JOURNAL OF NUCLEAR MEDICINE 7:773-780, 1966

The Problem of Display in the Visualization of Radioisotope Distributions

Stephen M. Pizer[1] and Henri G. Vetter

Boston, Mass.

INTRODUCTION

The best display for radioisotope scans[2] is still an open question. Although there is agreement that a certain amount of smoothing has a favorable effect, there is divergence of opinion on the desirable amount, and above all, considerable confusion on the fundamental issue of the meaningfulness of the data presented. The following discussion may shed some light on the question and will be applicable no matter what display technique, such as contour plotting, color recording, or stereo viewing, is selected for the final output.

When a scan is recorded, the primary information registered is the set of coordinates at which events were detected. The purpose of processing beyond this point is to supply a representation of the continuous activity distribution most likely *a posteriori* to be the cause of the set of recorded events. Superficial investigation suggests that, since the origin of each recorded event can only be ascertained in a probabilistic manner, a recorded dot represents the center of a two-dimensional probability distribution. Thus it should be replaced by a spatially distributed density reflecting this distribution, *viz.* the spread function of the collimator or imaging device. Some instruments feature such extended spot

---

[1]Physics Research Laboratory, Massachusetts General Hospital

[2]Except where specifically noted, we mean by "scan" the output of any radioisotope activity mapping device.

printers, usually in the form of a quasi-gaussian light spot which exposes a photographic film. The derivation on which this conventional approach is based can be found in Appendix II. Comprehensive explanation of all symbols and notation used in this paper can be found in Appendix I.

CRITIQUE OF CONVENTIONAL APPROACH

This approach is mathematically justified by equation 5, Appendix II, which gives the best estimate of the activity per unit area at any point, (x,y), called $\hat{A}(x,y)$. Careful examination shows that the underlying assumptions of this approach are incorrect. The fallacy lies in the fact that the dots are treated as if the events they represent originated in mutually exclusive target areas as viewed by the system. This is the case only if the dots are so few and far apart that the viewing areas about each as defined by the system's spread function do not intersect. However, in almost all scans, when the dots are closer together, more than one of them may reflect the same activity area. Thus the information about the activity distribution given by them can only be fully obtained if the events which they represent are considered jointly. That is, the conventional approach is incorrect because it treats the relationship between an event (disintegration) at (x,y) and any given dot as if it were independent of the partial information about the overall activity distribution given by all the dots in the scan. In the formal derivation the error is in the assumption stated by equation 1, namely that, given the set of dots, the best estimate of activity at (x,y) depends exclusively on the probability that an event occurred at (x,y). The best estimate, $\hat{A}(x,y)$, must also involve the joint probabilities of events in the surrounding area.

The following example will illustrate the fallacy. Assume the target to be scanned to be a point source in a zero-activity background. Assume successive scans with progressively higher levels of activity. As the number of dots increases, their density distribution will approach the system's spread function (by definition). The conventional processing will smooth the appearance of this spread function and in addition broaden it further, no matter how many counts we accumulate. However, it is clear that with both the exact knowledge of the spread function and that of all the dots recorded, we can infer with increasing certainty as the number of dots increases that the source causing the distribution is a point source. That is, we can say that the probability of activity at any point other than the center is very small, despite the high dot density displayed at that point, because of a knowledge of *all the dots*. A correct method of display should materialize this property by narrowing the image increasingly as the number of counts increases.

REVISED APPROACH

A correct approach to the problem, then, must be based on the JOINT *a posteriori* probability distribution of origins of events, given the whole set of dots—more fundamentally, on that of the whole activity distribution, given the whole set of dots—$p(\underline{A}|\underline{d})$.

At this point let us quickly define some terminology and concepts involved. Underlined symbols stand for arrays. $\underline{A}$ stands for the activity distribution visualized as a matrix of activity elements; $\underline{g}$ stands for the complete set of origins of the gamma-rays which caused the dots; $\underline{d}$ stands for the complete set of dots. We have a physical situation in which an activity distribution, $\underline{A}$, is probabilistically related to a dot distribution, $\underline{d}$, but where $\underline{A}$ causes $\underline{d}$. Because of this causality the probabilities involved that we must clearly distinguish are the following:

1. $p(\underline{A})$, the probability that the array $\underline{A}$ will occur, given no other knowledge, i.e. the *a priori* probability of $\underline{A}$;
2. $p(\underline{d}|\underline{A})$, the conditional probability of $\underline{d}$ given $\underline{A}$. It is an *a priori* conditional probability which embodies the statement "$\underline{A}$ causes $\underline{d}$";
3. $p(\underline{A}|\underline{d})$, the conditional probability of $\underline{A}$ given $\underline{d}$, which in the context of the causality involved means the probability that the observed array $\underline{d}$ has been caused by the hypothesized activity array $\underline{A}$. That is, $p(\underline{A}|\underline{d})$ is an *a posteriori* probability.

Our problem is to choose the best estimate of activity, denoted by $\hat{\underline{A}}$, given the set of dots, $\underline{d}$. $\hat{\underline{A}}$ must be based on $p(\underline{A}|\underline{d})$; in fact, it must be chosen as the maximum or minimum of some function of $p(\underline{A}|\underline{d})$. What that function is depends on one's definition of "best estimate", which in turn must depend on one's aims. For example, one may define $\hat{\underline{A}}$ as the activity distribution for which the *a posteriori* probability, $p(\underline{A}|\underline{d})$, is maximum, or one may define it as the activity distribution which minimizes the mean square error of the estimate.

The probability distribution, $p(\underline{A}|\underline{d})$, is derived in Appendix III, by conceptually proceeding from the dots to the origins of the events they represent to the activity distribution which produced these origins, as

$$p(\underline{A}|\underline{d}) = K_2\, p(\underline{A}) \sum_{\text{all } \underline{g}} p(\underline{d}|\underline{g}) p(\underline{g}|\underline{A})$$

$$= K_2\, p(\underline{A})/10 \sum_{\text{all } \underline{g}} p(\underline{d}|\underline{g}) \prod_{g_i \epsilon\, \underline{g}} A(x_i, y_i),$$

where the sum is taken over all possible sets of gamma-ray origins, the product is of the activities at the points denoted by all the members of that set, and where $p(\underline{A})$ is the *a priori* probability of the hypothesized activity distribution, $\underline{A}$; $p(\underline{d}|\underline{g})$ is the probability of the set of dots, $\underline{d}$, given the set of events, $\underline{g}$; and $p(\underline{g}|\underline{A})$ is the probability of the set of events, $\underline{g}$, given the activity distribution, $\underline{A}$.

DISCUSSION

The final formula does not automatically outline a specific practical processing procedure but gives a probability distribution on which the estimate must be based. Let us consider the different parts of equation 10 and the physical interpretations they suggest.

The $p(\underline{d}|\underline{g})$ term gives predominant weight in the estimate to those sets of gamma-ray origins which are more likely to have produced the set of dots. Thus, the effect of the term can be interpreted as generating a most probable set of

origins, with some uncertainty remaining as to the location of the members of that set.

The product term represents the relation between the activity distribution and the set of gamma-ray origins. It reflects the uncertainty due to the discrete nature of the emission process. Its effect is modified by the *a priori* assumptions about the activity distribution in the target. Thus the term $p(\underline{A})$ in equation 10.

The quantity, $p(\underline{A})$, is a factor which lowers the probability that the set of dots will be attributed to an unrealistic activity distribution. Thus it lowers the chance that the best estimate will be such a distribution. The specification of $p(\underline{A})$ will probably take the form of assumptions about the smoothness of the activity distribution and the type of detail relevant or probable. Practically since activity is assumed not to vary wildly from point to point, $p(\underline{A})$ will reflect this restriction and will always exclude high spatial frequencies. That is, the effect of $p(\underline{A})$ is a local smoothing on the estimate of activity.

The terms discussed above can be visualized as having two kinds of effects, one which emphasizes the most probable set of origins and can be considered as a shifting of the dots to obtain that set, and another which reflects uncertainties and can be considered as a smoothing. Accordingly the analysis suggests a break-up into two steps as an acceptable approximation, each of which can be achieved with reasonable means, analog or digital, and each of which could be enacted independently for a specific processing purpose.

1. Dot shifting: This procedure would find the most probable set of origins using $p(\underline{d}|\underline{g})$. This step is beneficial because it presents the eye with an unbiased set of origins in contrast to the set of dots, which should be shifted conceptually before being interpreted as activity. Each dot should be shifted by an amount and in a direction determined by the local dot density and dot density gradient. If we take the example of a high contrast boundary, we can easily see that for a dot recorded in the low density field but close to the boundary the probability it was caused by an event from the high density field is increased in accordance with the area of the higher density field seen by the detector at that time. The best estimate for this event's origin should therefore be shifted from the dot toward the boundary. We see that this processing will be gradient enhancing. We see also that in the process the size of a "hot spot" is shrunk while that of a "cold spot" is expanded, since the shift is always toward the area of higher density. The lack of this correction will account for some of the difference encountered in the detection of hot and cold spots in scans with conventional display.

2. Smoothing: This procedure would replace each discrete origin by an extended spot determined by the combined effect of the uncertainty of the origins due to the system's spread function and the uncertainty due to the inherent statistical nature of the emission process modified by the *a priori* assumptions about the target configuration. The shape of the spot will be circular to a first approximation. Notice that its size, which does not depend on the system's spread function alone, *may vary across the scan.* It is easy to see that this variation will always be in the direction of a smaller spot as the local density of events increases. Two factors are involved. The first is the uncertainty of the location of the shifted

dot due to the imaging spread function. The information on the basis of which the shift is computed increases with the local density, and thus the uncertainty of the shifted location decreases. The second factor is the uncertainty about the activity at a point due to the discrete nature of the emission process which requires that any estimate be based on the surrounding population of dots. The greater their density, the less is the area needed for that estimate. Thus both factors lead to a smaller spot as local density increases.

The smoothing procedure is clearly an approximation with respect to both the radial intensity distribution and the size of the spot since the *a priori* target assumptions are poorly known to begin with and the rigorous combination of both sources of uncertainty is impractical. As for the radial intensity distribution we can predict that, statistics being what they are, it can be taken as gaussian at very little risk. As for the size of the spot there is little to say *a priori* beyond the fact that it decreases as the local dot density increases. Empirical selection of a value seems indicated, keeping in mind that too much smoothing is irreversible and thus loses information while the eye can compensate for insufficient smoothing.

The two steps discussed above can be visualized as a means to approximate a directionally biased smoothing by means of an isotropic smoothing about a center combined with a shift of the center with respect to the dot. This could be generalized (if, warranted in quality by the results, and pricewise by processing economy over finding the best estimate directly) into a scheme by which more than one smoothing center would be assigned to each dot.

At this point it may be worth pointing out that the foregoing discussion, which applies directly to pictures obtained with a scintillation camera, is applicable to scans in the narrow sense of the term only to the extent that the line structure is fine enough to be ignored. If this is not the case, the scanning introduces a basic periodicity in the stepping direction. To reflect the additional uncertainty in the stepping direction, the smoothing in that direction must be broadened, leading to a non-isotropic spot.

CONCLUSIONS

We end up with two novel procedures affecting the output picture differently: variable smoothing, which achieves the aims of conventional smoothing without the drawback of having to compromise between too much smearing in the more dense area and insufficient smoothing resulting in mottling in the less dense area; and dot shifting, which has an edge enhancing effect and causes some overall shift in boundaries. Though the two procedures are somewhat interdependent, because of the different nature of their effects either one may be desirable even if the other is not implemented.

The commonly experienced dilemma encountered using a fixed size spot, that the most desirable smoothing in a less dense area can be obtained only at the expense of unsatisfactory smoothing in a more dense area, and vice versa, indicates that we can expect the effect of variable smoothing to be quite noticeable.

Because of the somewhat heuristic interpretation of equation 10 which leads to the specified procedure, no quantitative prediction can be made about the

suitable amount of shifting and smoothing and on the magnitude of the improvement thereby obtained. Only experimental results will give the answer and tell if the improvement due to shifting is worth the effort.

The specific practical suggestions resulting from this analysis are twofold:

1. A relatively simple post-scan processing (most likely achieved with a digital computer) could be based on dot shifting, followed by smoothing.

2. Variable smoothing could be achieved photographically by controlling the size of the luminous spot while maintaining the total light flux per spot constant. The most fruitful way to achieve such smoothing is in the form of post-scan processing, because the spot size at one location can then be determined as a function of all the surrounding dots. However, an inexpensive, on-line processing can be realized, which can be easily incorporated in any conventional photographic spot printer, by controlling the size of the spot by the output of a ratemeter. The absence of shifting may not affect the quality of the processing too much, if the collimator has been sensibly chosen to balance the statistical and geometric resolution effects and thus the amount of shifting is small compared to the size of the spot.

APPENDIX I

NOTATION

$A(x,y)$ represents the activity per unit area in a square area of side Δ centered at the point (x,y).

d_{xy} represents the occurrence of a dot in the element centered at the point (x,y).

g_{xy} represents the occurrence of a gamma-ray origin in the element centered at the point (x,y).

Underlined symbols represent arrays.

$\underline{A}$ represents the two-dimensional activity array. That is, the ij^{th} member of $\underline{A}$, A_{ij}, represents the activity per unit area in element ij.

$\underline{d}$ represents the dot array. That is, the i^{th} member of $\underline{d}$, d_i, represents the i^{th} dot by its co-ordinates, (u_i,v_i).

$\underline{g}$ represents the gamma-ray origin array. That is, the i^{th} member of $\underline{g}$, g_i, represents the i^{th} gamma-ray origin by its co-ordinates, (x_i,y_i).

The notation $g_i \epsilon \underline{g}$ means the i^{th} gamma-ray origin is a member of the array $\underline{g}$. Thus $\prod_{g_i \epsilon \underline{g}} A(x_i,y_i)$ means the product of the activity densities at the points (x_i,y_i), where (x_i,y_i) are the points where the gamma-ray origins, g_i, in $\underline{g}$ occur.

$p(\underline{B})$ means the probability that array $\underline{B}$ will occur. Similarly $p(\underline{b})$ is the probability that array $\underline{b}$ will occur and $p(b_i)$ is the probability that event b_i will occur.

$p(\underline{B}|\underline{h})$ is the conditional probability that array $\underline{B}$ will occur given the knowledge that array $\underline{h}$ has occurred. Similarly $p(\underline{B}|\underline{H})$, $p(\underline{B}|h_i)$, $p(b_i|\underline{H})$, etc.

APPENDIX II

DERIVATION FOR CONVENTIONAL APPROACH

The conventional approach is justified as follows. Divide the target into many equal and arbitrarily small elements. Let g_{xy} be the occurrence of an event in the target in the element centered at (x,y). Let d_i be the occurrence of a dot at (u_i,v_i) and $\underline{d} = (d_1,d_2,\ldots,d_n)$, the whole set of dots. Let $\hat{A}$ be the best *a posteriori* estimate of the activity distribution. Assume any activity distribution to be *a priori* equally probable. Then for any point (x,y),

$$\hat{A}(x,y) = F[p(g_{xy}|\underline{d})], \tag{1}$$

where F increases as its argument increases and $p(g_{xy}|\underline{d})$ is the *a posteriori* probability of an event at (x,y) given the set of dots. By Bayes's law

$$p(g_{xy}|\underline{d}) = p(\underline{d}|g_{xy})\frac{p(g_{xy})}{p(\underline{d})}, \tag{2}$$

where $p(\underline{d}|g_{xy})$ is the probability of receiving the set of dots given the fact that one of the events occurred at (x,y), and $p(g_{xy})$ and $p(\underline{d})$ are the *a priori* probabilities of an event at (x,y) and the set of dots, respectively. The fact that the elements are arbitrarily small implies that the probability of two events in the same element is negligible. Thus, assuming the d_i, $i = 1$ to N, to be *a priori* independent and equally probable,

$$p(\underline{d}|g_{xy}) = (p(d_i)^{N-1}) \sum_{i=1}^{N} p(d_i|g_{xy}). \tag{3}$$

Since $p(\underline{d}) = p(d_i)^N$, substituting equation 3 into equation 2 gives

$$p(g_{xy}|\underline{d}) = \sum_{i=1}^{N} p(d_i|g_{xy}) \left[\frac{p(g_{xy})}{p(d_i)}\right]. \tag{4}$$

Since the bracketed expression is independent of position,

$$\hat{A}(x,y) = G[\sum_{i=1}^{N} p(d_i|g_{xy})], \tag{5}$$

where G increases as its argument increases. The equation justifies the conventional approach.

APPENDIX III

DERIVATION FOR REVISED APPROACH

We will use the notation g_i to mean the occurrence of an event at (x_i,y_i) and $\underline{g}$ to mean a complete set of origins of the detected events. Similarly d_i and $\underline{d}$ refer to dots. Divide the target into many equal and arbitrarily small elements. Then the activity can be represented by a matrix $\underline{A}$, such that A_{ij} is the activity per unit area in the ij^{th} element. Our best estimate must be based on $p(\underline{A}|\underline{d})$, the *a posteriori* probability of activity given a set of dots.

By Bayes' law

$$p(\underline{A}|\underline{d}) = \frac{p(\underline{d}|\underline{A})\ p(\underline{A})}{p(\underline{d})}, \tag{6}$$

where $p(\underline{d}|\underline{A})$ is the conditional probability of a set of dots given an activity distribution and $p(\underline{A})$ and $p(\underline{d})$ are the *a priori* probabilities of activity and dots respectively.

$$p(\underline{d}|\underline{A}) = \sum_{\text{all } \underline{g}} p(\underline{d}|\underline{g})\ p(\underline{g}|\underline{A}), \tag{7}$$

where the sum is taken over all possible sets of events; $p(\underline{g}|\underline{A})$ is the probability of that set of events, given the activity; and $p(\underline{d}|\underline{g})$ is the probability of the set of dots, given that set of events.

By definition of activity the probability that an event will take place at (x,y) is proportional to the activity at (x,y) and independent of the activity anywhere else. That is,

$$p(\underline{g}|\underline{A}) = K_1 \prod_{g_i \epsilon \underline{g}} A(x_i,y_i). \tag{8}$$

The quantity $p(\underline{d}|\underline{g})$ depends on the system's spread function and the received set of dots and equals, assuming there are N detected events,

$$\sum_{\substack{\text{all} \\ \text{permutations}}} \frac{}{N!} \prod_{i=1}^{N} p(d_i,g_{j_i}), \tag{9}$$

where $j_i \neq j_k$ when $i \neq k$, and j_i is a positive integer $\leq N$. That is, the probability to receive $\underline{d}$, given the origins $\underline{g}$ is the sum of the probabilities of all possible correspondence sets between origins and dots where for each correspondence set the probability is the product over i of the individual probabilities of receiving a dot d_i from origin g_{j_i}, which is given by the system's spread function.

Substituting equation 8 into 7, and 7 into 6 and assuming any distribution of N dots is equally likely *a priori*, one gets

$$p(\underline{A}|\underline{d}) = K_2\, p(\underline{A}) \sum_{\text{all } \underline{g}} \left[p(\underline{d}|\underline{g}) \prod_{g_i \epsilon \underline{g}} A(x_i,y_i) \right], \tag{10}$$

where $p(\underline{d}|\underline{g})$ is given by equation 9.

ACKNOWLEDGEMENTS

The authors are indebted to Dr. Saul Aronow and Dr. Gordon Brownell of Massachusetts General Hospital for their helpful comments.

This research was partially supported by a National Science Foundation Graduate Fellowship and Public Health Service Grant 5T1 GM 889-10 and a grant from the Atomic Energy Commission.