

# Automated Estimation of the Parameters of Gibbs Priors to be Used in Binary Tomography

Hstau Y. Liao<sup>1,2</sup> and Gabor T. Herman<sup>3</sup>

*Center for Computer Science and Applied Mathematics  
Temple University  
Philadelphia, USA*

---

## Abstract

Image modeling using Gibbs priors was shown to be effective in image reconstruction problems. This motivated us to evaluate three techniques for estimating the priors: the heuristic method, the histogram method and Borges' method. We found that both the histogram and Borges' method accurately recovered the parameters needed to specify four different Gibbs distributions from training sets consisting of random sample images from those distributions. This was not the case for the heuristic method. We evaluated the usefulness of the estimated distributions as priors for binary tomography in two experiments: in one the images for the training set were taken from a Gibbs distribution determined by five parameters, and they were typical cardiac phantom images for the second one. We estimated the parameters using both the heuristic and Borges' method in both experiments. We used a modified Metropolis algorithm for the reconstructions. In the first experiment both estimation methods gave rise to excellent reconstructions, but this was not the case for the second experiment; however, even in this case typically less than four percent of the pixels were mis-classified in the reconstructions.

---

## 1 Introduction

The reconstruction of a binary image from a few projections (usually only two to four) is a problem of solving a very under-determined system of equations and generally results in a large class of solutions. By using appropriate prior

---

<sup>1</sup> This work was supported by the NIH grant HL28438 and the NLM contract on "Visible Human Image Processing Tools", as well as by the National Science Foundation grant DMS9612077.

<sup>2</sup> Email: [liaoh@hotmail.com](mailto:liaoh@hotmail.com)

<sup>3</sup> Email: [gaborherman@netscape.net](mailto:gaborherman@netscape.net)

information, the class of possible solutions can be limited to those which are reasonably typical of the class of images which contains the unknown image that one wishes to reconstruct. *Gibbs priors* describe the local character of an image, which can be viewed as a random sample from a *Gibbs distribution* defined by

$$(1) \quad \pi(f) = \frac{1}{Z} e^{-H(f)},$$

where  $\pi(f)$  is the probability of occurrence of the image  $f$ ,  $Z$  is the normalizing factor, and  $H(f)$  is the *energy* of  $f$ . The rules for calculating  $H(f)$ , for a binary image  $f$ , depend on the values assigned to a finite number (in the examples of our paper, this number is five) of parameters.

For certain types of Gibbs distributions it has been demonstrated that there are algorithms which recover an unknown image which is a typical sample from the distribution when provided with the projections of the image and with the values of the parameters of the Gibbs distribution [7]. A difficulty is that in a practical application the parameter values are not known to us. On the other hand, we usually have access to typical images of the application area. The problem that we address is the estimation of the parameters from such sample images for the purpose of reconstruction.

In the rest of this introductory section we make mathematically precise the ideas just introduced, both for Gibbs distribution parameter estimation and for binary tomography. In Section 2 we point out several considerations that can be made in order to reduce the number of parameters which need to be estimated for Gibbs priors. Section 3 gives some examples of images modeled by Gibbs distributions and their reconstructions from three projections. We describe the *Metropolis algorithm* which is used for generating random samples from a distribution. The Metropolis algorithm can be adapted to find images that approximately satisfy the projections in addition to having relatively high probability of occurrence according to the Gibbs distribution. Section 4 describes three published methods, from [1], [2] and [3], for the estimation of the parameters. Once we have estimated the parameters, in Section 5 we use the squared norm for measuring the success of the estimation process. In Section 6 we evaluate the estimated Gibbs priors from the point of view of their usefulness in binary tomography. Finally, in Section 7 we give the conclusions of the paper.

### 1.1 Parameter estimation for Gibbs priors

Let  $D$  be a fixed non-empty finite set which, for reasons which will become immediately obvious, we call the *domain*. In all our examples  $D$  will be a square subset of the square lattice (i.e.,

$$(2) \quad D = \{(i, j) \in \mathbb{Z}^2 \mid 0 \leq i, j < I\},$$

where  $\mathbb{Z}$  denotes the set of all integers), but the definitions developed here are applicable to an arbitrary  $D$ . A function  $f$  mapping from  $D$  into the set  $\{0, 1\}$

is called a *binary image*. We denote the set of all possible binary images over  $D$  as  $2^D$ . Any non-empty subset of  $D$  is called a *clique*. Given a clique  $q$ , a *configuration* (over  $q$ ) is defined as a function  $g$  mapping from  $q$  into the set  $\{0, 1\}$ . The set of all possible configurations over  $q$  will be denoted as  $2^q$ .

We define a *model* as a pair  $(Q, U)$  in which  $Q$  is a set of cliques and  $U$  is a function mapping the set  $G = \bigcup_{q \in Q} 2^q$  of all possible configurations over all cliques in  $Q$  into the real numbers. We refer to the value  $U(g)$  as the *potential* of the configuration  $g$ . For a model  $\mu = (Q, U)$ , the  $\mu$ -energy of a binary image  $f$  is defined as

$$(3) \quad H_\mu(f) = - \sum_{q \in Q} U[(f|q)],$$

where  $(f|q)$  denotes the restriction of  $f$  to the clique  $q$ ; i.e.,  $(f|q)$  is the configuration  $g$  over  $q$  such that, for all  $d \in q$ ,  $g(d) = f(d)$ . Any model  $\mu$  defines a *Gibbs distribution*  $\pi_\mu$  over  $2^D$  as follows. The probability assigned to a binary image  $f \in 2^D$  is

$$(4) \quad \pi_\mu(f) = \frac{1}{Z} e^{-H_\mu(f)},$$

where  $Z = \sum_{f' \in 2^D} e^{-H_\mu(f')}$ . (In the discussion below, it will usually be understood what  $\mu$  is at any particular stage. In such a case we will use  $H$  and  $\pi$  instead of  $H_\mu$  and  $\pi_\mu$ , respectively.)

We are now in position to state the general aim of this paper. Suppose that in some application of binary tomography we believe that the efficacy of the process can be improved by modeling the distribution of the binary images as they occur in that application by a Gibbs distribution. Let us assume that the size of the images (and, consequently, of  $D$ ) is fixed. We may also assume (for now) that the set  $Q$  of cliques of the model  $(Q, U)$  is also fixed. (For practical reasons one would try to keep the size of  $Q$  small.) In this case the model (and the resulting Gibbs distribution) is uniquely determined by the values of  $U$ . Typically, there are available to us a number of images which are considered representative samples of images in our application area. Our task is to estimate, based on these sample images, the parameters  $U(g)$ , where  $g \in G$ . (This process may fail in the sense that even with “the best” estimate of these parameters, the model  $(Q, U)$  may not be adequate for modeling the sample images. In this case, we may wish to increase the size of  $Q$ .)

To see the usefulness of image modeling by Gibbs distributions in binary tomography, consider the problem of reconstructing a binary image from its horizontal, vertical, and one of the two principal diagonal projections, under the assumption that the image is a sample from a known Gibbs distribution. If we view the image as a matrix of 0 and 1 entries, then the row (column) sum of the matrix corresponds to the horizontal (vertical) projection of the image. The inputs to the reconstruction problem are: the three projections of the image and the parameters that define the Gibbs distribution from which the image is assumed to be a sample.

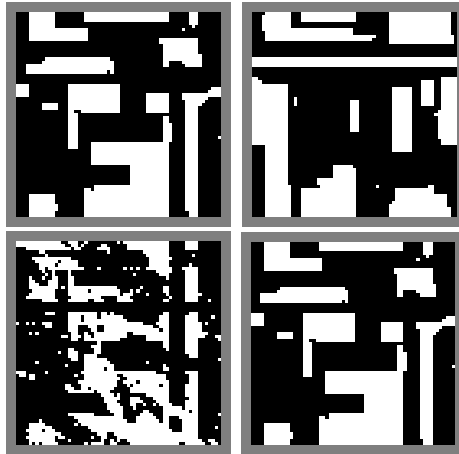


Fig. 1. Binary tomography using a Gibbs prior. The top-left image is a random sample from a particular Gibbs distribution. The top-right image is another random sample image from the same distribution. The bottom-left image is a binary image with exactly the same projections (in the three directions) as the one at the top-left. By combining both the prior information and the projection data into a reconstruction process, we obtained a perfectly reconstructed image shown at the bottom-right. The number of pixels whose color differs from the color of the corresponding pixel in the top-left image is 1,763 for the top-right image, 822 for the bottom-left image, and 0 for the bottom-right image. The images are all of size  $63 \times 63$ .

At the top-left of Figure 1 is a random sample image from a particular Gibbs distribution. At the top-right is another random sample image from the *same* distribution. At the bottom-left is a binary image with exactly the same projections (in the three directions) as the one at the top-left. Clearly these three images are very different from each other, which implies that neither one of the two types of information (the projections and the Gibbs parameters) by itself is adequate for the reconstruction. However, by combining the two types of information into a reconstruction process, we obtained a reconstructed image, shown at the bottom-right, which is exactly the same as the image at the top-left.

This demonstrates the worthwhileness of knowing the parameters which determine the Gibbs distribution from which the unknown image is assumed to be a random sample for the purpose of binary tomography.

## 1.2 Binary tomography on $M$ -grids

Following Kong and Herman [5], we define an  $M$ -grid, where  $M$  is an integer greater than 1, as a pair  $(\Gamma, V)$  where

- (i)  $V$  is an ordered set of  $M$  non-zero vectors  $(v_1, \dots, v_M)$  in a Euclidean space (of any dimension), no two of which are parallel, and
- (ii)  $\Gamma$  is the set of all linear combinations with integer coefficients of the

vectors in  $V$ ; thus

$$(5) \quad \Gamma = \left\{ \sum_{m=1}^M k_m v_m \mid (k_1, \dots, k_M) \in \mathbb{Z}^M \right\}.$$

From now on, the terms *grid* and *M-grid* ( $M \geq 2$ ) will be used interchangeably. The elements of  $V$  are called the *fundamental direction vectors* of the grid  $(\Gamma, V)$ ; these vectors define the “projection directions” that are used with the grid. If  $d$  is any grid point and  $v$  any fundamental direction vector of the grid, then the set of grid points  $l = \{d + kv \mid k \in \mathbb{Z}\}$  is called a *grid line*. All examples in this paper will be using the 3-grid  $(\Gamma, V)$  with  $V = \{(1, 0), (0, 1), (-1, 1)\}$  (so that  $\Gamma = \mathbb{Z}^2$ ).

For a fixed *M-grid*  $(\Gamma, V)$ , let the domain  $D$  be a finite non-empty subset of  $\Gamma$ . Given a binary image  $f$ , let  $L^{(m)}$  denote the set of all grid lines that are parallel to  $v_m$  ( $1 \leq m \leq M$ ) and intersect  $D$  (as a consequence, the number of elements in  $L^{(m)}$  is necessarily finite). The *projection* of  $f$  in the direction  $v_m$  is defined to be the function  $P_f^{(m)} : L^{(m)} \rightarrow \mathbb{N}_0$  ( $\mathbb{N}_0$  is the set of non-negative integers) whose values are the *line sums* along the grid lines  $l \in L^{(m)}$ : i.e.,

$$(6) \quad P_f^{(m)}(l) = \sum_{d \in (l \cap D)} f(d).$$

Two binary images  $f$  and  $f'$  are said to be *tomographically equivalent* if  $P_f^{(m)} = P_{f'}^{(m)}$ , for  $1 \leq m \leq M$ .

A *projection data set* on an *M-grid* is a collection  $(P^{(1)}, \dots, P^{(M)})$  of functions  $P^{(m)} : L^{(m)} \rightarrow \mathbb{R}$  ( $\mathbb{R}$  denotes the set of real numbers), for  $1 \leq m \leq M$ . Intuitively, the reconstruction problem of binary tomography is to find an  $f$  such that  $P_f^{(m)}$  is “consistent with”  $P^{(m)}$ , for  $1 \leq m \leq M$ . Since such a consistency is the same for all tomographically equivalent images, we can appeal to a Gibbs prior (if one is available to us) to specify the sought-after  $f$  more restrictively. In the following we assume that the Gibbs prior is determined by an energy function  $H$  as in (3). We distinguish between two cases.

*Noiseless case:* We know that  $P_f^{(m)}$  is exactly  $P^{(m)}$  ( $1 \leq m \leq M$ ). In this case it is reasonable to select, from the set of all  $f$  for which  $P_f^{(m)} = P^{(m)}$ , one that minimizes  $H(f)$ , since such an  $f$  has maximal prior probability among all the ones which are tomographically equivalent to it. A method for finding such an  $f$  in the case of a 2-grid is described in detail in [5].

*Noisy case:* In an actual application it is more likely that the  $P^{(m)}$  are only approximations (based on some physical measurements) of the  $P_f^{(m)}$ . In this case it is reasonable to look for an  $f$  which minimizes

$$(7) \quad H(f) + \alpha \sum_{m=1}^M \sum_{l \in L^{(m)}} \left| P_f^{(m)}(l) - P^{(m)}(l) \right|,$$

where the size of the parameter  $\alpha$  indicates our trust in the projection data set as compared to the prior.

For either of these approaches we need an  $H$ , which (provided that the set  $Q$  of cliques is fixed) is uniquely determined by the values of the potentials of all the configurations over all the cliques in  $Q$ . For the rest of this paper we will be concerned with the estimation of these potentials based on sample images.

## 2 Reducing the Number of Parameters

As we stated in the previous section, a Gibbs distribution is uniquely determined by a model  $\mu = (Q, U)$ , where the domain of  $U$  is the set  $G = \cup_{q \in Q} 2^q$  of all possible configurations over  $Q$ . In a very general sense, we can consider a partition  $G = \cup_{c=0}^C G_c$  (where  $G_c \cap G_{c'} = \emptyset$  for  $0 \leq c \neq c' \leq C$ ) of  $G$  such that for all  $g_1$  and  $g_2$  in  $G_c$ ,  $U(g_1) = U(g_2) = U_c$  ( $0 \leq c \leq C$ ). Consequently, the number of parameters to be estimated is reduced to  $C + 1$ . However, it is very easy to show that if a model  $\mu'$  is obtained from a model  $\mu$  by replacing  $U_c$  by  $U_c - U_0$  for  $0 \leq c \leq C$  then, for all images  $f$ ,  $\pi_{\mu'}(f) = \pi_{\mu}(f)$ . Hence we may assume, without loss of generality, that  $U_0 = 0$ , and therefore the number of parameters to be estimated is only  $C$ . If we do this, there is an interesting simplification of (3).

Given the domain  $D$ , let  $X$  and  $Y$  be two sets of configurations; i.e., if  $x \in X \cup Y$  then  $x : w \rightarrow \{0, 1\}$ , where  $w$  is a nonempty subset of  $D$ . Let  $N(X, Y)$  denote the *number of times an element of  $X$  appears in  $Y$* ; i.e., the number of all possible pairs  $(x, y)$  such that  $x \in X$ ,  $y \in Y$  and  $y$  restricted to the domain of  $x$  is equal to  $x$  (which implies, in particular, that the domain of  $x$  is a subset of the domain of  $y$ ). We note that with this definition  $N(X, Y) = \sum_{x \in X, y \in Y} N(\{x\}, \{y\})$ , where  $N(\{x\}, \{y\})$  is either one (if  $y$  restricted to the domain of  $x$  is equal to  $x$ ) or zero (otherwise). Then it follows that (3) can be re-written as

$$(8) \quad H(f) = - \sum_{c=1}^C N(G_c, \{f\}) U_c.$$

In practice it is desirable to deal with a few parameters rather than with many of them. We now discuss a general approach to producing reasonable partitions of  $G$ . Let  $T$  be a (necessarily finite) set of one-to-one mappings of an element of  $Q$  onto an element of  $Q$ . A configuration  $g$  over a clique  $q$  is said to be *T-equivalent* to a configuration  $g'$  over a clique  $q'$  if there exists a (possibly empty) sequence  $t_1, \dots, t_S$  of mappings such that

- (i) for  $1 \leq s \leq S$ ,  $t_s \in T$  or  $t_s^{-1} \in T$ ;
- (ii)  $q' = \{t_S \cdots t_1(d) \mid d \in q\}$ ; and
- (iii) for all  $d \in q$ ,  $g'(t_S \cdots t_1(d)) = g(d)$ .

Clearly,  $T$ -equivalence is an equivalence relation on  $G$ . The partitions that we use to reduce the number of parameters are based on the equivalence classes of such  $T$ -equivalences.

As an example, consider the domain  $D$  of (2) with the set of cliques  $Q$  defined as follows. For  $0 \leq i, j < I$ , define the clique

$$(9) \quad q_{(i,j)} = \{(i \oplus \epsilon_i, j \oplus \epsilon_j) \mid \epsilon_i, \epsilon_j \in \{-1, 0, 1\}\},$$

where  $\oplus$  denote addition in  $\mathbb{Z}_I$ ; i.e., addition modulo  $I$ . Let

$$(10) \quad Q = \{q_{(i,j)} \mid 0 \leq i, j < I\}.$$

As examples of one-to-one mappings, consider two mappings of  $q_{(1,1)}$  onto itself: one is a rotation  $\rho$  defined by  $\rho((i', j')) = (j', 2 - i')$ , for  $(i', j') \in q_{(1,1)}$ , and the other is a reflection  $\psi$  defined by  $\psi(i', j') = (2 - i', j')$ , for  $(i', j') \in q_{(1,1)}$ . For every  $(i, j) \in D$  we also define two one-to-one mappings  $\tau_{(i,j)}^h$  and  $\tau_{(i,j)}^v$  of  $q_{(i,j)}$  onto  $q_{(i \oplus 1, j)}$  and onto  $q_{(i, j \oplus 1)}$ , respectively, by  $\tau_{(i,j)}^h(i', j') = (i' \oplus 1, j')$ , for all  $(i', j') \in q_{(i,j)}$  and  $\tau_{(i,j)}^v(i', j') = (i', j' \oplus 1)$ , for all  $(i', j') \in q_{(i,j)}$ . The mappings  $\tau_{(i,j)}^h$  and  $\tau_{(i,j)}^v$  are, respectively, a horizontal translation and a vertical translation. If we now let  $T$  contain  $\rho$ ,  $\psi$ , and  $\tau_{(i,j)}^h$  and  $\tau_{(i,j)}^v$  for all  $(i, j) \in D$ , then we find that any configuration defined on a  $3 \times 3$  clique is  $T$ -equivalent to any other configuration which can be obtained from it by translations, rotations around the central pixel, and reflections in either horizontal or vertical central axis. One can reasonably argue that in many applications such  $T$ -equivalent configurations should be assigned the same potential.

We can go further with this idea. Some of these  $3 \times 3$  configurations may be considered to contain a particular local feature of one of the following types: a black region, a white region, a convex corner, a concave corner, or an edge. In Figure 2 we give examples of such special configurations. In fact, this set of examples is complete: e.g., a  $3 \times 3$  configuration in an image is a convex corner if, and only if, it is  $T$ -equivalent to one of the configurations identified as a convex corner in Figure 2 (where, for completeness, we may assume that the configurations are over the clique  $q_{(1,1)}$ ). The working definition of these five types of configurations is as follows. A configuration is one of these types only if there are  $k$  consecutive white pixels and  $8-k$  consecutive black pixels among the eight external pixels of the clique. If the central pixel is black, then the configuration is a black region, a convex corner, an edge, or a concave corner respectively, if, and only if,  $k = 0$ ,  $1 \leq k \leq 2$ ,  $k = 3$ , or  $4 \leq k \leq 5$ , respectively (see the upper row of Figure 2). By switching the color at each grid point of these local features, we will obtain the corresponding “opposite” local features (see the bottom row of the same figure). The opposite of a black region is a white region; the opposite of a convex corner is a concave corner (and vice-versa); the opposite of an edge is still an edge. Here and in the rest of this paper, white (black) pixels will represent grid points with value 1 (0).

A Gibbs distribution based on the five local features of this example can be defined by a model in which  $G = \cup_{c=0}^5 G_c$ , where  $G_0$  is the set of all configurations that do not contain to any of the above specified five local features,  $G_1$  contains the black regions,  $G_2$  the white regions,  $G_3$  the edges,  $G_4$  the convex corners, and  $G_5$  the concave corners. Unless otherwise stated,

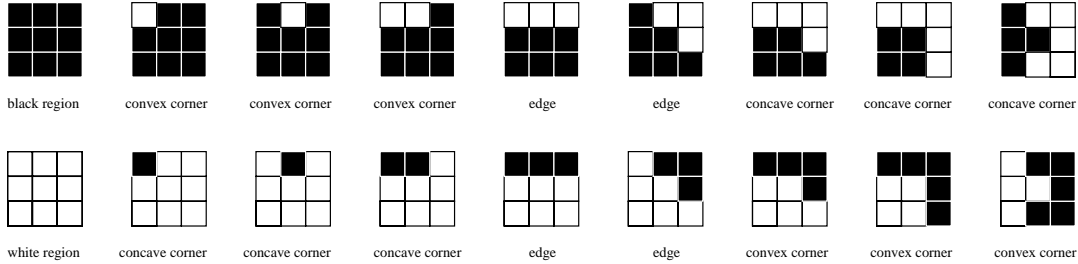


Fig. 2. Configurations of a  $3 \times 3$  clique that specifies the local features referred to as: a black region, a convex corner, a concave corner, an edge, and a white region.

all the models with which we work in the rest of the paper will be of this type. Their complete specification requires the five numbers  $U_1$ ,  $U_2$ ,  $U_3$ ,  $U_4$ , and  $U_5$  (recall that  $U_0 = 0$ ).

We choose this model with the five local features because it is simple and yet allows us to represent the type of images which have relatively large uniform regions. Later, in Section 6, we make reconstructions of cardiac cross-sectional images which are of this type.

### 3 Image Modeling and Reconstruction

A sample image from a given Gibbs distribution can be generated using a sufficiently long run of the *Metropolis algorithm* [8]. In fact, that algorithm can be used to generate samples from a distribution  $\gamma$  defined on binary images  $f$ , provided that  $\gamma(f) > 0$ , for all  $f \in 2^D$ . The algorithm starts with an arbitrary binary image, and in each step a point  $d$  of the current image  $f$  is randomly selected and its color is inverted, resulting in the image  $\tilde{f}$ . Then  $f$  is replaced by  $\tilde{f}$  with probability  $\min \left\{ 1, \gamma(\tilde{f})/\gamma(f) \right\}$ . To obtain a typical sample from a Gibbs distribution  $\pi$ , we simply set  $\gamma$  to  $\pi$ . For reconstruction, as we will see later, the same approach can be used by defining a distribution which takes into account the prior and the projection data.

#### 3.1 Modeling using Gibbs distributions

In this subsection we show samples from Gibbs distributions defined by the model described in the previous section. Specifically, we will consider various choices for the five potentials corresponding to the five types of configurations:  $U_1$  for black regions,  $U_2$  for white regions,  $U_3$  for edges,  $U_4$  for convex corners,  $U_5$  for concave corners.

We adopt the convention of specifying a particular Gibbs distribution of this type by  $(U_1, U_2, U_3, U_4, U_5)$ . Four random samples of size  $63 \times 63$  from different Gibbs distribution are shown in Figure 3. We can see for example that a higher  $U_3$ , which controls the “edginess”, gives rise to an “edgy” image (bottom-left); while a typical image from a distribution with higher  $U_4$  (for convex corners) has numerous small objects (top-right). Generation of



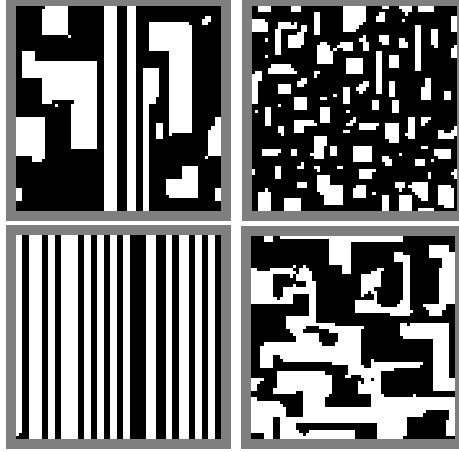


Fig. 3. Sample images ( $63 \times 63$ ) from various Gibbs distributions using parameters (1.2, 1.2, 1.2, 0.52, 0.2) (top-left), (1.2, 1.2, 1.2, 0.9, 0.2) (top-right), (1.2, 1.2, 1.4, 0.52, 0.2) (bottom-left), and (1.2, 1.2, 1.2, 0.52, 0.6) (bottom-right). Note that the difference between the top-left and any other image is caused by a change in only one parameter. With respect to the top-left image the top-right image has a higher  $U_4$  which results in more numerous but smaller objects; the bottom-left image is much more “edgy” since its  $U_3$  is higher; and in the bottom-right image we can see more concavities due to the higher value of  $U_5$ .

such sample images using the Metropolis algorithm is computationally feasible because we do not ever have to compute a probability  $\pi(f)$  (which would necessitate the impractically difficult calculation of the factor  $Z$ ) but only the ratio of two different  $\pi(f)$ ’s. Another fact which facilitates a low computational burden is that the change of the color of a point  $d$  only changes the configuration over cliques which are subsets of the closed neighbourhood of  $d$ , a concept which we now define.

Given a set of cliques  $Q$  of the domain  $D$ , the *closed neighbourhood*  $\kappa_d$  of  $d \in D$  with respect to  $Q$  is the subset of  $D$  defined by

$$(11) \quad \kappa_d = \{d' \mid \text{for some } q \in Q, d \in q \text{ and } d' \in q\}.$$

and the *neighbourhood*  $\sigma_d$  of  $d$  is defined as  $\kappa_d - \{d\}$ . To avoid repetitious discussion of trivial special cases, from now on we assume that a model  $(Q, U)$  is such that, for all  $d \in D$ ,  $\sigma_d$  is nonempty (and so it is a clique, although not necessarily a clique of  $Q$ ). As an example, consider the  $D$  of (2) and let  $Q$  be as defined in (10). Then  $\kappa_{(2,2)}$  comprises  $(2, 2)$  itself and its 24 surrounding

points, namely

$$\begin{aligned}
& (0, 4) \ (1, 4) \ (2, 4) \ (3, 4) \ (4, 4) \\
& (0, 3) \ (1, 3) \ (2, 3) \ (3, 3) \ (4, 3) \\
& (0, 2) \ (1, 2) \ (2, 2) \ (3, 2) \ (4, 2) \\
& (0, 1) \ (1, 1) \ (2, 1) \ (3, 1) \ (4, 1) \\
& (0, 0) \ (1, 0) \ (2, 0) \ (3, 0) \ (4, 0)
\end{aligned}$$

Let  $g$  be a configuration over a clique  $q$  and let  $d \in D - q$ . Then we denote by  $g_k^d$  ( $k \in \{0, 1\}$ ) the configuration over  $q \cup \{d\}$  for which  $g_k^d(d) = k$  and  $(g_k^d|_q) = g$ . We define the *local interaction vector* for  $d$  and  $g$  to be  $a^d(g) = (a_1^d(g), \dots, a_C^d(g))$ , where  $a_c^d(g) = N(G_c, \{g_1^d\}) - N(G_c, \{g_0^d\})$ , for  $1 \leq c \leq C$ .

Using this notation, together with (4) and (8) it is easy to derive that in a Metropolis step (as described at the beginning of this section)

$$(12) \quad \frac{\pi(\tilde{f})}{\pi(f)} = e^{[1-2f(d)] \sum_{c=1}^C a_c^d(f|\sigma_d)U_c}.$$

For the example of the last section, the value of  $\sum_{c=1}^C a_c^d(f|\sigma_d)U_c$  is uniquely determined by the 24-dimensional vector of 0's and 1's provided by the 24 points surrounding  $d$  (this is due to the equivalence under horizontal and vertical translations). Hence, prior to running the Metropolis algorithm, these  $2^{24}$  possible values can be precalculated and stored in a table. During the running of the algorithm, the needed value is obtained from the table by a simple look-up based on the colors of the points surrounding  $d$  [9].

To insure that the algorithm has been run long enough (*burn in*) to provide a typical sample of the distribution, we made the following experiment. We initialized the Metropolis algorithm with two different images: a blank black image and another which is completely white. Then both cases were run until they stabilized with images of similar energy and similar number of white pixels. The time for the stabilization process is measured in *cycles*: in each cycle,  $63 \times 63 = 3,969$  pixels are randomly selected from the image. In particular, it requires  $2 \cdot 10^4$  cycles for the distributions of Figure 3 to reach stabilization. However, all the samples in the same figure were produced by running the algorithm for  $5 \cdot 10^5$  cycles.

### 3.2 Reconstruction using Gibbs priors

Suppose that we wish to reconstruct the top-left image in Figure 1 given its horizontal, vertical, and the NW-SE diagonal projections (i.e., we are considering a 3-grid). In the noiseless case, the existence of an image satisfying the projections is assured. However, as is illustrated in Figure 1 at the bottom-left (by an image which is tomographically equivalent to the one at the top-left), the three projections without further information are not sufficient for recon-

struction. The principle that we are using to achieve reconstruction is that if we add the information that the image belongs to a particular Gibbs distribution, then the likely solutions will be those which are “close” to the original image [7].

Now consider the noisy case of Subsection 1.2; i.e., the minimization of (7). Define the non-zero valued distribution  $\gamma$  on binary images:

$$(13) \quad \gamma(f) = \frac{1}{Z'} e^{-\beta[H(f) + \alpha \sum_{m=1}^M \sum_{l \in L^{(m)}} |P_f^{(m)}(l) - P^{(m)}(l)|]},$$

where  $\beta > 0$  and  $Z'$  is the normalizing factor. Clearly for a fixed  $\beta$ , the problem of searching for the minimum of (7) is equivalent to the problem of finding the maximum of (13). Note also that the argument which maximizes (13) is the same for any  $\beta > 0$ . In physics, the method of allowing a system of many particles (whose probability distribution resembles (13)) to find a configuration of low energy by slowly increasing  $\beta$  (which corresponds to lowering the temperature) is called *annealing*. Here we use *simulated annealing* by applying the Metropolis algorithm to  $\gamma$  and varying the factor  $\beta$  from a low to a high value. Unlike the standard simulated annealing [4], in our version we keep in the memory the image with the highest  $\gamma$  for a given  $\beta$ , to be used as the initial configuration for the next  $\beta$  in the annealing schedule.

## 4 Parameter estimation

In Figure 1 we have illustrated that if we have the projections of a binary image and the Gibbs distribution from which the image is a sample, then by applying the Metropolis algorithm to (13) with an appropriate annealing schedule, we may obtain a very good reconstruction. Hence, given a typical sample collection of images in a certain application area (we refer to this collection as the *training set* and denote it by  $F$ ), it is worthwhile to estimate a Gibbs distribution which may give rise to such a sample collection. Since a Gibbs distribution is defined by a model  $(Q, U)$ , we need to discover both the set of cliques  $Q$  and the corresponding potentials defined by  $U$ . Here we assume that the set of cliques  $Q$  it is given, as well as the partition  $\{G_c \mid 0 \leq c \leq C\}$  of  $G$ , and therefore the only remaining task is to estimate the  $U_c$  for  $1 \leq c \leq C$  (recall that  $U_0 = 0$ ). We discuss a number of previously proposed methods for doing this.

### 4.1 A heuristic approach

This method [2] defines, for  $1 \leq c \leq C$ ,

$$(14) \quad U_c = \kappa [\ln(N(G_c, F)/|G_c| + 1) - \ln(N(G_0, F)/|G_0| + 1)]$$

(recall that  $N(G_c, F)$  is the number of times a configuration in  $G_c$  appears in the training set), where  $|X|$  denotes the number of elements in the set  $X$  and the constant  $\kappa$  is determined by an additional criterion; in this paper we attempt to select  $\kappa$  so that the expected number of white pixels in a sample

image from the resulting distribution equals the average number of white pixels in the images of the training set. We note that this constant need not be considered separately in the reconstruction process since it can be absorbed in the  $\beta$  of the annealing schedule.

In the context of Figure 2, a higher potential using this definition implies that the local feature of the corresponding type occurs with a higher frequency in the sample collection.

#### 4.2 The histogram method

This method is based on the previously defined notion of a local interaction vector. The theory behind the method can be found in [3].

Let  $\Omega = \{(d, g) | d \in D \text{ and } g \text{ is a configuration over } \sigma_d\}$ . Let us partition  $\Omega$  by the condition that two items belong to the same class of the partition if, and only if, they share the same local interaction vector  $a^d(g)$ . Let  $\Omega = \cup_{b=1}^B \Omega_b$  be this partition and let  $a^b = (a_1^b, \dots, a_C^b)$  be the unique local interaction vector for the elements of  $\Omega_b$ . We also define, for  $k \in \{0, 1\}$  and  $1 \leq b \leq B$ ,  $\Omega_b^k = \{g_k^d | (d, g) \in \Omega_b\}$ .

Given the training set  $F$ , the histogram method aims at satisfying, in the least-squares sense, the system of equations

$$(15) \quad \sum_{c=1}^C a_c^b U_c = \ln \frac{N(\Omega_b^1, F)}{N(\Omega_b^0, F)},$$

for  $1 \leq b \leq B$ . (A hint of the reasonableness of this requirement is provided by (12).) This estimator is well-defined only if  $N(\Omega_b^k, F) \neq 0$  for  $k \in \{0, 1\}$  and  $1 \leq b \leq B$ , which is troublesome since the condition is likely to be violated if we are given a small number of sample images. The elimination of an equation from (15) if either  $N(\Omega_b^0, F) = 0$  or  $N(\Omega_b^1, F) = 0$  can lead to rank-deficiency and preclude a unique solution to the least-squares problem. The advantage of the estimate is that it can be given in closed form and is very easy to calculate. In 1999, Borges [1] proposed two “improvements” to this method which we now discuss.

#### 4.3 Borges’ variant

Theoretical justification of this variant can be found in [1]. The first improvement is the capability to consider cases when either  $N(\Omega_b^1, F) = 0$  or  $N(\Omega_b^0, F) = 0$  by replacing the right hand side of (15) by

$$(16) \quad \Xi(N_1, N_0) = \begin{cases} 0, & \text{if } N_1 = N_0, \\ \frac{1}{N_0+1} + \dots + \frac{1}{N_1}, & \text{if } N_1 > N_0, \\ -\frac{1}{N_1+1} - \dots - \frac{1}{N_0}, & \text{if } N_1 < N_0, \end{cases}$$

where, to simplify notations,  $N_1$  and  $N_0$  denote respectively  $N(\Omega_b^1, F)$  and  $N(\Omega_b^0, F)$ . The second improvement is provided by an estimate of the mean

error in the  $\Xi(N(\Omega_b^1, F), N(\Omega_b^0, F))$ . Based on this estimate, we minimize the linear combination of the squared differences, in which the coefficient given to the squared difference arising from the  $b^{th}$  equation in (15) is  $we_b^2$ , where

$$(17) \quad we_b = \left[ \frac{\pi^2}{3} - \frac{1}{N(\Omega_b, F) + 1} \sum_{k=0}^{N(\Omega_b, F)} \Xi^2(k, N(\Omega_b, F) - k) \right]^{-\frac{1}{2}}.$$

For  $N(\Omega_b, F) > 400$ , a good approximation to  $we_b$  is  $\sqrt{3 + 0.05N(\Omega_b, F)}$ .

## 5 Evaluation as Modelers

### 5.1 Figure of merit

Suppose that we select a Gibbs distributions (say one of the four illustrated in Figure 3), and we generate some random samples and use them as the training set for a parameter estimation method. The *figure of merit* used in this paper to measure the success of the estimation method is the squared norm:

$$(18) \quad \epsilon = \sum_{c=1}^C (U_c - \tilde{U}_c)^2,$$

where  $U_c$  and  $\tilde{U}_c$  (for  $1 \leq c \leq C$ ) are the actual and estimated values, respectively, of the potentials determining the Gibbs distribution.

### 5.2 Experimental results

In Section 3.1, we generated one random sample from each of four Gibbs distributions. Here we take many samples as the input to the estimation processes and report on the behavior of the figure of merit  $\epsilon$  as a function of the number of samples taken.

The variation of  $\epsilon$  for the four distributions is shown in Figure 4. Clearly Borges' method is able to recover the parameters in all the four cases; and so can the histogram method, except that the latter requires, in general, more samples than Borges' method to reach, within a certain tolerance, the original parameters. The heuristic method failed to converge to a similar low value of  $\epsilon$ . Figure 5 shows typical sample images from distributions estimated by the heuristic method; they do not resemble those from the corresponding original distributions. We do not report on the heuristic method for the distribution (1.2, 1.2, 1.2, 0.52, 0.6) because its expected number of white pixels is 2110, while the same item for the distributions defined by (14) applied to the training set is never greater than 2000 for any  $\kappa$ . Thus, our proposed way of selecting the  $\kappa$  in (14) is not guaranteed to produce a result. However, as pointed out already, this is not essential for reconstruction, since the  $\kappa$  can be absorbed into the  $\beta$  of the annealing schedule.

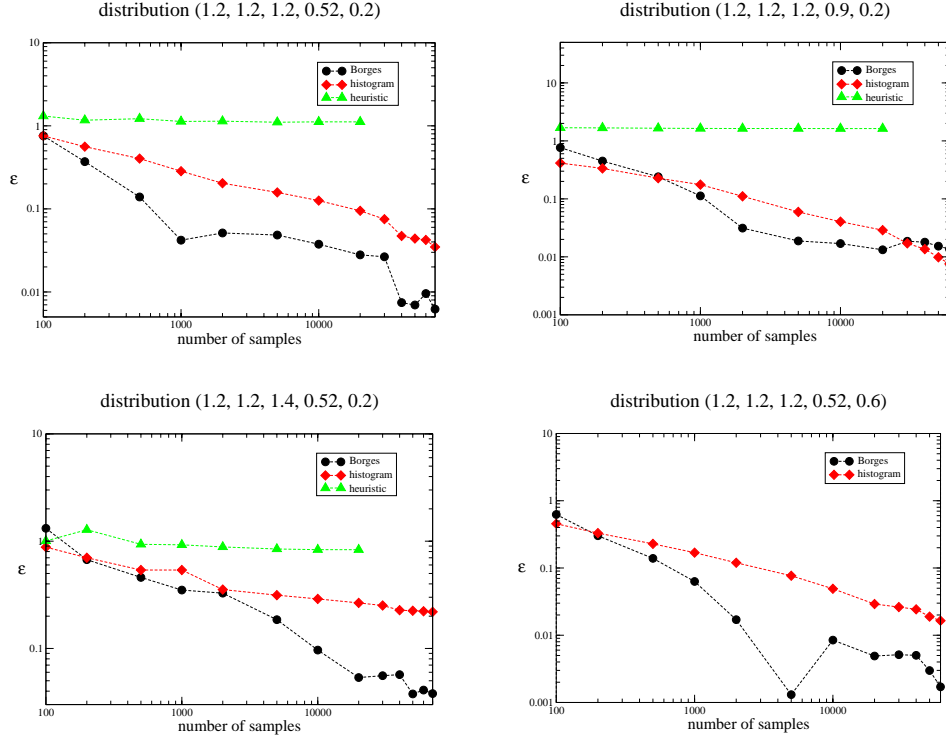


Fig. 4. Figure of merit  $\epsilon$  of (18) as a function of the number of samples for the four distributions in Figure 3.

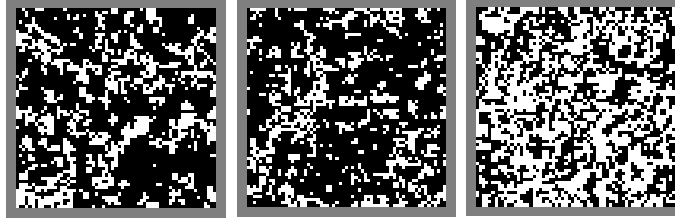


Fig. 5. Typical samples from the estimated distribution using heuristic method corresponding to the distribution (1.2, 1.2, 1.2, 0.52, 0.2) (left), (1.2, 1.2, 1.2, 0.9, 0.2) (center) and (1.2, 1.2, 1.4, 0.52, 0.2) (right).

## 6 Evaluation as Priors for Binary Tomography

### 6.1 Introduction: Figure of merit

In this section we discuss our main application of image modeling using Gibbs priors: binary tomography. For a given binary phantom image, we measure the quality of its reconstructed image by using the number of points at which the two images differ, and we denote this number by  $\zeta$ . We report on two experiments: in one, called Experiment I, the phantom images are from the Gibbs distribution (1.2, 1.2, 1.2, 0.52, 0.2), and in the second one, called Experiment II, we reconstruct cardiac cross-sectional images.



Fig. 6. Testing set of images which are samples from the distribution  $(1.2, 1.2, 1.2, 0.52, 0.2)$ .

### 6.2 Training set, testing set and projection data for Experiment I

The results from the previous section suggest that the estimated parameters can be made to be arbitrarily close (as measured by  $\epsilon$ ) to the original parameters as the size of the training set increases. However, in practice we are not likely to be able to obtain an indefinitely large number of samples; therefore, for this experiment and for Experiment II we restricted the number of samples to be one thousand. As a consequence, the estimated parameters are  $(1.18, 1.19, 1.19, 0.50, 0.17)$  by Borges' method and  $(0.73, 0.49, 0.30, 0.30, 0.19)$  by the heuristic method. Neither in this experiment nor in Experiment II do we report on the histogram method; since, as we have seen in the previous section, its output is indistinguishable from that of Borges' method when the number of samples is large.

For the testing set we generated an additional (to the training set) ten sample images from the same distribution (see Figure 6). For each sample image in the testing set, we took the (noiseless) projections in the three directions (horizontal, vertical, and NW-SE diagonal) as the input to the reconstruction process, together with the estimated parameters. All the images are of size  $63 \times 63$ .

### 6.3 Training set, testing set and projection data for Experiment II

Our phantoms represent cardiac cross-sectional images and consist of three geometrical objects of statistically variable size, shape and location: an ellipse representing the left ventricle, a circle representing the left atrium, and the difference between two circular sectors representing the right ventricle. Such a geometrical description defines a function over the subset  $[0, 63] \times [0, 63]$  of  $\mathbb{R}^2$  into  $\{0, 1\}$  and can be digitized by restricting the domain of the function to the  $D$  defined in (2) with  $I = 63$  (see Figure 7). Before we can apply the estimation methods of Section 4, we need to decide on the set of cliques  $Q$  and the partition of  $G = \cup_{q \in Q} 2^q$  for this kind of cardiac images. It is far from obvious what those choices should be. In Experiment II we investigate



Fig. 7. Digitized noisy cardiac phantom images in the testing set.

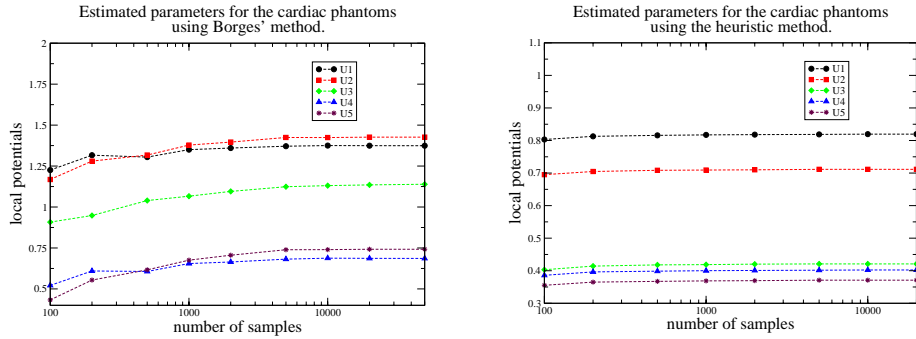


Fig. 8. Estimated parameters for the noisy cardiac phantom images.

whether our simple model using only five parameters is good enough for the reconstruction of such images.

For this model and for the digitized images described in the last paragraph, Borges' method failed to converge. An intuitive explanation for this is the following. If, for some  $b$  ( $1 \leq b \leq B$ ) the numbers  $N(\Omega_b^0, F)$  and  $N(\Omega_b^1, F)$  are such that one is always zero and the other increases proportionally with the size of the training set, in order to satisfy (15) for that  $b$ , at least one of the parameters will have to increase without bound. We found that adding a little noise to the digitized images will assure that this will not occur. Noise was added by inverting the color of each pixel with probability 0.001; the images in Figure 7 were created in this way. The resulting estimated distributions were (0.82, 0.71, 0.42, 0.40, 0.37) when using the heuristic method and (1.35, 1.38, 1.07, 0.65, 0.68) when using Borges' method. In Figure 8 we report on the history of the estimated values of the parameters (by the two methods) as we increased the size of the training set. Typical sample images from the estimated distributions based on one thousand training images are shown in Figure 9. Although they do not at all resemble the cardiac phantom images, past experience [2] indicates that they may nevertheless be quite useful for reconstruction.

Physically collected projection data correspond to line integrals through



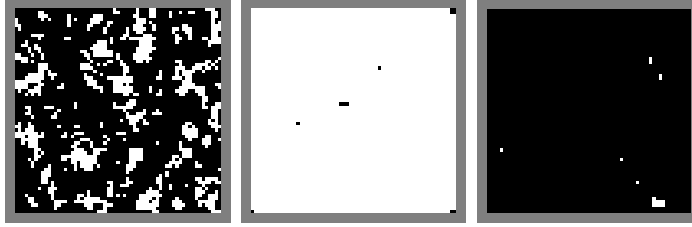


Fig. 9. Typical samples from the estimated distribution using the heuristic method (left) and using Borges’ method (center and right).

the geometrically defined phantoms. These line integrals maybe perceived as “noisy” versions of the line sums, as defined in (6), of the noisy discretized images (of the type shown in Figure 7) that we are trying to reconstruct. The training set is also used to model the nature of the noise in the projection data set. The line sums of the digitized images are always integers, and hence there are only finitely many possible such line sums. Suppose that one of these values is  $p$ ; there will be a number of rows in the noisy cardiac phantom images in the training set for which the line sums is  $p$ , for each of these rows we can calculate the corresponding line integral through the underlying geometrically defined phantom. Our noise model is predicated on the assumption that the “noisy” version of a horizontal line sum with value  $p$  is a random sample from a Gaussian distribution whose mean and standard deviation are those of the set of line integrals calculated as stated in the previous sentence. Similar calculations are done for horizontal rows with line sums other than  $p$  and, separately, for columns and for diagonals.

Just as in Experiment I, the testing set consists of ten additional binary images, generated by the same method as used to generate those of the training set. Figure 7 shows the images in the testing set. The projection data set for these images were generated using the noise model described in the previous paragraph.

#### 6.4 Experimental details

Once we have the projections data set and the estimated priors, we make the reconstructions by maximizing (13) using the Metropolis algorithm. We then determine the quality of our reconstructions by using the figure of merit  $\zeta$ .

For (13), we need to specify  $\alpha$  (the weighting factor between the prior and the projection data) and the Metropolis algorithm needs an annealing schedule (see Subsection 3.2). To find an optimum  $\alpha$  we start with a gross annealing schedule and find the  $\alpha$  which minimizes the average  $\langle \zeta \rangle$  of the  $\zeta$ ’s over the ten images in the testing set. The  $\zeta$  for one image is determined in the following way. Each image in the testing set is reconstructed five times; the average of the  $\zeta$ ’s over the five reconstructions is what determines the  $\zeta$  for that image. We then refine the annealing schedule and search again for an optimum  $\alpha$  within a neighbourhood of the  $\alpha$  previously found. This process



Fig. 10. Phantoms number 3, 7 and 9 (from left to right) in the top row together with their respective reconstructed images using Borges' estimated parameters (shown at the middle row) and heuristics method's parameters (shown at the bottom row).

is repeated until no significant improvement in  $\langle \zeta \rangle$  is observed.

To specify an annealing schedule, we introduce the following function defined on  $\lambda$  which denotes the cycle number (see the end of Subsection 3.1) in the Metropolis algorithm:

$$(19) \quad W_{n_1, n_2}^{step}(\lambda) = \begin{cases} 0, & \text{if } \lambda \leq n_1 \cdot step, \\ n_2 - n_1, & \text{if } \lambda > n_2 \cdot step, \\ \left\lceil \frac{\lambda}{step} \right\rceil - n_1, & \text{otherwise,} \end{cases}$$

where for a real number  $r$ ,  $\lceil r \rceil$  denotes the minimum integer greater than or equal to  $r$ .

In Experiment I, the optimal  $\alpha$  was estimated to be respectively 1.25 and 3.8 for the priors estimated by the heuristic and Borges' method. The corresponding annealing schedules were respectively  $\beta(\lambda) = 0.96 \cdot W_{0,1}^{5 \cdot 10^4}(\lambda) + 0.096 \cdot W_{1,11}^{5 \cdot 10^4}(\lambda)$  and  $\beta(\lambda) = 0.3 \cdot W_{0,1}^{5 \cdot 10^4}(\lambda) + 0.025 \cdot W_{1,13}^{5 \cdot 10^4}(\lambda)$ . In Experiment II, the optimal  $\alpha$  was estimated to be respectively 0.74 and 1.8 for the heuristic method and Borges' method. The corresponding annealing schedules were respectively  $\beta(\lambda) = 1.34 \cdot W_{0,1}^{5 \cdot 10^4}(\lambda) + 0.34 \cdot W_{1,9}^{5 \cdot 10^4}(\lambda)$  and  $\beta(\lambda) = 0.6 \cdot W_{0,1}^{5 \cdot 10^4}(\lambda) + 0.025 \cdot W_{1,4}^{5 \cdot 10^4}(\lambda) + 0.01 \cdot W_{4,16}^{5 \cdot 10^4}(\lambda)$ .

	Experiment I		Experiment II	
	heuristic	Borges	heuristic	Borges
1	3	0	72	155
2	1	0	34	69
3	5	0	98	200
4	4	0	55	60
5	8	0	130	102
6	1	0	58	182
7	0	0	197	130
8	5	5	68	86
9	5	0	30	35
10	1	0	177	200
$\langle \zeta \rangle$	3	1	92	122

Table 1

Comparison between the heuristic method and Borges' method in the two experiments. The numbers 1 to 10 in the first column correspond to the images which are the following: in Experiment I these are the samples shown in Figure 6 from the distribution  $(1.2, 1.2, 1.2, 0.52, 0.2)$ , while in Experiment II they are the cardiac phantom images shown in Figure 7. The rest of entries are values of  $\zeta$  defined as the number of points at which the phantom and its corresponding reconstructed image differ. The last row reports the average  $\langle \zeta \rangle$  over the ten phantoms rounded to the nearest integer.

### 6.5 Experimental results

The comparison between the heuristic approach and Borges' method in the two experiments is shown in Table 1. In Figure 10 we show the reconstructed images for phantom number 9 which is the one most accurately reconstructed by both methods. We also show the results for phantom number 3, which is typically accurate for the heuristic method and the least accurate for Borges' method, and for phantom number 7 whose reconstruction by Borges' method is typically accurate but is the least accurate for the heuristic method.

## 7 Conclusions

In Section 5 we applied three methods for estimating Gibbs priors. Both Borges' method and the histogram method showed successful results for the proposed model with five parameters and converged to the original parameters

as the size of the training set increased. This was not the case for the heuristic method. However, in Section 6 we showed that the parameters estimated by all three methods lead to almost perfect reconstructions.

It was also demonstrated in Section 6 that despite the fact that the proposed model does not describe adequately the ensemble of the cardiac cross-sectional images (see Figure 9), the percent of mis-classified pixels in a typical reconstruction is still below four, which may well be sufficiently small to provide useful information regarding the cardiac structures which are being imaged.

## References

- [1] Borges C.F., *On the estimation of Markov random field parameters*, IEEE Trans. Pattern Anal. Mach. Intell. **21** (1999), 216–224.
- [2] Carvalho B.M., Herman G.T., Matej S., Salzberg C. and Vardi E., *Binary tomography for triplane cardiography*, in Information Processing in Medical Imaging, A. Kuba, M. Samal and A. Todd-Pokropek (eds.), Springer, Berlin, (1999), 29–41.
- [3] Derin H. and Elliot H., *Modeling and segmentation of noisy and textured images using Gibbs random fields*, IEEE Trans. Pattern Anal. Mach. Intell. **9** (1987), 39–55.
- [4] Duda R.O., Hart P.E., Stork D.G., “Pattern Classification”, Chapter 7. John Wiley & Sons, New York, 2001.
- [5] Kong T.Y. and Herman G.T., *Tomographic equivalence and switching operations*, in Discrete Tomography: Foundations, Algorithms and Applications, G.T. Herman and A. Kuba (eds.), Birkhäuser, Boston, 1999.
- [6] Levitan E., Chan M. and Herman G.T., *Image-modeling Gibbs priors*, CVGIP: Graph. Models Image Proc. **57** (1995), 117–130.
- [7] Matej S., Herman G.T. and Vardi A., *Binary tomography on the hexagonal grid using Gibbs priors*, Int. J. Imag. Sys. and Tech. **9** (1998), 126–131.
- [8] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E., *Equations of state calculations by fast computing machines*, J. Chem. Phys. **21** (1953), 1087–1092.
- [9] Vardi E., Herman G.T. and Kong T.Y., *Speeding up stochastic reconstructions of binary images from limited projection directions*, Lin. Alg. and its App. Special issue on Discrete Tomography (to appear).