*, vol. 19, pp. 601–615, June 2000.

[70] J. Qi and R. M. Leahy, “A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data,” *IEEE Trans. on Medical Imaging*, vol. 18, pp. 293–305, April 1999.

[71] J. Qi and R. M. Leahy, “Resolution and noise properties of MAP reconstruction for fully 3-D PET.,” *IEEE Trans. on Medical Imaging*, vol. 19, pp. 493–506, May 2000.

[72] K. Wienhard, M. Dahlbom, L. Eriksson, C. Michel, T. Bruckbauer, U. Pietrzyk, and W.-D. Heiss, “The ECAT EXACT HR: performance of a new high resolution positron scanner,” *J. Comput. Assist. Tomogr.*, vol. 18, pp. 110–118, Jan.-Feb. 1994.

[73] M. Yavuz and J. A. Fessler, “Statistical image reconstruction methods for randoms-precorrected PET scans,” *Medical Image Analysis*, vol. 2, pp. 369–378, December 1998.

[74] P. Huber, *Robust Statistics.* New York: John Wiley & Sons, 1981.

[75] L. Kaufmann, “Implementing and accelerating the EM algorithm for positron emission tomography,” *IEEE Trans. on Medical Imaging*, vol. MI-6, no. 1, pp. 37–51, 1987.

[76] E. *Ü*. Mumcuoğlu, R. Leahy, and S. R. Cherry, “Bayesian reconstruction of PET images: Methodology and performance analysis,” *Phys. Med. Biol.*, vol. 41, pp. 1777–1807, September 1996.

[77] M. T. Chan, R. M. Leahy, E. Ü.. Mumcuoğlu, S. R. Cherry, J. Czernin, and A. Chatziioannou, “Comparing lesion detection performance for PET image reconstruction algorithms: a case study,” *IEEE Trans. on Nuclear Science*, vol. 44, pp. 1558–1563, August 1997.

[78] J. Llacer, E. Veklerov, L. R. Baxter, S. T. Grafton, L. K. Griffeth, R. A. Hawkins, C. K. Hoh, J. C. Mazziotta, E. J. Hoffman, and C. E. Metz, "Results of a clinical receiver operating characteristic study comparing filtered back-projection and maximum likelihood estimator images in FDG PET studies," *Journal of Nuclear Medicine*, vol. 34, pp. 1198–1203, July 1993.

[79] G. Herman, *Image Reconstruction from Projectons: The Fundamentals of Computerized Tomography.* New York: Academic Press, 1980.

[80] G. C. McKinnon and R. H. T. Bates, "Towards imaging the beating heart usefully with a conventional CT scanner," *IEEE Trans. on Biomedical Engineering*, vol. BME-28, pp. 123–127, February 1981.

[81] J. G. Sanderson, "Reconstruction of fuel pin bundles by a maximum entropy method," *IEEE Trans. on Nuclear Science*, vol. NS-26, pp. 2685–2688, April 1979.

[82] T. Inouye, "Image reconstruction with limited angle projection data," *IEEE Trans. on Nuclear Science*, vol. NS-26, pp. 2666–2684, April 1979.

[83] G. H. Glover and N. J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT resonstruction," *Med. Phys.*, vol. 8, pp. 799–807, Nov.-Dec. 1981.

[84] S. K. Chang, "The reconstruction of binary patterns from their projections," *Communications of the ACM*, vol. 14, pp. 21–25, January 1971.

[85] A. Shliferstein and Y. T. Chien, "Switching components and the ambiguity problem in the reconstruction of pictures from their projections," *Pattern Recognition*, vol. 10, no. 5-6, pp. 327–340, 1978.

[86] A. Kuba, "The recognition of two-directionally connected binary patterns from their two orthogonal projections," *Comput. Vision Graphics and Image Process.*, vol. 27, pp. 249–265, September 1984.

[87] S. K. Chang and C. K. Chow, "The reconstruction of three-dimensional objects from two orthogonal projections and its application to cardiac cineangiography," *IEEE Trans. on Computers*, vol. C-22, pp. 18–28, January 1973.

[88] M. Soumekh, "Binary image reconstruction from four projections," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1280–1283, April 11-14 1988.

[89] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects – parts I and II: Performance analysis and robustness analysis," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. ASSP-32, pp. 886–906, August 1984.

[90] Y. Bresler, J. A. Fessler, and A. Macovski, "A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 840–858, August 1989.

[91] G. T. Herman and A. Kuba, eds., *Discrete Tomography: Foundations, Algorithms and Applications.* Birkhauser Boston, Cambridge, MA, 1999.

[92] S. Geman and D. McClure, “Bayesian image analysis: An application to single photon emission tomography,” in *Proceedings of the Statistical Computing Section, American Statistical Association*, (Washington, DC), pp. 12–18, 1985.

[93] B. Hunt, “Bayesian methods in nonlinear digital image restoration,” *IEEE Trans. on Computers*, vol. c-26, no. 3, pp. 219–229, 1977.

[94] S. Geman and D. McClure, “Statistical methods for tomographic image reconstruction,” *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5–21, 1987.

[95] H. Derin, H. Elliott, R. Cristi, and D. Geman, “Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 707–719, November 1984.

[96] C. A. Bouman and B. Liu, “Multiple resolution segmentation of textured images,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 99–113, February 1991.

[97] H. Hart and Z. Liang, “Bayesian image processing in two dimensions,” *IEEE Trans. on Medical Imaging*, vol. MI-6, pp. 201–208, September 1987.

[98] Z. Liang and H. Hart, “Bayesian image processing of data from constrained source distributions–I: Non-valued, uncorrelated and correlated constraints,” *Bull. Math. Biol.*, vol. 49, no. 1, pp. 51–74, 1987.

[99] G. T. Herman and D. Odhner, “Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography,” *IEEE Trans. on Medical Imaging*, vol. 10, pp. 336–346, September 1991.

[100] G. T. Herman, A. R. De Pierro, and N. Gai, “On methods for maximum a posteriori image reconstruction with normal prior,” *J. Visual Comm. Image Rep.*, vol. 3, pp. 316–324, December 1992.

[101] T. Hebert and S. Gopal, “The GEM MAP algorithm with 3-D spect system response,” *IEEE Trans. on Medical Imaging*, vol. 11, pp. 81–90, March 1992.

[102] F. Heitz, P. Perez, and P. Bouthemy, “Multiscale minimization of global energy functions in some visual recovery problems,” *Comput. Vision Graphics and Image Process.*, vol. 59, pp. 125–134, January 1994.

[103] D. Snyder and M. Miller, “The use of sieves to stabilize images produced with the EM algorithm for emission tomography,” *IEEE Trans. on Nuclear Science*, vol. NS-32, pp. 3864–3871, October 1985.

[104] E. Veklerov and J. Llacer, “Stopping rule for the MLE algorithm based on statistical hypothesis testing,” *IEEE Trans. on Medical Imaging*, vol. MI-6, no. 4, pp. 313–319, 1987.

[105] T. Hebert, R. Leahy, and M. Singh, “Fast MLE for SPECT using an intermediate polar representation and a stopping criterion,” *IEEE Trans. on Nuclear Science*, vol. 35, pp. 615–619, February 1988.

[106] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems.* New York: Winston and Sons, 1977.

[107] J. Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society B*, vol. 36, no. 2, pp. 192–236, 1974.

[108] R. Kindermann and J. Snell, *Markov Random Fields and their Applications*. Providence: American Mathematical Society, 1980.

[109] R. Kashyap and R. Chellappa, “Estimation and choice of neighbors in spatial-interaction models of images,” *IEEE Trans. on Information Theory*, vol. IT-29, pp. 60–72, January 1983.

[110] D. Pickard, “Inference for discrete Markov fields: The simplest nontrivial case,” *Journal of the American Statistical Association*, vol. 82, pp. 90–96, March 1987.

[111] R. Dubes and A. Jain, “Random field models in image analysis,” *Journal of Applied Statistics*, vol. 16, no. 2, pp. 131–164, 1989.

[112] K. Sauer and C. Bouman, “Bayesian estimation of transmission tomograms using segmentation based optimization,” *IEEE Trans. on Nuclear Science*, vol. 39, pp. 1144–1152, 1992.

[113] C. A. Bouman, “Cluster: an unsupervised algorithm for modeling Gaussian mixtures.” Available from http://www.ece.purdue.edu/~bouman, April 1997.

[114] P. Aigrain, H. Zhang, and D. Petkovic, “Content-based representation and retrieval of visual media: A state of the art review,” *Multimedia Tools and Applications*, vol. 3, pp. 179–202, November 1996.

[115] M. Swain and D. Ballard, “Color indexing,” *International Journal of Computer Vision*, vol. 7, pp. 11 – 32, November 1991.

[116] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query by image and video content: The QBIC system,” *Computer*, vol. 28, pp. 23 – 32, September 1995.

[117] A. Pentland, R. W. Picard, and S. Sclaroff, “Photobook: Content-based manipulation of image databases,” *International Journal of Computer Vision*, vol. 18, pp. 233 – 254, June 1996.

[118] J. Y. Chen, C. A. Bouman, and J. P. Allebach, “Multiscale branch and bound image database search,” in *SPIE/IS&T Conference on Storage and Retrieval for Image and Video Databases V*, vol. 3022, (San Jose, CA), pp. 133 – 144, February 13 - 14 1997.

[119] T. Frese, C. A. Bouman, and J. P. Allebach, “A methodology for designing image similarity metrics based on human visual system models,” in *SPIE/IS&T Conference on Human Vision and Electronic Imaging II*, vol. 3016, (San Jose, CA), pp. 472 – 483, February 10 - 13 1997.

[120] B. Scassellati, S. Alexopoulos, and M. Flickner, “Retrieving images by 2D shape: a comparison of computation methods with human perceptual judgments,” in *SPIE/IS&T Conference on Storage and Retrieval for Image and Video Databases*, vol. 2185, (San Jose, CA), pp. 2 – 14, February 7-8 1994.

[121] Y. Rubner, L. J. Guibas, and C. Tomasi, “The earth mover’s distance, multi-dimensional scaling, and color-based image retrieval,” in *ARPA Image Understanding Workshop*, (New Orleans, LA), pp. 661–668, May 1997.

[122] J. MacCuish, A. McPherson, J. E. Barros, and P. M. Kelly, “Interactive layout mechanisms for image database retrieval,” in *SPIE Conference on Visual Data Exploration and Analysis III*, vol. 2656, (San Jose, CA), pp. 104 – 115, Jan. 31 - Feb. 2 1996.

[123] R. N. Shepard, “Toward a universal law of generalization for psychological science,” *Science*, vol. 237, pp. 1317 – 1323, September 1987.

[124] J. B. Kruskal, *Multidimensional scaling.* Beverly Hills, CA: Sage Publications, 1978.

[125] C. L. Krumhansl, “The psychological representation of musical pitch in a tonal context,” *Cognitive Psychology*, vol. 11, pp. 346–374, July 1979.

[126] C. R. Cavonius, M. Mueller, and J. D. Mollon, “Difficulties faced by color-anomalous observers in interpreting color displays,” in *SPIE Conference on Perceiving, Measuring, and Using Color* (M. H. Brill, ed.), vol. 1250, (Santa Clara, CA), pp. 190 – 195, 1990.

[127] A. R. Rao and G. L. Lohse, “Identifying high level features of texture perception,” *Computer Vision Graphics and Image Processing:Graphical Models and Image Proc.*, vol. 55, pp. 218 – 233, May 1993.

[128] M. Vanrell, J. Vitrià, and X. Roca, “A multidimensional scaling approach to explore the behavior of a texture perception algorithm,” *Machine Vision and Applications*, vol. 9, no. 5-6, pp. 262 – 271, 1997.

[129] T. F. Cox and M. A. A. Cox, *Multidimensional scaling.* Monographs on statistics and applied probability, London: Chapman & Hall, 1st ed., 1994.

[130] D. A. Rabenhorst, J. A. Gerth, and C. N. Mills, *BMDP/DIAMOND for Windows: User’s Guide.* Los Angeles, CA: BMDP Statistical Software, Inc., 1995. Currently available through SPSS, Inc., Chicago, IL.

[131] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *IEEE Trans. on Information Theory*, vol. 38, pp. 587–607, March 1992.

[132] M. R. Luettgen, W. C. Karl, A. S. Willsky, and R. R. Tenney, “Multiscale representations of Markov random fields,” *IEEE Trans. on Signal Processing*, vol. 41, December 1993. special issue on Wavelets and Signal Processing.

[133] R. W. Dijkerman and R. R. Mazumdar, “Wavelet representations of stochastic processes and multiresolution stochastic models,” *IEEE Trans. on Signal Processing*, vol. 42, pp. 1640–1652, July 1994.

[134] M. R. Luettgen and A. S. Willsky, “Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination,” *IEEE Trans. on Image Processing*, vol. 4, pp. 194–207, February 1995.

[135] R. W. Dijkerman, R. R. Mazumdar, and A. Bagchi, "Reciprocal processes on a tree - modeling and estimation issues," *IEEE Trans. Automatic Control*, vol. 40, pp. 330–335, February 1995.

[136] P. Perez and F. Heitz, "Restriction of a Markov random field on a graph and multiresolution statistical image modeling," *IEEE Trans. on Information Theory*, vol. 42, pp. 180–190, January 1996.

[137] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using a hierarchical Markovian model," in *Fifth International Conference on Computer Vision*, (Cambridge, MA), pp. 169–174, June 20-23 1995.

[138] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recognition*, vol. 32, pp. 591–604, April 1999.

[139] D. Tretter, C. A. Bouman, K. Khawaja, and A. Maciejewski, "A multiscale stochastic image model for automated inspection," *IEEE Trans. on Image Processing*, vol. 4, pp. 507–517, December 1995.

[140] K. W. Khawaja, A. A. Maciejewski, D. Tretter, and C. A. Bouman, "Automated assembly inspection using a multiscale algorithm trained on synthetic images," *IEEE Robotics and Automation Magazine*, vol. 3, pp. 15–22, June 1996.

[141] S. Krishnamachari and R. Chellappa, "Multiresolution Gauss-Markov random field models for texture segmentation," *IEEE Trans. on Image Processing*, vol. 6, pp. 251–267, February 1997.

[142] W. W. Irving, L. M. Novak, and A. S. Willsky, "A multiresolution approach to discrimination in SAR imagery," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 33, pp. 1157–1169, October 1997.

[143] A. B. Frakt, W. C. Karl, and A. S. Willsky, "A multiscale hypothesis testing approach to anomaly detection and localization from noisy tomographic data," *IEEE Trans. on Image Processing*, vol. 7, pp. 825–837, June 1998.

[144] T.-I. Hsu and R. Wilson, "A two-component model of texture for analysis and synthesis," *IEEE Trans. on Image Processing*, vol. 7, pp. 1466–1476, December 1998.

[145] C.-T. Li and R. Wilson, "Image segmentation based on a multiresolution Bayesian framework," in *IEEE International Conference on Image Processing*, vol. 3, (Chicago, IL.), pp. 761–765, October 4-7 1998.

[146] R. D. Nowak and R. G. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Trans. on Image Processing*, vol. 8, pp. 666–678, May 1999.

[147] A. J. Kim, H. Krim, and A. S. Willsky, "Segmentation and compression of SAR imagery via hierarchical stochastic modeling," in *IEEE International Conference on Image Processing*, vol. 3, (Santa Barbara, California), pp. 488–491, October 26-29 1997.

[148] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Data Compression Conference*, (Snowbird, Utah), pp. 221–230, 1997.

[149] E. P. Simoncelli and R. W. Buccigrossi, "Embedded wavelet image compression based on a joint probability model," in *IEEE International Conference on Image Processing*, vol. I, (Santa Barbara, California), pp. 640–643, October 26-29 1997.

[150] C. Graffigne, F. Heitz, P. Perez, F. Preteux, M. Sigelle, and J. Zerubia, "Hierarchical Markov random field models applied to image analysis: a review," in *SPIE Conference on Neural, Morphological and Stochastic Methods in Image and Signal Processing*, vol. 2568, (San Diego, CA), pp. 2–17, July 10-11 1995.

[151] H. Lucke, "Which stochastic models allow Baum-Welch training?," *IEEE Trans. on Signal Processing*, vol. 44, pp. 2746–2756, November 1996.

APPENDIX

A.6 A Review of Multiscale Random Process Liturature

A.6.1 Continuous State Models

In [12], Basseville, Benveniste, Chou, Golden, Nikoukhah and Willsky develop a linear signal processing technique for signals on trees. A distance measure on the tree is defined as the number of nodes traversed on the shortest path between two nodes. This measure is used to define isometries as mappings that preserve all distances on the tree. Two classes of stationarity are defined and investigated. The first and stronger class is given by isotropic processes whose covariance function is invariant to isometries on the tree, i.e. it depends only on the distance between two nodes in the tree. Isotropic systems are defined such that the linear system response commutes with isometries on the tree. These systems are non-causal in scale. The second, weaker class of stationarity is obtained by requiring the covariance of associated processes to be invariant only to translations on the tree, i.e. isometries that shift an infinite path on the tree by one node. Processes and systems associated with this weaker class of stationarity are called stationary. The system response of stationary systems is analyzed, specifically it is shown that the input/output response for two nodes only depends on the distance of the two nodes from their first common tree ancestor.

Parameterizations of processes and Levinson-Schur type algorithms are developed for both isotropic and stationary processes. Multiscale AR(p) processes are introduced as parameterizations for isotropic processes. A p-th order AR process has inputs only from nodes at the same or coarser scales with distance less or equal p. It is shown that the parameterization of such processes by partial correlation coefficients is more convenient than using the AR coefficients, and generalizations of the Levinson and Schur recursions for these processes are derived. For stationary systems, it is shown that any such system can be written as a combination of two basic systems, a coarse-to-fine interpolation and a forward-average shift. These systems are shown to correspond to classical separable 2-D systems. Requiring these systems to be causal and have finite-memory leads to the MAR state representation. Several estimation algorithms for MAR processes are developed. For the case, where the MAR system parameters depend only on scale, the Haar wavelet transform is applied to the states and observations at each scale. This results in two decoupled white noise dynamic systems at each node. For the general case, a Kalman-like generalization of the Rauch-Tung-Striebel (RTS) algorithm is used for state estimation. [cont, lin, wav, Mar 92]

In [131], Simoncelli, Freeman, Adelson and Heeger propose shift-invariant multiscale transforms as an alternative to wavelet transforms which typically are not shift invariant. The authors define shift-invariance in space, orientation and scale. Specifically, a transform is considered shift-invariant with respect to a subband, if the subband response at any parameter value (e.g. location for space invariance) can be interpolated from the transform coefficients of that subband. It is shown, that in order to obtain shift invariant self-inverting transforms, it is generally necessary to relax orthogonality. This leads to overcomplete transform representations. Examples of shift-invariant transforms are proposed. [cont, lin, March 92]

In [13, 14], Basseville, Benveniste and Willsky develop linear system theory for multiscale autoregressive (MAR) processes as the multiresolution counterpart to LTI systems for time series. In part I [13], isotropic processes on trees are defined as processes whose covariance is only a function of the distance between tree nodes. Causal white noise driven models that produce isotropic processes are developed as a class of AR models on trees. Levinson and Schur recursions are developed for isotropic processes on dyadic trees. In part II [14], lattice structures for whitening and modeling are developed for isotropic processes on dyadic trees. It is shown, that parameterization by reflection coefficients leads to a more elegant representation as compared to the coefficient representation of these processes. Whitening and modeling filters are specified in terms of reflection coefficients and a recursive filtering technique is presented. [cont, lin, Aug 92]

In [15], Chou, Golden and Willsky develop a class of linear multiscale stochastic models based on a modification of the wavelet synthesis equation. The wavelet synthesis equation is modified by a allowing a vector of scaling coefficients at each location. Before applying the synthesis filter, the coarse scale scaling coefficients are multiplied by a diagonal matrix, i.e. are filtered separately at each location in a spatially homogeneous way. Furthermore, instead of interpolating noise realizations of detail coefficients, the noise is directly added to the interpolated scaling coefficients. The resulting model is Markovian in scale and allows for data measurements at all scales. The model is applied to a smoothing problem for a first-order Gauss-Markov process. [cont, lin, Dec 93]

In [132], Luettgen, Karl, Willsky and Tenney show how MRF's can be represented exactly on trees. The proposed representations are based on a generalization of midpoint deflection for Brownian motion (1-D case) which is generalized to a midline deflection construction for the 2-D case. For representation of a 1-D MRF, each tree node corresponds to an interval of the fine scale signal. The nodes contain the two endpoints as well as the center point of the interval. The node's children are then refinements of this interval, with the left child containing the interval from the left boundary point to the center. This model can be modified by including the 2 center points of each interval in order to avoid overlap of data points in adjacent tree nodes. For the 2-D case, the model is generalized such that each node contains the boundary of a square block as well as the horizontal and vertical midlines. In order to reduce the complexity of the 2-D model, an approximate model is proposed in which the 2-D block boundaries are approximated by scaling coefficients of their 1-D wavelet transform. Since the proposed MRF representation is scale recursive, it allows for fast recursive realization algorithms. Results are demonstrated for sample path generations of textures. [cont, Dec 93]

In [20], Luettgen, Karl and Willsky propose a non-iterative multiscale algorithm for the computation of optical flow. The problem formulation by Horn and Schunck is interpreted in an estimation theoretic framework. The Horn and Schunck gradient prior term is replaced by a "similar" prior term which is realized by a multiresolution MAR model. The multiscale model is the standard MAR model where the state at each node depends only on the parent state. The state vector at each node corresponds to the optical flow, whereas the output vector corresponds to the observed brightness. The MAR output equation is chosen to satisfy the brightness constraint as in the approach by Horn and Schunck. The noise term in the state equation is chosen to result in a prior similar to that of Horn and Schunck. The reconstruction algorithm is recursive, Kalman-like and results in constant complexity per image pixel, regardless of image size. Results for different reconstruction sequences are presented. [cont, lin, Jan 94]

In [133], Dijkerman and Mazumdar use wavelet representations to obtain multiresolution representations of stochastic processes. The decorrelation properties of the wavelet transform are investigated. An AR model for the wavelet coefficients is developed. Wavelet coefficients are modeled as linear functions of coarse scale coefficients and causal coefficients within scale. This model is used to obtain approximate multiscale representations of random processes with specified fine scale correlation. The approximation of the correlation can be evaluated by visual inspection of the difference between the correlation matrices or by using Bhattacharya distance. [cont, lin, July 94]

In [19], Claus and Chartier extend and generalize the work in [13, 14]. The Levinson and Schur recursions for isotropic processes are generalized from dyadic to q^{th} order trees. The Schur recursion is generalized and formulated more compactly. [cont, lin, Aug 94]

In [134], Luettgen and Willsky develop a fast likelihood calculation algorithm for noisy observations in a stochastic MAR multiresolution model. In the MAR model, each node state depends only on the parent state and is white noise driven. The observations are assumed to be independent white conditioned on the node state. Since the observations are not independent in general, fast likelihood calculation is problematic. The proposed algorithm relies on pre-whitening the data followed by fast likelihood calculation on the independent residuals. The whitening is performed by first applying Kalman filtering resulting in partial whitening on paths from the tree root to the leaves. The remaining correlations are removed by using the multiscale state model to subtract the conditional expectation of the observations between nodes that are still correlated. The overall algorithm consists of an upward and a downward sweep and has constant per pixel computational complexity. The proposed algorithm is applied to a texture discrimination task and compared to a GMRF model. [cont, lin, Feb. 1995]

In [135], Dijkerman, Mazumdar and Bagchi introduce reciprocal random processes on N-ary trees. Reciprocal processes are defined such that the conditional probability of a value of a tree node, given the values of all other nodes in the tree is only dependent on the values of the parent node and the children nodes. Nearest neighbor models (NNM) are introduced and it is shown that a zero-mean Gaussian process on a binary tree is reciprocal if and only if it admits a NNM. A recursive description of such processes is derived and it is shown that in the case of zero boundary conditions on the truncated tree, reciprocal processes reduce to autoregressive processes. The proposed framework is applied to a 1-D smoothing problem. [cont, lin, Feb. 95]

In [136], Perez and Heitz investigate the Markovian properties of MRF transformations typically used in multiresolution image modeling. The locality of subsampled MRFs is studied by considering the restriction of the original fine scale MRF to a subset of its original site. Locality is examined by determining the minimal neighborhood with respect to which the restricted field is an MRF. The authors show that most subsampling schemes as well as typical renormalization group transformations result in loss of locality. For hierarchical MRFs defined on quadtrees and pyramidal structures, it is shown that the restriction to a single resolution level results in loss of locality, making separate simulation for each level intractable. [cont, jan 96]

In [35], Irving, Fieguth and Willsky propose an overlapping tree approach to address the problem of blockiness with standard quadtree multiscale models. A linear mapping is used to transform a given fine scale process into an overcomplete, overlapping tree structure. Specifically, each fine scale node of the tree corresponds to only a single pixel of the original process. However, multiple fine scale nodes can correspond to the same pixel in the original process, i.e. the linear transform effectively copies single pixels into multiple different nodes, resulting in an overcomplete representation at fine scale. At the coarser scales, nodes correspond to overlapping spatial averages of pixels. The spatial overlap at each scale and some additional constraints uniquely specify the fine scale forward transform. Multiscale modeling and estimation are performed using the standard MAR framework where each node depends only on its tree parent. The observations are assumed to be conditionally white given the state of the node. Note that for estimation applications, the original process observations are transformed into the tree representation. This results in multiple copies of a single observations in different nodes. Although these nodes contain the same observations, they are modeled as independent by the MAR framework. To compensate for this, the MAR noise variances in each node are multiplied by the number of nodes that contain the same observation. The results of the estimation or modeling are transformed back into the original process domain. The back-transform is obtained by averaging values of fine scale nodes corresponding to the same pixel location. Since the averaging is performed over nodes corresponding to the same location, the authors claim no loss in resolution. There is a degree of freedom in choosing the backtransform weights for nodes corresponding to the same pixel. Experimental results are provided for stationary and non-stationary texture synthesis as well as for texture denoising. [cont,lin,Nov 1997]

In [21], Fieguth, Karl and Willsky develop a multiresolution algorithm for surface estimation as an alternative to commonly used variational methods. The multiresolution model is the standard MAR model, i.e. scale recursive such that the state at each node is a linear function of the parent node state. The observations are assumed to be independent for each node given the state. The paper focuses on specifying a multiresolution model of this form to approximate a commonly used variational model that requires the solution to an Euler-Lagrange pde. To improve the model, the overlapping tree model [35] is applied. The surface reconstruction is performed using an earlier published recursive Kalman-like algorithm. [cont, lin, May 98]

In [22], Frakt and Willsky propose a computationally efficient algorithm to determine the parameters of an internal MAR multiscale stochastic model such that the finest scale of the MAR process matches a specified correlation function. The complexity of the proposed algorithm is $O(N^2)$ where N denotes the number of signal samples at fine scale, i.e. N^2 equals the size of the specified fine scale correlation function. [cont, lin, 98]

In [23], Daoudi, Frakt and Willsky unite the multiscale autoregressive framework (MAR) with wavelet models for general wavelet bases. In order to use the MAR framework with bases other than the Haar basis, state augmentation is performed such that the scaling coefficients at each node only depend on scaling and wavelet coefficients in the direct parent node. Specifically, nodes contain overlapping information. A second state augmentation is then performed in order to satisfy internality, i.e. so that the state at each node is a linear function of the states in the leaves of its subtree. Specifically, at each node, the children nodes are augmented to contain a range of scaling coefficients that are adjacent to the node. The augmented scaling coefficients are then propagated down the tree where the information is split between children in order to keep the state dimension bounded. This satisfies the necessary and sufficient condition for internality that each state is a linear function of the states in the two children. The fine scale signal is extracted by simply taking the scaling coefficient at the correct leaf position instead of averaging of the same coefficient in the overcomplete representation. The authors point out, that both wavelet and scaling coefficients are predicted from both coarse scale wavelet and scaling coefficients. The model observes the tree constraint and thus leads to recursive algorithms as developed earlier for the MAR framework. [cont, lin, April 99]

A.6.2 Discrete State Models

In [96], Bouman and Liu propose a multiresolution texture segmentation algorithm. Texture labels are modeled as discrete MRF's at multiple resolutions. Coarse scale labels correspond to fine scale blocks with constant labels. The MRF parameters are chosen to be identical at all scales. The image observations at the finest scale are modeled as causal AR processes given the texture label. Coarse scale log-likelihood expressions are computed as the decimated fine scale log-likelihood. The segmentation algorithm proceeds in a coarse-to-fine scale fashion. At each scale, the algorithm iteratively computes the MAP segmentation and uses the result to initialize the segmentation at the next finer scale. In order to perform unsupervised segmentation, the algorithm is extended to include parameter estimation routines. Specifically, the number of textures and their parameters are estimated using a modified AIC criterion. The estimation of texture parameters is performed by alternating texture segmentation and parameter estimation. [discr, Feb 91]

In [27, 38], Kato, Berthod and Zerubia propose a 3-D MRF for modeling the label pyramid in image segmentation tasks. The model is based on a multiresolution pyramid for the segmentation labels. The coarse scale models are obtained by assuming a constant segmentation label for all fine scale pixels within the support of the coarse scale block. The 3-D MRF neighborhood system consists of a standard MRF neighborhood within the same scale plus the quadtree parent and the four children at the adjacent scales. The segmentation is performed at all resolutions is parallel using multitemperature annealing. The multitemperature annealing assigns higher temperature to the coarser resolution levels, thereby using the higher SNR at coarse scales to increase robustness to local minima. The 3-D MRF model was first proposed in [27] without the multitemperature annealing. [discr, May 93/Jan96]

In [102], Heitz, Perez and Bouthemy develop a multiresolution relaxation algorithm to minimize a single global energy function in image segmentation tasks. The model uses a multiresolution label pyramid assuming a Markov random field (MRF) model with an 8-point neighborhood at each scale. The coarse scales are based on a nested subspace approach where each coarse scale label corresponds to a block at fine scale that has constant labels. The central contribution of the paper is the computation of the coarse scale MRF energy functions to match the fine scale energy function. The authors show that the fine scale energy function can be rewritten for the coarse scale using 1 through 4 site clique energy terms associated with the same 8-point neighborhood used at fine scale. Observations are only used at the finest scale. The model is applied to estimating discrete valued long range motion in an image sequence. The coarse to fine reconstruction algorithm uses ICM optimization at each scale. The coarse scale reconstructions are used to initialize the next finer scales. This algorithm extends the work by Bouman and Liu to include a systematic computation of the coarse scale MRF energy functions. [discr, Jan 1994]

In [28], Bouman and Shapiro propose a hierarchical multiresolution MRF model for image segmentation. The approach is based on a multiscale Markov random field (MSRF) consisting of a sequence of Markov Random fields forming a Markov chain in scale. The authors propose a sequential MAP (SMAP) estimator that minimizes the expected size of the largest misclassified region in the image. This is in contrast to the more conservative MAP estimator that maximizes the probability that all pixels are classified correctly. The SMAP estimator computes the image classification in a coarse to fine scale fashion. At each scale, the best segmentation is computed given the segmentation at the previous coarser scale and the observed data. The optimization at each scale can be expressed as a sum of a coarse scale log-likelihood and a log prior expression given by the coarse-to-fine scale transition probabilities of the Markov chain in scale. The coarse scale log-likelihoods account for all possible label changes on the way down to the finest scale where observations are available. Assumptions for the prior model are that the labels at each scale are independent given the labels at the next coarser scale. Furthermore, each label is assumed to depend only on a small window at the next coarser scale. Two choices of coarse scale neighborhoods are considered, one corresponding to a quadtree model and one corresponding to a pyramidal model. For the quadtree model, the coarse scale neighborhood consists of only the tree parent. For this model, the coarse scale log-likelihoods can be computed using a fine-to-coarse scale recursion. In the pyramidal model, each label depends on a 3-point window at the next coarser scale. For this model, a fine-to-coarse likelihood recursion is not possible. Thus, a hybrid model is used, where the segmentation at each scale is computed using a pyramidal model for the coarser scale dependencies and a quadtree model for the finer scales to recursively compute the log-likelihood expressions. For unsupervised segmentation, a two-pass EM procedure is proposed to estimate the prior parameters. The computational overhead of the parameter estimation is reduced by subsampling the image at high resolutions. [discr, Mar 94]

In [137, 138], Kato, Zerubia and Berthod propose hyperparameter estimation techniques for the 3-D MRF model presented in [27, 38]. The parameter estimation is performed by using expectation maximization (EM) techniques. The hyperparameters are obtained by iterating MAP segmentation and ML parameter estimation. In [138], the authors introduce two different techniques based on adaptive simulated annealing (ASA) and iterative conditional estimation (ICE). Extending the algorithms in [27, 38] to use the hyperparameter estimation results in an unsupervised segmentation algorithm where only the number of classes is assumed to be known. [discr, Jun 95].

In [34], Laferte, Heitz, Perez and Fabre provide an extension to SMAP[28] by allowing observations at multiple resolutions. As in SMAP, this is done for both quadtree and a pyramidal graph structures. The authors present a non-iterative, Viterbi-like algorithm for exact MAP estimation on the quad-tree. [seg, 95]

In [139] and [140] Tretter, Bouman, Khawaja and Maciejewski use a hierarchical stochastic tree model for object detection and estimation of object parameters such as location, scale and orientation in automated inspection tasks. The proposed state model consists of an object tree where each node corresponds to a subassembly of the object under inspection. Each node's state contains the location, orientation and scale of the subassembly. The states are modeled using Gaussian coarse-to-fine scale transition probabilities between parent and child nodes. The data term for each node is based on the wavelet transform of a 2-D monochrome image of the entire object. Based on the node's state, the measured data points are compared to an affine transform of mean and variance statistics of the known subassembly obtained from training data. The estimation of the states is defined using a sequential MAP (SMAP) cost function. The SMAP estimation is carried out using a fast tree search strategy that identifies and evaluates multiple candidate paths of object relationships in scale. Training and parameter estimation for the model are performed using a similar search in conjunction with the EM algorithm.

In [141], Krishnamachari and Chellappa develop two methods to approximate subsampled GMRF's with coarse scale GMRF's. The first method optimizes the parameters of the coarse scale GMRF to minimize the Kullback-Leibler distance between the pdf's of the subsampled GMRF and the coarse scale GMRF. The second method optimizes the coarse scale parameters based on the local conditional distribution invariance. The authors propose a texture segmentation algorithm using this model. Texture regions are modeled using a fine scale GMRF whose parameters are assumed to be known. Based on the parameters at fine scale, GMRF parameters at coarser resolutions are estimated using the proposed framework. The segmentation is then computed in a coarse-to-fine scale fashion. Starting with the coarsest scale, the segmentation is computed separately at each scale using ICM. The segmentation at each scale is then propagated as initialization to the next finer scale. In addition, a confidence measure is attached to each location of the segmentation at each scale. At finer scales, only pixels with low confidence coarse scale segmentation are processed. [discr, Feb 97].

In [142], Irving, Novak and Willsky apply a multiresolution AR model to SAR target discrimination. High resolution SAR data is organized into a quadtree structure by downsampling and applying a log-detection transform. The data is modeled using a white noise driven AR model in scale where the value at each node depends on the direct ancestors at coarser scales. Training data is used to obtain the AR coefficients and a model for the residual noise for the two classes to be discriminated, man made objects and natural clutter. The AR coefficients are determined using a least squares approach. The residuals are modeled as Gaussians for the man-made case and as Rayleigh for the natural clutter. Classification is performed using a hypothesis test based on a multiresolution discriminant. The discriminant is computed as the log ratio of the conditional pdf's of the residuals given the two coarsest scales. [discr, oct 97]

In [143], Frakt, Karl and Willsky propose a computationally efficient multiscale algorithm for hypothesis testing in tomography applications. The proposed algorithm proceeds from coarse to fine scale. At each scale, a composite hypothesis test is performed and all but one composite hypothesis are discarded. The remaining hypothesis is split into four smaller composite hypotheses at the next finer scale. For the tomography application, the hypothesis at each coarse scale location is the presence of an anomaly which is contained entirely within a region corresponding to the coarse scale location. The regions corresponding to adjacent coarse scale locations overlap. The first hypothesis test at the coarsest scale identifies a single region. At the next scale, this region is split into four and the best region is selected. The decision statistics are computed as affine transforms of the tomographic projection data. The coefficients of the affine transform are optimized to maximize the worst case difference between the likelihood means for the cases where the hypothesis is true and false. The projection data is modeled as jointly Gaussian. Anomalies are modeled as square patches that are independent from the background which is modeled as a Gaussian MRF. The proposed approach is compared to a simpler approach that uses hypotheses corresponding to anomalies that take up the entire region corresponding to a coarse scale location. [disc, June 98]

In [144], Hsu and Wilson propose a multiresolution model for texture analysis and synthesis. Textures are modeled as transformations of a prototype patch plus a stochastic residual component such as an MRF. The prototype transformation consists of a translation followed by an affine coordinate transform and spatial windowing. The selection of an appropriate analysis scale is formulated as a rate-distortion problem. The rate is computed as the expected information rate of the set of prototype transforms and the distortion is computed as the expected mean square error between transformed prototype and ground truth. The authors propose a texture analysis by synthesis algorithm based on a multiresolution Fourier transform (MFT) representation of the texture patches. [disc, Oct 1998]

In [145], Li and Wilson develop a spatially non-homogeneous multiscale MRF for texture segmentation. The model is based on both region and boundary processes which contribute to the MRF potential functions at each scale. The MRF neighborhood consists of the 4-point neighborhood at the same scale as well as the quadtree parent at the next coarser scale. The coarse to fine scale unsupervised segmentation algorithm simultaneously optimizes both boundary and region processes. At each scale, a Gibbs sampling scheme is used to identify the boundary sites. The segmentation result at the current scale is then used as coarse scale condition and initialization for reconstruction at the next finer scale. [discr, Oct 1998]

In [29], Comer and Delp propose a multiresolution AR approach to texture segmentation. The observed data is modeled using a multiresolution AR process on a Gaussian pyramid. The observation at each node is modeled as an AR process using arbitrary coefficients at coarse scale and the causal coefficients at the same scale. The label pyramid is modeled using a 3-D MRF model (Kato, Zerubia, Berthod). The multiscale MPM reconstruction and parameter estimation are performed simultaneously using an EM/Gibbs sampler like approach. [discr, March 99]

In [30], Laferte, Perez and Heitz develop non-iterative inference algorithms for a discrete multiscale model with a causal Markovian structure. The quadtree model assumes that nodes are conditionally independent given the quadtree parent. Observations are assumed to be independent and to be available at all resolutions. Three recursive estimation algorithms are derived, a Viterbi-like algorithm for exact MAP estimation, a Welch-like algorithm for MPM estimation and an extension of SMAP for the case where observations are available at all nodes. Furthermore, two different EM techniques for parameter estimation are developed. Results for supervised and unsupervised image classification are presented. [discr, nonlin, 00]

A.6.3 Hybrid and Nonlinear Continuous State Models

In [24], Yang and Wilson propose an adaptive multiresolution method for adaptive image restoration. The prior model uses a quadratic smoothness constraint on the difference between the segmentation and the interpolated segmentation at the previous coarser scale. The segmentation is computed using Hopfield neural networks to solve the optimization. The optimization is performed from coarse to fine scale. The segmentation at each scale is computed separately conditioned on the previous coarser scale solution. The model is extended to be adaptive by relaxing the prior term in the vicinity of edges. [cont,non-lin, jul 95]

In [16], Nowak provides a multiscale hidden Markov model (MHMM) approach for Bayesian image analysis assuming Gaussian or Poisson noise processes. The model is based on the Haar wavelet decomposition of both the signal (mean) and the observations. The fine scale observations are assumed to be spatially independent. The development is based on examining the conditional pdf of a node observation given the observation at the parent node and the signal at all nodes, all in terms of the scaling coefficients. For the Gaussian case, this conditional node likelihood depends on the Haar wavelet coefficient of the signal. For the Poisson case, it is a function of the ratio of the signal's scaling coefficients of child and parent node. The signal dependencies, i.e. the Haar coefficient for the Gaussian case and the ratio of child/parent scaling coefficients are defined as canonical parameters. The approach focuses on modeling these canonical parameters. One advantage of this representation is that it leads to a simple factorization of the overall likelihood which in turn leads to elegant recursive estimation algorithms. Two classes of prior models are proposed. First, independent parameter models are formulated based on mixture modeling of the canonical parameters. Specifically, the canonical parameters at each node are modeled independently using distributions that are conjugate to the conditional likelihood, i.e. mixtures of Gaussian densities for the Gaussian case and mixtures of Beta densities for the Poisson case. Second, a multiscale HMM (MHMM) model with a causal structure in scale is developed. Each node is assigned a hidden state such that given the hidden state, the canonical (prior) parameters for different nodes are independent. The state transition probabilities are specified for parent to child state transitions. An upward/downward algorithm is used to calculate the posterior state probabilities. These are then used to recursively compute the posterior density as in the independent parameter case. Applications to Bayesian image denoising and edge detection are examined. For the Poisson case, the model is limited to the Haar wavelet basis. [cont,nlin, 98]

In [17], Crouse, Nowak and Baraniuk develop multiscale HMM models using mixtures of Gaussians to model wavelet coefficients. The model is motivated by the non-ideal decorrelation properties of the wavelet transform, specifically clustering of small wavelet coefficients and persistence of large and small coefficient values across scales. Wavelet coefficients are modeled as mixtures of Gaussians. Each node is assigned a hidden state such that the state probabilities correspond to the mixture component probabilities. Three models are considered, an independent parameter model (IM) with independent node states, a Markov chain model where states are connected within scale (1-D) and a hidden Markov tree (HMT) model where the state variables are connected across scale. The authors point out, that for the HMT model, Markovianity is assumed only for the states but not directly imposed on the wavelet coefficients. The parameters of the HMM models are obtained as the ML estimates and are computed using the EM algorithm. In order to increase the robustness of the model, tying between tree nodes is performed, i.e. wavelet coefficients at different locations are tied by assuming the same distribution. Applications to signal estimation as well as signal detection and classification are discussed. [cont, nonlin, May 98]

In [36], Nowak extends the independent parameter model using Gaussian mixtures presented in [17] to average over circular shifted versions of the wavelet transform in order to reduce blockiness. The resulting model is shown to be stationary with respect to the fine scale signal. The autocorrelation is piecewise linear with a power spectrum that decays with approximately 1/f. In contrast, the model using only a single shift of the wavelet transform results is non-stationary with piecewise constant correlation. The proposed model is applied to denoise a 1/f Gaussian process. [cont, 98]

In [25], Nowak and Kolaczyk propose a multiresolution MAP estimation algorithm for Poisson inverse problems. The approach is based on the independent parameter multiresolution prior model earlier proposed in [16]. Splitting coefficients consisting of parent-child ratios of Haar scaling coefficients are modeled as mixtures of Beta densities. The MAP reconstruction is computed using the EM algorithm. For a prior model consisting of a two component mixture of beta densities where one component is a point mass, the EM updates are simple analytical expressions.

To formulate the algorithm, the likelihood and prior expressions are formulated in terms of the unobservable data, i.e. the emissions from each image pixel. The prior model is derived by expressing the conditional parent-child relationships of the unobservable data. These distributions depend on the splitting factors of the emission means. These splitting factors are modeled as mixtures of Beta distributions. The MAP reconstruction is computed using the EM algorithm. The E-step calculates the expectation of the unobservable data given the observed projection data and a current estimate of the image. The M-step maximizes the conditional pdf of the splitting factors given the unobservable data. [cont, nlin, Nov 98]

In [146], Nowak and Baraniuk develop a wavelet shrinkage approach for Poisson noise. For Poisson measurements, the noise variance of the wavelet coefficients is signal dependent. The authors propose a wavelet filter that adapts to the signal to noise ratio of each coefficient. The filter is designed by cross-validation using multiple time frames of photon data. More specifically, the signal resulting from summing all but one data frames is filtered, the result is compared to the remaining data frame and the filter is optimized with respect to a quadratic cost function called predictive sum of squares (PRESS). [cont, nlin, May 99]

A.6.4 Specialized Applications

Image Coding

In [147], Kim, Krim and Willsky use a multiresolution AR model to develop a segmentation driven compression algorithm for SAR images. The model is based on a quadtree multiresolution pyramid where each level contains a log-magnitude image representation of the observed image. The multiresolution AR model linearly predicts the image value at each node based on its coarse scale ancestors in the quadtree. The AR coefficients for each node are determined to provide the best least-squares fit to the measured data for a window centered around the node. Segmentation is performed based on the AR coefficients for each node. Specifically, the conditional likelihood of the AR coefficient vector given the node's class is modeled as multivariate Gaussian. This model is then used to perform ML classification of the AR coefficients. After segmenting the image, image coding is performed by predicting each node's value from its coarse scale ancestors using the AR coefficients for the node's class. The prediction errors are further compressed using soft-thresholding and a wavelet packet basis for transmission of each resolution level of the error image. [seg/comp, Oct 97]

In [148], LoPresto, Ramchandran and Orchard propose an image coding technique based on a hierarchical mixture modeling of wavelet coefficients. The wavelet coefficients are modeled as being drawn from an independent Generalized Gaussian distribution field with spatially slowly varying variances. The model variance for each wavelet coefficient is computed as the ML estimate given a neighborhood of quantized wavelet coefficients of the same orientation. Specifically, the neighborhood can include coefficients at the next coarser scale and the causal coefficients at the same scale. The image coding is then performed using rate-distortion entropy/quantization coding. The coder for each wavelet coefficient is selected from a lookup table based on the estimated variance of the wavelet coefficient. The coders are optimized off-line for each variance level. [compr, wav, 97]

In [149], Simoncelli and Buccigrossi use a hierarchical wavelet model for image compression. It is demonstrated, that while neighboring wavelet coefficients are typically nearly uncorrelated, their magnitudes can be strongly correlated. Based on this result, the authors propose a coding strategy based on linear prediction of wavelet coefficients from coefficients that are adjacent in scale, space and orientation. The neighborhood of wavelet coefficients is selected off-line, using a greedy algorithm to maximize compression. [compr,lin-mod, Oct. 97]

A.6.5 Other

In [150], Graffigne, Heitz, Perez, Preteux, Sigelle and Zerubia provide a review and a classification scheme for hierarchical MRF models. Models are classified into explicit hierarchical based and induced hierarchical based. In explicit hierarchical models the hierarchy is integrated in the model definition, examples are quadtree-models with parent-child interaction or families of stochastic processes defined on various sets of sites. In induced hierarchical models such as renormalization techniques and sequences of constrained configuration spaces, the hierarchy is implied by a transformation of the original MRF. The two main categories, explicit and induced, are divided into subclasses based on the nature of the space on which the order relationship operates. A review of the existing work in each category is presented. Several models are compared experimentally based on a binary image segmentation task. [other, 95]

In [151], Lucke investigates the graph structures necessary to allow for Baum-Welch training and Viterbi algorithm optimization for HMM's. The author proves that an algorithm similar to Baum's exists whenever the graph structure is chordal, i.e. all cycles of length greater than 3 have an edge that joins two non-consecutive vertices of that cycle. [other, 96]

A.7 Proof of Theorem

First note that $X_s^{(n)}$ is a circularly stationary random process for each resolution $0 \leq n \leq L$. This results from the assumption that $X_s^{(0)}$ is circularly stationary together with the form of the recursions (2.1) and (2.2). Next, notice that

$$\begin{bmatrix} X_{\partial_s}^{(n+1)} \\ Z_s^{(n+1)} \end{bmatrix} = Q \begin{bmatrix} X^{(n)}_{(-p+2(s-w))\mathrm{mod}2^N} \\ X^{(n)}_{(-p+2(s-w)+1)\mathrm{mod}2^N} \\ \vdots \\ X^{(n)}_{(p+2(s+w))\mathrm{mod}2^N} \end{bmatrix}$$

where Q is a fixed matrix, where p is such that the support of the kernels h and g in (2.1) and (2.2) is within $[-p, p]$ and the centered window ∂s is of length $2w + 1$. Based on the equation above and the fact that $X_s^{(n)}$ is a circularly stationary process, the distribution of $[X_{\partial_s}^{(n+1)}, Z_s^{(n+1)}]$ does not depend on s. Using the assumption that $\nu^* = \arg\max_\nu E[\log p_z(Z)]$ exists and is unique, we write

$$\begin{aligned}
\max_\nu E[\log p_z(Z)] &= \max_\nu E\left[\sum_{(s,n)\in S} \log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(Z_s^{(n)}|X_{\partial s}^{(n)}) + \mathrm{const}\right] \\
&= \max_\nu \sum_{(s,n)\in S} E\left[\log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(Z_s^{(n)}|X_{\partial s}^{(n)})\right] + \mathrm{const} \\
&= \sum_{(s,n)\in S} \max_{\nu_s^{(n)}} E\left[\log p_{z_s^{(n)}|x_{\partial s}^{(n)}}(Z_s^{(n)}|X_{\partial s}^{(n)})\right] + \mathrm{const} \\
&= \sum_{(s,n)\in S} \max_{\nu_s^{(n)}} E[f_n(Z_s^{(n)}, X_{\partial s}^{(n)}, \nu_s^{(n)})] + \mathrm{const}\ .
\end{aligned}$$

Since neither the f_n nor the distribution of $[X_{\partial_s}^{(n+1)}, Z_s^{(n+1)}]$ depend on s,

$$(\nu_s^{(n)})^* = \arg\max_{\nu_s^{(n)}} E[f_n(Z_s^{(n)}, X_{\partial s}^{(n)}, \nu_s^{(n)})]$$

is not a function of s. This proves the theorem.

A.8 Computational Complexity

The number of multiplications for optimizing the wavelet graph model with respect to a single scaling coefficient $x_i^{(l)}$ is proportional to the size of the set $\tilde{S}_i \triangleq \bigcup_{n>l} \tilde{S}_i^{(n)}$, where the $\tilde{S}_i^{(n)}$ are as defined in (2.40). For a general wavelet transform, the size $|S_i^{(n)}|$ of the sets $S_i^{(n)}$ as defined in (2.39) is proportional to $n - l$.

Define $S_i \triangleq \bigcup_{n>l} S_i^{(n)}$, then

$$|S_i| = \sum_{n=l+1}^{L} |S_i^{(n)}| \;\; \propto \;\; \sum_{n=l+1}^{L} (n-l) \tag{A.7}$$

$$\propto \;\; (L-l)^2 \leq (\log N^{(l)})^2 \; . \tag{A.8}$$

Based on (2.41), the size of the set $\tilde{S}_i$ is upper bounded by $|S_i||\partial s|$ where $|\partial s|$ denotes the fixed size of the prediction window. Therefore, $|\tilde{S}_i| \propto (\log N^{(l)})^2$ and consequently steps (2.42)-(2.45) can be executed in order $(\log N^{(l)})^2$ multiplications. Thus, the complexity for a full update of $\hat{x}^{(l)}$ is order $N^{(l)}(\log N^{(l)})^2$. For a Haar wavelet decomposition, this reduces to $N^{(l)} \log N^{(l)}$ since $S_i^{(n)}$ contains only a single node at each scale n. The complexities for the 1-D and the 2-D case are the same.

A.9 Tree-structured Nonlinear Classifier

In this section, we describe the agglomerative clustering method used to obtain the classifiers $\mathcal{K}^{(n)}[\cdot]$ and the parameter vectors $\theta_k^{(n)}$ for class k. For the classifier at scale n, we assume a training set $\{x_{\partial s}^{(n)}/\sigma_n, \; z_s^{(n)}/\sigma_n\}$. We then define a normalized training set $\{v_s, w_s\}$ with samples $v_s = V(x_{\partial s}^{(n)} - x_s^{(n)})/\sigma_n$ and $w_s = z_s^{(n)}/\sigma_n$ where V is the matrix that eliminates the zero center component of $x_{\partial s}^{(n)} - x_s^{(n)}$ and is the identity otherwise. Thus, v_s has one fewer component than $x_{\partial s}^{(n)}$. Since the training is performed separately for each scale n, we simplify the notation by omitting the dependence on n. Our objective is to form a classification tree for v_s such that each tree node is associated with a MMSE linear predictor to predict w_s from v_s. We first perform a vector quantization (VQ) of $\{v_s\}$ with a pre-specified, fixed number of clusters K, currently $K = 150, 100, 100, 50, 50$ for scales $n = 0$ through $n = 4$. The distance metric for the VQ is the Euclidean distance. The number of iterations M for the VQ is set to a constant value, currently $M = 5$. The next step is to compute the MMSE linear predictors for the VQ clusters. Let $c_1, \ldots, c_K$ denote the K clusters. Define $\mu_{c_k,v}$ and $\mu_{c_k,w}$ as the means and $C_{c_k,vv}$ and $C_{c_k,ww}$ as the covariance matrices of the samples in cluster c_k with respect to v and w. Let $C_{c_k,vw}$ denote the cross-covariance of the samples in cluster c_k. We then compute a MMSE linear predictor $[A_{c_k}, b_{c_k}]$ for cluster c_k as

$$\begin{aligned} \tilde{A}_{c_k} &= C^t_{c_k,vw} C^{-1}_{c_k,vv} && \text{(A.9)} \\ b_{c_k} &= \mu_{c_k,w} - \tilde{A}_{c_k} \mu_{c_k,v} && \text{(A.10)} \\ A_{c_k} &= \tilde{A}_{c_k} V \ . && \text{(A.11)} \end{aligned}$$

Further, the total prediction error ε_{c_k} of cluster c_k over the training set is obtained as

$$\varepsilon_{c_k} = \mathrm{trace}(C_{c_k,ww} - 2\tilde{A}_{c_k} C_{c_k,vw} + \tilde{A}_{c_k} C_{c_k,vv} \tilde{A}^t_{c_k}) \ . \quad \text{(A.12)}$$

Our objective is now to successively merge pairs of clusters in a greedy fashion to obtain a binary tree. Consider merging two clusters c_k and c_l into a new cluster $c_m = c_k \cup c_l$. Based on the statistics of the merged cluster c_m, we can compute the linear predictor $[A_{c_m}, b_{c_m}]$ and prediction error ε_{c_m} using (A.9)-(A.12). The increase in total prediction error on the training set due to the merging is then $\Delta\varepsilon_{c_k,c_l} = \varepsilon_{c_m} - (\varepsilon_{c_k} + \varepsilon_{c_l})$. Merging clusters strictly based on minimum increase in prediction error is not very sensible for small clusters whose linear predictor may be over-parameterized such that $\varepsilon_{c_k} = 0$. To merge small clusters in a meaningful way, we introduce a regularization term based on the distance between clusters c_k, c_l and the merged cluster c_m

$$\rho_{c_k,c_l} = \alpha(|c_k||\mu_{c_k,v} - \mu_{c_m,v}|^2 + |c_l||\mu_{c_l,v} - \mu_{c_m,v}|^2) \quad \text{(A.13)}$$

where α is a small regularization constant and $|c_k|$ denotes the number of samples in cluster k. We then define the cost function M_{c_k,c_l} for the merging of c_k and c_l as

$$M_{c_k,c_l} = \Delta\varepsilon_{c_k,c_l} + \rho_{c_k,c_l} \ . \quad \text{(A.14)}$$

Starting with the initial VQ partitioning, we now merge the clusters into a binary tree. At any given stage, we combine the two clusters c_k, c_l whose merging results in the smallest M_{c_k,c_l}. This results in a binary tree where each node is associated with its optimal linear prediction filter for w_s. The leaves of the tree are the VQ clusters. In order to not overfit the classification model, we perform optimal tree pruning[43, 44] using a second data set for cross-validation. The pruning set $\{\tilde{v}_s, \tilde{w}_s\}$ is classified into the tree by assigning each data sample to the closest VQ cluster and to all of its parents in the tree. The pruning is performed based on the prediction error only. The pruned nodes are marked as such but are not removed from the tree.

The covariance parameter R_k for each tree node is computed as a linear combination of the conditional covariance of the pruning samples in class k and the expected conditional covariance over the entire pruning set. Let $R_{1,k}$ be the conditional covariance matrix of the pruning samples in class k

$$R_{1,k} = \frac{1}{|c_k|} \sum_{\{s:\tilde{v}_s \in c_k\}} (\tilde{w}_s - A_{c_k}\tilde{v}_s - b_{c_k})(\tilde{w}_s - A_{c_k}\tilde{v}_s - b_{c_k})^t \tag{A.15}$$

where $|c_k|$ is the number of pruning samples that fall into node k. Further, let use define R_2 as the expected conditional covariance over the entire pruning set $R_2 = \frac{1}{N_p} \sum_{\{k:c_k \in \mathcal{L}\}} |c_k| R_{1,k}$ where $\mathcal{L}$ is the set of tree leaves after discarding the pruned nodes and N_p is the total number of samples in the pruning set. We then compute the covariance parameter R_k for node k as

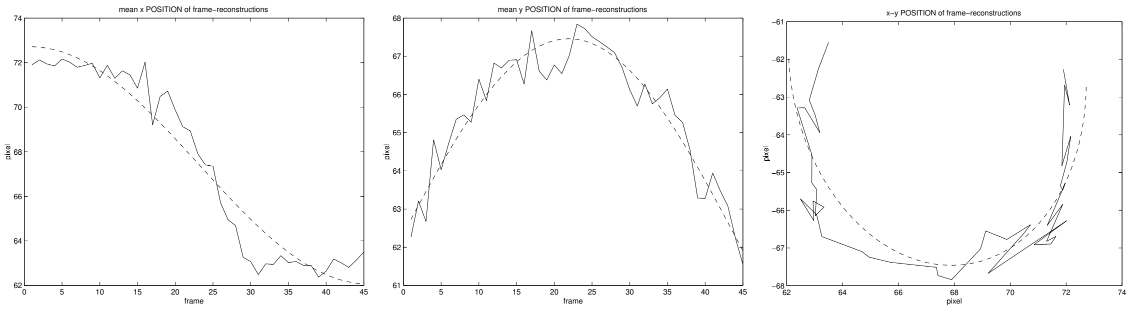
$$R_k = (1 - \delta)R_{1,k} + \delta R_2 \tag{A.16}$$

where δ is a small constant. The term δR_2 is added to impose a lower limit on the R_k. For our experimentation, we use a fixed value $\delta = 0.001$ except for one case where the training set of discrete images requires a larger value of $\delta = 0.3$.

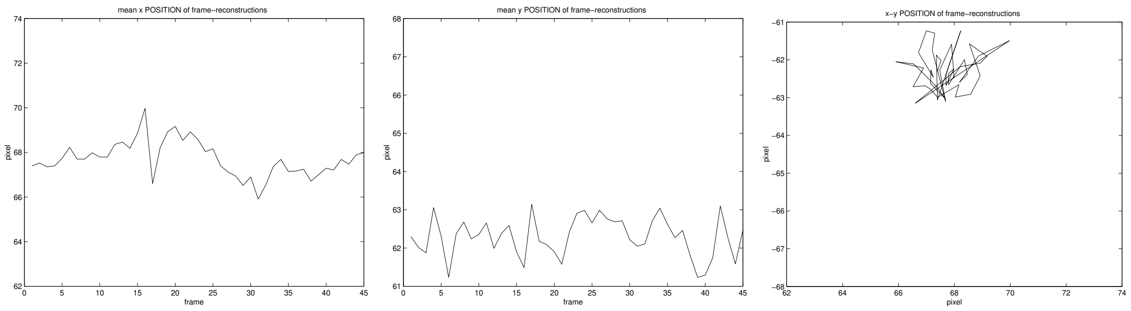
In order to perform the classification (2.46) at runtime, we first find the VQ cluster with minimum Euclidean distance to $V(\hat{x}_{\partial s}^{(n)} - \hat{x}_s^{(n)})/\sigma_n$. We then follow the tree upwards until we reach the first node that is not marked pruned. This node corresponds to the class $k_{n,s}$ and contains the associated parameter vector $\theta_{k_{n,s}}^{(n)}$.

A.10 Precise Estimation of Line Source Position

An accurate estimate of the initial line source position s_0 is obtained as follows. We first obtain an approximate estimate $\hat{s}_0$ based on the projection data. Specifically, since the data is taken at the zero gantry rotation, we have good horizontal resolution allowing us to estimate the horizontal position quite accurately from the sinogram



(a) before calibration



(b) after calibration

Fig. A.10. Circular blurring effect before and after calibration. Plotted are the reconstructed coordinates of a point source for reconstructions of single frames at different gantry angles. Shown in (a) is a result for an initial, inaccurate estimate of s_0. The dashed lines indicate the fitted circle. Shown in (b) is the result after correcting the estimate of the initial position of the line source based on the fitted circle in (a).

data. To obtain an estimate for the vertical position, we calculate the average vertical position from the frame image reconstructions for each source position. Based on the initial estimate $\hat{s}_0$, we compute a first empirical system kernel $h(\theta, \Delta t)$. We then use this kernel to reconstruct a second data set of a point source taken for a full set of F frames with gantry angles $\alpha_f = f * 180/F$, $0 \leq f \leq F-1$. We separately compute reconstruction images x_f for each frame using the model $h(\theta - \alpha_f, \Delta t)$ and then estimate the reconstructed source position in each frame as the center of gravity in x_f computed over a window around the expected position. Let $\xi_f = (\xi_{f,1}, \xi_{f,2})$,

$0 \leq f \leq F-1$ denote the sequence of estimated source positions which should lie on a circle.

To estimate the radius and phase of the circle, we fit two sine waves $\bar{\xi}_{f,1}$ and $\bar{\xi}_{f,2}$ of equal amplitude and different phase to the sequences $\xi_{f,1}$ and $\xi_{f,2}$

$$\bar{\xi}_{f,1} = \alpha \sin(f\pi/F + \phi_1) + \beta_1 \tag{A.17}$$

$$\bar{\xi}_{f,2} = \alpha \sin(f\pi/F + \phi_2) + \beta_2 \tag{A.18}$$

where α, ϕ_1, ϕ_2, β_1 and β_2 are determined to obtain a least squares fit to $\xi_{f,1}$ and $\xi_{f,2}$. Figure A.10(a) shows an example of the sequences $\xi_{f,1}$ and $\xi_{f,2}$ and the corresponding fits. Based on the fit, we directly obtain the offset ϵ_s as the first point on sine waves

$$\epsilon_{s,1} = \bar{\xi}_{0,1} = \alpha \sin(\phi_1) \tag{A.19}$$

$$\epsilon_{s,2} = \bar{\xi}_{0,2} = \alpha \sin(\phi_2) \tag{A.20}$$

Thus, we obtain a corrected estimate $\hat{s}'_0$ as

$$\hat{s}'_0 = \hat{s}_0 - \epsilon_s. \tag{A.21}$$

This more accurate estimate is then be used to re-compute a more precise $h(\theta, \Delta_t)$. The new system kernel $h(\theta, \Delta_t)$ is then used to repeat the frame reconstructions of the second data set. By estimating the new radius and phase of the resulting circle we can evaluate the accuracy of the model. If necessary, the calibration procedure can be repeated iteratively, until the estimated radius of the circle is below a specified threshold. Figure A.10(b) shows the position estimates as a function of frame after calibration initialized with the offset shown in Fig. A.10(a).

VITA

VITA

Thomas Frese was born on December 29, 1970 in Herdecke, Germany. He grew up in Witten, Germany, where he graduated from high school in 1990. In 1991, he enrolled in the Electrical Engineering program at Ruhr-Universität Bochum, Germany, where he received the degree of Diplom-Ingenieur in 1996. From 1994 through 1995 he was an exchange student in the Electrical and Computer Engineering program at Purdue University, West Lafayette, Indiana. In 1996, he joined the Ph.D. program in Electrical and Computer Engineering at Purdue University.