Sidelobe elimination for generalized synthetic discriminant functions by a two-filter correlation and subsequent postprocessing of the intensity distributions

Mario Montes-Usategui, Juan Campos, Ignasi Juvells, and Santiago Vallmitjana

One of the most important problems in optical pattern recognition by correlation is the appearance of sidelobes in the correlation plane, which causes false alarms. We present a method that eliminate sidelobes of up to a given height if certain conditions are satisfied. The method can be applied to any generalized synthetic discriminant function filter and is capable of rejecting lateral peaks that are even higher than the central correlation. Satisfactory results were obtained in both computer simulations and optical implementation.

1. Introduction

The design of filters that use an optical correlator for pattern recognition has undergone great development during the past few years. The matched filter,[1] introduced by VanderLugt in 1964, consists of the Fourier transform of the object to be recognized. This filter gives the maximum signal-to-noise ratio in the correlation plane but is unable to discriminate between similar objects. The discrimination capabilities and the light efficiency can be enhanced with a phase-only filter,[2] whose main drawback is its sensitivity to noise.

The response of these filters depends on the scale, the orientation, and in general any deformation in the input pattern. A possible solution is to expand the reference objects into a set of orthogonal functions that are invariant to one of these deformations. For instance, rotation invariance can be achieved with a circular-harmonic expansion[3] (CHE). The information about the object to be recognized and the object to be discriminated against can be introduced simultaneously by means of synthetic discriminant filters[4] (SDF's).

The authors are with the Laboratori d'Optica, Departament de Física Aplicada i Electrònica, Universitat de Barcelona, Diagonal 647, Barcelona 08028, Spain.

Received 1 June 1993; revised manuscript received 4 November 1993.

0003-6935/94/143050-10$06.00/0.

The main shortcoming in the last two designs is the appearance of sidelobes caused by the loss of information in the circular-harmonic-expansion case and the lack of control over the whole correlation plane in the SDF approach. Several partial solutions to this problem can be found in the literature.[5,6]

The SDF filters are designed so that the central correlation with the training set takes prespecified values. This can be though of as a set of K equations with N unknowns, where K is the number of objects in the training set and N is the number of components in the filter. Such a system has infinite solutions, of which the SDF is a particular one, formed by a linear combination of the training images. The N–K degrees of freedom in the system may be used to optimize a performance criterion giving a family of filters known as generalized SDF's.[7] The minimum average correlation energy (MACE) filter[8] is a particular solution that minimizes the energy in the correlation plane, thus providing sharp peaks and reducing the sidelobes. There are other filters such as entropy optimized filters[9] or maximum-discrimination filters[10] that are also designed to control the whole correlation plane and to produce sharp correlations.

In this study we present a method to eliminate sidelobes, provided certain conditions are met. The process consists of the use of two filters and the subsequent postprocessing of the correlation distributions. The two filters are obtained by a new filter being established that modifies the whole correlation plane except the central point. This new filter is then either added to or subtracted from the filter being corrected.

The remainder of the paper is organized as follows. In Section 2 we introduce the method and develop the mathematical expression of the filter. In Section 3 the validity conditions are discussed and a performance analysis is carried out. The results of a computer simulation of the method are presented in Section 4, together with several examples. In Section 5 we present the results of the optical implementation of the method, and in Section 6 some remarks and conclusions are made.

2. Method

A. Theoretical Considerations

If we perform the correlation of an input image with two differently designed filters restricted to produce the same central correlation, we can postprocess the output distributions in order to improve the results. For example, if we binarize them by applying a given threshold and then multiply the output planes pixel by pixel (which is equivalent to processing them with the logical operation AND), we can eliminate the sidelobes that are more than the threshold value that are not common to both planes, and we can maintain the central peak.

The existence of false alarms in such a procedure depends on the appearance of common sidelobes. The method we present ensures, in certain conditions, that no sidelobe is common to both correlation planes.

The idea of an image being processed by means of two digital or optical processes and of the final image being obtained by the pointwise multiplication of each single output has been applied in the past in various problems. The procedure is similar to operations commonly used in mathematical morphology. Casasent *et al.*,[11] using the correlation with two different filters, detected simple geometrical shapes immersed in high clutter and noise. The procedure involved the binarization of the correlations and the pixel-by-pixel multiplication of both results. The method was an optical implementation of the morphological hit–miss transform. More recently, Crowe *et al.*[12] proposed the utilization of a similar method to reduce the sidelobes appearing in imaging systems owing to the finite size of the pupil, thus improving the spatial resolution.

We restrict our study to the case in which we obtain real-valued correlations. The general case of having complex distributions (such as those produced by circular-harmonic-expansion filters) is treated briefly in Section 6.

The method can be applied to a wide variety of filters (which is referred to as the base filters in what follows) and consists of addition and subtraction of a new filter (called the correcting filter), designed so that the following conditions are fulfilled:

(I) The filter is orthogonal to every image in the training set (the filter and the images are treated as vectors with the usual lexicographic ordering). This requirement ensures that the central correlations produced by the base filter remain unchanged.

(II) The correlation between the correcting filter and the images produces a constant plane with a predefined value. This condition is accomplished only in an approximate form by means of a Lagrange minimization process.

The output distributions obtained with the new filters have two terms:

$$(h_b + h_c) * x_i = h_b * x_i + h_c * x_i, \qquad (1)$$

$$(h_b - h_c) * x_i = h_b * x_i - h_c * x_i, \qquad (2)$$

where h_b is the base filter, h_c is the correcting filter, x_i is one of the images in the training set, and the symbol $*$ means correlation.

The expression $h_b * x_i$ is real and may take positive and negative values. The term $h_c * x_i$ in Eq. (1), which is constant over the whole plane, increases the positive sidelobes and decreases the negative ones. Conversely in Eq. (2), $h_c * x_i$ increases the negative sidelobes and decreases the positive ones. A suitable choice of the value of the constant plane and the threshold ensures that no sidelobe is common to both binarized correlations. These considerations are discussed in detail in Section 3.

Although we put the emphasis on generalized SDF filters because of the practical importance of these designs, the method can be applied to other filters with minor modifications. The only condition that is necessary for the base filter is to show a good discrimination between the true and the false classes since the only point that is not changed is the central correlation. Therefore the method can be used in conjunction with filters that are already designed to avoid sidelobes, such as the entropy optimized filters[9] or the maximum-discrimination filters,[10] to achieve higher discrimination capabilities.

B. Filter Design

Let $x_1(j), x_2(j), \ldots, x_k(j)$ be the k training images of N components $(j = 1, \ldots, N)$, and let $X_1(w), \ldots, X_k(w)$ denote their Fourier transforms. Let $H_c(w)$ be the Fourier transform of the correcting filter.

Condition (I) in Subsection 2.A can be written as follows:

$$\sum_{w=1}^{N} H_c(w) X_i^*(w) = 0, \qquad i = 1, \ldots, k. \qquad (3)$$

Condition (II), as mentioned above, can be achieved only approximately by minimization of the following error function:

$$E = \frac{1}{K} \sum_{i=1}^{K} \sum_{w=1}^{N} |D(w) - H_c(w) X_i^*(w)|^2, \qquad (4)$$

where $D(w)$ represents the Fourier transform of the desired shape for the correlation between the images and the filter (a plane in our case). Expression (4) is therefore a measure of the mean error between the correlations obtained and those desired. This filter is a particular case of the minimum-squared-error synthetic discriminant function[13] design introduced by Vijaya Kumar *et al.*

Expression (3) can be written compactly as

$$S^{+}h_c = 0, \tag{5}$$

where h_c is the N-dimensional vector whose components are $H_c(w)$:

$$h_c = [H_c(1), H_c(2), \ldots, H_c(N)]^T. \tag{6}$$

S is an $N \times K$ matrix formed by the Fourier transforms of the images arranged in columns:

$$S = \begin{bmatrix} X_1(1) & X_2(1) & \cdots & X_K(1) \\ X_1(2) & X_2(2) & \cdots & X_K(2) \\ \vdots & \vdots & \vdots & \vdots \\ X_1(N) & X_2(N) & \cdots & X_K(N) \end{bmatrix}. \tag{7}$$

The superscript + means the conjugate transpose of the matrix. Finally, 0 represents the K-dimensional vector with all its components zero.

We can express Eq. (4) with the same formalism by defining the $N \times N$ diagonal matrix,

$$P_i = \begin{bmatrix} X_i(1) & 0 & \cdots & 0 \\ 0 & X_i(2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & X_i(N) \end{bmatrix}, \tag{8}$$

and the N-dimensional vector,

$$d = [D(1), D(2), \ldots, D(N)]^T. \tag{9}$$

With such definitions the error can be written as

$$E = \frac{1}{K}\sum_{i=1}^{k}[(d - P_i^{*}h_c)^{+}(d - P_i^{*}h_c)]$$

$$= d^{+}d - h_c^{+}r - r^{+}h_c + h_c^{+}Ph_c, \tag{10}$$

where

$$r = \frac{1}{K}\sum_{i=1}^{K}(P_i d), \tag{11}$$

$$P = \frac{1}{K}\sum_{i=1}^{K}(P_i^{*}P_i). \tag{12}$$

Conditions (I) and (II) can be accomplished simultaneously by minimization of

$$L[h_c] = (d^{+}d - h_c^{+}r - r^{+}h_c + h_c^{+}Ph_c) - 2\lambda^{+}(S^{+}H_c) \tag{13}$$

by means of a Lagrange optimization process. In expression (13), λ denotes a K-dimensional complex vector containing the Lagrange multipliers.

By calculating the gradients with respect to the filter components and the Lagrange multipliers and by setting them to zero, we obtain the following expression for the correcting filter (the mathematical details can be found elsewhere[13]):

$$h_c = [I - P^{-1}S(S^{+}P^{-1}S)^{-1}S^{+}]P^{-1}r. \tag{14}$$

3. Necessary Conditions and Performance Analysis

Let us suppose we have designed a generalized SDF that solves a two-class problem. The prespecified values for the correlation with classes A and B are called p_0 and p_1, respectively, and with no loss of generality we suppose that $p_0 > p_1$. The criterion for classification of an image as a member of one of the two classes is the following: If the correlation intensity at the center exceeds a given threshold, the image is classified as belonging to class A; otherwise, the image is assigned to class B.

As commented on in Section 2, the method for elimination of sidelobes involves the correlations with two filters that have an opposite effect. The first, which we call the positive filter, reduces the negative sidelobes and enhances the positive ones. The second filter, called negative, reduces the positive peaks and enhances the negative ones. Our goal is to determine the proper settings for the threshold and the constant plane resulting from the correlations between the images and the correcting filter in such a way that the only point in the output intensity distributions that passes the threshold in both cases is the central peak.

The equations that ensure the above statement can be written as follows:

$$\theta(|x| + c)^2 < p_0^2, \tag{15}$$

$$(|x| - c)^2 < \theta p_0^2, \tag{16}$$

$$c^2 < \theta p_0^2, \tag{17}$$

$$p_1^2 < \theta p_0^2, \tag{18}$$

where θ is a factor between zero and one that represents the threshold, c is the value of the constant plane, and x is the height of the maximum sidelobe to be suppressed.

The necessity of conditions (15)–(18) is discussed in the following considerations. Maximum sidelobe x increased by constant c may become higher than value p_0 of the correlation for class A. This situation is illustrated in Fig. 1. In Fig. 1(a) the positive sidelobe ($x > 0$) is increased by positive constant c. The resulting intensity can be seen in Fig. 1(b), in which it appears higher than the intensity of the central correlation p_0^2. For negative sidelobes we have an equivalent situation. In Fig. 1(c) the negative sidelobe ($x < 0$) is increased (in absolute value) by negative constant $-c$, and the resulting intensity

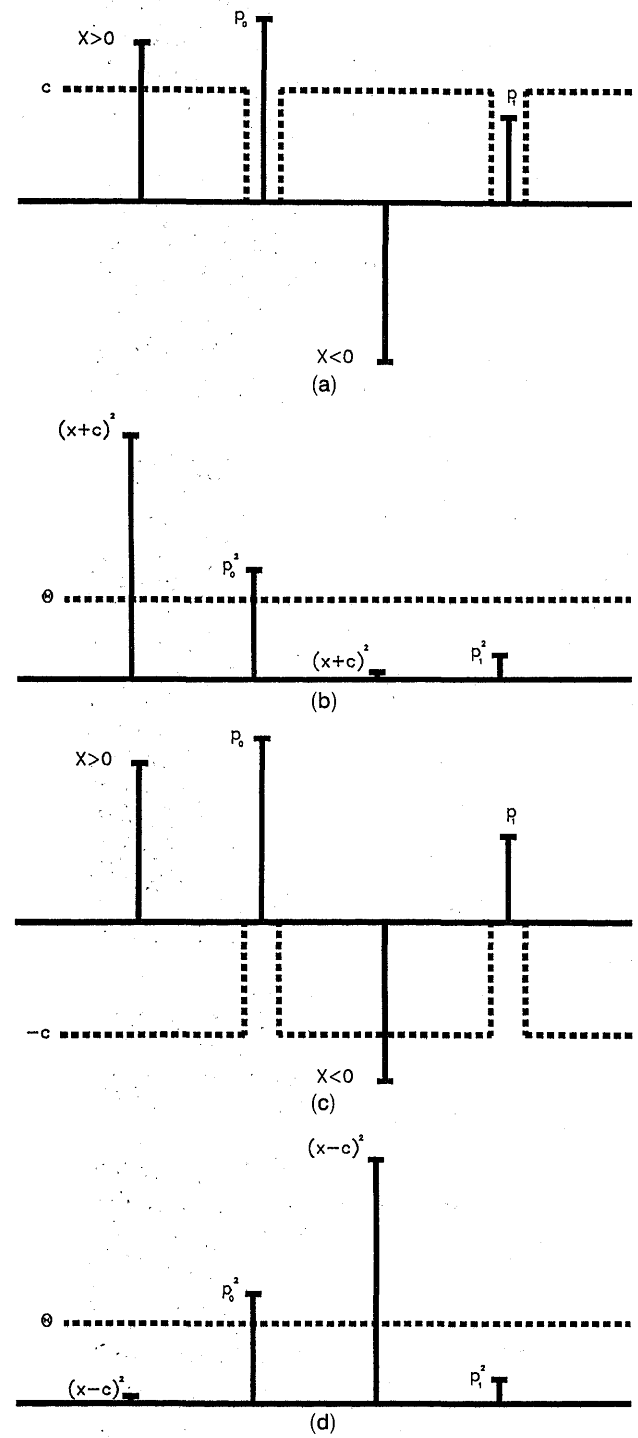

Fig. 1. (a) Amplitude of the correlation plane produced by the base and the correcting filters when both are added (positive filter). The dashed curves represent the constant amplitude correlation value given by the correcting filter. The solid lines represent the correlation peaks (p_0 and p_1) and the sidelobes ($x > 0$ and $x < 0$) obtained with the base filter. (b) Intensity distribution for the positive filter. (c), (d) Same as (a) and (b) for the case of the negative filter.

[Fig. 1(d)] is also higher than the central correlation for class A. Inequality (15) is then necessary to guarantee that p_0 passes the threshold. The expression $(|x| + c)^2$ is the intensity of the increased peak no matter whether the maximum sidelobe is positive or negative. In other words, this factor takes into account the case $(x - c)^2$ for negative peaks and $(x + c)^2$ for positive ones.

Inequality (16) represents the condition for the decreased maximum sidelobe to be eliminated in the binarization; i.e., when the absolute value of the sidelobe is decreased, the resulting intensity has to be lower than the limit marked by the threshold value [see Figs. 1(a) and 1(b) for $x < 0$ and Figs. 1(c) and 1(d) for $x > 0$]. The condition is written assuming that the threshold is applied on p_0, but the maximum value in the output intensity distribution might be different. For instance, if we had two maximum sidelobes, both with the same absolute value but with different sign, and if the increased term $(|x| + c)^2$ were higher than p_0^2, the threshold value would be $\theta(|x| + c)^2$ because when one decreases, the other increases (the case represented in Fig. 1). By setitng a more restrictive condition, such as Eq. (16), we can cover all the cases.

In addition, we need to ensure that the sidelobes lower than x disappear. We have to treat two cases separately:

(a) If $c < |x|/2$, then $(|x| - c)^2 > (|x| - |x|/2)^2 > c^2$. In this situation, inequality (16) guarantees the elimination of every sidelobe in the range from zero to x [see Figs. 2(a) and 2(b)].

(b) If $c > |x|/2$, the former condition is not assured, and small sidelobes may surpass the threshold. This situation is depicted in Figs. 2(c) and 2(d).

Inequality (17) is then necessary to take into account case (b). Finally, inequality (18) is required for a correct classification of class B.

In the typical case in which $p_0 = 1$ and $p_1 = 0$, inequalities (15)–(18) become

$$|x| - \theta^{1/2} < c < \theta^{-1/2} - |x|, \tag{19}$$

$$c < \theta^{1/2}. \tag{20}$$

As can be observed in the above expressions, the value of c is not completely determined by the conditions, but there is a range of permissible values that is necessary for the reliability of the method. This necessity is caused by, as mentioned above, the correlation planes with the correcting filter not being exact planes but approximate versions obtained by means of a minimization process.

The maximum sidelobe that fulfills expression (19) is obtained when

$$|x| - \theta^{1/2} = \theta^{-1/2} - |x|; \tag{21}$$

whence

$$|x| = (\theta + 1)\theta^{-1/2}/2. \tag{22}$$

On the other hand, by considering inequality (20) and

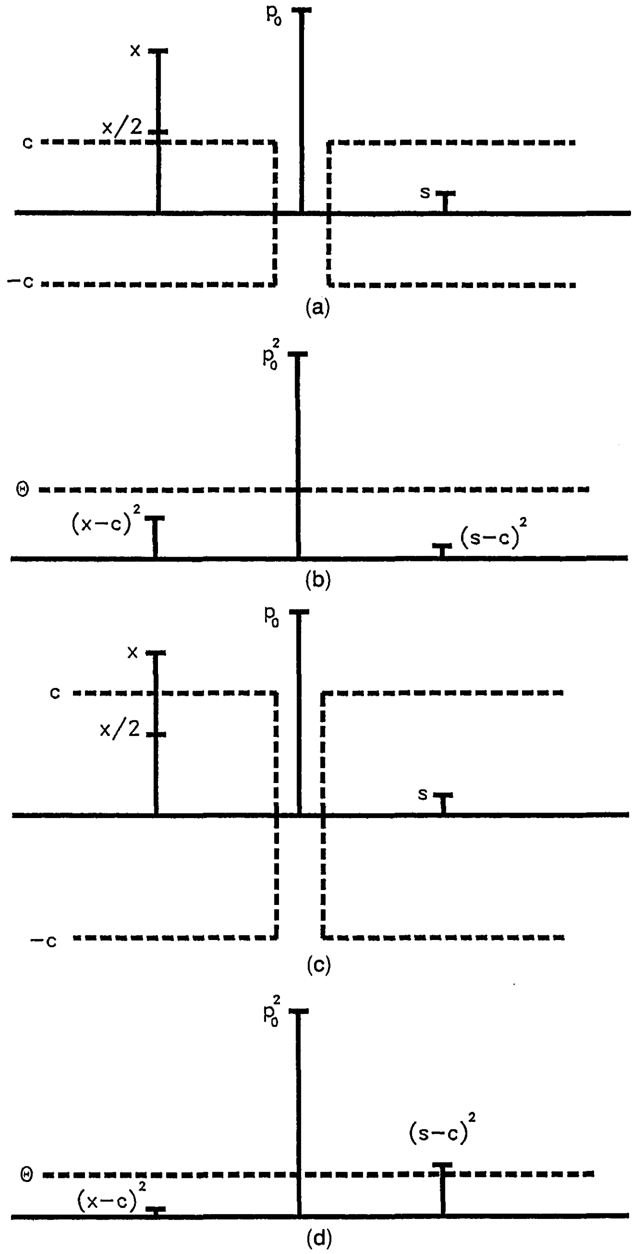

Fig. 2. Illustration of the necessity of inequality (17): (a) Case $c < x/2$. The solid lines represent the amplitude of the correlation with the base filter, which produces a high (peak x) and a small (peak s) sidelobe. (b) Intensity distribution corresponding to (a). If $c < x/2$, sidelobes lower than x are suppressed by Eq. (16). (c) The same as (a) when $c > x/2$. (d) Intensity distribution corresponding to (c). When $c > x/2$, small sidelobes increased by the constant may surpass the threshold value.

the leftmost part of inequality (19), or

$$c < \theta^{1/2},$$

$$c > |x| - \theta^{1/2},$$

we obtain the maximum $|x|$ when

$$|x| - \theta^{1/2} = \theta^{1/2}, \qquad (23)$$

whence $|x|$ is

$$|x| = 2\theta^{1/2}. \qquad (24)$$

Therefore the maximum sidelobe can be written as

$$|x|_{\max} = \min[2\theta^{1/2}, (\theta + 1)\theta^{-1/2}/2]. \qquad (25)$$

The optimum threshold, θ, would be

$$2\theta^{1/2} = (\theta + 1)\theta^{-1/2} \Rightarrow \theta = 1/3,$$

and finally $|x|_{\max} \approx 1.15$ (115% of the central correlation) in amplitude or $|x|^2_{\max} \approx 1.32$ (132% of the central correlation) in intensity.

In practice it is not possible to reach this limit because the range of permissible values for c in inequality (19) is reduced to a single point. However, in most practical situations the method enables the elimination of sidelobes higher than the central peak, which cannot be corrected by binarization of the output intensity produced by the base filter with a single threshold.

Figure 3 represents the permissible variation of constant c (shaded area) when the threshold value is fixed ($p_0 = 1, p_1 = 0$, and $\theta = 0.5$). The graph shows that for small variations in c, large sidelobes can be eliminated, but for wide variations the height of the sidelobe must be small. By varying the threshold and by fixing the desired height of the sidelobe to be suppressed, we obtain Fig. 4. ($|x| = 1$). As can be observed, there is a value of θ for which the permissible range of c is maximum because of the monotonic behavior of the restrictions in inequalities (19) and (20). This optimum threshold can be calculated by use of the following equation:

$$\theta_{op}^{1/2} = \theta_{op}^{-1/2} - |x|. \qquad (26)$$

Consequently we can determine all the parameters

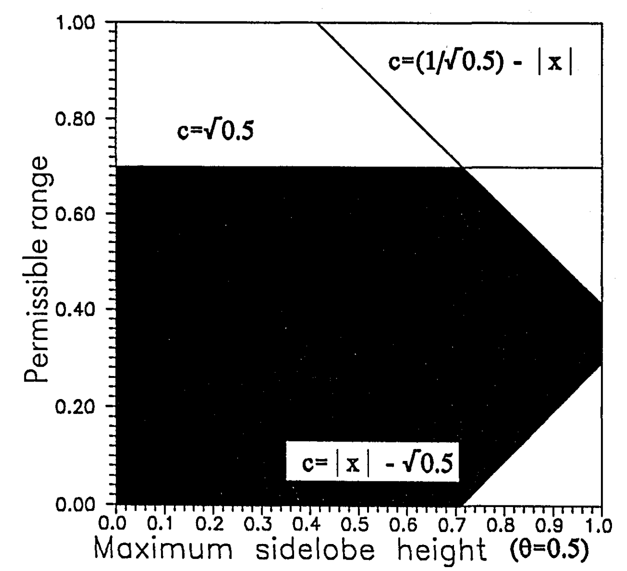

Fig. 3. Permissible range for constant c as a function of the maximum sidelobe to be eliminated ($p_0 = 1, p_1 = 0, \theta = 0.5$).

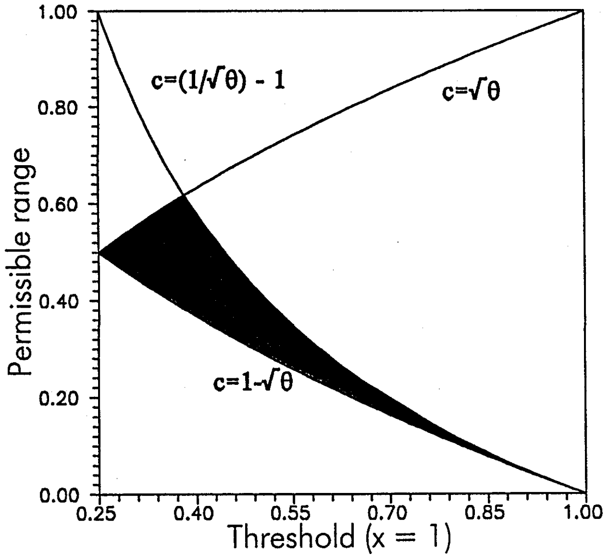

Fig. 4. Permissible range for constant c as a function of the threshold ($p_0 = 1, p_1 = 0$, and $|x| = 1$).

envisaged by the method by selecting the maximum height of the sidelobe to suppress, by calculating the optimum threshold by means of Eq. (26), and finally by choosing constant c as the midpoint value of the range given by inequality (19).

The validity of the method is determined by the extent to which the actual correlations between the images and the correcting filter are constant planes with the expected accuracy. The possibility of the sidelobes being eliminated in the range from 0 to a prespecified value, $|x|$, depends on whether the minimization procedure is capable of producing a correcting filter so that the correlations obtained with the images in the training set satisfy inequality (19). Therefore the performance of the method is determined by the deviations in the correlation distribution from the expected plane; in other words, the maximum sidelobe that can be suppressed depends on the range of variation of the points in the correlation plane. On the other hand, the height of the sidelobes depends on the similarity between images in different classes; i.e., the more similar the images with different conditions are, the larger the expected values of the sidelobes become.

The method gives more accurate correlations planes, i.e., the correcting power is higher, when the images in the training set are more similar and therefore when higher sidelobes appear. To demonstrate this property, let us assume we have two N-dimensional images whose Fourier transforms are $X_1(w)$ and $X_2(w)$.

The expression for the error in Eq. (4) can then be written as

$$E = \frac{1}{2}\left[\sum_{w=1}^{N} |D(w) - H_c(w)X_1^*(w)|^2 + \sum_{w=1}^{N} |D(w) - H_c(w)X_2^*(w)|^2\right] \quad (27)$$

or in vectorial form as

$$E = \tfrac{1}{2}[(d - P_1^*h_c)^+(d - P_1^*h_c) + (d - P_2^*h_c)^+(d - P_2^*h_c)]. \quad (28)$$

Let us suppose that P_1 and P_2, which are diagonal matrices, can be inverted (i.e., there is no frequency for which the Fourier transform of the images is zero), and let us define filters h_1 and h_2 so that

$$h_1 = (P_1^*)^{-1}d, \quad (29a)$$

$$h_2 = (P_2^*)^{-1}d; \quad (29b)$$

h_1 and h_2 represent the filters that give exactly the desired shape when they are correlated with images X_1 and X_2, respectively. The assumption that P_1 and P_2 are invertible is not so restrictive, and similar requirements are needed in other filter designs. In particular, in MACE filters, the matrix that represents the average energy of the images in the training set must also be full rank.

Because h_c is the optimum filter, by substituting h_c for h_l in Eq. (28) we have

$$\begin{aligned} E &\le \tfrac{1}{2}[(d - P_1^*h_1)^+(d - P_1^*h_1) + (d - P_2^*h_1)^+(d - P_2^*h_1)] \\ &= \tfrac{1}{2}[(d - P_2^*h_1)^+(d - P_2^*h_1)] \\ &= \tfrac{1}{2}[(P_2^*h_2 - P_2^*h_1)^+(P_2^*h_2 - P_2^*h_1)] \\ &= \tfrac{1}{2}\{[P_2^*(h_2 - h_1)]^+P_2^*(h_2 - h_1)\}, \end{aligned} \quad (30)$$

and from Eqs. 29(a) and 29(b),

$$P_2^*h_2 - P_1^*h_1 = 0. \quad (31)$$

If we express image X_2 as a function of X_1, we can write

$$P_2 = P_1 + \Delta, \quad (32)$$

and therefore

$$(P_1^* + \Delta^*)h_2 - P_1^*h_1 = 0, \quad (33)$$

$$P_1^*(h_2 - h_1) = -\Delta^*h_2 = -\Delta^*(P_1^* + \Delta^*)^{-1}d. \quad (34)$$

As P_1 is a full-rank matrix, if Δ tends to zero, then $(h_2 - h_1)$ tends to zero, and in consequence the error in inequality (30), which depends on this difference, becomes increasingly small:

$$\Delta \to 0, P_2 \to P_1 \rightleftarrows (h_2 - h_1) \to 0 \rightleftarrows E \to 0. \quad \text{Q.E.D.}$$

We carried out an experimental verification of this property. The details are given in Section 4.

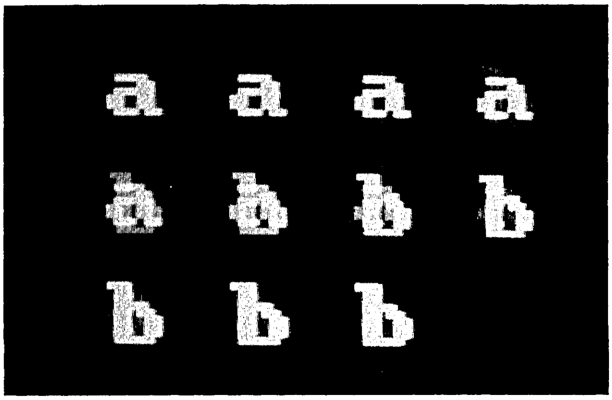

Fig. 5. Sequence of images used in the simulation. Letter a is ab(0), letter b is ab(10), and the intermediate patterns are ab(i), with $i = 1, \ldots, 9$.

4. Computer Experiments

In this section we present the results of several experiments carried out by means of a computer simulation in order to test the suitability of the method in practical situations. We performed a study of the dependence between the value of the expected sidelobes and the correcting capabilities of our method by using the images depicted in Fig. 5. A set of ten correcting filters was designed, each of them calculated by use of a pair of images from the sequence ab(0)–ab(i); namely, filter 1 was calculated with ab(0) and ab(1), filter 2 with ab(0) and ab(2), and so on. The measure of the similarity between image ab(i) and ab(0) was calculated with the following expression:

$$S[\mathrm{ab}(0), \mathrm{ab}(i)] = \frac{|[\mathrm{ab}(0) * \mathrm{ab}(i)]_{(0,0)}|^2}{|[\mathrm{ab}(0) * \mathrm{ab}(0)]_{(0,0)}|^2}, \qquad (35)$$

where the symbol $*$ means correlation.

The error function in Eq. (4) as well as the deviation from a perfect plane ($c = 0.35$) for each single image was computed for every correcting filter, and the results are represented in Fig. 6. The graph shows the dependence between these deviations from the expected shape and the similarity measure given by Eq. (35). For very similar images such as ab(0)

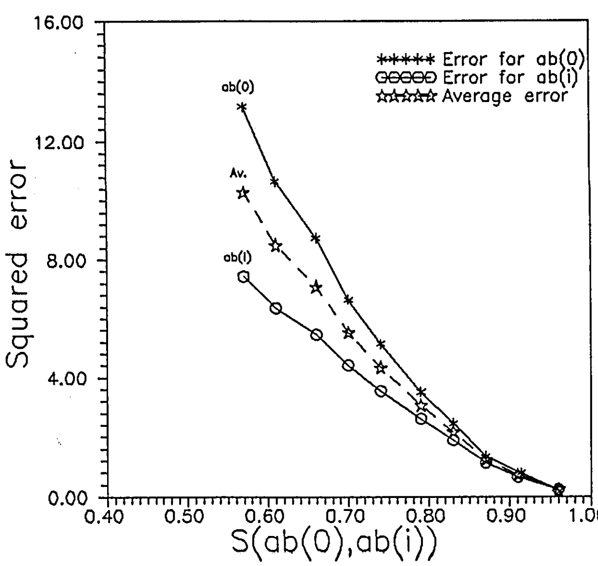

Fig. 6. Squared error as a function of similarity between images.

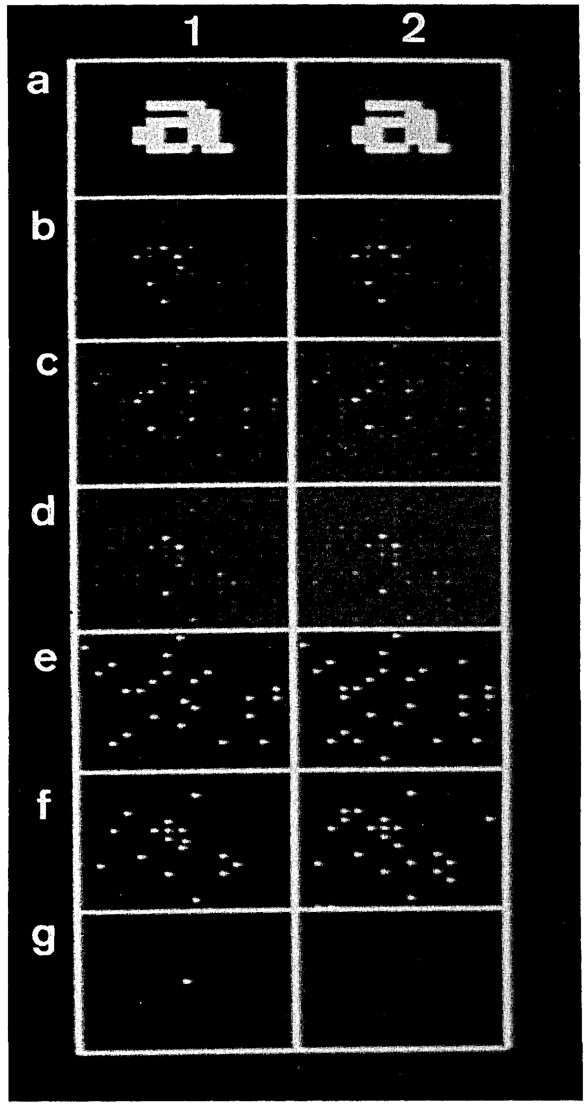

Fig. 7. a1, a2, Images used to design the MACE filter. The imposed values were 1 for image a1 and 0 for image a2. b1, b2, Intensity of the correlation betweeen the MACE filter and images a1 and a2, respectively. c1, c2, Intensity of the correlation between the positive filter and images a1 and a2. d1, d2, Intensity of the correlation between the negative filter and images in a1 and a2. e1, e2, Same as images c1 and c2 binarized with $\theta = 0.36$. f1, f2, Same as images d1 and d2 binarized with $\theta = 0.36$. g1, Result of pixel-by-pixel multiplication of images e1 and f1. g2, Result of pixel-by-pixel multiplication of images e2 and f2.

Fig. 8. Images used to design the filters. The imposed values for the central correlations were 1 for S, 1 for C, and 0 for E.

and ab(1) $\{S[\text{ab}(0), \text{ab}(1)] = 0.96\}$ the error is small ($E = 0.25$), and conversely, when the similarity between images is small $\{S[\text{ab}(0), \text{ab}(1)] = 0.57$ for ab(0) and ab(10)$\}$, the error is high ($E = 10.3$). The result shows the expected behavior; i.e., when large sidelobes are more likely to appear, owing to the similarity between images with different constraints, the procedure we propose is more powerful because of a smaller variation with respect to the desired plane.

The increasing correcting power enables the elimination of sidelobes, even if they are higher than the central peak, as illustrated in Fig. 7. In Fig. 7, two images with a similarity $S = 0.90$ were used to build a MACE filter by imposition of images (a1) and (a2) to give values of 1 and 0, respectively. The MACE filter is an antisidelobe design, but in this situation it gives several lateral peaks, the largest of which has a value of 126% of the central correlation (in intensity), as shown in images (b1) and (b2).

In order to eliminate the sidelobe, we prepared the correcting filter using the following parameters:

$$|x| = (1.26)^{1/2} = 1.12,$$

$$\theta_{op} = 0.36,$$

$$c = 0.53.$$

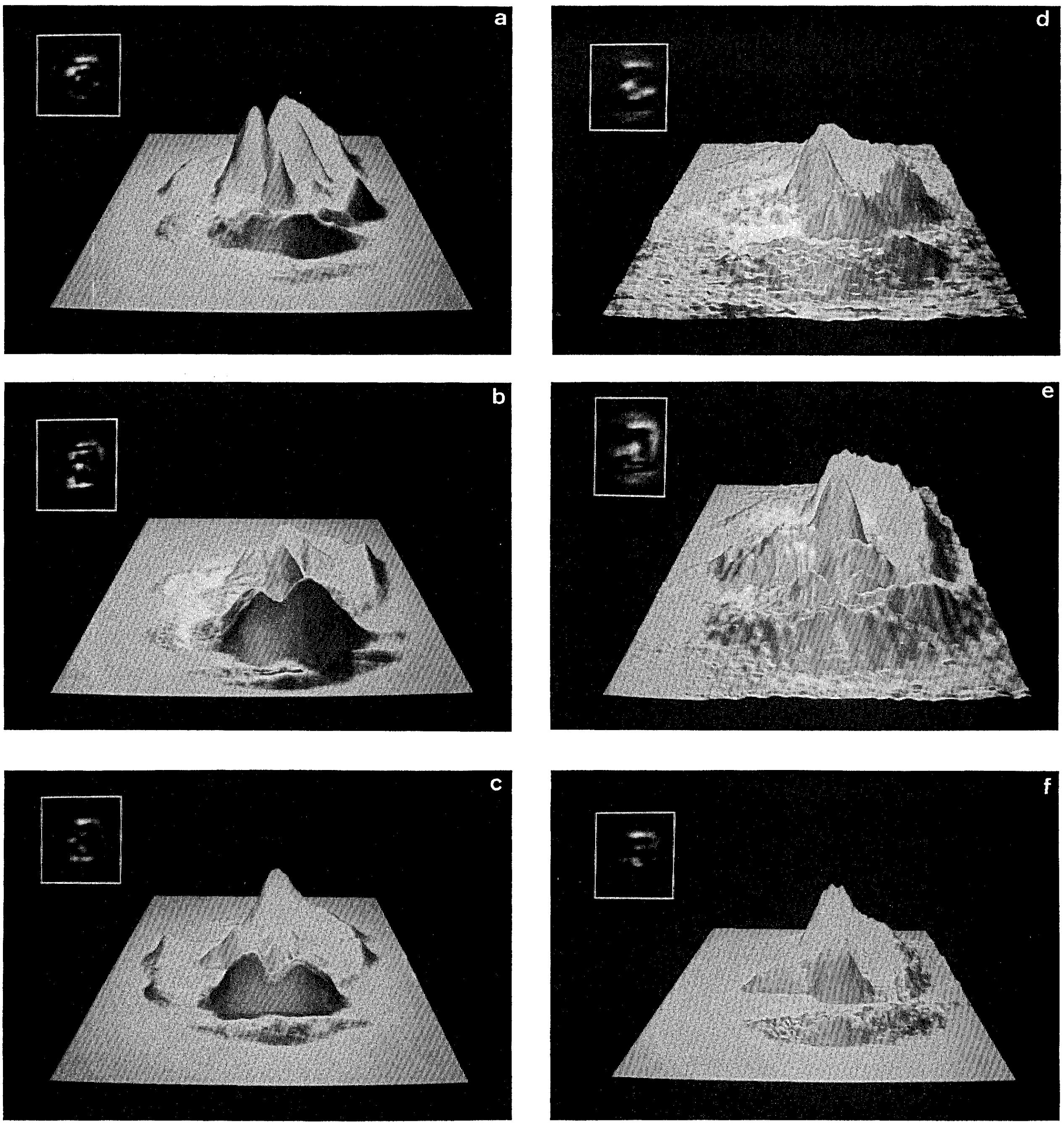

Fig. 9 continued.

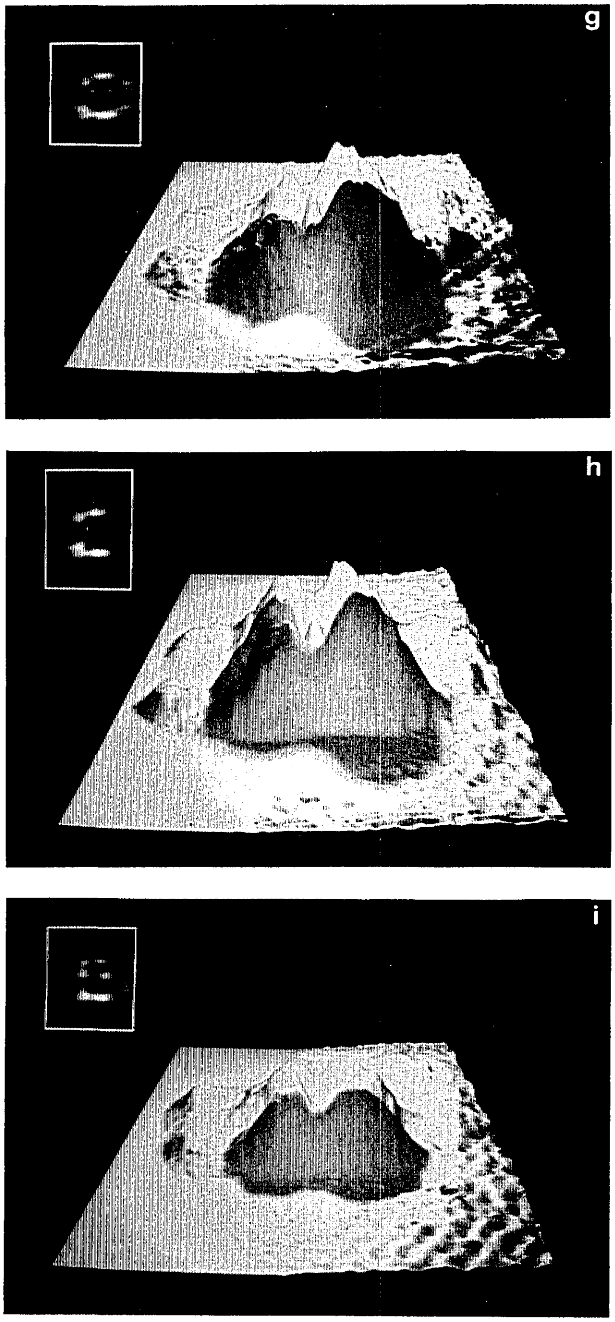

Fig. 9. (a), (b), (c) Intensity of the correlation between the composite filter and letters S, C, and E, respectively. (d), (e), (f) Same as (a), (b), and (c) with the positive filter. (g), (h), (i) Same as (a), (b), and (c) with the negative filter.

The results of the correlations between the positive and the negative filters are shown in images (c1)–(d2) of Fig. 7, and the binarized results are shown in images (e1)–(f2). After pixel-by-pixel multiplication of both binarized planes, images (g1) and (g2) were obtained. As can be observed, all the sidelobes are suppressed, and a perfect detection of the central correlations is possible.

The method is then capable of producing a significant increase in the discriminant abilities of the SDF filters, including those such as the MACE design, which are specifically designed to avoid the appearance of sidelobes. However, there is no noise resistance included in the minimum-squared-error SDF filter design, so the procedure is highly sensitive to noisy inputs.

The possible solutions to this problem are the same as those used to introduce noise resistance in the MACE design: trade-off filters[14] and the modification used in minimum noise and correlation energy filters.[15] A further study on the suitability of these solutions will be carried out in the future.

5. Optical Results

The method proposed for elimination of sidelobes was tested by use of a convergent correlator.[16] This setup has an advantage in that it permits easy matching between the scales of both the input image and the filter.

The filters were built by means of computer-generated holograms codified by Burkhardt's method,[17] displayed on a laser printer, and photoreduced. The holograms were sandwiched to avoid uncontrolled phases owing to thickness variations in the photographic film. A low-power He–Ne laser provided the coherent illumination. Finally, a CCD

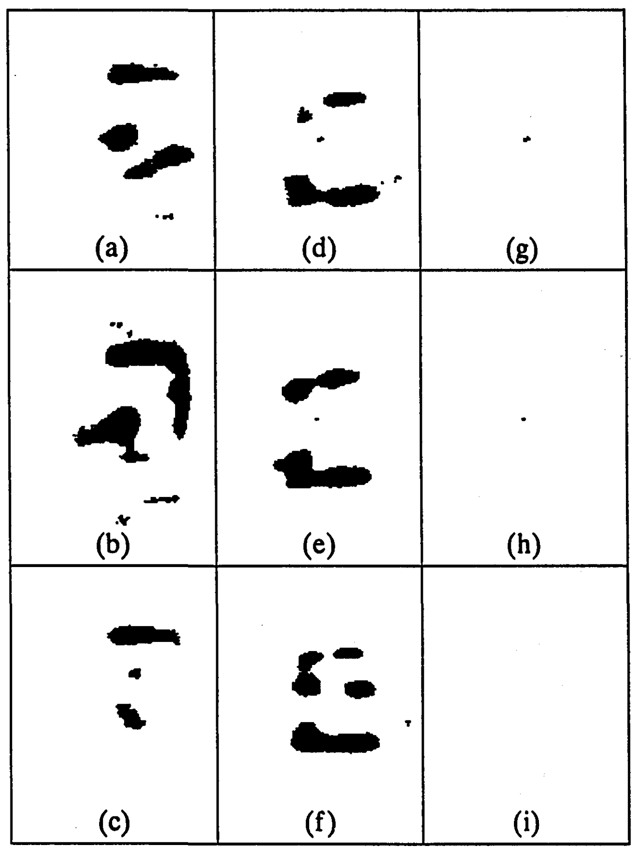

Fig. 10. (a), (b), (c) Intensity of the correlation between the positive filter and letters S, C, and E, respectively, binarized with $\theta = 0.4$. (d), (e), (f) Same as (a), (b), and (c) with the negative filter. (g) Result of pixel-by-pixel multiplication of images (a) and (d). (h) Result of pixel-by-pixel multiplication of images (b) and (e). (i) Result of pixel-by-pixel multiplication of images (c) and (f).

camera and a frame grabber were used to capture the resulting correlation distributions.

The images that were used in the design of the filters are shown in Fig. 8. The values imposed for the correlation at the origin were 1, 1, 0, for S, C, and E, respectively. The correlations between the three letters and a composite filter are shown in Figs. 9(a)–9(c), in which large sidelobes can be observed.

The correlations obtained with the positive and the negative filters are depicted in Figs. 9(d)–9(i). As can be seen, the effects of the two filters are opposite; the sidelobes reduced by the first one are enhanced by the other and vice versa, as expected. By binarizing the results obtained with the positive and the negative filters with a threshold of 0.40, we obtain the images in Figs. 10(a)–10(f). Finally, in Figs. 10(g)–10(i) the results of pixel-by-pixel multiplication of binarized images are represented.

The results observed in the optical implementation were satisfactory and showed good agreement with previous computer simulations. Therefore the method seems to be suitable for application in a practical situation.

6. Final Remarks and Conclusions

The existence of lateral peaks is one of the most important problems in optical pattern recognition by means of correlation. In this study we present a method that eliminates every sidelobe within a given range, provided certain conditions are fulfilled. The method has the following properties:

- It can be applied to a wide variety of filters.
- The method ensures the elimination of the sidelobes if certain conditions are satisfied.
- Sidelobes higher than the central correlation can be suppressed.
- The method is more powerful when higher sidelobes are expected.

The procedure was tested with simulation and optical implementation and gave satisfactory results in both cases. However, some limitations were found. In Section 4 the noise sensitivity of the filter was pointed out. A possible way to reduce this drawback is by means of a compromise filter.[14] The underlying idea of this type of design is to introduce a new term in the error function to be minimized, which represents the output variance of the noise. A parameter is used to balance the importance of the two conflicting goals. This solution represents a trade-off between the noise resistance and the minimization with respect to the desired shape, which thus affects the height of the sidelobes that can be eliminated by the procedure. A similar idea, which also involves a compromise between the two magnitudes to be minimized, is applied in minimum noise and correlation energy filters.[15] A further study of the optimum selection for this trade-off is needed.

On the other hand, the method we present is valid only for filters that produce correlations distributions with real values. The procedure may be generalized to the case in which complex distributions are obtained, by use of a battery of filters, each producing a constant complex-valued plane whose phases are distributed over the entire unit circle.

The greater the number of filters used, the greater the height of the lateral peaks that can be eliminated, because the directions of the opposing vectors approach 180° with respect to the direction of the sidelobe. However, an increasing number of filters implies a more complex procedure. The study of the minimum number of filters required for elimination of sidelobes with a given height is currently in progress.

This work has been supported in part by the Spanish Comisión Interministerial de Ciencia y Tecnologia under project ROB91-0554.

References

1. A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory **10,** 139–145 (1964).
2. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23,** 812–816 (1984).
3. Y. N. Hsu and H. H. Arsenault, "Optical character recognition using circular harmonic expansion," Appl. Opt. **21,** 4016–4019 (1982).
4. C. F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition," Appl. Opt. **19,** 1758–1761 (1980).
5. A. Metioui and L. Leclerc, "Sidelobe reduction methods in optical pattern recognition," J. Opt. (Paris) **21**(4), 161–170 (1990).
6. A. Metioui, H. H. Arsenault, and L. Leclerc, "Methods for reducing sidelobes associated with composite filters," Opt. Commun. **71,** 332–336 (1989).
7. Z. Bahri and B. V. K. Vijaya Kumar, "Generalized synthetic discriminant functions," J. Opt. Soc. Am. A **5,** 562–571 (1988).
8. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," Appl. Opt. **26,** 3633–3640 (1987).
9. M. Fleisher, U. Mahlab, and J. Shamir, "Entropy optimized filter for pattern recognition," Appl. Opt. **29,** 2091–2098 (1990).
10. G. Zalman and J. Shamir, "Maximum discrimination filter," J. Opt. Soc. Am. A **8,** 814–821 (1991).
11. D. Casasent, R. Schaefer, and R. Sturgill, "Optical hit–miss morphological transform," Appl. Opt. **31,** 6255–6263 (1992).
12. D. G. Crowe, J. Shamir, and T. W. Ryan, "Sidelobe reduction in optical signal processing," App. Opt. **32,** 179–183 (1993).
13. B. V. K. Vijaya Kumar, A. Mahalanobis, S. Song, S. R. F. Sims, and J. F. Epperson, "Minimum squared error synthetic discriminant functions," Opt. Eng. **31,** 915–922 (1992).
14. P. Refregiér, "Filter design for optical pattern recognition: multicriteria optimization approach," Opt. Lett. **15,** 854–856 (1990).
15. G. Ravichandran and D. Casasent, "Minimum noise and correlation energy optical correlation filter," Appl. Opt. **31,** 1823–1833 (1992).
16. A. VanderLugt, *Optical Signal Processing* (Wiley, New York, 1992), Chap. 3, pp. 71–140.
17. C. B. Burckhardt, "A simplification of Lee's method of generating holograms by computer," Appl. Opt. **9,** 1949 (1970).