

A METHOD FOR RECONSTRUCTING LABEL IMAGES FROM A FEW PROJECTIONS, AS MOTIVATED BY ELECTRON MICROSCOPY

Hstau Y. Liao and Gabor T. Herman

Discrete Imaging and Graphics Group
Department of Computer Science
The Graduate Center, CUNY, New York, USA

ABSTRACT

Our aim is to produce a tessellation of space into small voxels and, based on only a few tomographic projections of an object, assign to each voxel a label that indicates one of the components of interest constituting the object. Traditional methods are not reliable in applications, such as electron microscopy in which (due to the damage by radiation) only a few projections are available. We postulate a low level prior knowledge regarding the underlying distribution of label images, and then directly estimate the label image based on the prior and the projections. We use a relatively efficient approximation to a global search for the estimation.

1. INTRODUCTION

In computerized tomography (CT) many possible density values within the object being reconstructed may occur, and for this reason a large number of projections is necessary to ensure an accurate reconstruction of the density distribution. In many situations, however, the ultimate aim is not the density distribution itself but rather a distribution of labels that correspond to one of the components (such as protein or ice in biological macromolecules) constituting the object, and there are good reasons (such as damage by radiation) why only a few projections can be collected. Making use of the knowledge that the reconstruction should contain only a few values (labels) to make up for the lack of the availability of the number of projections typically required in CT is the central theme of *discrete tomography* (DT) [1]. In DT a labeling of the components of the object can be directly produced based on the projections. To achieve the same in CT, the reconstructed density distribution needs to be segmented into components by an additional process.

Our main application is in electron microscopy of macromolecules. Electron microscopy is capable of imaging complex biological macromolecules in order to further the understanding of their functions.

Our long-term aim is to produce, based on the micrographs, a tessellation of space into small voxels, each labeled as containing ice, protein, or RNA. Traditional approaches using methods of CT would first assign, based on the projections, to each voxel a *gray value* (which is related to the atomic density of that voxel) and

This work was supported by the NIH grant HL70472 and by the NSF grant DMS0306125.

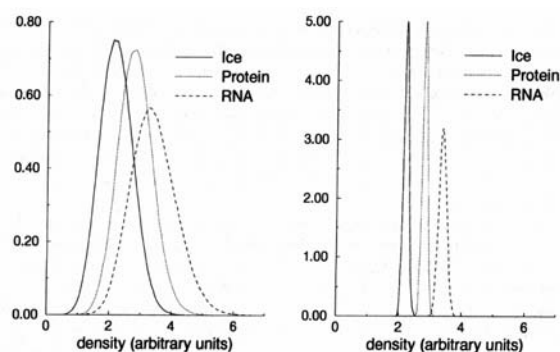


Fig. 1. Histograms of the densities corresponding to volumes sampled using voxels of edge length equal to 2.5 Å (left) and 7.5 Å (right), obtained from volumes composed of only ice, only protein, or only RNA; from [3].

then would segment this gray value image to obtain the label image. However, since typically the number of projections required in CT is much larger than that in DT, a reconstructed gray value image from only a few projections is likely to be very inaccurate, leading to an incorrect segmentation. A particular difficulty arises if thresholding [2] is used for segmentation (which is a common current practice) because at a resolution of 2.5 Å or better the density distributions corresponding to different labels greatly overlap (see Figure 1).

We postulate a low level Gibbs prior on the underlying distribution of label images, and then directly estimate, using an approximation to the *maximum a posteriori probability* (MAP) approach, an optimum label image based on the prior and the measured projections. Here we report on our early experiments aimed at evaluating this kind of approach.

2. THE PROPOSED MODEL

Let \mathbf{X} be a set of *label images* each one of which is an I -dimensional vector $\mathbf{x} = (x_1, \dots, x_I)^t$, where $x_i \in X$ (the finite set of *labels*), for $1 \leq i \leq I$. We assume that there is a prior distribution that assigns to every label image \mathbf{x} a probability $\pi(\mathbf{x})$. Typically this prior distribution is a *Gibbs distribution*, which means that

$$\pi(\mathbf{x}) = Z^{-1} \exp[-H(\mathbf{x})], \quad (1)$$

where $\pi(\mathbf{x})$ is the probability of occurrence of the image \mathbf{x} , Z is the normalizing factor, and $H(\mathbf{x})$ is referred to as the “energy” of

\mathbf{x} (see, e.g., [4]).

Let \mathbf{Y} be a set of *gray value images* each one of which is a J -dimensional vector $\mathbf{y} = (y_1, \dots, y_J)^t$, where $y_j \in Y$ (the set of *gray values*, $Y \subseteq \mathbb{R}$), for $1 \leq j \leq J$. We assume that there is a conditional distribution that, given a label image \mathbf{x} , assigns a probability $\phi(\mathbf{y}|\mathbf{x})$ to every gray value image \mathbf{y} . (Since \mathbf{Y} is not necessarily finite, it would be more precise to say that $\phi(\mathbf{y}|\mathbf{x})$ is the probability density function defining the conditional distribution of the gray value image \mathbf{y} given the label image \mathbf{x} . For the sake of brevity, we will continue to refer to a notation such as $\phi(\mathbf{y}|\mathbf{x})$ as a ‘‘probability’’ rather than a ‘‘probability density function.’’) Initially we work with the special case in which $J = I$ and, for every label x , there is a distribution that assigns (independently) a probability $\varphi(y|x)$ to every gray value y . Consequently, $\phi(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^J \varphi(y_j|x_j)$.

Let \mathbf{W} be a set of *measurement vectors* each one of which is a K -dimensional vector $\mathbf{w} = (w_1, \dots, w_K)^t$, where $w_k \in W \subseteq \mathbb{R}$, for $1 \leq k \leq K$. We assume that there is a conditional distribution that, given a gray value image \mathbf{y} , assigns a probability $\chi(\mathbf{w}|\mathbf{y})$ to every measurement vector \mathbf{w} . For the preliminary experiments reported below, we consider only the special case in which $\chi(\mathbf{w}|\mathbf{y})$ has a product form; i.e., given a gray value image \mathbf{y} , there is (for $1 \leq k \leq K$) a conditional probability $\psi_k(w_k|\mathbf{y})$ of the k^{th} measurement being w_k , and $\chi(\mathbf{w}|\mathbf{y}) = \prod_{k=1}^K \psi_k(w_k|\mathbf{y})$.

The *likelihood* $\rho(\mathbf{w}|\mathbf{x})$ of measuring \mathbf{w} given a label image \mathbf{x} , assuming (as reasonable) that \mathbf{w} does not depend directly on \mathbf{x} , is the integral (or sum if \mathbf{Y} is finite)

$$\rho(\mathbf{w}|\mathbf{x}) = \int_{\mathbf{Y}} \chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x})d\mathbf{y}. \quad (2)$$

3. THE ESTIMATION APPROACH

According to the MAP criterion, the estimate of the label image is \mathbf{x}_{MAP} , which is the \mathbf{x} in \mathbf{X} that maximizes the product of $\pi(\mathbf{x})$ and $\rho(\mathbf{w}|\mathbf{x})$; i.e.,

$$\mathbf{x}_{MAP} = \arg \max_{\mathbf{x}} [\pi(\mathbf{x})\rho(\mathbf{w}|\mathbf{x})]. \quad (3)$$

Due to the non-linearity and the non-convexity of the objective function (that is inside the brackets in (3)), the MAP estimator(s) cannot be obtained using local search techniques. Furthermore, the objective function is defined on labels, rather than on continuous variables; therefore, we need to use a combinatorial optimization technique that involves having to evaluate repeatedly the value of (2) for various \mathbf{x} .

Except for Monte Carlo techniques, which are inherently slow, we are not aware of general approaches for the computation of the integral on the right hand side of (2). Therefore, we decided to investigate the existence of alternative approaches that are based on criteria other MAP, but that can be efficiently implemented, and that at the same time deliver good reconstructions. In the alternative approach on which we report in this paper, we aim at finding the label image $\hat{\mathbf{x}}$ that maximizes the product of $\pi(\mathbf{x})$ and the term inside the likelihood function of (2); i.e.,

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \left\{ \pi(\mathbf{x}) \max_{\mathbf{y}} [\chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x})] \right\}. \quad (4)$$

Let

$$G(\mathbf{x}, \mathbf{y}) = \chi(\mathbf{w}|\mathbf{y})\phi(\mathbf{y}|\mathbf{x}). \quad (5)$$

In [5], we proposed as the solution a local maximizer of $F(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{x})G(\mathbf{x}, \mathbf{y})$, which is obtained (starting from a gray value image $\mathbf{y}^{(0)}$) by alternately maximizing $F(\mathbf{x}, \mathbf{y})$ with respect to \mathbf{x} and with respect to \mathbf{y} . The reconstructions were not statistically significantly better than those produced by an earlier of our methods [6]. Here we intend to improve our previous methods by using an approximation to global search.

If we let $\tilde{\mathbf{y}}(\mathbf{x}) = \arg \max_{\mathbf{y}} [G(\mathbf{x}, \mathbf{y})]$, then both $\tilde{\mathbf{y}}(\mathbf{x})$ and $G(\mathbf{x}, \tilde{\mathbf{y}}) = \max_{\mathbf{y}} [G(\mathbf{x}, \mathbf{y})]$ are functions of \mathbf{x} . Therefore, (4) can be rewritten as

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{ \pi(\mathbf{x})G[\mathbf{x}, \tilde{\mathbf{y}}(\mathbf{x})] \}. \quad (6)$$

For the maximization, we apply *simulated annealing* via the *Metropolis algorithm* [4]. The Metropolis algorithm is a sampling algorithm, in which, starting with an arbitrary image, in each iterative step a new image \mathbf{x}' is generated from the current one \mathbf{x} . After a sufficiently large number of steps, an image produced in this way can be considered to be a typical sample from a given distribution $\gamma(\mathbf{x})$. Simulated annealing, often used as a combinatorial optimization technique, successively applies a sampling algorithm to the distribution $[\gamma(\mathbf{x})]^\beta$, where β is a positive real number. Initially, β is set to a low value, so that the graph of $[\gamma(\mathbf{x})]^\beta$ is nearly flat. After the sampling converges to the stationary distribution at the current β , β is increased to a higher value according to an *annealing schedule*. A theorem of [7] gives an annealing schedule that guarantees that the process converges (almost surely) to a global maximum. In practice, however, the annealing schedule in [7] is too slow and faster schedules are used instead.

In each step of the Metropolis algorithms in our method, a pixel of the current image \mathbf{x} is randomly selected and its label is changed, resulting in the image \mathbf{x}' . Then \mathbf{x} is replaced by \mathbf{x}' with probability $\min \{1, \gamma(\mathbf{x}')/\gamma(\mathbf{x})\}$. We define a *cycle* as I such steps. Note that only the ratio $\gamma(\mathbf{x}')/\gamma(\mathbf{x})$ is needed in each step, which eliminates the need to compute the normalizing factor, such as the Z in (1). Because our objective function $\pi(\mathbf{x})G[\mathbf{x}, \tilde{\mathbf{y}}(\mathbf{x})]$ is positive valued, we can in principle use simulated annealing to find $\hat{\mathbf{x}}$ by setting $\gamma(\mathbf{x})$ to be the objective function. This implies that in each step of the Metropolis algorithm, the ratio $\pi(\mathbf{x}')G[\mathbf{x}', \tilde{\mathbf{y}}(\mathbf{x}')] / \pi(\mathbf{x})G[\mathbf{x}, \tilde{\mathbf{y}}(\mathbf{x})]$ must be calculated, which is troublesome, due to the typically enormous number of steps required in simulated annealing. In our experiments, for the type of Gibbs distribution we are dealing with, we have an efficient way to compute the ratio $\pi(\mathbf{x}')/\pi(\mathbf{x})$ based on a look-up table [8]. However, we are currently unable to compute directly the ratio $G[\mathbf{x}', \tilde{\mathbf{y}}(\mathbf{x}')] / G[\mathbf{x}, \tilde{\mathbf{y}}(\mathbf{x})]$ without computing separately both the numerator and the denominator. Nonetheless, we note that $\tilde{\mathbf{y}}(\mathbf{x}') \approx \tilde{\mathbf{y}}(\mathbf{x}'')$ when \mathbf{x}' and \mathbf{x}'' differ at only a few pixels. Therefore, our proposed estimation approach is to use the same $\tilde{\mathbf{y}}$ in many steps, and only update it when the current label image and the one that gave rise to the last update differ at more than a given number of pixels.

4. EXPERIMENTAL DETAILS

In order to test this general approach in the context of tomographic reconstruction of label images from a few projections, we performed some initial experiments. We picked some computer generated phantom images (which are the label images) and their corresponding gray value images (one gray value image for each phantom). We reconstructed the label images based on the projections by the method in [5] and our new method. An indicator that tells us

how good these reconstructions are was used to evaluate the two approaches. The inputs to the reconstruction process were computer simulated projection data based on the generated gray value images. In particular we made the following choices.

We considered 2D images of size $I = 63 \times 63$ and the set of labels $X = \{\text{black, white}\}$. All the images were typical samples from the distribution $\pi(\mathbf{x})$ used in Experiment 1 of [9], which assigns higher probability to images that have relatively large uniform white regions over a black background. In order to evaluate our approach, we sampled 50 label images using the Metropolis algorithm. Next, for each one of these label images, we randomly picked a gray value image (with $J = I = 63 \times 63$). We assumed a product structure for $\phi(\mathbf{y}|\mathbf{x})$: for a label image, a gray value image was produced by sampling at each pixel according to the probability $\varphi(y_j|x_j)$. The $\varphi(y_j|x_j)$, for $1 \leq j \leq J$, were Gaussian distributed with mean and variance equal to μ_{x_j} , where $\mu_{\text{black}} = 4$ and $\mu_{\text{white}} = 9$, so that

$$\begin{aligned} \phi(\mathbf{y}|\mathbf{x}) &= \prod_{j=1}^J \varphi(y_j|x_j) \\ &= \exp \left\{ - \sum_{j=1}^J \left[\ln Z_j + \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}} \right] \right\} \end{aligned} \quad (7)$$

where $Z_j = (2\pi\mu_{x_j})^{-\frac{1}{2}}$. There is no particular reason for choosing these values of mean and variance, except for the fact that they clearly reflect the idea of overlapping gray values in higher resolution electron microscopy, as shown by the left histogram in Figure 1.

The projections forming the measurement vector \mathbf{w} were simulated as follows. There were eight projections using parallel lines that encountered pixel centers, such that the tangent of the angle between these lines and the “positive horizontal direction” were 0, infinity, -1 , 1 , -0.5 , 0.5 , -2 , and 2 . We formed a K -dimensional column vector $\mathbf{z} = (z_1, \dots, z_K)^t$ from the sums of the pixel values in the gray value image along the K lines in all the projections. Initially, we chose $\psi_k(w_k|\mathbf{y})$, for $1 \leq k \leq K$, to be Gaussian distributed with mean z_k and variance equal to $0.01 \cdot z_k$; i.e.,

$$\begin{aligned} \chi(\mathbf{w}|\mathbf{y}) &= \prod_{k=1}^K \psi_k(w_k|\mathbf{y}) \\ &= \exp \left\{ - \sum_{k=1}^K \left[\ln Z'_k + \frac{(w_k - z_k)^2}{0.02 \cdot z_k} \right] \right\}, \end{aligned} \quad (8)$$

where $Z'_k = (0.02\pi z_k)^{-\frac{1}{2}}$. By substituting (8) and (7) into (5), we get that

$$G(\mathbf{x}, \mathbf{y}) = C \exp \left[- \sum_{k=1}^K \frac{(w_k - z_k)^2}{0.02 \cdot z_k} - \sum_{j=1}^J \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}} \right], \quad (9)$$

where $C = C(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^K \frac{1}{Z'_k} \prod_{j=1}^J \frac{1}{Z_j}$.

We found an efficient algorithm for computing $\tilde{\mathbf{y}}(\mathbf{x})$ if a further approximation is made: the replacement of the variances $0.02 \cdot z_k$ in (8) and (9) by $0.02 \cdot w_k$, for $k = 1, \dots, K$. As a result of this approximation, the $G(\mathbf{x}, \mathbf{y})$ of (9) in our original problem statement is now defined by

$$G(\mathbf{x}, \mathbf{y}) = C \exp \left[- \sum_{k=1}^K \frac{(w_k - z_k)^2}{0.02 \cdot w_k} - \sum_{j=1}^J \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}} \right], \quad (10)$$

where C is defined as after (9) but with $Z'_k = (0.02\pi w_k)^{-\frac{1}{2}}$. Maximizing this new $G(\mathbf{x}, \mathbf{y})$ with respect to \mathbf{y} is the same as maximizing what is inside the brackets in (10) with respect to \mathbf{y} . To

Table 1. Quality of reconstruction

Method	% of correct classifications
Thresholding the gray value image	86.2±0.6
Method in [5]	94.5±2.1
Our method:	
Experiment A	97.3±1.0
Experiment B	97.4±1.0

find the required $\tilde{\mathbf{y}}(\mathbf{x}) = \arg \max_{\mathbf{y}} [G(\mathbf{x}, \mathbf{y})]$, we need to find the \mathbf{y} that minimizes

$$\sum_{k=1}^K \frac{(w_k - z_k)^2}{0.02 \cdot w_k} + \sum_{j=1}^J \frac{(y_j - \mu_{x_j})^2}{2\mu_{x_j}}. \quad (11)$$

In this minimization problem, the w_k 's and x_j 's are fixed and each of the z_k 's is a linear combination of the y_j 's. So (11) is a quadratic in \mathbf{y} , and the minimization task closely resembles that in Equation (3.1) of Chapter 11 in [10]. Hence, as an estimate of $\tilde{\mathbf{y}}(\mathbf{x})$, we can take the gray value image produced by a row-action iterative reconstruction algorithm with relatively fast convergence properties, an *Algebraic Reconstruction Technique* (ART), which is an adaptation of the one described in [10], pp. 190-192. The relaxation parameter was set to 1, and the data were accessed in the order described in [5].

For the estimation of $\hat{\mathbf{x}}$, we use simulated annealing, as described in the previous section. The β of the annealing schedule was increased from 0.5 to 1.5 in increments of 0.05. For each β , we ran 5,000 cycles of the Metropolis algorithm, during which we updated $\tilde{\mathbf{y}}(\mathbf{x})$, using ART with ν cycles through the data, when the current label image and the one that gave rise to the last update differed at more than ρ pixels. For the starting value $\beta = 0.5$, ν and ρ were set to, respectively, 5 and 100 (about 2.5% of the total number of pixels). As β increases, the label image changes less rapidly and therefore, with the same computational burden, we can estimate $\tilde{\mathbf{y}}(\mathbf{x})$ with higher accuracy by gradually increasing ν and decreasing ρ . At each increment of β , ν was increased by 5, while ρ was decreased by 5 until a minimum of $\rho = 10$ was reached. The values of ν and ρ and their rate of change were not optimized. We term the experiment just described as *Experiment A*. In a separate experiment, called *Experiment B*, we considered fewer control parameters and performed the update of $\tilde{\mathbf{y}}(\mathbf{x})$ with a fixed $\nu = 5$ at every 100 cycles of the Metropolis algorithm, which was run 50,000 cycles for each β . We note that both Experiments took about the same CPU time.

5. EXPERIMENTAL RESULTS

In Table 1 we duplicate, from [5], the quality of the estimators based on our previous method and the quality of the label images obtained by thresholding directly the gray value images at the optimum threshold level (which, in our case, is 6.4683). We also indicate the quality of the estimators based on our new proposed approach. The reconstruction quality is measured by the average (over the 50 images of the experiment) percentage of correctly classified labels in the estimates. Note that in terms of misclassified labels, the average percentage of these given by our new approach is less than half of that by the method from [5]. The significance of the difference of the reconstruction qualities between the new method (either Experiment A or Experiment B) and the earlier

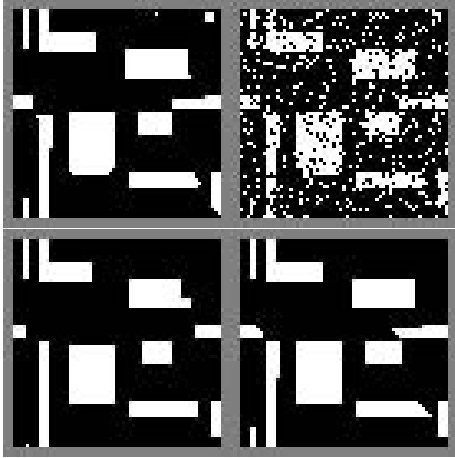


Fig. 2. One of the 50 label images (top-left) along with its estimators based on the method in [5] (bottom-left) and the estimator based on Experiment B of our method (bottom-right). The top-right label image is produced by thresholding (at the optimum threshold level) the gray value image. The number of misclassified labels are 407 (top-right), 127 (bottom-left) and 42 (bottom-right).

method was measured using the pairwise t -test. We found that the difference is significant at the 10^{-20} level: this means that if the two methods were equally good, then the probability of observing differences as large as or larger than what we have observed as a consequence of nothing but statistical fluctuations is less than 1 in 10^{20} . Regarding the superiority of the method of Experiment B as compared to that of Experiment A, the significance is only at the 0.25 level, which is usually considered not significant. Figure 2 shows actual reconstructions by the three main methods.

6. CONCLUSIONS AND DISCUSSION

Table 1 demonstrates that our approach (which is based on an approximation to a global search) to reconstructing label images from a few projections outperforms that in [5] (which is based on a local search algorithm).

Except for one of the 50 label images in the two experiments, the estimator found by our algorithms achieves both a higher value of the objective function being maximized (see (6)) and a larger number of correctly classified labels than the estimator proposed in [5]. This suggests that the criterion stated in (4), as an alternative one of the original MAP criterion (3), is reasonable. However, we think that even better quality of reconstructions can be achieved by the MAP estimators, for which search is hampered by the non-convexity and non-linearity of the Gibbs distribution, as well as the complicated nature of the likelihood function of (2) that involves an integration over all the gray value images. Thus, the discovery of an efficient algorithm for finding the true MAP estimator is an important open problem.

7. REFERENCES

[1] G.T. Herman and A. Kuba, Eds., *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhäuser, Boston, 1999.

[2] P.K. Sahoo, S. Soltani, A.K.C. Wong, and Y.C. Chen, “A survey of thresholding techniques,” *Computer Vision, Graphics, and Image Processing*, vol. 41, pp. 233–260, 1996.

[3] J.M. Carazo, C.O. Sorzano, Rietzel E., Schröder R., and R. Marabini, “Discrete tomography in electron microscopy,” in *Discrete Tomography: Foundations, Algorithms and Applications*, G.T. Herman and A. Kuba, Eds., pp. 405–416. Birkhäuser, Boston, 1999.

[4] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer, Berlin, 1995.

[5] H.Y. Liao and G.T. Herman, “Tomographic reconstruction of label images from a few projections,” *Electronic Notes in Discrete Mathematics, 9th International Workshop Combinatorial Image Analysis*, vol. 12, Available at <http://www.elsevier.com/jeing/31/29/24/71/23/show/Products/notes/index.htm>, 2003.

[6] H.Y. Liao and G.T. Herman, “Reconstruction of label images from a few projections as motivated by electron microscopy,” in *Proc. IEEE 28th Annual Northeast Bioeng. Conf.*, Philadelphia, PA, 2002, pp. 205–206, IEEE.

[7] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.

[8] E. Vardi, G.T. Herman, and T.Y. Kong, “Speeding up stochastic reconstructions of binary images from limited projection directions,” *Linear Algebra Appl.*, vol. 339, pp. 75–89, 2001.

[9] H.Y. Liao and G.T. Herman, “Automated estimation of the parameters of Gibbs priors to be used in binary tomography,” *Discrete Appl. Math.*, to appear. For a preliminary version see *Electronic Notes in Theoretical Computer Science*, vol. 46, Elsevier, <http://www.elsevier.nl/jeing/31/29/23/86/27/48/46026.pdf>, 2001.

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