Two direct methods for reconstructing pictures from their projections—A comparative study

by GABOR T. HERMAN

State University of New York
Buffalo, New York

INTRODUCTION

There are situations in the natural sciences and medicine (e.g., in electron microscopy or X-ray photography) in which it is desirable to estimate the gray levels of a picture at individual points from the sums of the gray levels along straight lines (projections) at a few angles. Usually, in such situations, the picture is far from determined and the problem is to find the “most representative” picture of the class of pictures which have the given projections.

In recent years, there has been a very active interest in this problem. A number of algorithms have been proposed for solving it. The algorithms are applicable in a large and varied number of fields. This is perhaps best demonstrated by the reference list, where we see that such algorithms have been published or discussed in journals on computer science, theoretical biology, molecular biology, cellular biology, biophysics, medical and biological engineering, roentgenology, radiology, chemistry, optics, crystallography, physics, electrical engineering, as well as in general science journals. The most important use of the algorithms is the reconstruction of objects (e.g., biological macromolecules, viruses, protein crystals or parts of the human body) from electronmicrographs or X-ray photographs.

Most of the effort in this area has so far been concentrated in developing algorithms and in applying these algorithms for actual reconstructions. So far, there has been only limited attention paid to the relative merits of the algorithms. The present paper is a report on the first of a series of comparative studies of various algorithms for reconstructing pictures from their projections. The two techniques that we shall compare are one of the algebraic reconstruction techniques (ART) of Gordon, Bender & Herman (1970) and a summation technique, which has been independently discovered by a number of people, but the most detailed previous study of which has been given by Vainshtein (1971a, b).

Both these techniques are direct techniques in the sense that the reconstruction is being done entirely in real space (density space), without the use of Fourier transforms. This is worth pointing out, since the first applicable method for reconstructing pictures from their projections (discovered independently by DeRosier & Klug [1968] and Tretiak, Eden & Simon [1969]) depended on the fact that the Fourier transform of a projection is a section of the Fourier transform of the picture. Thus, reconstruction was achieved by taking Fourier transforms of the projections, interpolating in Fourier space to get an approximation of the Fourier transform of the picture, and then taking an inverse Fourier transform to get an approximation of the picture itself. Detailed descriptions of such techniques were given by Crowther, DeRosier & Klug (1970) and Ramachandran (1971).

Crowther, DeRosier & Klug (1970) state that “for general particles and methods of data collection, the method of density space reconstruction is computationally impracticable on presently available computers.” However, they appear to be wrong on this point. Both the methods discussed in this paper have been implemented, and precise timings will be given on a number of examples to show that they are not computationally impracticable. Even more interestingly, expanding the mathematics on which the Fourier techniques are based, Ramachandran & Lakshminarayanan (1971a, b) succeeded to produce another direct method, called the convolution method, for the reconstruction of pictures from their projections. They demonstrate on a number of idealized pictures that their technique is faster than the Fourier techniques and that it also gives a better reconstruction. Detailed comparative
studies between ART, the convolution method and the Fourier techniques are presently under way, and will be reported on in later publications.

In the next section, we shall briefly describe the problem of reconstruction of objects from their projections and discuss the criteria we shall use for evaluating algorithms for solving this problem. This is the first exhaustive discussion of such criteria, it will form the basis of later comparative studies as well. The following two sections will give brief descriptions of ART and the summation method, respectively. In particular, we shall give a version of the summation method which is superior to previously published versions. In the final section, we shall give a report on the results of our experiments.

A basic difference between the two techniques is that the summation method takes a fixed length of time to run, while ART is an iterative technique such that its result can be improved by repeated iterations. One iteration of ART costs about the same as the summation technique. In rough terms, the results of our experiments indicate the following.

If the number of projections is small, one iteration of ART seems to perform somewhat worse than the summation method. As the number of projections increases, the relative performance of one iteration of ART as opposed to the summation method improves, and eventually it produces superior results. The break even point seems to be at eight projections. However, by repeated iterations, the performance of ART can be improved and it eventually surpasses the summation method in all our experiments. The improvement in ART is especially impressive with eight or more projections. In such cases, we obtain, with five iterations of ART, results which are far superior to the results of the summation method.

METHOD OF COMPARISON

First we wish to describe briefly the problem of reconstructing pictures from their projections. A more detailed description of the problem can be found, for example, in Frieder & Herman (1971).

We assume that $f(x, y) = 0$ outside a square region in the first quadrant of a rectangular coordinate system (see Figure 1). Further, we assume $0 \leq f(x, y) \leq 1$ everywhere. Following Rosenfeld (1969), we shall call a function with such properties a picture function. ($f(x, y) = 0$ means white, $f(x, y) = 1$ means black, with other gray levels in between.)

Let $\theta$ be an angle, $-90^\circ < \theta \leq 90^\circ$, and suppose that $l$ is a line making an angle $\theta$ with the positive $x$ axis. Suppose, further, that $l$ is divided into equal line segments and that we draw lines perpendicular to $l$ from the points separating these segments. Thus, we get a number of infinite bands, partitioning the $(x, y)$-plane. We shall refer to these bands as the rays of the projection associated with $\theta$. Suppose there are $r_\theta$ rays which actually intersect the rectangle. For $1 \leq k \leq r_\theta$, let $R_{f,k,\theta}$ denote the integral of $f(x, y)$ in the $k$th band which intersects the rectangle. (The integral need only be carried out in the shaded region in Figure 1, since $f(x, y) = 0$ elsewhere.) We shall refer to $R_{f,1,\theta}, \ldots, R_{f,r_\theta,\theta}$ as the ray-sums of the projection associated with $\theta$.

The problem of reconstructing pictures from their projections can now be stated as follows.

Give an algorithm which, given

1. the position of the square region within which $f(x, y) \neq 0$;
2. angles $\theta_1, \theta_2, \ldots, \theta_m$, together with the widths and position of rays of the projections associated with these angles;
3. the ray-sums of the projections associated with $\theta_1, \theta_2, \ldots, \theta_m$;

will produce a function $f'$ which is a good approximation of $f$.

One can only hope for an approximation, since it has been proved, even for the limiting case when the ray
widths tend to zero, that a non-trivial picture cannot be uniquely determined from its projections (Frieder & Herman [1971], Theorem 1). The algorithm ought to be such that for pictures that we are likely to be interested in, they will provide a good enough approximation. This idea is discussed in some detail by Frieder & Herman (1971). For the purpose of the present paper, the following should suffice.

As will be seen in the next two sections, reconstruction algorithms tend to be somewhat heuristic. Hence, an analytical estimate of their performance is somewhat difficult to obtain. Even if there was a method to do it, we would be faced with the problem that the functions \( f \), to which the algorithms may be applied in practice, form a small but not clearly defined subset of the set of all picture functions \( f \). Hence, analytical methods, which give equal weight to all picture functions, may give much worse estimates than what one would obtain in practice. Therefore, it appears reasonable to evaluate and compare reconstruction techniques on the basis of their performance on selected typical pictures and test patterns. This is, indeed, what we are going to do in this paper.

Since our algorithms are to be implemented on a digital computer, it is reasonable to demand that the output \( f' \) be a digitized quantized picture function (Rosenfeld [1969], Section 1.1). Indeed, in all our examples, \( f' \) will be defined on a 64\( \times \)64 matrix, and can assume as many values between 0 and 1 as the accuracy of the computer allows. In pictorial representation, we shall use only 16 equally spaced gray levels between 0 and 1, inclusive.

For convenience, we shall use as test data only picture functions which are of the same kind, i.e., 64\( \times \)64 digitized pictures with 16 equally spaced quantized gray levels. When working out the ray-sums, we shall sum the values of \( f(x, y) \) at those points which lie within the ray. At first sight, it may appear that this may change the problem in an essential way, but we shall show in the Appendix that this is not so.

In all our experiments, the rays are defined in such a way, that it is true for at least one edge of the 64\( \times \)64 matrix that each of the points on that edge is a midpoint of a ray. The reasons for this are discussed in Frieder & Herman (1971) and, more briefly, in the Appendix.

We have decided to use four test patterns (Plate A). The first two of these are binary valued picture functions (black and white) which have been used by Vainshtein in his demonstrations of the summation method. (A1 comes from Vainshtein [1971a] and A2 from Vainshtein [1971b].) We are not aware of any other previously published test patterns on which the summation method was demonstrated. The situation is better concerning ART, Gordon, Bender & Herman (1970), Herman & Rowland (1971) and Bellman, Bender, Gordon & Rowe (1971), all contain a number of test patterns showing the operation of ART. We have decided to use two of these for the present comparative study, both half-tone (i.e., all 16 gray levels occur in them). One is the face of a little girl, Judy (A3), the other is the photograph of some stomata (A4). The first of these has been used by Gordon, Bender & Herman (1970) and Gordon & Herman (1971), the second has been used by Herman & Rowland (1971). Since in previous studies these pictures were not digitized at 64\( \times \)64 points, we kept the previous digitization (49\( \times \)49 and 50\( \times \)50 points) and inserted them into a white frame. This makes comparison with previously published tests concerning these patterns easier.
For each of the four pictures, we have carried out
reconstruction operations using four different sets of
projection angles. The first three sets have been used by
Vainshtein in his demonstration of the summation
method. In all three sets the projection angles are
equally spaced between \(-90^\circ\) and \(+90^\circ\) (with \(90^\circ\) being
one of the angles), and there are 4, 8 and 30 angles,
respectively. We used these sets of angles, so that our
results can be directly compared with the results of
Vainshtein (1971a, b).

The fourth set of projection angles comes from a
small range. This is particularly important in electron-
microscopic applications, because of the restricted
range of the tilting stage. We have, in fact, used the
same set of projection angles that have been used by
Bender, Bellman & Gordon (1970) in their reconstruc-
tion (using ART) of ribosomes from electronmicro-
graphs. We have rotated these angles so that none of
them is near \(0^\circ\) or \(90^\circ\). The reason for this is that both
ART and the summation technique would tend to give
unusually good results if the edges in a test pattern
(like A1) are parallel to one of the projection angles.
The fourth set of projection angles which we used was
actually \(35^\circ\), \(44^\circ\), \(62^\circ\), \(80^\circ\), \(98^\circ\) (\(= -82^\circ\)) and
\(116^\circ\) (\(= -64^\circ\)). Thus, all these views come from a
range of \(81^\circ\) as opposed to a full range of \(180^\circ\).

Four pictures, with four sets of projections each,
provide us with 16 reconstructions to be carried out
with each of the techniques we are investigating. Then
we are faced with the problem of evaluating the success
of the reconstructions. There is no standard way of
doing this. We have decided to use four different sets of
criteria: overall nearness of the original and recon-
structed pictures, resolution of fine detail, visual
evaluation, and cost of reconstruction. We shall now
discuss these criteria in some detail.

**Overall nearness**

(1) Gordon, Bender & Herman (1970) and Gordon
& Herman (1971) used the root mean square distance
\(\delta\) as a measure of the overall nearness of the original
and reconstructed pictures.

\[
\delta = \left[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (f(x_{i,j}, y_{i,j}) - f'(x_{i,j}, y_{i,j}))^2 \right]^{1/2},
\]

where \(x_{i,j}\) and \(y_{i,j}\) are the \(x\) and \(y\) coordinates of the
\((i, j)\)th point in the \(n \times n\) matrix. The validity of such
a measure has been questioned by Crowther & Klug
(1971), but the discussion by Frieder & Herman
(1971, especially Theorem 2) shows that \(\delta\) is a very
reasonable measure for the overall performance of a
reconstruction technique.

(2) Ramachandran & Lakshminarayanan (1971a, b)
use another measure \(R\) of overall nearness in their
comparison of the Fourier techniques with their con-
volution method. This is adopted from X-ray crys-
tallography, and we shall work it out as well as the \(\delta\)
measure for all our experiments.

\[
R = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |f(x_{i,j}, y_{i,j}) - f'(x_{i,j}, y_{i,j})|}{\sum_{i=1}^{n} \sum_{j=1}^{n} f(x_{i,j}, y_{i,j})}.
\]

Thus, \(R\) is the mean relative error of the reconstruction.

(3) For a number of applications one wants to make
the assumption that \(f'\) is a binary valued picture
function. This is obviously so if we know that \(f\) is
binary valued. Even if \(f\) is not binary valued, we may
wish to contour \(f'\) to get some idea of the overall shape
of the object which is in the picture \(f\). Let

\[
f(x, y) = \begin{cases} 1, & \text{if } f(x, y) \geq 0.5, \\ 0, & \text{if } f(x, y) < 0.5, \end{cases}
\]

and let \(f'\) be similarly defined. We shall refer to \(f\) and
\(f'\) as the contoured versions of \(f\) and \(f'\). (Naturally,
for a binary valued picture function \(f, f' = f\).) We shall also
work out the mean relative error, \(\tilde{R}\), for the contoured
versions of \(f\) and \(f'\).

\[
\tilde{R} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |\tilde{f}(x_{i,j}, y_{i,j}) - \tilde{f}'(x_{i,j}, y_{i,j})|}{\sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{f}(x_{i,j}, y_{i,j})}.
\]

The measure \(\tilde{R}\) has also been used by Chang &
Shelton (1971) in comparing two algorithms for binary
pattern reconstructions.

(4) A property of the output of a reconstruction
should be that it is consistent with the information
available to the algorithm. In other words, the ray-
sums of the reconstructed picture should be the same
as the ray-sums of the original. Even though there are
algorithms which achieve this (Gordon & Herman
[1971]), and, for binary valued pictures only, Chang
[1970], Chang & Shelton [1971]), they usually tend
to be slow for practical applications. Both ART and
the summation technique will produce pictures where
the ray-sums will only approximate the given data.
One appropriate measure of the overall nearness of the
reconstruction is the root mean square distance \(p\)
between the ray-sums on which the reconstruction is based and the ray-sums of the reconstructed picture.

\[ p = \left( \frac{1}{\rho_{a, b}} \sum_{i=1}^{m} \sum_{j=1}^{n} (R_{f,j,k} - R'_{f,j,k})^2 \right)^{1/2}, \]

where \( m \) is the number of projections and \( R_{f,j,k} \) denotes the ray-sum in the \( k \)’th ray of the \( l \)’th projection for the reconstructed picture \( f' \).

(5) A simple way of deciding the overall success of a reconstruction is to ask whether the reconstructed pattern resolves the original. Based on a suggestion of Harris (1964), Frieder & Herman (1971) used the following criterion for answering this question. A reconstruction resolves a picture, if the output is nearer (in the sense of \( \delta \)) to the original than to uniform gray. Since the mean gray level of input and output is the same both for ART and for the summation technique, the distance from uniform gray of the output is nothing but the standard deviation of the output. In our experiments, we shall say that the output of the reconstruction resolves the original if, and only if, the standard deviation of the output is greater than \( \delta \).

Resolution in fine detail

The major objection of Crowther & Klug (1971) to \( \delta \) as a measure of difference between the reconstruction and the original is that “it reaches a low value once the large scale features are correct and is relatively insensitive to errors in the fine details.” As it was pointed out by Frieder & Herman (1971), this objection is not quite valid. However, it is reasonable to ask what is the guaranteed maximum error in the average grayness of a region of certain size in the reconstructed picture. With this idea in mind, Frieder & Herman (1971) introduced the notion of \((a, l, \epsilon)\)-resolution. They said that two picture functions \( f \) and \( g \) with \( 4a^2 \) non-zero area \((a, l, \epsilon)\)-resolve each other if, for every square region of size \( b \), the mean grayness of the pictures in that region differs by less than \( \epsilon \).

The problem with \((a, l, \epsilon)\)-resolution is that it is difficult to calculate. For example, if we wanted to find the minimal \( \epsilon \) such that an original and a reconstructed 64X64 picture \((32, 8, \epsilon)\)-resolve each other, we would have to work out the mean gray value of nearly 6000 regions, each with 64 points. We have therefore devised another method, which is a good approximation to \((a, l, \epsilon)\)-resolution, but which is computationally much simpler.

For \( 0 \leq l \leq 6 \), let

\[ \epsilon(l) = \max_{0 \leq e \leq 2^{l-1}} \left[ \sum_{0 \leq i \leq 2^{l-1}} \sum_{0 \leq j \leq 2^{l-1}} \left| f(x_{i,2^l-1}, y_{j,2^l-1}) - f'(x_{i,2^l-1}, y_{j,2^l-1}) \right| \right] \]

In other words, \( \epsilon(l) \) is the maximum difference between \( f \) and \( f' \), when they are both digitized as \((64/2^l) \times (64/2^l)\) pictures. Thus, \( \epsilon(0) \) is the maximum difference between the values at any point of \( f \) and \( f' \), while \( \epsilon(6) \) is the difference between the mean gray values of \( f \) and \( f' \).

For every \( l \), \( \epsilon(l) \) indicates the point by point reliability of the reconstruction when both the original and the reconstructed picture are digitized as \((64/2^l) \times (64/2^l)\) pictures. For example, if \( \epsilon(2) = .01 \), and the gray level in the 16X16 digitized version of the reconstructed picture is .253 at a certain point, then we know that the gray level in the 16X16 digitized version of the original is between .243 and .263.

The notion expressed above is easily generalized for test patterns which are digitized as \( 2^n \times 2^n \) matrices, for any \( k \). Also, \( \epsilon(l) \) is computationally easy to calculate, it only requires a repeated increase of the roughness of digitization by a factor of two.

For each of our experiments, we shall give the values of \( \epsilon(0), \epsilon(1), \ldots, \epsilon(5) \). \( \epsilon(6) \) is always zero.

An alternative way of representing our results regarding resolution in fine detail is to state the size of detail which is reliable within a certain error. For any \( \epsilon > 0 \), \( l(\epsilon) \) will denote the maximum number \( n \), such that \( n \) is a power of two and the \( n \times n \) digitization of the original and the reconstructed pictures differ by less than \( \epsilon \) at all points. Thus, \( l(0.1) = 32 \) will mean that the original and the reconstructed pictures digitized at 32X32 points will differ at each point by less than 0.1, but if they are digitized at 64X64 points, then they differ by more than 0.1 at least at one point. So, \( l(\epsilon) \) is roughly the resolution of the reconstruction if the error tolerance is \( \epsilon \). For each of our experiments, we shall work out \( l(0.01), l(0.02), l(0.05), l(0.10), l(0.25) \text{ and } l(0.50) \). Clearly, there is no point in calculating \( l(\epsilon) \) for \( \epsilon > 0.5 \).

Visual evaluation

The methods mentioned in (a) and (b) above include all computationally possible ways of measuring the success of reconstructions which are known to us. We left out measuring resolution by the use of Fourier transforms (see \((a, l, \epsilon)\)-resolution in Fourier space in the paper by Frieder & Herman (1971)), because of
its computational difficulty, and its inappropriateness for comparing two direct methods. Even though the mathematical measures we used should provide quite conclusive evidence of the relative merits of two reconstruction techniques, there has not yet been enough experience with them to know the correlation between the values of the measures and visual acceptability of the reconstruction. For this reason, we shall give for each of our experiments the reconstructed picture, which can then be compared with the original. When the test-pattern is a binary valued picture function, we shall also give the contoured version of the reconstructed picture.

Cost of reconstruction

This is obviously an important measure. In practical applications, especially three-dimensional reconstruction, one has to carry out a large number of reconstructions, and cost can be a prohibitive factor.

To insure that our cost comparisons are valid, we have incorporated both ART and the summation method in the same general program written in FORTRAN. They make use of the same set of service subroutines, e.g., the one for finding out which points lie in a particular ray.

Time, rather than cost is given, since it is easier to obtain. However, cost appears to be a more stable measure between installations. Since all experiments have been run on a CDC 6400, at a cost of $47.5 per hour, the reader can easily work out the actual cost of the runs. (Rate may vary from installation to installation.) Run time is given in seconds.

DESCRIPTION OF ART

The method we shall describe now is one of the algebraic reconstruction techniques of Gordon, Bender & Herman (1970), the one which was called in that paper the direct additive method. Other papers relevant to this method are Bender, Bellman & Gordon (1970), Crowther & Klug (1971), Bellman, Bender, Gordon & Rowe (1971), Frieder & Herman (1971) and Herman & Rowland (1971).

The basic idea of the method is the following. Starting from a blank picture, the ray-sums of all the projections are satisfied one after the other by distributing the difference between the desired ray-sum and the actual ray-sum equally amongst all the points in the ray. While satisfying the ray-sums of a particular projection, the process usually disturbs the ray-sums of previously satisfied projections. However, as we repeatedly go through satisfying all the projections, the disturbances get smaller and smaller, and eventually the method converges to a picture which satisfies all the projections. Because we start with a uniform blank, and while satisfying each of the projections we make as uniform changes as possible, the final product tends to be as smooth as a picture satisfying the given projections can possibly be. For practical applications, this seems to be a desirable property of a reconstruction algorithm. Roughly speaking, our reconstructed picture will show only features which are forced upon it by the projections, rather than features which are introduced by the reconstruction process.

Mathematically, let $f'$ be the partially reconstructed picture just before we wish to satisfy the projection associated with the angle $\theta$. Let $(x_{i,j}, y_{i,j})$ lie on the $k$'th ray of that projection, and let $N_{k,\theta}$ be the number of points on that ray. Then $f'$ is changed into $f''$ by the rule

$$f''(x_{i,j}, y_{i,j}) = f'(x_{i,j}, y_{i,j}) + (R_{f',k,\theta} - R_{f,k,\theta})/N_{k,\theta}.$$

If the value of $f''$ obtained in this way is negative, it is rounded to zero, if its value is greater than one, it is rounded to one.

The process of satisfying all projections one after the other exactly once is referred to as one iteration. The accuracy of ART increases with the number of iterations, and we report on the results of experiments after 1, 5 and 25 iterations.

DESCRIPTION OF THE SUMMATION METHOD

This method has been described by Vainshtein (1971a, b), Gordon, Bender & Herman (1970), Gaarder & Herman (1971), Ramachandran & Lakshminarayanan (1971b). The most detailed discussion of its properties is contained in the work of Vainshtein (1971a, b).

Roughly speaking, the idea is to distribute each of the ray-sums equally amongst all the points in the corresponding ray. If there are $m$ projections, this will result in a picture whose total density is $m$ times what it should be. We therefore subtract from the value at each point $(m-1)d$, where $d$ is the mean density of the picture. (The mean density can be worked out from the ray-sums.) Rounding negative values to zero and values greater than one to one, we get our first approximation to the input.

This procedure has a lot to recommend it. It is conceptually and computationally simple. It can be implemented by a photo-summation device without the use of a digital computer. It seems to provide us with a smooth picture. In fact, if the projections and rays
satisfy certain conditions, the result of the algorithm will be a picture which satisfies the projections and for which the variance of gray levels is smaller than for any other picture satisfying the projections (see Gaarder & Herman [1971]).

However, the algorithm has some rather peculiar properties which make it often useless in practice. It appears that as a result of rounding to zero and one, the average density of the picture increases. Hence, we get, with 30 projections, the picture in A5 corresponding to the test pattern in A1. Vainshtein (1971a) recommends that the output should be contoured at a level so that the total density of input and output are the same. However, this advice can only be followed for binary-valued picture functions. To improve the quality of the output for arbitrary picture functions, we carried out the following process.

First of all, if a ray-sum is zero, we know that the value of the function at all points on that ray is zero. The first step in our modified algorithm is to mark all points which lie on rays with ray-sums equal to zero.

Then we carry out the process described above, but equally distributing the ray-sums only amongst the unmarked points in the ray. After subtracting \((m-1)d\) from the value at each point, we round all negative values to zero.

At this stage, the mean density \(\bar{d}\) of the reconstructed picture will usually be greater than \(d\). So we multiply the value of the reconstructed function at each point by \(d/\bar{d}\), making the mean density of the reconstruction equal to that of the original. We should at this point round all values greater than 1 to 1, but there was no need for this in any of the 16 experiments that we tried.

We found that this modified summation method gave much better results than the simpler technique described at the beginning of the section. For example, for the test-pattern A1 and 30 projections, we get the following comparisons.

<table>
<thead>
<tr>
<th>(\delta)</th>
<th>(\epsilon(0))</th>
<th>(\epsilon(2))</th>
<th>(\epsilon(4))</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple method</td>
<td>0.32</td>
<td>1.0</td>
<td>1.0</td>
<td>0.44</td>
</tr>
<tr>
<td>modified method</td>
<td>0.18</td>
<td>0.75</td>
<td>0.57</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Clearly, the large improvement is worth the small additional cost. In all the experiments reported on in the next section, we used the modified summation method.

A final comment. ART and the summation method have been referred to (Crowther & Klug [1971]) as "very similar." The descriptions above clearly indicate that this is not so, and, as can be seen from the next section, the power of the two methods is quite different. It is therefore invalid to draw conclusions about the behavior of one of these methods from the observed behavior of the other one. The two methods could be combined in an iterative summation method, where after each iteration the difference between the desired ray-sums and actual ray-sums is simultaneously equally distributed amongst the points on the rays. This method should give similar results to ART, but it requires a considerably larger memory in a digital computer implementation.

RESULTS OF THE EXPERIMENTS

We summarize the results of our experiments in the following tables. Table I contains results on \(\delta, R, \bar{R}, \rho\), whether the reconstruction resolves the original, and time, \(t\). Table II contains \(\epsilon(0), \epsilon(1), \epsilon(2), \epsilon(3), \epsilon(4)\) and \(\epsilon(5)\). Table III contains \(l(0.01), l(0.02), l(0.05), l(0.10), l(0.25)\) and \(l(0.50)\). These are marked as 1 percent, 2 percent, 5 percent, 10 percent, 25 percent, and 50 percent in the table.

In all the tables, the numbers 4, 8, 30 and 6 on the top refer to the number of projections. The first three sets are equally spaced, while the last set consists of the angles 35°, 44°, 62°, 80°, 98° and 116°.

The vertical arrangement of the four numbers in each of the entries of the table refer to the result by the summation method, followed by the results by ART after 1, 5 and 25 iterations.

As can be seen from these tables, if the number of projections is small, one iteration of ART seems to perform somewhat worse than the summation method. As the number of projections increases, the relative performance of one iteration of ART as opposed to the summation method improves, and eventually it produces superior results. The break-even point seems to be at eight projections. However, by repeated iterations, the performance of ART can be improved and it eventually surpasses the summation method in all our experiments. The improvement in ART is especially impressive with eight or more projections. In such cases, we obtain, with five iterations of ART, results which are far superior to the results of the summation method. These conclusions are clearly demonstrated on the plates showing the reconstructed pictures.

Plate B contains reconstructions of the wrench (A1), Plate C contains reconstructions of \(\rho(R)\) (A2), Plate D contains reconstructions of Judy (A3), and Plate E contains reconstructions of the stomata (A4). Plates F and G contain the contoured versions of the reconstructions of the wrench and \(\rho(R)\).
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>WRENCH (A1)</th>
<th>r (R) (A2)</th>
<th>JUDY (A3)</th>
<th>STOMATA (A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
</tr>
<tr>
<td></td>
<td>.22 .15 .13 .27</td>
<td>.24 .21 .13 .24</td>
<td>.17 .13 .10 .22</td>
<td>.20 .17 .11 .26</td>
</tr>
<tr>
<td></td>
<td>.17 .07 .02 .21</td>
<td>.21 .15 .05 .19</td>
<td>.14 .10 .03 .20</td>
<td>.18 .15 .03 .23</td>
</tr>
<tr>
<td></td>
<td>.11 .03 .00 .15</td>
<td>.19 .10 .00 .15</td>
<td>.13 .10 .03 .19</td>
<td>.18 .12 .03 .22</td>
</tr>
<tr>
<td>( R )</td>
<td>.60 .50 .48 .91</td>
<td>1.1 .99 .87 .11</td>
<td>.32 .30 .28 .54</td>
<td>.42 .40 .37 .62</td>
</tr>
<tr>
<td></td>
<td>.82 .62 .35 1.1</td>
<td>1.2 1.0 .54 1.2</td>
<td>.41 .32 .23 .54</td>
<td>.48 .40 .25 .63</td>
</tr>
<tr>
<td></td>
<td>.49 .16 .03 .69</td>
<td>.91 .56 .08 .79</td>
<td>.29 .20 .07 .45</td>
<td>.40 .28 .06 .55</td>
</tr>
<tr>
<td></td>
<td>.26 .05 .00 .41</td>
<td>.67 .29 .00 .49</td>
<td>.27 .19 .06 .42</td>
<td>.38 .25 .05 .51</td>
</tr>
<tr>
<td>( \rho )</td>
<td>.03 .05 .05 .08</td>
<td>.08 .07 .04 .09</td>
<td>.11 .10 .10 .20</td>
<td>.17 .15 .15 .22</td>
</tr>
<tr>
<td></td>
<td>.06 .01 .01 .09</td>
<td>.08 .05 .01 .09</td>
<td>.16 .10 .06 .19</td>
<td>.17 .13 .05 .21</td>
</tr>
<tr>
<td></td>
<td>.02 .00 .00 .05</td>
<td>.05 .02 .00 .04</td>
<td>.12 .08 .03 .19</td>
<td>.16 .10 .01 .18</td>
</tr>
<tr>
<td></td>
<td>.01 .00 .00 .02</td>
<td>.04 .01 .00 .02</td>
<td>.11 .07 .02 .19</td>
<td>.16 .10 .01 .19</td>
</tr>
<tr>
<td>( a )</td>
<td>24 24 23 25</td>
<td>22 22 21 23</td>
<td>15 17 17 18</td>
<td>14 17 20 19</td>
</tr>
<tr>
<td></td>
<td>19 17 15 20</td>
<td>15 15 12 17</td>
<td>15 12 13 14</td>
<td>13 13 14 18</td>
</tr>
<tr>
<td>( t )</td>
<td>7.3 3.6 1.2 8.3</td>
<td>5.9 5.0 1.5 6.3</td>
<td>2.9 1.6 .87 3.1</td>
<td>2.5 2.2 .88 4.0</td>
</tr>
<tr>
<td>( \rho )</td>
<td>2.8 .92 .01 2.8</td>
<td>1.8 1.8 .03 2.0</td>
<td>.54 .35 .16 .94</td>
<td>.42 .52 .12 1.0</td>
</tr>
</tbody>
</table>

In all plates, the first row gives the reconstructions based on 4 equally spaced projections, the second row gives the reconstructions based on 8 equally spaced projections, the third row gives the reconstructions based on 30 equally spaced projections and the fourth row gives the reconstructions based on the 6 projections from the 81° range. Within each row, the first picture is the result obtained by the summation method, which is followed by the results obtained by ART after 1, 5 and 25 iterations, respectively.

There are a number of observations worth making regarding the results of our experiments.

First of all, the basic measures \( \delta \) and \( R \) give, with hardly any exceptions, the same ordering between different experiments for the same picture. There seems to be little reason to use the one rather than the other. The orderings based on these measures are also in very strong correspondence with the orderings based on any of the resolution measures.

Another interesting point is that there is no instance when a picture reconstructed by ART does not resolve the original after five iterations.

The third observation is in reference to Table III. Very little information can be gained from a digitization with fewer than 8X8 points. In all our experiments the summation method produced a picture whose 8X8 version differs from the 8X8 original at some point by at least 0.1. In fact, for the wrench (A1), even 30 projections can only produce a picture, where the difference between the 8X8 pictures is more than .25, at least at one point.

It has been claimed for ART (Frieder & Herman [1971]) that for an \( n \times n \) resolution it requires approximately \( n \) equally spaced projections. Table III bears
<table>
<thead>
<tr>
<th>$p(r)$ (R) (A2)</th>
<th>JUDY (A3)</th>
<th>STOMATA (A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
</tr>
<tr>
<td>$\varepsilon$ (0)</td>
<td>0.84 0.79 0.75 0.81</td>
<td>0.82 0.68 0.62 0.83</td>
</tr>
<tr>
<td>$\varepsilon$ (1)</td>
<td>0.78 0.67 0.69 0.75</td>
<td>0.80 0.67 0.69 0.83</td>
</tr>
<tr>
<td>$\varepsilon$ (2)</td>
<td>0.81 0.54 0.21 0.65</td>
<td>0.90 0.80 0.50 0.70</td>
</tr>
<tr>
<td>$\varepsilon$ (3)</td>
<td>0.90 0.80 0.00 0.80</td>
<td>0.90 0.80 0.00 0.80</td>
</tr>
<tr>
<td>$\varepsilon$ (4)</td>
<td>0.82 0.68 0.00 0.80</td>
<td>0.83 0.67 0.00 0.80</td>
</tr>
<tr>
<td>$\varepsilon$ (5)</td>
<td>0.83 0.68 0.00 0.80</td>
<td>0.90 0.80 0.00 0.80</td>
</tr>
</tbody>
</table>

From the collection of the Computer History Museum (www.computerhistory.org)
<table>
<thead>
<tr>
<th></th>
<th>WRENCH (A1)</th>
<th>$p$ (R) (A2)</th>
<th>JUDY (A3)</th>
<th>STOMATA (A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
<td>4 8 30 6</td>
</tr>
<tr>
<td><strong>1%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>2 1 1 1</td>
<td>1 2 1 1</td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>1</td>
<td>2 4 1</td>
<td>2 1 4 1</td>
<td>2 2 4 2</td>
<td>2 2 8 2</td>
</tr>
<tr>
<td>1</td>
<td>4 64 1</td>
<td>2 4 32 2</td>
<td>2 4 8 2</td>
<td>2 2 8 2</td>
</tr>
<tr>
<td><strong>2%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 1</td>
<td>1 2 2 1</td>
<td>2 2 2 1</td>
<td>2 2 2 1</td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>2 2 1 1</td>
<td>1 2 2 1</td>
</tr>
<tr>
<td>1</td>
<td>4 8 1</td>
<td>2 2 8 2</td>
<td>2 4 16 2</td>
<td>2 4 16 2</td>
</tr>
<tr>
<td>2</td>
<td>8 64 2</td>
<td>2 4 64 2</td>
<td>4 4 16 2</td>
<td>2 4 16 2</td>
</tr>
<tr>
<td><strong>5%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 2 2 2</td>
<td>2 2 2 2</td>
<td>2 2 2 1</td>
<td>2 2 2 1</td>
</tr>
<tr>
<td>1</td>
<td>2 2 2 1</td>
<td>2 2 4 2</td>
<td>2 4 2 2</td>
<td>2 2 2 1</td>
</tr>
<tr>
<td>2</td>
<td>8 16 2</td>
<td>4 4 16 2</td>
<td>4 4 16 2</td>
<td>4 4 16 2</td>
</tr>
<tr>
<td>4</td>
<td>8 64 2</td>
<td>4 8 64 4</td>
<td>4 4 16 2</td>
<td>4 4 16 4</td>
</tr>
<tr>
<td><strong>10%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 4 2</td>
<td>4 4 4 2</td>
<td>2 2 2 2</td>
<td>4 4 4 2</td>
</tr>
<tr>
<td>2</td>
<td>4 4 2</td>
<td>4 4 4 2</td>
<td>4 4 4 4</td>
<td>4 4 4 2</td>
</tr>
<tr>
<td>4</td>
<td>8 32 2</td>
<td>4 8 32 4</td>
<td>4 8 32 4</td>
<td>4 8 32 4</td>
</tr>
<tr>
<td>8</td>
<td>16 64 4</td>
<td>4 8 64 8</td>
<td>4 8 32 4</td>
<td>8 8 32 4</td>
</tr>
<tr>
<td><strong>25%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4 4 2</td>
<td>8 8 8 4</td>
<td>8 4 8 4</td>
<td>8 8 8 4</td>
</tr>
<tr>
<td>4</td>
<td>8 8 2</td>
<td>8 8 16 4</td>
<td>4 8 8 4</td>
<td>8 8 8 4</td>
</tr>
<tr>
<td>8</td>
<td>16 64 4</td>
<td>8 8 32 8</td>
<td>8 16 64 4</td>
<td>8 8 64 4</td>
</tr>
<tr>
<td>8</td>
<td>32 64 8</td>
<td>8 16 64 8</td>
<td>8 16 64 4</td>
<td>8 8 64 8</td>
</tr>
<tr>
<td><strong>50%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8 8 4</td>
<td>8 8 16 8</td>
<td>16 64 64 8</td>
<td>16 16 32 8</td>
</tr>
<tr>
<td>8</td>
<td>16 16 4</td>
<td>8 16 16 8</td>
<td>16 16 16 8</td>
<td>8 16 32 8</td>
</tr>
<tr>
<td>16</td>
<td>32 64 8</td>
<td>16 16 64 16</td>
<td>16 64 64 8</td>
<td>16 32 64 16</td>
</tr>
<tr>
<td>32</td>
<td>64 64 16</td>
<td>16 32 64 16</td>
<td>16 64 64 16</td>
<td>16 32 64 16</td>
</tr>
</tbody>
</table>

Plate D

Plate E

From the collection of the Computer History Museum (www.computerhistory.org)
this out. For the binary valued pictures with \( n \) equally spaced projections the error at \( n \times n \) resolution is less than 0.05 at all points, while for the half-tone pictures it is less than 0.1. (This is after 25 iterations, sometimes sooner.)

For the case of 30 projections, for all four test patterns, the error at each point in the 64\( \times \)64 reconstructed picture after 25 iterations of ART is less than .175, and in the case of the binary valued pictures it is actually less than .015 (see Table II).

For perfect contouring we need an error which is less than .5 at each point. Thus, we see that contoured 16\( \times \)16 reconstructions of all the pictures are perfect, with all four sets of angles, when ART is used with 25 iterations. For the summation method, the same statement holds only for 4\( \times \)4 reconstructions.

At two points, our results do not strictly correspond to results in earlier publications.

The first of these is the bad quality of the contoured output of the summation method (Plates F and G). The contoured output obtained by Vainshtein (1971a) in the same experiment appears to be superior. The reason for this difference is that Vainshtein contoured at a level which will make the contoured input and output to have the same density, while we made the half-tone output to have the same density as the input, and then contoured at .5. The reader can easily judge for himself, from Plates B and C, what the effect of contouring at different levels would have been.

The other point concerns the performance of ART on the half-tone pictures using 6 projections from the 81° range. The results appear to be much worse than what have previously been obtained either for Judy (Gordon, Bender & Herman [1970]) or the stomata (Herman & Rowland [1971]). There are two reasons for this. One is that we are putting a fairly wide white frame around the picture. This frame does not contribute to the total density, but the fact that it goes all the way around the picture cannot possibly be detected from projections taken in a small range. Since ART is intended to produce a picture as smooth as possible while satisfying the projections, it will smooth out the picture into that part of the original white frame which is not uniquely determined from the projections. This will also bring with itself a general lowering of density in the middle portion of the picture, causing some distortions in trying to satisfy some of the projections.

The second reason for the bad quality of the reconstructions of the stomata, as opposed to earlier experiments with angles within the same range, is that previously all the projection angles clustered around the horizontal, while due to the rotation discussed in Section two, now they cluster around the vertical. The stomata seem to have many horizontal bands, thus, reconstructions from projections from a small horizontal range produce better results, than from a small vertical range. (This is not the case with Judy.) The countermeasure to being a victim of the orientation of the object relative to the small range of views is to average a number of independent reconstructions (see Herman & Rowland [1971]). The following table gives the value of \( \delta \) when the Judy and the stomata test patterns are recon-
structured, with and without a white band around them, with six views from two small ranges, as well as the average in each case. All results are based on ART after 25 iterations.

<table>
<thead>
<tr>
<th>Angles</th>
<th>Angles</th>
</tr>
</thead>
<tbody>
<tr>
<td>-45°, -36°, -18°, 6°, 18°, 36°</td>
<td>35°, 44°, 62°, 80°, 98°, 116°</td>
</tr>
</tbody>
</table>

| Judy | .14 | .16 | .13 |
| Judy | .19 | .19 | .16 |
| stomata | .16 | .23 | .17 |
| stomata | .17 | .22 | .17 |

As can be seen from this table, averaging considerably improves the quality of the reconstruction, for the 64×64 pictures. Using a smaller frame can also be helpful in this respect. Thus, the bad results we obtained for half-tone pictures with six projections from a small range can be avoided by a careful choice of the square in which \( f(x, y) \) is assumed to be possibly non-zero (it should be as small as possible), or by averaging. Naturally, the same comments apply to the summation method as well.

ACKNOWLEDGMENTS

This research has been supported by NSF Grant GI998. The author is most grateful to Mr. Stuart Rowland for his constant help with programming and in other matters. The following colleagues have also been helpful both by enlightening discussions and in producing the actual pictures: R. Bender, G. Frieder, R. Gordon and J. Rowe. The pictures have been produced by the courtesy of Optronics International, Chelmsford, Massachusetts.

REFERENCES AND BIBLIOGRAPHY

S H BELLMAN R BENDER R GORDON J E ROWE JR
ART is science, being a defense of algebraic reconstruction techniques for three-dimensional electron microscopy

J Theor Biol 32 pp 205-216 1971
R BENDER S H BELLMAN R GORDON
ART and the ribosome: a preliminary report on the three-dimensional structure of individual ribosomes determined by an algebraic reconstruction technique
J Theor Biol 29 pp 483-487 1970
S K CHANG
The reconstruction of binary patterns from their projections
Comm ACM 14 pp 21-25 1971
S K CHANG G L SHELTON
Two algorithms for multiple-view binary pattern reconstruction
IEEE Transactions on Systems Man and Cybernetics 1 pp 90-94 1971
R A CROWTHER
Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs
Phil Trans Roy Soc Lond B 261 pp 221-230 1971
R A CROWTHER L A AMOS J T FINCH
D J DEROISIER A KLUG
Three dimensional reconstructions of spherical viruses from electron micrographs
R A CROWTHER D J DEROISIER A KLUG
The reconstruction of a three-dimensional structure from projections and its application to electron microscopy
R A CROWTHER A KLUG
ART and science, or, conditions for 3-D reconstruction from electron microscope images
J Theor Biol 32 199-203 1971
D J DEROISIER
Three dimensional image reconstruction of helical structures
Berichte der Bunsen-Gesellschaft 74 pp 1127-1128 1970
D J DEROISIER
Three-dimensional image reconstruction of helical structures
D J DEROISIER
Reconstruction of three-dimensional images from electron micrographs
Biophysical Society Abstracts Fifteenth Annual Meeting February 15-18 New Orleans p 218a 1971
D J DEROISIER A KLUG
Reconstruction of three dimensional structures from electron micrographs
Nature 217 pp 130-134 1968
D J DEROISIER P B MOORE
Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry
J Mol Biol 52 pp 355-369 1970
J T FINCH A KLUG
Three-dimensional reconstruction of the stacked-disk aggregate of tobacco mosaic virus protein from electron micrographs
Phil Trans Roy Soc Lond B 261 pp 211-219 1971
G FRIEDER G T HERMAN
Resolution in reconstructing objects from electron micrographs
J Theor Biol 33 pp 180-211
N T GAARDER G T HERMAN
Algorithms for reproducing objects from their x-rays

From the collection of the Computer History Museum (www.computerhistory.org)
APPENDIX

Justification of the use of digitized rather than continuous test patterns

A basic assumption of any picture reconstruction process is that the picture is reasonably smooth relative to the size of the ray widths. It is clearly unreasonable to expect an algorithm to reconstruct detail which is much finer than the width of the rays.

Since our digitization was done approximately on the same scale as the ray width (i.e., the distance between two adjacent points in the matrix on which the digitized picture is defined is of the same order of magnitude as the ray-widths), we may conclude that values of the digitized picture at neighboring points are highly correlated. Furthermore, by our smoothness assumption, if we divide the continuous picture into squares with centers at the points of digitization, at all points in the squares the gray value of the picture will be near to the digitized value at the center.

The way we have chosen the position and width of the rays is such that relative positioning of the squares and the rays will be as shown in Figure 2.

In the digitized version, the contribution to the indicated ray in Figure 2 will come from the value at the center.
Figure 2

The center of the square, which is supposed to equal the total grayness in the square. In the continuous version, however, regions \(a\) and \(b\) in the square do not contribute to this ray, while regions \(a'\) and \(b'\), which are outside the square, do contribute. However, the area of \(a\) is the same as that of \(a'\) and area of \(b\) is the same as that of \(b'\). Hence, if the picture function is smooth relative to the ray widths, the ray-sums of the digitized and continuous versions of the picture are approximately the same.

To demonstrate that this argument works in practice, we took ray-sums of a continuous version of the test pattern in AI, with a genuinely circular boundary. The following shows the order of differences in our measurements when we used the continuous and discrete test patterns. The discrete pattern is not binary valued, rather it is the \(64 \times 64\) digitization of the continuous pattern with 16 gray levels. All results refer to the experiments with four projections, ART and five iterations.

<table>
<thead>
<tr>
<th></th>
<th>(\delta)</th>
<th>(R)</th>
<th>(\rho)</th>
<th>(\epsilon(1))</th>
<th>(\epsilon(3))</th>
<th>(\epsilon(5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuous test pattern</td>
<td>.1642</td>
<td>.4461</td>
<td>7.515</td>
<td>.5755</td>
<td>.2315</td>
<td>.0168</td>
</tr>
<tr>
<td>digitized test pattern</td>
<td>.1633</td>
<td>.4429</td>
<td>7.230</td>
<td>.5658</td>
<td>.2303</td>
<td>.0161</td>
</tr>
</tbody>
</table>

It appears that there is no essential difference between the results of the two experiments. Other tests gave similar results (see, e.g., Herman & Rowland [1971]). Since the use of digitized test patterns makes experimentation considerably simpler, it seems advisable to use them, rather than continuous test patterns.