Sensitivity Matrix Coded Aperture Imaging Method

Berthold K.P. Horn
2003 August 8th

The standard coded aperture imaging approach assumes:

- sources at infinity;
- isolated point sources in specific directions;
- masks with flat power spectrum (except for DC);
- thin masks free of vignetting;
- collimation to block rays from “out of view” directions;
- visible sources cast full cycle of mask pattern on detectors;
- use of correlation with the mask pattern to generate an image.

In practice, typically:

- sources are not infinitely far away;
- sources may not be point sources;
- masks have finite thickness;
- collimation may be undesirable or imperfect;
- sources may project only a partial mask shadow on detectors.

Using the standard correlation decoding method leads to artifacts in these circumstances. Some of these artifacts may be ameliorated using mask/anti-mask techniques. However, one can do even better.

Also, when sources are not infinitely far away, an opportunity exists to gather more information by taking additional measurements using different positions for the detector array. This extra information may be useful for recovering additional parameters of the source distribution, such as distances to sources, or simply to provide finer resolution.

However, the traditional mask correlation method does not provide an effective way of dealing with this new information, or with the departure from ideal conditions met in practical applications. Further, only a very small number of possible masks have the special properties required for correlation decoding. Only masks with certain number of elements are usable, and these have the property that they are approximately half open.

With nearby sources, and multiple detector array positions, different mask arrangements may be preferrable, It may even be advantageous to have three-dimensional masks (e.g. rods at various distances from the detector array).

Basic Concept

For computational purposes, the source distribution must be described by a finite set of numbers. In the ideal case mentioned above, this is simply the set of intensities of infinitely distant point sources in a fixed number of specific directions at equal angular intervals.

The description of the source distribution can instead be the set of coefficients of some analytic approximation such as a series expansion (e.g. power series, or orthonormal series). It could also, for example, be the intensities of extended sources distributed to fill space near the detector array. The task is to estimate the unknown parameters in the model of the source distribution assumed. The number of unknowns may be the order of the series expansion, or may be the number of distinct sources assumed.

The detector count rates are the measurements. In the case of a single exposure, the number of measurements equals the number of detectors. With multiple exposures — with different positions of the detector array — the total number of measurments is the product of the number of detectors and the number of exposures.

To obtain useful results, one would normally ensure that the number of measurements is equal to, or greater than, the number of unknowns. Least squares methods may be used in the overdetermined case. If, on the other hand, fewer measurements than unknowns are available, then the problem may be "regularized" by adding some penalty term (e.g. discouraging source distributions that have rapid spatial variations).

Method of Solution

Each parameter of the unknown source distribution may be taken in isolation and the detector response calculated that results when only that parameter is non-zero. For example, if the model is a set of point sources distributed in some pattern in space, then for each point source position one determines how each detector responds. The quantities of interest are the "sensitivities," the ratios of detector count rates to the source intensity.

In the case of point sources, each of these is proportional to the solid angle at a source of the part of a detector element not obstructed by opaque parts of the mask. In the case of an extended source, the desired result is a weighted average of solid angles for different positions in the source, where the weights correspond to the fractional intensity of that part of the source.

Depending on the complexity of the model of source distributions, this sensitivity may be obtained using analytic methods involving geometric reasoning, numerical integration, Monte Carlo simulation, or may even be measured experimentally.

If multiple exposures are to be used with different detector array positons, then this exercise is repeated for each detector array position. Detector responses from multiple exposures can be concatenated into a single measurement vector.

Importantly — if we ignore saturation — detector responses vary linearly with each source distribution parameter, and reponses to different source distribution parameters add up. So the relationship between unknown source parameters and detector measurements form a system of linear equations. We can use the linear system to process experimental data once we have analysed the model to determine the set of sensitivities.

If we have arranged to have as many measurements as there are unknown source parameters, then we simply invert the matrix of sensitivities (i.e. solid angles) and use it to process sets of measurements. If instead there is a mismatch between number of unknowns and number of measurements, pseudo-inverses may be used. If the matrix is too large to make an explicit inverse feasable, then iterative methods can be used to solve the equations.

Sanity Check: Applying the Method to Ideal Coded Masks

As a check on how this works, consider the idealized case described at first. Here we assume a set of infintiely distant sources arranged in specific direction determined by the spacing between the mask and the detector array and the size of the detector and mask elements. Each source projects one cycle of the mask pattern onto the detector array.

As we step from source to source, the shadow of the mask is shifted cyclically by one detector element. Each column of the sensitivity matrix is proportional to the mask value (0 for a closed element and +1 for an open element). The coefficients from a circulant matrix, a matrix whose columns are cyclically shifted versions of one cycle of the mask pattern.

The inverse of this special matrix turns out the be its transpose, with 1's replaced with one value, a, and the 0's replaced with another, b (where a and b have similar magnitude and opposite sign and depend on the number of open and closed elements in the mask).

Consequently, the rows of the matrix used in processing the detector array measurements have the same pattern as the mask itself (except for the replacement of 1's by *a* and 0's by *b*). This corresponds to processing the measurements by correlation with a sequence derived in a simple fashion from the mask. Here *a* and *b* are chosen such that the correlation is 1 when the pattern is aligned, and 0 when it is not.

Limitations of Ideal Coded Mask Method

Note that the sources are assumed to lie in special directions that happen to align the shadows of the mask elements exactly with the detector elements. Sources in other positions split shadows of mask elements across detectors. The simple decoding scheme does not apply in this case. Inferior results are obtained in practice when this is ignored and the ideal source model is used in a situation where the sources in fact do not lie in the ideal directions.

Note also that the mask pattern repeats cyclically to achieve the desired cyclical shift in detector exposures. As a result, several source directions can potentially produce the same pattern on the detector array. To avoid this ambiguity, additional means, such as collimation, must be employed to prevent radiation from sources that are "too far off-axis" from reaching the detector. An unavoidable side effect of such collimation is a drop-off in response with "off axis" angle that causes the observed detector measurements to differ from the ideal ones.

Further, the mask has two repetitions of the basic pattern and must be blocked beyond that on either side to prevent ambiguous patterns from appearing. This means, however, that there are "off axis" source directions that lead to incomplete shadows of the basic mask pattern. The ideal correlation decoding method cannot deal with either the effects of collimation or such partial exposure.

Nearby Sources

If sources are not infinitely far away, then the shadows of mask elements on the detector array will be magnified and no longer cover exactly one detector element. If the sources happen to all lie in a plane at a fixed distance from the mask and detector arrays, then one can compensate for this by making the detector element spacing correspondingly larger. But this trick can't be used if the source distance is unknown, or sources appear at a variety of distances.

Using the correlation decoding method leads to artifacts in this situation, since the data does not correspond to the ideal case. These artifacts can be ameliorated by subtracting data obtained with the complement of the mask from data obtained with the mask. This corresponds to using a mask with mask element values -1 and $+1$ instead of 0 and 1.

The method presented here, based on setting up and solving a system of linear equations, can deal with this situation in an *exact* fashion. For each assumed source position, the detector measurements are computed based on the solid angle at the source of the part of the detector not blocked by mask elements. With nearby sources, open elements in the mask may illuminate more than one detector element, which means that entries in the matrix that would have been zero in the ideal case of distant sources are no longer exactly zero.

The inverse of the new matrix of sensitivities can not be obtained analytically since it no longer has the perfect circulant structure. The numerical inverse can, however, still be used to obtain *exact* solutions for source intensities in this situtation.

Multiple Exposures

There isn't, of course, much point in taking multiple exposures from different detector array positions in the ideal case of sources infintitely far away, since the direction of incident rays are not affected. When sources are nearby, however, additional information can be obtained by "viewing" the scene from more than one position. Also, a single exposure cannot be used to recover the three-dimensional distribution of sources (unless additional information is provided). Multiple exposures, however, makes this possible.

The added information from multiple exposures can instead be used in other ways, such as providing higher resolution of an arrangement of sources at a fixed distance. The key limitation is always that there be at least as many measurements as unknowns. The quality of the result will depend on how near independent the measurements are.

Extended Sources

The ideal coded aperture model treats the sources as composed of a finite number of isolated point sources each positioned so as to produce a shadow that has mask elements falling exactly on detector elements. There are situtations were sources may be extended in space and/or may not be aligned to produce mask element shadows exactly matching detector elements. Also, even if the actual sources are point sources, it may be advantageous to model them as extended so as to avoid the harsh aliasing present in the point source model when not aligned with the mask and detectors.

The ideal coded aperture method does not deal with extended sources or “misaligned” sources (other than to approximate them with point sources in the ideal positions). The new method can deal with extended sources, and sources in arbitrary positions.

Similarly, rather than representing the source distribution as a sum of localized sources, point-like or not, it is possible to represent it as a series, orthononormal or otherwise. Even irregular arrangement of detectors can be accomodated.

Conditioning and Noise Amplification

The sensitivity matrix will be singular if the measurements are not independent. If some measurements are closely related, then the matrix may be ill conditioned. So the practical utility of the method revolves around properties of the sensitivity matrix, which in turn depends on the data acquistion strategy and the source pattern model.

One way to get a feeling for the quality of such a method is to consider the magnitudes of the elements of the sensitivity matrix, and, more importantly, the magnitude of the elements of its inverse.

In the case of the ideal coded aperture method the entries of the sensitivity matrix are all (fixed multiples of) 1 or 0. The elements of the inverse matrix are all (fixed multiples of) two values: a and b. The signal amplification is proportional to a times the number of open cells in the mask, while the amplificition of noise is proportional to the square-root of the sum of squares of coefficients in a row, which comes to approximately a times the square root of the number of elements in one cycle of the mask. This means that the ratio of signal amplification to noise amplification is inveresly proportional to the square root of the number of cells in one cycle of the mask pattern.

With non-ideal imaging, such as when a set of sources is nearby, one finds that the sensitivities no longer just have two values because of the imperfect alignment of shadows of mask elements and detector cells. The inverse matrix has coefficients that are larger in magnitude than before. The inverse matrix does, of course, compute the *exact* result in the absence of noise, but it amplifies noise somewhat more than was the case in the ideal imaging situation. There is no way around this: one either accepts artifacts due to inexact reconstruction or has to accept increased noise.

With poor choice of source pattern model or imaging strategy, the coefficients of the inverse matrix may get enormous, trying desparately to differentiate source patterns that happen to produce similar detector measurements. Poor results will be obtained, for example, if one expects to increase resolution of the source pattern beyond what is reasonable, or expects depth resolution in a situation where there is little effect on detector signal with variations in depth.

While it may be hard sometimes to predict *a priori* how poorly conditioned the sensitivity matrix will be, it is always possibly to look at the magnitude of coefficients in the inverse. Generally, measurements that seem to repeat or nearly repeat measurements already taken do not contribute and should be avoided. Beyond this simple heuristic, there is little at this point to guide in the design of a source pattern model and a data collection strategy.

What is so Special about the Ideal Coded Aperture Method?

The original ideal coded aperture method has many restrictions and practical limitations as pointed out above. However, it has exceptionally low noise amplification. The inverse matrix contains coefficients that are all about the same magnitude and are all relatively small.

Attempts to go beyond the severe restrictions of the ideal coded aperture methods tend to be subject to increased noise amplification, unless carefully designed or heavily overdetermined.

Experiments

The new method has been implemented for the case of a one-dimensional mask and one-dimensional detector array. In this case, the size of the sensitivity matrix is such that direct inversion is feasable. Detector fluxes can be computed using either exact, geometric methods, or approximated using Monte Carlo simulation of randomly chosen rays emitted from the sources. When using geometric calculation of detector fluxes, exact inverses are obtained even when sources are not infinitely far away, when sources are not lined up with mask and detector elements, and when sources are extended.

Multiple exposures with the detectors array and mask arrays in different positions can be used to recover either higher resolution “images” of a set of sources at a fixed distance, or sources arranged at more than one distance.

The numerical values of the coefficients of the inverse matrix show a wide variation in magnitude, depending on the problem to be solved and details of the alignment and specific dimensions. When the magnitude of these coefficients is high, reconstruction can greatly amplify noise in the detector measurements.

It is hoped that good designs for source patterns and mask and detector arrangements can be guided by study of the magnitude of the coefficients of the inverse matrix.

Summary: Capabilities of the New Method

The new method can handle:

- nearby sources;
- extended sources;
- sources not specially lined up with mask and detector elements;
- irregular detector arrangements;
- unconventional masks;
- series representation of source pattern;
- mismatch between number of measurements and number of unknowns;
- multiple exposures;
- recovery of depth information;
- recovery of images with increased resolution.