Genetic algorithm for optical pattern recognition

Uri Mahlab and Joseph Shamir

Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

H. John Caulfield

Center for Applied Optics, University of Alabama in Huntsville, Huntsville, Alabama 35899

Received October 29, 1990

A genetic algorithm is used to generate binary reference functions for optical pattern recognition and classification. Procedures based on the properties of convex functions can be implemented directly on hybrid electro-optical systems. Computer simulations demonstrate the efficiency of this novel approach.

Most of the traditional procedures for the construction of synthetic discriminant functions for optical pattern recognition are derived by analytic formulations.[1] An alternative approach[2,3] is based on iterative numerical techniques that are used to implement nonlinear functions of the data involved in the generation of the required synthetic discriminant function. The recently introduced concept of an entropy-optimized filter[3,4] led to optical correlators of excellent discrimination. The entropy-optimized filters were first constructed by using a gradient descent algorithm, which became cumbersome when additional constraints were introduced, and then simulated annealing methods[5] had to be applied.

The above-mentioned algorithms are suitable for conventional serial computations, and they do not exploit the massive parallelism available in optical systems even if a large fraction of the computations are performed optically.[6] The purpose of this Letter is to introduce into this field the genetic algorithm[7,8] (GA), which replaces serial procedures with a parallel approach.

A short review of an iterative procedure for spatial filter generation in a 4-f optical pattern recognition system is followed by the adaptation of GA to such an architecture. The promising results of the computer simulation experiments described below indicate the power of the new method and lay the groundwork for a proposed electro-optical implementation.

The complex amplitude distribution over the output plane of a coherent optical correlator is given by

$$c(x_0, y_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)h^*(x + x_0, y + y_0)\mathrm{d}x\mathrm{d}y, \tag{1}$$

where $h(x, y)$ is the spatial filter function and $f(x, y)$ is the input function. In a joint-transform correlator $h(x, y)$ is used as a reference function, while in a 4-f correlator its Fourier transform is the filter.

We intend to discriminate between two sets of patterns, $\{f_n^D(x, y)\}$ and $\{f_n^R(x, y)\}$, by generating a filter that produces a strong and narrow peak for patterns of the first class and a uniform distribution for patterns of the second class. In a most general sense we may define a distribution function over the output plane by the relation

$$\Phi(x, y) = \frac{\mathcal{L}[c(x, y)]}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{L}[c(x, y)]\mathrm{d}x\mathrm{d}y}, \tag{2}$$

where $\mathcal{L}$ is a nonlinear operator over $c(x, y)$ such that $\mathcal{L}[c(x, y)]$ is a nonnegative quantity on (x, y). Here we take $\mathcal{L}$ to be the absolute value operator suitable for intensity detection. The distribution Φ has all the properties of a probability density for which one may define a general entropy function,[9] given by

$$S = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Psi[\Phi(x, y)]\mathrm{d}x\mathrm{d}y, \tag{3}$$

where Ψ is a strictly convex function.[10] The entropy function is maximized for a uniform distribution of Φ and minimized when a single narrow peak is obtained. Converting all functions into digitized form, $x \to m$, $y \to n$, we represent the various functions as two-dimensional matrices of $N \times N$ pixels requiring an array of $(2N - 1) \times (2N - 1)$ pixels in the correlation plane. A single steep peak over the correlation plane at some point, denoted by (k, l), is represented by the ideal distribution

$$\Phi^D(m, n) = \begin{cases} 1 & \text{at } (m = k, n = l) \in \text{(domain of } \Phi), \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

while a uniform distribution due to a rejected pattern has the form

$$\Phi^R(m, n) = \frac{1}{(2N - 1)^2}, \qquad \forall (m, n). \tag{5}$$

For each proper convex function Ψ the general entropy function has its upper and lower bounds.[9] Taking for the present example the convex function

$$\Psi(x) = x^2, \tag{6}$$

we obtain

0146-9592/91/090648-03$5.00/0

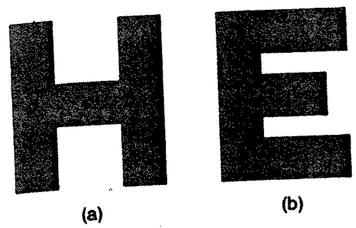

Fig. 1. Input training set: (a) the pattern to be detected (the letter H), (b) the pattern to be rejected (the letter E).

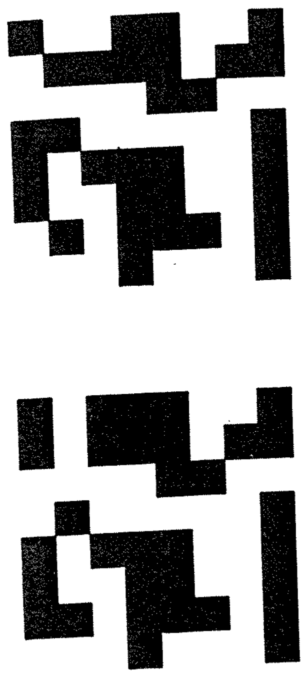

Fig. 2. Two different members of the population of reference functions. Black represents +1, white represents −1.

$$S^D_{\min} = -1, \qquad S^R_{\max} = \frac{-1}{(2N-1)^2}, \tag{7}$$

where the indices D and R denote the square entropy due to patterns from the D and R subsets, respectively.

If we follow Ref. 3, a cost function defined by the relation

$$M = \sum_{f_n^D} S^D - \sum_{f_n^R} S^R \tag{8}$$

has its ideal minimum determined by the ideal values given in Eqs. (7).

Considering M to be a functional of a specific filter function, we seek a reference function $h_{\mathrm{GEF}}(i, j)$ that will minimize the cost function,

$$M_{\min} = M[h_{\mathrm{GEF}}(i, j)]. \tag{9}$$

Since currently available spatial light modulators (SLM's) operate best in a binary mode, we restrict our actual filter function to this mode. However, for higher performance,[6] with a proper system modification, we can use the effective values $h(i, j) \in \{-1, 1\}\ \forall\ i, j$. To use a GA one should have the following features[7]: (i) a chromosomal representation of solutions to the problem, generally binary; (ii) an evaluation function that gives the fitness of the population; or (iii) combination rules (genetic operators) to produce new structures from old ones—reproduction, crossover, and mutation.

If we regard the cost function in Eq. (8) as a fitness value for a given spatial filter function, our process lends itself quite readily for implementation by GA, where each binary filter function constitutes a member of the population. The algorithm used is summarized as follows:

(1) Start: Select at random a population of m members (binary functions) $\{h_1, h_2, \ldots, h_m\}$ and evaluate the values of the cost functions, M_i $\{i = 1, 2, \ldots, m\}$. Compute the average value of the cost function $\theta = 1/m \sum_{i=1}^{m} M_i$. Set a discrete time parameter t to zero. Define a probability P for a mutation to occur and set it to some $P_{\max}$.

(2) Crossover/mutate: Select the function h_l that corresponds to the minimal cost function, M_l. Pick from the population a function h_j at random. The two functions, h_l and h_j, are the parents to be used for generating an offspring function. Select a random integer k between 0 and n, where n is the dimension of the vectors h representing the filter functions. Create the offspring function, h_c, by taking the first k elements from one of the parents, randomly, and the remaining $n - k$ elements from the other parent. Induce a mutation (inverting the sign of the elements) with probability P on each element of the offspring vector h_c. Evaluate the offspring cost function, M_c.

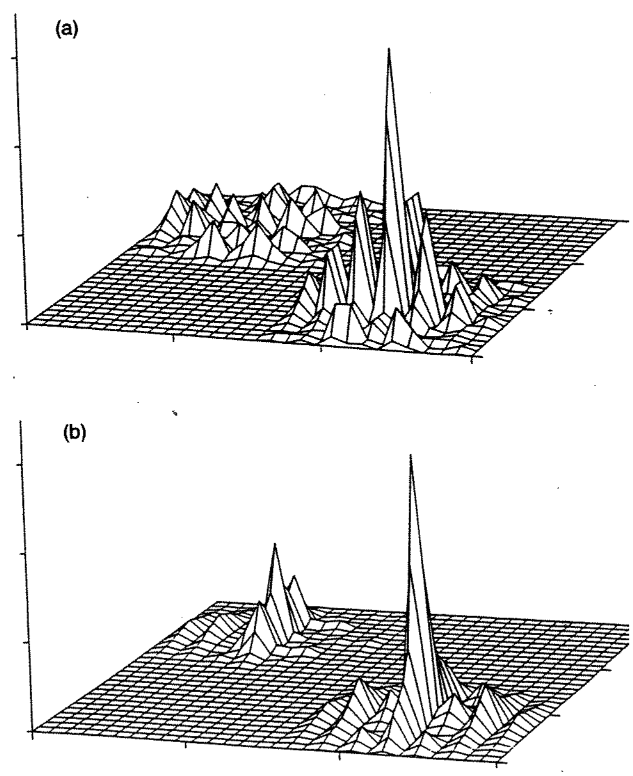

Fig. 3. Output correlation intensity with (a) the refere function produced by GA (discrimination ratio of 1:7.1) (b) the original function as reference (discrimination rati 1:2.7).

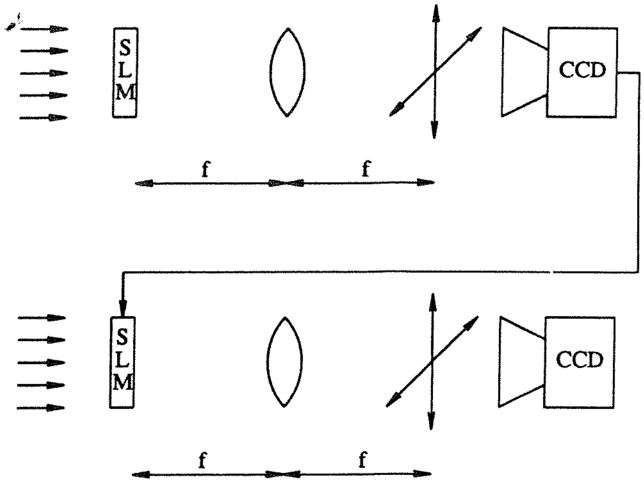

Fig. 4. Joint-transform correlator architecture for electro-optical implementation of learning algorithm. One of the reference functions and the training set are presented on the upper SLM, and the control computer (not shown) analyzes the output signal detected by the lower charge-coupled-device (CCD) camera.

(3) Reproduce: Pick at random a function h_d from the population subject to the constraint $M_d \geq \theta$. Replace h_d in the population with h_c and update the average value of the cost function, $\theta \rightarrow \theta + (1/m)(M_c - M_d)$.

(4) Setting parameters: Set the new parameters, $t \rightarrow t + 1$ and $P \rightarrow P_{\max}(1/t)^r$. Terminate the procedure when adequate discrimination is achieved (a predetermined value of M). If $P > P_{\min}$, go to step (2); otherwise go to step (1). Selection of the parameters r, $P_{\min}$, and $P_{\max}$ depends on the particular problem at hand. In most cases tested the procedure was terminated much before $P_{\min}$ was reached.

In our simulation experiments the input pattern was as shown in Fig. 1. A function $h(x, y)$ was generated to detect pattern (a) and to reject pattern (b). Two members of the final population are shown in Fig. 2. These filters yield a discrimination ratio of 1:7.1 as shown in the correlation results of Fig. 3(a). For comparison, Fig. 3(b) shows the correlation plane for a binarized matched filter having a discrimination ratio of 1:2.7.

To improve the filter generation process we propose an electro-optical learning architecture as shown in the schematic diagram of Fig. 4. Owing to the limitations of the joint-transform correlator,[11] all the pixel values of the candidate binary filters are stored in the computer memory and are presented sequentially to the system, following the algorithm described. The selected filter function is presented on the upper SLM, side by side with the patterns of the training set, and the correlation peaks are detected over the output plane by the charge-coupled-device camera. The cost function is calculated and the genetic operations are performed by the controlling computer.

(After the original submission of this Letter the electro-optical system was implemented as described and also in a 4-f architecture.)

This research was performed within the Technion Advanced Opto-Electronics Center established by the American Technion Society, New York.

References

1. B. V. K. Vijaya Kumar, Z. Bahri, and L. Hassebrook, Proc. Soc. Photo-Opt. Instrum. Eng. **960,** 18 (1988).
2. R. Kallman, Appl. Opt. **25,** 1032 (1986).
3. M. Fleisher, U. Mahlab, and J. Shamir, Appl. Opt. **29,** 2091 (1990).
4. U. Mahlab and J. Shamir, Opt. Lett. **14,** 146 (1989).
5. P. J. M. van Luarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications* (Reidel, Dordrecht, The Netherlands, 1987).
6. J. Rosen, U. Mahlab, and J. Shamir, Opt. Eng. **29,** 1101 (1990).
7. D Lawrence, *Genetic Algorithm and Simulated Annealing* (Kaufmann, Los Altos, Calif., 1987).
8. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, Mass., 1987).
9. J. Ziv and M. Zakai, IEEE Trans. Inf. Theory **IT-19,** 275 (1973).
10. A. W. Robert and D. E. Veberg, *Convex Function* (Academic, New York, 1973).
11. D. A. Gregory and J. A. Loudin, Proc. Soc. Photo-Opt. Instrum. Eng. **1053,** 198 (1989).