EVALUATION OF IMAGE-RECOVERY ALGORITHMS ON THE BASIS OF TASK PERFORMANCE

Kenneth M. Hanson#
Thomson CGR - Service Etudes Probatoires
283, rue de la Minière
78530 BUC, FRANCE

Nous présentons une méthode nouvelle d'évaluation des algorithmes de restauration d'images, fondée sur une estimation numérique de l'accomplissement d'une tâche spécifiée. Dans cette méthode, nous utilisons une technique de simulation de Monte Carlo pour la génération de scènes appropriées à l'application désirée, la saisie de données, la restauration puis l'analyse des images ainsi obtenues. La qualité de l'algorithme de restauration est estimée au moyen d'un indice mesurant la qualité de l'exécution de la tâche. L'utilité de cette approche est mise en évidence dans une étude de la technique de reconstruction algébrique (ART) d'une image à partir de ses projections, au cours de laquelle est évaluée l'efficacité de l'addition d'une contrainte de non-négativité pour une tâche de détection. La comparaison entre les résultats de détection obtenus à partir des images reconstruites par ART avec et sans contrainte montre que cette contrainte améliore en général la détectabilité, en particulier dans le cas où l'on dispose d'un nombre limité de projections non bruitées.

A new method for evaluating image-recovery algorithms is presented, which is based upon the numerical assessment of the performance of a specified imaging task. In this method a Monte Carlo technique is used to simulate the full imaging sequence including the generation of the scenes appropriate to the desired application, subsequent data taking, image recovery, and analysis of the final images. The evaluation of the algorithm is made on the basis of a task performance index, which measures how well the task is performed. The usefulness of this approach is demonstrated by a study of the algebraic reconstruction technique (ART) for the recovery of an image from its projections. The effectiveness of the nonnegativity constraint in the ART algorithm for the performance of the task of detection is assessed. Comparison between the detection results achieved with images reconstructed with unconstrained and constrained ART shows that the nonnegativity constraint is generally useful in improving detectability, especially when the projection data consist of a limited number of noiseless projections.

INTRODUCTION

It is often necessary to choose between alternative image-recovery algorithms for a particular imaging application. This choice becomes critically important when the available data are limited and/or are noisy. Existing conventional measures of the fidelity of reconstructed images, such as minimum variance with respect to the original image, do not always seem to be correlated with the usefulness of images and so do not help one select an algorithm. In the approach to algorithm evaluation presented here, the point of view taken is that the overall purpose of the imaging procedure is to provide certain specific information about the object or scene under investigation. Therefore that image-recovery algorithm should be chosen that best conveys the desired information.

The present method employs a Monte Carlo technique to simulate scenes and the corresponding measurements. The specified tasks are then performed using the reconstructed scenes and the accuracy of the task performance is evaluated. The advantage of this numerical approach is that it readily handles complex imaging situations, nonstationary imaging characteristics, and nonlinear reconstruction algorithms. Its major disadvantage is that it provides an evaluation that is valid only for the specific imaging situation investigated.

PROPOSED METHOD

While the effects of image noise on task performance can be predicted for a variety of tasks [1], the same can not be said of the effects of artifacts. They appear as patterns and do not often behave like stationary noise. The masking effects of measurement noise are truly random in nature. Artifacts created by inadequacies in the geometry of the measurements often manifest themselves as seemingly unpredictable irregularities that look like noise, but in a strict sense, they are not. Because their pattern is usually determined by the scene being imaged, to test how well an algorithm handles artifacts, it is necessary to vary the scene in a realistic way. For example, objects are normally randomly placed in the scene relative to the discretely sampled measurements as well as the reconstruction grid. A single realization of a simple scene is inadequate to judge a reconstruction algorithm. It is necessary to obtain a statistically meaningful average of the response of an algorithm to many realizations of the ensemble of scenes with which it must cope. The proposed method is based upon computer simulation of scenes appropriate to the desired application, subsequent data taking and analysis of the data. A Monte Carlo technique, one which employs pseudorandom numbers to generate its results, is used in this simulation because it provides the abovenoted variations within the ensemble.

The method proposed consists of the following steps:

a) Define the class of scenes to be imaged including as much complexity as exists in the intended application. Variations in scene from one realization to another should be fully specified.

b) Define the geometry of the measurements. The deficiencies in the measurements such as blur, uncertainties in the geometry, and uncertainties in the measurements (noise) should be specified. Variations with position might be included.

c) Define clearly the task to be performed. The task might be simple detection, for example. Alternatively, it could be object detection or discrimination between two types of objects, or something more complex, such as multiple discrimination, parameter estimation, etc.

On leave from Los Alamos National Laboratory, MS P940, Los Alamos, New Mexico 87545, USA

d) Define the method of task performance. This should be consistent with the intended application. If the task is to be performed by computer, then the intended analysis algorithm may be used. If the task is to be performed by a human observer, some approximation to the human should be used. Alternatively, a maximum likelihood algorithm (ideal observer) may be employed to define the best possible performance (under the prevailing assumptions made about the extent of auxillary information).

e) Create a representative scene and the corresponding measurement data by means of a Monte Carlo simulation technique. All the variations in scene content and the uncertainties in the measurements are included by means of pseudorandom selection of the uncertain parameters.

f) Reconstruct the scene with the algorithm being tested.

g) Perform the specified task using the reconstructed images.

h) Repeat steps e) through g) a sufficient number of times to obtain the necessary statistics on the task performance.

i) Evaluate the task performance. For binary discrimination tasks (of the yes-no variety), a receiver operating characteristic (ROC) curve [2] may be generated. For parameter estimation tasks, the standard measure of rms error might be employed.

EXAMPLE - EVALUATION OF NONNEGATIVITY CONSTRAINT

As a concrete example, we have tested the usefulness of the nonnegativity constraint in the algebraic reconstruction technique (ART) [3]. The scene is assumed to consist of a number of non-overlapping, high-contrast and low-contrast discs placed on a zero background. The low-contrast discs have an amplitude that is 0.1 times that of the high-contrast ones. For these examples, 10 discs of each variety are randomly placed within a circle of reconstruction, which has a diameter 128 pixels in the reconstructed image. The diameter of all the discs is 8 pixels. The first image of the series of images generated for these tests is shown in Fig. 1. In this computed tomographic (CT) problem, the measurements are assumed to consist of a specified number of parallel projections, each containing 128 samples.

The ART algorithm with 10 iterations has been used in this example to reconstruct the original scene. Variable relaxation (or damping) factors are used to attenuate each update of the reconstruction. The choice of the relaxation factor is a complex issue, which will not be discussed in detail here. Suffice it to say that, for the examples presented here, the algorithm started with a relaxation factor of 0.2 for 100 views and 1.0 for the other cases, which involved limited numbers of projections. The relaxation factors were multiplied by 0.8 after each iteration, resulting in a final factor that is about 7 times smaller than the initial one. This provides regularization in the estimation procedure, which converges to a least- squares solution in the limit that the relaxation factor approaches zero [4]. The result of reconstructing Fig. 1 from 12 noiseless views spanning 180° is shown in Fig. 2. The seemingly random fluctuations in the background are artifacts produced by the limited number of projections and arise mainly from the high-contrast discs. It appears that the nonnegativity constraint improves the reconstruction considerably in that it has reduced the confusion caused by the fluctuations in the background. However, some of the low-contrast discs have not been reproduced. Also, there still remain many fluctuations in the background that may mislead one to suspect the presence of discs in places where none

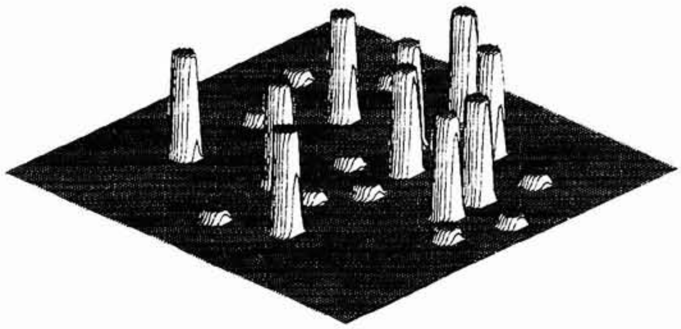

Figure 1. The first randomly generated scene consisting of 10 high-contrast and 10 low-contrast discs.

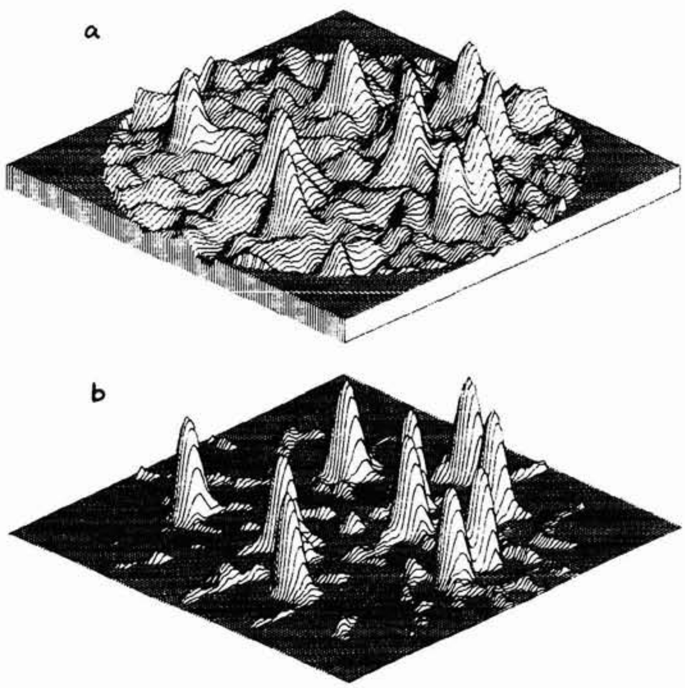

Figure 2. Reconstructions of Fig. 1 from 12 noiseless parallel projections subtending 180° obtained with the ART algorithm (a) without and (b) with the nonnegativity constraint.

exist in reality. A statistically significant comparison between reconstructions with and without the nonnegativity constraint must be made to assess its value.

It will be assumed that the task to be performed is simple detection of the low-contrast discs. It is assumed that the position of a possible disc is known beforehand as is the background. To perform the stated task of detection, it is assumed that the sum over the area of the disc provides an appropriate decision variable. This is an approximation to the matched filter, which is known to be the optimum decision variable when the image is corrupted by additive uncorrelated Gaussian noise [2]. After reconstruction, the sums in regions where the low-contrast objects are known to exist are calculated as well as those over regions where none exist. These two data sets may be displayed as histograms in this decision variable as shown in Fig. 3. A disc will be said to be present in each location where the value of the decision variable is above a chosen threshold. The probability that the presence of a disc is correctly detected, called the true-positive probability, is estimated as the area under the dashed histogram above the threshold. The probability of falsely stating a disc to be present, the false-positive probability, is the area under the solid curve above the threshold. As the threshold is lowered to increase the true-positive rate, the false-positive rate also increases. The choice of the threshold value depends on the relative costs associated with correctly and incorrectly detecting discs. The relationship between the two distributions is often characterized by the detectability index d', which is the ratio of the difference between the mean values of each distribution to the square root of their average variance. This is sometimes called the signal-to-noise ratio (SNR) for detection. For the histograms shown in Fig. 3, d' is 0.87.

The same results may be displayed as an ROC curve [2], which shows the variation of true-positive probability with false-positive probability as the decision threshold changes. The ROC curve summarizes binary detection task performance. Figure 4 shows the ROC curve generated directly from the distributions in Fig. 3. Comparison between the ROC curves produced by unconstrained ART and constrained ART shows that the nonnegativity constraint has dramatically enhanced the performance for this detection task. The area under the ROC curve for no constraint is 0.736 compared with 0.961 with the constraint. These areas correspond to effective indices for detectability d_A of 0.89 and 2.49, respectively, which are calculated to be the same as d' under the assumption of Gaussian-shaped frequency histograms. Thus the use of the nonnegativity constraint has increased the detectability by 180% in this case of a limited number of views.

Table I tabulates the results obtained under varying data-taking conditions. To refer the noise level to the magnitude of the projections, the peak projection value for the low-contrast discs is 0.80. The nonnegativity constraint is seen to be generally useful. The constraint is particularly helpful when the data are limited by the geometry. It has little effect when detectability is limited by the noise in the data.

DISCUSSION

We have presented a new method that can be used to test the effectiveness of reconstruction algorithms. This method is based on a Monte Carlo simulation of the complete imaging process from the composition of the original scene to the final interpretation of the reconstructed scene. The accuracy with which a specified task is performed is the goal of the simulation. This method is in accordance with the notion that an algorithm can only be properly evaluated by trying it out on a statistically meaningful sample of trials. A major advantage of the Monte Carlo technique is that new effects may be easily added. On the other hand, only the overall effect of all the conditions is observed. It is difficult to determine the relative contributions of individual effects. The Monte Carlo simulation technique is particularly useful in situations that do not lend themselves to analytic analysis. An example of this is the typical problem of the effect of discrete sampling on signal analysis as, for example, in the problem of the detection of small objects. In such a case, what makes sense is to average the detectability over all possible positions of the object

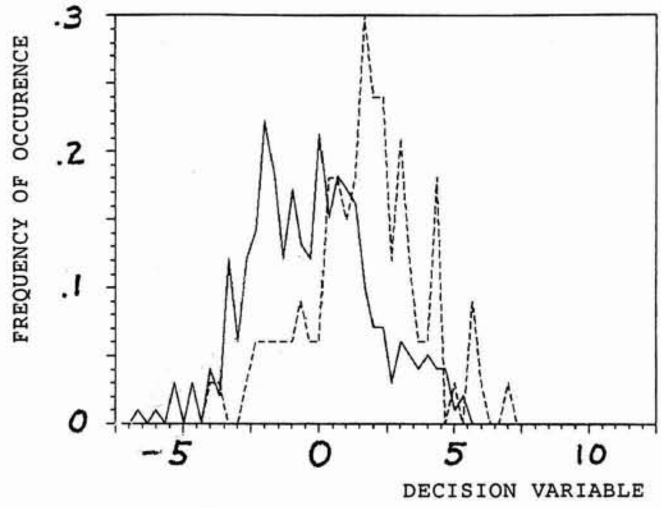

Figure 3. The frequency distributions of the decision variable (the sum over a circular region) evaluated where a low-contrast disc is known to exist (dashed line) and where none exists (solid line) for ART reconstructions without the nonnegativity constraint.

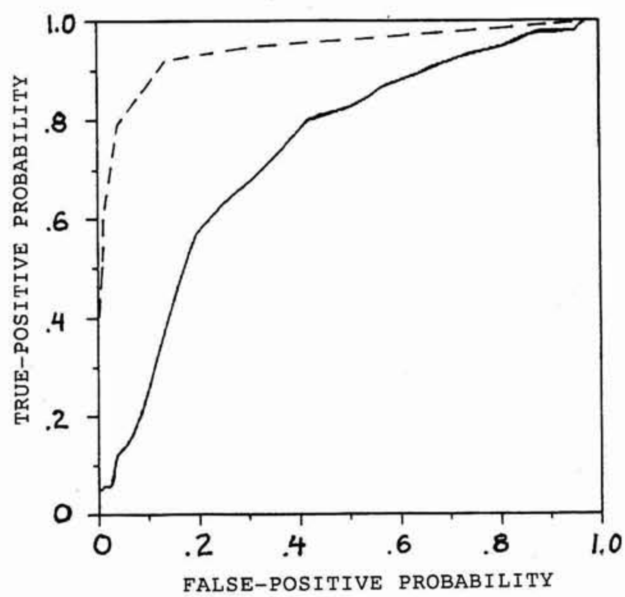

Figure 4. The receiver operating characteristic (ROC) curve derived from the frequency distributions shown in Fig. 3, that is, for unconstrained ART (solid line) and the same for reconstructions obtained with the nonnegativity constraint (dashed line). The nonnegativity constraint is seen to markedly improve task performance since its ROC curve is always significantly above that for no constraints.

TABLE I

Summary of the effect of the nonnegativity constraint on the detectability index d_A, determined from the area under the ROC curve, for various kinds of projection data. Results were obtained from tests made on the CT reconstructions of 10 randomly generated scenes for various kinds of deficiencies in the data.

no. proj.	Δθ (deg.)	rms noise	d_A without constraint	d_A with constraint	Improvement
100	180	8	1.98	2.00	+1%
100	180	4	4.10*	3.94*	-4%
8	180	0	0.45	0.79	+76%
12	180	0	0.89	2.49	+180%
16	180	0	1.94	5.51*	+184%
16	90	0	1.19	2.49	+109%
32	90	0	1.25	3.57	+186%
16	180	2	1.63	2.74	+69%

* indicates d′ calculated from mean and variances of frequency distributions because of statistical inaccuracy of d_A

relative to the discrete measurements and reconstruction. The Monte Carlo method is perfect for this.

We have seen that the nonnegativity constraint is generally useful for the specific problem addressed here - detection of low-contrast discs in the presence of high-contrast discs using CT reconstructions. This constraint is particularly helpful when the data consist of limited number of noiseless projections. When the data are degraded by additive noise, the nonnegativity constraint does not improve detectability. The effectiveness of the nonnegativity constraint was found to depend on the choice of relaxation parameters used in the ART algorithm. An optimal selection needs investigation.

There are many possible extensions to this preliminary effort. Alternative choices for the decision variables could be pursued. For example, a weighted sum of the reconstruction values over a local region could be used. The optimal weights might be determined from the reconstructions provided and could be dependent upon the reconstruction procedure. As the detection task specified in the present example is truly simple and not very closely related to most real problems, more complex and interesting tasks could be explored.

Clearly this approach of random simulation is generally applicable to test any or all aspects of the entire imaging chain from scene generation to the final method of task performance.

REFERENCES

1. K. M. Hanson, "Variations in task and the ideal observer," Proc. SPIE **419**, 60-67 (1983).
2. A. D. Whalen, **Detection of Signals in Noise** (Academic, New York, 1971).
3. R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques for three-dimensional electron microscopy and x-ray photography," J. Theor. Biol. **29**, 471-481 (1970).
4. Y. Censor, P. P. B. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems," Numer. Math. **41**, 83-92 (1983).
5. K. M. Hanson, "The detective quantum efficiency of CT reconstruction: The detection of small objects," Proc. SPIE **173**, 291-298 (1979).