COMBINATORIAL DESIGN OF NEAR–OPTIMUM MASKS FOR CODED APERTURE IMAGING

Axel Busboom, Harald Elders–Boll, and Hans Dieter Schotten

Institut für Elektrische Nachrichtentechnik
RWTH Aachen, D–52056 Aachen, Germany
busboom@ient.rwth-aachen.de

ABSTRACT

In coded aperture imaging the attainable quality of the reconstructed images strongly depends on the choice of the aperture pattern. Optimum mask patterns can be designed from binary arrays with constant sidelobes of their periodic autocorrelation function, the so–called URAs. However, URAs exist for a restricted number of aperture sizes and open fractions only. Using a mismatched filter decoding scheme, artifact–free reconstructions can be obtained even if the aperture array violates the URA condition. A general expression and an upper bound for the signal–to–noise ratio as a function of the aperture array and the relative detector noise level are derived. Combinatorial optimization algorithms, such as the Great Deluge algorithm, are employed for the design of near–optimum aperture arrays. The signal–to–noise ratio of the reconstructions is predicted to be only slightly inferior to the URA case while no restrictions with respect to the aperture size or open fraction are imposed.

1. INTRODUCTION

Coded aperture imaging (CAI) [1, 2] has evolved as a standard technique for imaging high energy photon sources and has found numerous applications, e. g., in X– and gamma–ray astronomy and in nuclear medicine. As its major advantage over conventional imaging systems using rastering or pinhole collimators, CAI provides a better photon collection efficiency while preserving high angular resolution.

In coded aperture imaging a mask of transparent and opaque elements is placed in front of a position sensitive detector. The image recorded by the detector can be thought of as the superposition of many pinhole camera images pertaining to the individual aperture openings. With a suitable aperture choice it is possible to reconstruct the original image from the detector data and to improve the signal–to–noise ratio (SNR) with respect to a single pinhole aperture.

An important parameter in the design of the aperture array is the *open fraction* ρ which is defined as the ratio of the transparent to the total aperture area. In the literature, several approaches to the computation of the optimum value of the open fraction have been taken (e. g., [3, 4, 5]). While the different approaches lead to different numerical results, it seems to be agreed upon that the optimum open fraction is a function of the *relative detector noise level* ξ, i. e., the ratio of unmodulated detector background noise to the total intensity of the sources to be imaged.

This work was supported in part by grant FKP0015901B5A awarded by the German government (Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie) and by INTRAVIS GmbH, Herzogenrath, Germany.

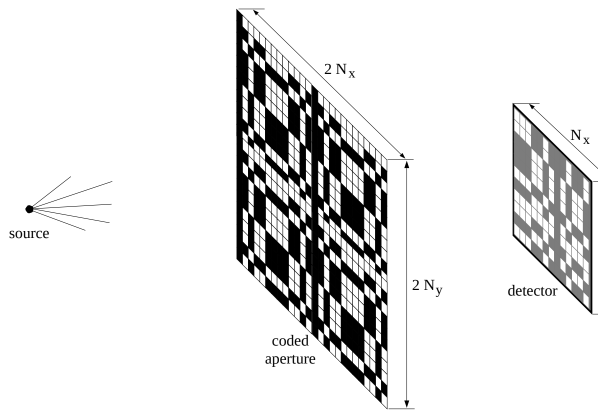

Figure 1. Coded aperture imaging system.

The most commonly used aperture patterns are Uniformly Redundant Arrays (URAs) [6], i. e., arrays with constant sidelobes of their periodic autocorrelation function. If URAs are used as coded masks, the source distribution can be reconstructed from the coded image by correlating it with the aperture array itself. We will refer to this decoding scheme as matched filtering. If any other arrays than URAs are used, then matched filtering will lead to artifacts, i. e., systematic errors in the reconstruction. Unfortunately, URAs exist for a limited number of array sizes and open fractions only.

In this contribution, we propose a coded imaging setup which does not impose any restrictions to the aperture size or open fraction. We use mismatched filtering in order to obtain an artifact–free reconstruction for almost arbitrary aperture arrays. Our approach to the aperture design is based upon stochastic combinatorial optimization techniques, such as Simulated Annealing or the Great Deluge algorithm. For comparison, we also examine simple gradient search strategies. While these algorithms almost always fail to find the true optimum array, they were found to yield aperture masks with near–optimum performance in all cases.

2. SYSTEM MODEL

For our analysis we assume the standard arrangement used by most authors (e. g., [3, 6]) depicted in Figure 1. The aperture consists of a 2×2 mosaic of a basic pattern of $N_x \times N_y$ elements. The detector has the same size and the same number of elements as one basic aperture pattern.

With this setup, it is possible to reconstruct an area of $N_x \times N_y$ source pixels, the so–called *fully coded field of view* (FCFV). The detector image $d(i,\ j)$ can be expressed as

$$d(i,\ j) = [s(N_x - i,\ N_y - j) * a(i,\ j)] + n(i,\ j) + b \quad (1)$$

where $i \in \{0,\ 1,\ \ldots,\ N_x - 1\}$, $j \in \{0,\ 1,\ \ldots,\ N_y - 1\}$, $s(i,\ j)$ is the source distribution to be imaged, $a(i,\ j)$ is the basic aperture array consisting of ones (transparent elements) and zeros (opaque elements), and the $*$ symbol denotes the two–dimensional periodic convolution. In high energy photon imaging, the predominant noise source is typically the quantum noise due to the counting statistics at the detector. The term $n(i,\ j)$ models this quantum noise and is assumed to be Poisson distributed with zero–mean. The term b is the signal independent detector background noise which is assumed to be constant over the entire detector area. Note that, particularly in astronomical applications, this detector noise may be by some orders of magnitude larger than the sources to be imaged. For the reconstruction, we assume that we can (at least approximately) subtract the detector noise b. If this is not the case in practice, b will yield a constant offset in the reconstruction. As a quite general decoding scheme, we assume that, after background subtraction, the detector image is filtered using some two–dimensional decoding filter $g(i,\ j)$, i. e.,

$$\begin{aligned} \hat{s}(i,\ j) &= [d(N_x - i,\ N_y - j) - b] * g(i,\ j) \\ &= s(i,\ j) * [a(N_x - i,\ N_y - j) * g(i,\ j)] \\ &+ \ n(N_x - i,\ N_y - j) * g(i,\ j). \end{aligned} \quad (2)$$

If $g(i,\ j)$ is chosen as the correlational inverse of $a(i,\ j)$, i. e., such that $a(N_x - i,\ N_y - j) * g(i,\ j)$ equals a two–dimensional Dirac pulse, then an artifact–free reconstruction, i. e., a reconstruction free of systematic errors, is obtained. Note that the correlational inverse can be easily computed, e. g., in the discrete Fourier domain [7]. It exists unless $a(i,\ j)$ has DFT coefficients equal to zero. Under the aforementioned assumptions, it can be shown [8] that the signal–to–noise ratio of the reconstructed image is given by

$$\mathrm{SNR} = \sqrt{\frac{I_T}{N\,(\rho + \xi)\,E_g}} \quad (3)$$

where $I_T = \sum s(i,\ j)$ is the total source intensity within the FCFV, $N = N_x N_y$ denotes the total number of elements of one basic aperture pattern, $\rho = \sum a(i,\ j)/N$ is the open fraction of the aperture, $\xi = b/I_T$ denotes the relative detector noise level, and $E_g = \sum g^2(i,\ j)$ is the signal energy of the decoding filter g.

The objective of aperture design now is to choose $a(i,\ j)$ such that the SNR is maximized for a given aperture size $N_x \times N_y$ and detector noise level ξ. We will first assume the open fraction ρ to be fixed. Maximization of the SNR is then equivalent to minimizing E_g. Let $A(i,\ j)$ and $G(i,\ j)$ denote the discrete Fourier transforms of $a(i,\ j)$ and $g(i,\ j)$, respectively. Since g was assumed to be the correlational inverse of a, we have $G(i,\ j) = 1/A^*(i,j)$, and with Parseval's theorem

$$E_g = \sum_{i,\ j} g^2(i,\ j) = \frac{1}{N} \sum_{i,\ j} \frac{1}{|A(i,\ j)|^2}. \quad (4)$$

Furthermore, for a fixed open fraction we know that $A(0,\ 0) = \rho N$ and $\sum |A(i,\ j)|^2 = N \sum a^2(i,\ j) = \rho N^2$. Subject to these constraints, the expression in Equation (4) obviously takes on its minimum if and only if

$$|A(i,\ j)|^2 = \frac{\rho(1-\rho)N^2}{N-1} \quad \forall (i,\ j) \neq (0,\ 0). \quad (5)$$

The periodic autocorrelation function of arrays which satisfy (5), is given by

$$\tilde{\varphi}_{aa}(i,\ j) = \begin{cases} \rho N & \text{if } i = j = 0, \\ \dfrac{\rho N(\rho N - 1)}{N-1} & \text{otherwise.} \end{cases} \quad (6)$$

This demonstrates that URAs, i. e., binary arrays with constant sidelobes of their periodic autocorrelation function, are optimum with respect to maximization of the SNR of the decoded images. For URAs the decoding filter g has, except for a scaling factor and a constant offset, the same structure as the aperture array. The reconstruction can therefore be interpreted as a matched filtering. So far we have assumed the open fraction to be fixed. The optimum value which should be chosen for ρ, depends on the relative detector noise level ξ. According to the optimization criterion used here, it converges to 0.5 for large ξ, i. e., if the detector noise is large with respect to the imaged sources in the field of view, and to zero (single pinhole aperture) for small ξ. Note, however, that different optimization criteria may lead to different numerical results for the optimum open fraction (e. g., [3, 4, 5]).

A survey of known construction methods for URAs can be found in [9]. Unfortunately, URAs are known only for a very limited number of aperture sizes and open fractions. For a large number of aperture sizes, the nonexistence of URAs is proven [9]. In practice the aperture size may be constrained, e. g., by the available detector size and resolution, and cannot always be chosen arbitrarily. It may therefore be necessary to design coded masks of a given size and open fraction for which no URA exists. In this case the aperture design task becomes a very high–dimensional combinatorial optimization problem, its objective being the maximization of a "quality function" $Q = ((\rho + \xi)\,E_g)^{-1}$ where both ρ and E_g depend on the aperture pattern. The analytical results which have been obtained for URAs, constitute an upper bound for the SNR. Note that this bound can only be attained for aperture sizes and open fractions for which URAs actually exist.

Due to the large dimensionality of the optimization problem, an exhaustive search for the best aperture array is obviously unrealistic. The approach taken in this paper is to utilize heuristic optimization algorithms, e. g., Simulated Annealing or the Great Deluge Algorithm, to design good, but not necessarily optimum, aperture arrays.

3. COMBINATORIAL SEARCH ALGORITHMS

The most straightforward search strategy we have explored is a simple gradient search ("Hill Climber"). Starting at some random array, each array element is temptatively toggled (from one to zero or vice versa) and the resulting change in the quality function is computed. The modification with leads to the largest increase is accepted. This search is repeated until no further rise of the quality function is possible. In the following, we will refer to this as the GS1 algorithm.

In our application, evaluation of the quality function Q is computationally very expensive since it involves computing the DFT or FFT of the aperture array. This fact suggests a modification to the GS1 algorithm: In each step, a single array element is selected at random. If inverting this element increases Q, then the change is immediately accepted, otherwise a new array element is chosen at random. This GS2 algorithm was found to work significantly faster than GS1, particularly for larger arrays, without any noticeable loss in the final quality function.

GS1 and GS2 share the disadvantage that they terminate in some local maximum of the quality function which may not be the desired solution. Therefore, various heuristic search strategies such as "Simulated Annealing" [10], "Threshold Accepting" [11], the "Great Deluge Algorithm" [12] and the "Record–to–Record Travel" [12] have been developed. These algorithms have in common that they do accept, under certain conditions, a change even if it leads to a temporary reduction of the quality function. Large decreases are less likely to be accepted than small ones, and during the execution of the algorithm the conditions for accepting a decrease become more and more stringent ("cooling" process in Simulated Annealing). The algorithms have converged when the quality function cannot be further increased and when no reduction of the quality function is acceptable any more. The heuristic search algorithms have the chance to travel through more local optima than the gradient search strategies. The probability for them to find "better" maxima, though usually not the global maxima, is therefore increased.

4. RESULTS

Applied to the aperture array design problem, we have found all of the mentioned heuristic search algorithms to perform similarly well. The Great Deluge Algorithm (GD) was selected for closer examination since it appeared to yield slightly better results in some cases [13]. Surprisingly good results were also obtained with the GS2 algorithm. However, for larger aperture sizes, the GD algorithm clearly outperformed GS2.

In the GD algorithm, convergence is controlled by a so–called "water level" W which, at any time, is less than or equal to the current value of the quality function Q. A change is accepted if and only if the new value of Q is still above the water level. By gradually increasing W the algorithm is forced to converge. Crucial aspects in the application of the GD algorithm are the choice of the initial water level W_0 and the scheme for increasing W in each step. W_0 should be small enough to initially accept almost any change. We have estimated the mean μ and the standard deviation σ of the quality function from a number of randomly chosen aperture arrays and chosen $W_0 = \mu - c\sigma$ where $c = 1$ was found to yield sufficiently small initial water levels. The best results were obtained when the water level was increased adaptively according to

$$W_{i+1} = W_i + d\ (Q_i - W_i),\ \ 0 < d < 1 \qquad (7)$$

in the i–th step. This scheme ensures that the water level rises quickly when the current value of the quality function is much higher than the water level while it rises only slowly when no more large increases of the quality function occur. The parameter d allows to control the convergence speed of the Great Deluge algorithm. If it is large ($c \approx 0.1$), the algorithm converges quickly, but does not yield significantly better results than GS2. If d is smaller ($10^{-5} \ldots 10^{-3}$), convergence is slow, but the results become significantly better.

Figure 2 shows some of the results obtained with the GD algorithm for one–dimensional apertures. In this example, the detector noise was assumed to be predominant ($\xi \to \infty$). The values on the vertical axis of Figure 2 are coding gains, i. e., the signal–to–noise ratios according to Equation (3) normalized to the SNR of a single pinhole aperture. The solid line shows the analytically computed upper bound. Note that in computation of the bound we have assumed that URAs exist for all aperture sizes. Therefore, the bound can only be attained in cases where a Uniformly Redundant Array actually exists. These URAs are marked with $\times$ signs in the figure. The circles mark the gains of the arrays found by the GD algorithm. For $N \leq 20$, the results of an exhaustive search are also displayed (+ signs). Note that for these small lengths, the GD algorithm always succeeded in finding the optimum arrays. For $N = 43$, it failed to find the existing URA for the first time.

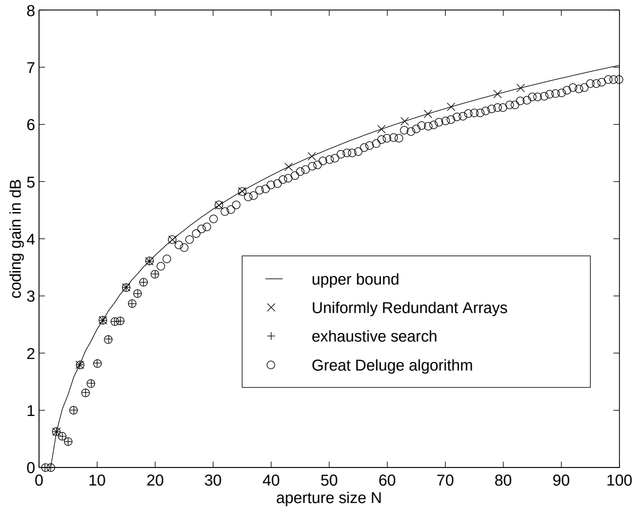

Figure 2. Comparison of one–dimensional apertures designed using the GD algorithm to the upper bound ($\xi \to \infty$).

As a more realistic case, we have designed a two–dimensional aperture array of size 41×43 for $\xi \to \infty$. For this size, a Uniformly Redundant Array with open fraction $\rho \approx 0.5$ does exist which allows a quantitative assessment of the heuristically designed apertures. For a fair comparison of the GD and GS2 algorithms, we have allowed both algorithms to compute 36 hours on a SPARC 20 workstation. During this time, the GS2 algorithm performed many gradient searches starting from different random arrays. The best result of the gradient searches was taken as the final result. The GD algorithm performed only one search starting from a single random array. Table 1 shows the coding gains achieved by the two search strategies and the coding gain of the URA. Even though both algorithms failed to find the URA, their results are only marginally inferior to the optimum aperture.

Aperture	Gain	Transmission ρ
GS2	12.89 dB	0.503
GD	12.97 dB	0.494
URA	13.22 dB	0.500

Table 1. Results for 41×43 aperture and $\xi \to \infty$.

Figures 3 and 4 show the aperture array found by the GD algorithm and its periodic autocorrelation function, respectively. Figure 4 demonstrated that the autocorrelation sidelobes of the found arrays are reasonably flat, hence, the array is a good approximation to the URA.

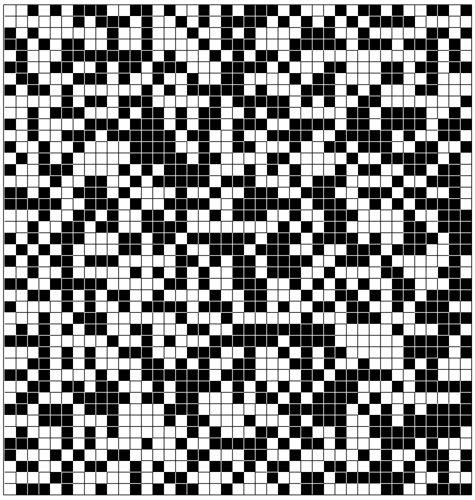

Figure 3. 41×43 aperture array found using the Great Deluge Algorithm for $\xi \to \infty$.

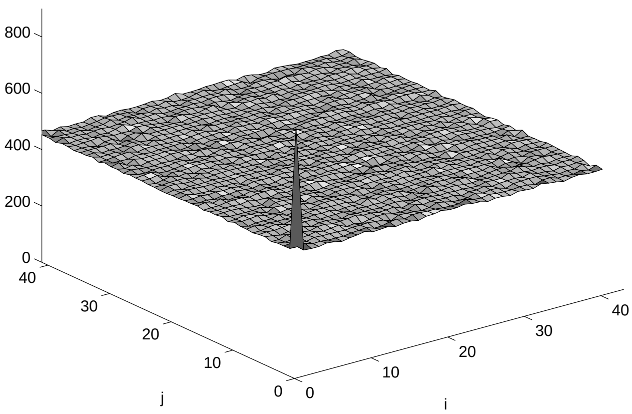

Figure 4. Periodic autocorrelation function of the array shown in figure 3.

An interesting observation we made was that the open fractions of the heuristically designed aperture arrays were in most cases slightly larger than the theoretically expected values, particularly for smaller background levels. Since this was also true for arrays found by exhaustive search for small aperture sizes, we suspect that the theoretical values tend to underestimate the optimum open fractions. Note that the theoretical results were obtained under the assumption that URAs exist for arbitrary aperture sizes and open fractions. Therefore, they do not necessarily represent the optimum open fraction for non–uniformly redundant arrays.

The GS2 and GD algorithms were found to perform better when the random arrays they started with had approximately the optimum open fraction for the given background level rather than an open fraction of 50 % and when in each step the probabilities for toggling a one or a zero element were identical.

In conclusion, our analysis has shown that using aperture arrays designed by combinatorial search algorithms in conjunction with a mismatched filter decoding scheme, coded aperture imaging systems can be designed whose performance is predicted to be only marginally inferior to URA–based systems. The approach does not impose any restrictions with respect to the feasible aperture sizes or open fractions. In addition, it offers the opportunity to easily incorporate additional constraint into the aperture design. For example, self–supporting masks can be generated which do not have isolated closed elements and thus can be realized without the need for an additional support grid.

REFERENCES

[1] L. Mertz and N. O. Young, "Fresnel transformations of images," in *Proc. Int. Conf. Optical Instruments and Techniques,* K. J. Habell, editor (Chapman and Hall, London, 1961), p. 305.

[2] R. H. Dicke, "Scatter–hole cameras for X–rays and gamma rays," Astrophys. J. **153,** L101–L106 (1968).

[3] J. Gunson and B. Polychronopulos, "Optimum design of a coded mask X–ray telescope for rocket applications," Mon. Not. R. astr. Soc. **177,** 485–497 (1976).

[4] E. E. Fenimore, "Coded aperture imaging: predicted performance of uniformly redundant arrays," Appl. Opt. **17,** 3562–3570 (1978).

[5] J. J. M. in 't Zand, J. Heise, and R. Jager, "The optimum open fraction of coded apertures," Astron. Astrophys. **288,** 665–674 (1994).

[6] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," Appl. Opt. **17,** 337–347 (1978).

[7] M. Antweiler and H. D. Lüke, "Mismatched–Filterung periodischer Binärarrays," Frequenz **48,** 58–65 (1994).

[8] A. Busboom, H. D. Schotten, and H. Elders–Boll, "Coded aperture imaging with multiple measurements," J. Opt. Soc. Am. A, to be published.

[9] L. Bömer and M. Antweiler, "Two–dimensional binary arrays with constant sidelobes in their PACF," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Institute of Electrical and Electronics Engineers, Glasgow, 1989), pp. 2768-2771.

[10] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220,** 671–680 (1983).

[11] G. Dueck and T. Scheuer, "Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing," J. Comput. Phys. **90,** 161–175 (1990).

[12] G. Dueck, "New optimization heuristics — the great deluge algorithm and the record–to–record travel," J. Comput. Phys. **104,** 86–92 (1993).

[13] M. Walessa, "Synthese von Sequenzen und Arrays mit gutem inversen Filter mittels stochastischer Optimierungsverfahren," IENT, RWTH Aachen, Aachen, Germany (Diploma Thesis, 1996).