Maximum likelihood for target location in the presence of substitutive noise

Pascuala Garcia-Martinez, Philippe Réfrégier, Henri H. Arsenault, and Carlos Ferreira

We consider the optimal likelihood algorithm for the estimation of a target location when the images are corrupted by substitutive noise. We show the relationship between the optimal algorithm and the sliced orthogonal nonlinear generalized (SONG) correlation. The SONG correlation is based on the application of a linear correlation to corresponding binary slices of both the input scene and the reference object with appropriate weight factors. For a particular case, we show that the optimal strategy is a function of only the number of pixels for which the gray values in the noisy image match the ones of the reference image when the substitutive noise is uniformly distributed. This is exactly what a particular definition of the SONG correlation does. © 2001 Optical Society of America

OCIS codes: 100.5010, 100.4550, 070.4340.

1. Introduction

Nonlinear digital filtering has had an impressive growth in the past three decades.¹ However, one of the main problems in digital image filtering applications is the bulk of the image data and the large number of operations per pixel. Usually the operations involve several additions, multiplications, comparisons, and nonlinear function evaluations. So the computational complexity required is relatively high. For many applications real-time image processing is required. Optics, specifically, parallel processing, can provide an alternative for such cases in that operations carried out at the speed of light and the inherent parallelism of optical processors are advantages offered by the use of optics. Convolution and correlation are well-known examples of operations that may be performed with device technologies such as high-speed spatial light modulators and detectors. Nonlinear filtering involve the combination of nonlinear functions with common linear correlation or convolution.

0003-6935/01/233855-06\$15.00/0

© 2001 Optical Society of America

An image can be decomposed by means of elementary binary decompositions, and because those slices are binary, the complexity of the process is reduced significantly. Threshold decomposition is a widely used binarization method that has been applied to nonlinear image processing.² Recently Garcia-Martinez and Arsenault³ have proposed an orthogonal binary-sliced image decomposition. The authors showed how the slices obtained with threshold decomposition can be defined in terms of the sliced orthogonal nonlinear generalized (SONG) elementary binary functions.³ The motivation of finding different binary decompositions is to define new nonlinear correlations to achieve higher discrimination capabilities for pattern recognition compared with conventional linear filtering. Two nonlinear correlations have been defined by means of linear correlations between binary decompositions of the input scene and the target. The threshold-decomposition point of view leads to the morphological correlation,^{4,5} whereas the SONG decomposition leads to the SONG correlation.^{3,6} Both nonlinear correlations have been implemented optically with a joint transform correlator (JTC): Each pair of elementary binary joint input slices (one slice from the target and one from the input scene) are placed next to each other in the input plane. For each pair, the joint power spectrum is performed. The sum of the joint power spectra of all the slices is stored and then fed back at the input plane for a second Fourier transform to produce the morphological correlation or the SONG correlation. Other papers describe a way to binarize the joint power spectrum in a JTC to achieve pattern-

P. Garcia-Martinez (e-mail: pascuala.garcia@uv.es) and C. Ferreira are with the Departament d' Òptica, Universitat de València, 46100 Burjassot, Valencia, Spain. P. Réfrégier is with the Physics and Image Processing Group, Fresnel Institute, Domaine Universitatire de Saint-Jérôme, 13397 Marseille Cedex 20, France. H. Arsenault is with the Centre d'optique, photonique et laser, Universite Laval, Ste-Foy, Quebec, G1K 7P4, Canada.

Received 14 August 2000; revised manuscript received 3 April 2001.

recognition robustness to noise and to distortion.^{7,8} Those nonlinear correlations are extremely selective for pattern recognition compared with common linear filtering methods.

Furthermore, the SONG correlation is robust when the image is corrupted by substitutive noise. This noise is often considered in communication theory.¹ This noise is viewed as an impulsive noise or outliers from a statistical point of view. The noise will destroy part of an image, whereas other parts remain unaffected. So an optimum detection solution may be to find a technique for which only this unaffected part will be considered. This is indeed what a specific definition of the SONG correlation³ states, because the SONG correlation counts the number of pixels of an image that remain unaffected by the noise. Although we have shown recently that the SONG correlation is robust to substitutive noise,^{3,6} we had not demonstrated that it is optimal. This will be accomplished in this paper by use of the maximum-likelihood criteria.

For optical pattern recognition, one can determine the filter by optimizing heuristic criteria or by using a decision theory approach. The decision theory approach consists of the maximization of a likelihood function or the minimization of a probability of error. Classical linear filters are optimum in the presence of some particular additive noise sources.^{9,10} However, such filters are suboptimum when different noise models are considered. Nonlinear global filtering techniques are a natural extension of basic linear filters when different noise sources are considered.¹¹ Moreover, nonlinear filters based on decision theory can be applied to the well-known nonlinear JTCs.^{7,8,12}

In this maximum-likelihood approach, we shall determine the statistical model for substitutive noise and we shall obtain the maximum-likelihood expression for target location. We show that this function is related to the nonlinear SONG correlation. As an example, we show that the optimum strategy is a function only of the number of pixels for which the gray values in the noisy image match the gray values in the reference image when the substitutive noise is random and uniformly distributed. This is exactly what a particular definition of the SONG correlation states. Experimental results are given to illustrate this result.

2. SONG Decomposition and SONG Correlation

A two-dimensional image f(x, y) with discrete gray levels can be decomposed into a sum of disjoint elementary images $e_m[f(x, y)]$ satisfying the orthogonality property

$$e_m[f(x, y)]e_n[f(x, y)] = 0 \quad \text{if } m \neq n, \\ e_m[f(x, y)]e_n[f(x, y)] = 1 \quad \text{if } m = n. \quad (1)$$

Each subimage $\{e_m[f(x, y)]\}$ represents a gray-level slice of the object.⁶ The SONG decomposition of f(x, y) is³

$$f(x, y) = \sum_{i=0}^{Q-1} F_i e_i [f(x, y)], \qquad (2)$$

where the coefficients F_i are weights and Q is the total number of gray levels in the image.

The elementary binary images have the property

$$e_i[f(x, y)] = \begin{cases} 1 & f(x, y) = i \\ 0 & \text{otherwise} \end{cases}$$
(3)

Note that each object point has only one gray level, so each unshifted *i* slice is disjoint and therefore orthogonal to all the others. For the standard gray-scale image representation, $F_i = i$.

The standard correlation between two objects g(x, y) and f(x, y) can be written as

$$g(x, y) \otimes f(x, y) = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} G_i F_j e_i [g(x, y)]$$
$$\otimes e_j [f(x, y)], \qquad (4)$$

where \otimes denotes the linear correlation and the coefficients are equal to the gray levels, i. e., $G_i = i$ and $F_j = j$. Now the problem with this correlation is that it puts higher weights on brighter parts of the targets, but there is usually no reason that brighter parts of a target should be more important than the others; that is the reason that many pattern-recognition techniques binarize both the reference and the target, thus giving equal weights to all the gray levels by settings all the weights equal to unity or some other constant value. This is convenient when the target may be segmented from the scene of which it is a part, but in highly cluttered or noisy scenes it is often not feasible to segment potential targets from the scene.

The coefficients G_i and F_j may be arranged into a matrix with rows and columns (i, j). We now generalized this matrix by replacing the product coefficients G_i and F_j by generalized weights W_{ij} . So the correlation expression Eq. (4) becomes

$$\Omega_{gf}(x, y) = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} W_{ij} e_i [g(x, y)] \otimes e_j [f(x, y)].$$
(5)

This is what we have called SONG correlation.³ Setting different values on the terms of the matrix Wallows the matrix to represent various known correlation types such as standard matched filtering, binary filtering, morphological correlation,⁵ but it also allows us to define new kinds of correlation. The SONG correlation that we have proposed is obtained by setting

$$W_{ij} = 0 \quad \text{for } i \neq j,$$

$$W_{ij} = 1 \quad \text{for } i = j.$$
(6)

In this case, the double sum reduces to the single sum and the correlation becomes

$$\Omega_{gf}^{P}(x, y) = \sum_{i=0}^{Q-1} e_i[g(x, y)] \otimes e_i[f(x, y)].$$
(7)

In this expression, only the gray levels in the two images having the same values are correlated together after having their values set equal to unity, so the correlation is a sum of correlations between binary images. This is the SONG correlation as used in our previous papers.^{3,6} In the following, we shall consider for simplicity objects located such that their correlation peaks appear at the origin (0, 0); because the correlation operation is shift invariant, it is trivial to generalize to the case of targets located at point (*x*, *y*) and to the presence of multiple targets because the correlation is also additive.

Images are rarely digitized to fewer than 256 gray levels, but targets of interest in the scene usually cover a smaller range of levels, say, 64 for the purpose of discussion. It is clear that if 64 gray levels are maintained, the expression shown in Eq. (7) will require the summing of 64 correlations. In fact, such a large number of correlations is not required: It is a simple matter to add neighboring gray levels together to reduce the number of gray levels to a more manageable number, say, 4 or 8, which are the numbers used in our previous experimental results.^{3,6}

One of the additional advantages of defining the SONG correlation as a sum of linear correlations is the possibility of implementing this operation optically⁶ with a JTC. The JTC carries out a Fourier transformation, which is a linear operation, so the sum of Eq. (7) is performed on the joint power spectrum, which is the intensity output detected in an intermediate step of the JTC operation. This idea has been used for the optical implementation of non-linear morphological correlation.⁵

The SONG correlation is extremely discriminating because we are counting the number of points in the reference object and in the image that have the same grav levels at the same locations. However, owing to the selectivity, if the aspect of the object changes a little bit because of a slight rotation or a change in scale, the SONG correlation can be more sensitive. To alleviate this we have satisfactorily applied the circular harmonic decomposition to the SONG correlation to achieve rotation-invariant pattern recognition.13 A change of illumination or gray-level changes will affect the SONG correlation. A complete solution to this problem is beyond the scope of this paper, but we are aware of the importance of the question which is the reason that we are carefully studying the gray-level quantization and the optimum weights to control the tolerance to illumination.

As was shown in Ref. 3, the correlation Ω^P is extremely good when the images are corrupted by substitutive noise because the pixels that remain unaffected are the only information that is considered in the correlation process. Although the results are there,³ we have not given the exact mathematical theory for that location estimation. In Section 3 we show the relation between the optimal approach for location estimation with substitutive noise and the SONG correlation.

3. Noise Robustness of the SONG Correlation Against Substitutive Noise

We now consider the performance of this SONG correlation for images that are degraded by substitutive noise. This noise is known as impulsive noise or outliers from a statistical point of view. In this section we analyze the optimum likelihood algorithm for object location estimation, so the statistics of the noise are required.

A. Noise Model and Likelihood Expression

Let f_i be the reference image. We assume that this reference has been corrupted by substitutive noise. This noise is modeled in the following way.

1. For each pixel the value of the gray level is modified with probability p and not modified with probability q (so p + q = 1). One may introduce a Bernoulli random variable b_i whose probability law is $P_B(0) = q$ and $P_B(1) = p$.

2. If the gray-level value is modified, a new random value x_i is affected to the value of this pixel. The probability density function of this random variable x_i will be denoted $P_x(x)$.

The statistical model for the corrupted image g_i is

$$g_i = (1 - b_i)f_i + b_i x_i.$$
 (8)

If there is no *a priori* information on the reference object f_i one can consider that $P(f_i) = \text{constant.}$ Moreover, using the Bayes law one can show with the *a posteriori* probability $P[f_i|g_i]$ that the reference object present in the scene is equivalent to maximizing the *a priori* probability, $P[g_i|f_i]$, which corresponds to the likelihood of the observed image g_i :

$$L = P[g_i|f_i]. \tag{9}$$

Moreover, because the random value for each pixel is assumed to be chosen independently of the others, the logarithm of the likelihood (log likelihood) is

$$l = \sum_{i=1}^{N} \ln(P[g_i|f_i]),$$
(10)

where $P[g_i|f_i]$ is the probability that at pixel *i* the gray-level value is g_i for the noisy image; we know that it is equal to f_i for the reference image. Using Bayes relations we show that the log-likelihood expression is (see Appendix A)

$$l = \sum_{i=1}^{N} \delta[g_i - f_i]G(f_i) + B_g,$$
(11)

where the functions $G(f_i)$ and B_g , are defined in Appendix A. We have used the definition of the Kronecker function, $\delta(y)$. The expression shown in Eq. (11) has two distinct responses that are due to the

Kronecker function: one when the corresponding pixels in the observed image and in the reference image have the same gray-level values at the same positions and another when they are different. We will apply in Subsection 3.B this expression to the location estimation problem.

B. Location Estimation Problem

1. General Case

Now assume that the reference image has been translated by an unknown quantity k and that we want to estimate this value from the noisy image g_i . The image model is now

$$g_i = (1 - b_i) f_{i-k} + b_i x_i.$$
(12)

The log likelihood is now dependent on the hypothesis for the unknown location k:

$$l(k) = \sum_{i=1}^{N} \ln(P[g_i|f_{i-k}]).$$
(13)

The maximum likelihood k_{ML} of k is obtained by choosing the value of k that maximizes l(k).

It is straightforward to generalize the previous results of Eq. (11):

$$l(k) = \sum_{i=1}^{N} \delta[g_i - f_{i-k}]G(f_{i-k}) + B_g.$$
(14)

Note that B_g is independent of k and that $k_{\rm ML}$ is obtained by maximizing

$$l'(k) = \sum_{i=1}^{N} \delta[g_i - f_{i-k}] G(f_{i-k}).$$
(15)

Comparing the maximum-likelihood expression, Eq. (15), with the general SONG correlation in Eq. (5), we see that both expressions are similar if the correlation weight factors are identified with the function $G(f_{i-k})$ and if we perform the correlations using only the diagonal terms of the SONG correlation matrix. So, the optimum solution for the location estimation in a maximum-likelihood sense in the presence of substitutive noise is a nonlinear correlation in which only the corresponding gray levels for the two images (g and f), having the same gray levels at the same positions and after having their values weighted by the function $G(f_{i-k})$, are correlated together. It is worth noting that $G(f_{i-k})$ depends on p (or q) in a nontrivial way. So if p is unknown, it is difficult to estimate it. We will study the particular case of when the probability density function is a uniformly distributed random function.

2. Case of a Uniformly Distributed Random Variable

To go beyond the previous result, we need to consider a particular probability density function for the random variable x_i . Let us consider the case where x_i is



Fig. 1. Input scene with the reference object, located in the lower part of the figure, and another object to be rejected.

a uniformly distributed random variable. For that case $P_x(x) = a$ and so

$$G(f_i) = \ln(q + pa) - \ln(pa), \tag{16}$$

and maximizing l'(k) is equivalent to maximizing

$$\tilde{l}(k) = \sum_{i=1}^{N} \delta[g_i - f_{i-k}], \qquad (17)$$

which is equivalent to maximizing the number of pixels whose value perfectly matches the pixel of the reference. This operation is done by the definition of the SONG correlation given in Eq. (7).

4. Analysis of the Experimental Results

We compare the noise robustness of the particular definition of Eq. (7) for the SONG correlation in the presence of substitutive noise, with linear correlation, nonlinear morphological correlation⁵ and the phase-only filter. The input scene is shown in Fig. 1. It consists of two objects, the reference object being the one in the lower part of the image. The images have eight gray levels.

We corrupt the noise-free image shown in Fig. 1 with substitutive noise in a uniformly random way. This consists of randomly picking one pixel from the image with a probability of changing that pixel value or to leave the original image as it was. The noisy image is shown in Fig. 2.

To estimate the amount of noise in the image, we define a parameter for the signal-to-noise ratio as

$$SNR = N_i / N_c - 1, \qquad (18)$$

where N_i is the number of pixels of the original image and N_c is the number of pixels corrupted. If the image is a noise-free image, then $N_c = 0$, so SNR $= \infty$. In contrast, if the image is totally corrupted, then $N_i \cong N_c$, so SNR $\cong 0$. Then low values of signal to noise will mean that the image is highly degraded.



Fig. 2. Input scene highly degraded with uniform distribution random noise.

We define the discrimination capability (DC) as

$$DC = 1 - \frac{CrossCorr}{AutoCorr}.$$
 (19)

A high value of DC means that the value of the cross correlation is low compared with the autocorrelation, which means that good discrimination and good noise robustness are achieved. However, a low value of the ratio means that the energy of the cross correlation has almost the same value as that of the autocorrelation.

Table 1 shows the DC of the Ω^P correlation and of other common detection methods. What is represented in the table is how the noise level affects the cross correlation; so, we measure the discrimination capability, which in this case is connected with the noise robustness.

From Table 1, when images are highly degraded, only the Ω^P correlation is able to detect the reference object (DC = 0.94); note that for the phase-only filter, the matched filters and the nonlinear morphological correlation, the value is DC = 0, which corresponds to a false alarm detection or a nondetection of the target. In contrast, the Ω^P correlation is stable for discrimination over the whole range of noise.

Those results can be understood from the definition of the Ω^P correlation. It implies the counting of all the pixels that have the same gray levels at the same positions for both the input image and the reference object. Then if the image is highly degraded to the extent that the gray-scale information of only a few

Table 1. Discrimination Capability (DC) of Several Pattern-Recognition Operations and the New SONG Correlation When Different Noise Degrees Are Considered

Signal-to-Noise Ratio	Ω^P	Linear Correlation	Morphological Correlation	Phase-Only Filter
5.70	0.97	0.10	0.40	0.95
1.70	0.97	0.13	0.36	0.93
0.57	0.96	0.01	0.29	0.70
0.30	0.95	0	0.18	0.45
0.13	0.94	0	0	0



Fig. 3. (a) Three-dimensional plot of the SONG (Ω^{P}) correlation. (b) Three-dimensional plot of the common phase-only filter. This method is not able to detect the reference object.

pixels remains, the Ω^P correlation will still give a signal. This is not the case for other linear filtering techniques. Those filters have been shown to be highly discriminant but not particularly robust against substitutive noise. Similar considerations apply to matched filters and for the nonlinear morphological correlation.⁵

Figure 2 is the input scene of Fig. 1 degraded with a SNR = 0.13. The visual pattern information is wiped out by the noise, but if some pixels of the image remain unaffected, we will still get a high signal when the Ω^P correlation is applied. Figure 3 shows the correlation outputs for noisy image detection. Figure 3(a) shows the three-dimensional plots for the SONG output correlation plane. Note that we detect the image with the same DC as for the noise-free image. However, in the correlation plot of Fig. 3(b) for the phase-only filter, the reference object cannot be extracted from the noise. For the classical matched filter, the morphological correlation yields similar, poor results.

5. Final Remarks

In conclusion, for location estimation with substitutive noise, the optimum strategy is in general not only a function of the number of pixels for which the gray values in the noisy image match the ones of the reference image but also of other complex functions. However, those functions may be related with the correlation weights in the SONG correlation definition. When the substitutive noise is uniformly distributed. the optimum strategy for location estimation is dependent only on the number of pixels for which the gray values in the noisy image match the ones in the reference image. This is also what a particular definition of the SONG correlation states. Obviously, counting the number of image pixels that remain unaffected by substitutive noise is the same as counting the gray values in the observed image that match the corresponding pixels of the reference object. Different probability density functions for the noise are under consideration to establish different weights for the SONG correlation.

Appendix A

Consider the expression for the log-likelihood expression [see Eq. (9)] that uses the Bayes relations:

$$P[g_i|f_i] = P[g_i, b_i = 0|f_i] + P[g_i, b_i = 1|f_i].$$
 (A1)

Furthermore,

$$P[g_i, b_i | f_i] = P[g_i | f_i, b_i] P_B[b_i],$$
(A2)

and because $P_B(0) = q$ and $P_B(1) = p$, one can write

$$P[g_i|f_i] = qP[g_i|f_i, b_i = 0] + pP[g_i|f_i, b_i = 1].$$
(A3)

Because $b_i = 1$ means that the gray-level value of f_i has been changed into a random variable x_i , then

$$P[g_i|f_i, b_i = 1] = P_X(g_i).$$
 (A4)

However, $b_i = 0$ means that the gray-level value of f_i has not been changed. In that case we must have $g_i = f_i$ and

$$P[g_i|f_i, b_i = 0] = \delta(g_i - f_i),$$
 (A5)

where $\delta(y)$ is the Kronecker symbol:

$$\delta(y) = \begin{cases} 1 & \text{if } y = 0\\ 0 & \text{otherwise} \end{cases}.$$
 (A6)

One obtains

$$P[g_i|f_i] = q\delta[g_i - f_i] + pP_X[g_i]$$
(A7)

and

$$l = \sum_{i=1}^{N} \ln(q\delta[g_i - f_i] + pP_X[g_i]).$$
(A8)

We note that

$$\ln(q\delta[g_{i} - f_{i}] + pP_{X}[g_{i}]) = \delta[g_{i} - f_{i}]\ln(q + pP_{X}[g_{i}]) + (1 - \delta[g_{i} - f_{i}])\ln(pP_{X}[g_{i}]), \quad (A9)$$

which can also be written as

$$\ln(q\delta[g_{i} - f_{i}] + pP_{X}[g_{i}]) = \delta[g_{i} - f_{i}][\ln(q + pP_{X}[f_{i}]) - \ln(pP_{X}[f_{i}])] + \ln(pP_{X}[g_{i}]).$$
(A10)

Let us define

$$G(f_i) = \ln(q + pP_X[f_i]) - \ln(pP_X[f_i]) \quad (A11)$$

and

$$B_{g} = \sum_{i=1}^{N} \ln(pP_{X}[g_{i}]), \qquad (A12)$$

 $\mathbf{S0}$

$$l = \sum_{i=1}^{N} \delta[g_i - f_i] G(f_i) + B_g.$$
 (A13)

This research has been supported by grants from the Natural Sciences Engineering Research Council of Canada and from the Funds pour la Formation de Chercheure et l'Aide a la Recherche (FCAR) of Quebec.

References

- I. Pitas and A. N. Venetsanopoulos, Nonlinear Digital Filters: Principles and Applications (Kluwer Academic, Boston, 1990).
- J. P. Fitch, E. J. Coyle, and N. C. Gallagher, Jr., "Median filtering by threshold decomposition," IEEE Trans. Acoust. Speech Signal Process. ASSP-32, 1183–1188 (1984).
- 3. P. Garcia-Martinez and H. H. Arsenault, "A correlation matrix representation using sliced orthogonal nonlinear generalized decomposition," Opt. Commun. **174**, 503–515 (2000).
- P. Maragos, "Morphological approaches to image matching and object detection," in *ICASSP-89: 1989 International Conference on Acoustic, Speech and Signal Processing* (Institute of Electrical and Electronics Engineers, New York, 1989), Vol 3, pp. 1568–1571.
- P. Garcia-Martinez, D. Mas, J. Garcia, and C. Ferreira, "Nonlinear morphological correlation: optoelectronic implementation," Appl. Opt. 37, 2112–2118 (1998).
- P. Garcia-Martinez, H. H. Arsenault, and S. Roy, "Optical implementation of the sliced orthogonal nonlinear generalized correlation for images degrades by nonoverlapping background noise," Opt. Commun. 173, 185–193 (2000).
- B. Javidi, "Nonlinear joint power spectrum based optical correlation," Appl. Opt. 28, 2358–2367 (1989).
- Ph. Réfrégier, V. Laude, and B. Javidi, "Nonlinear joint transform correlation: an optimum solution for adaptive image discrimination and input noise robustness," Opt. Lett. 19, 405– 407 (1994).
- G. L. Turin, "An introduction to matched filters," IRE Trans. Inf. Theor. IT-6, 311–319 (1960).
- A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory IT-10, 139–145 (1964).
- Ph. Réfrégier, V. Laude, and B. Javidi, "Basic properties of nonlinear global filtering techniques and optimal discriminant solutions," Appl. Opt. 34, 3915–3923 (1995).
- Ph. Réfrégier and F. Goudail, "Decision theory approach to nonlinear joint-tranform correlation," J. Opt. Soc. Am. A 15, 61-67 (1998).
- P. Garcia-Martinez, H. H. Arsenault, and C. Ferreira, "Improved rotation invariant pattern recognition using circular harmonics of binary gray-level slices," Opt. Commun. 185, 41-48 (2000).