UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/1058-6

WINTER COLLEGE ON OPTICS

9 - 27 February 1998

Pattern Recognition Filters

P. Réfregier

Ecole National Superieure de Physique de Marseille, France

Filter design for optical pattern recognition: heuristic and statistical approaches.

Ph. Réfrégier,
Signal and Image Laboratory,ENSPM,
Domaine universitaire de Saint-Jérôme
13 397 Marseille cedex 20 France

Abstract

Filtering methods for pattern recognition are reviewed. In particular, the heuristic and the statistical approaches are analyzed for different situations. The stability of the different techniques are discussed and methods of regularization are described for both the heuristic and the statistical approaches.

Contents

1 Introduction

1.1 The need of filter design

An important problem in optical correlation for pattern recognition, object location or detection, is the efficiency of the algorithms involved in these signal processing tasks. Indeed, without a particular attention to this problem, only trivial tasks can be resolved with optical architectures which implement the classical linear matched filter. However, for that trivial tasks, there exists classical numerical solutions which require simple hardware and which can be more efficient that the matched filter method.

In fact, the spatial matched filter which is optimal in additive gaussian noise, does not lead to acceptable performance for difficult pattern recognition or target location. For example, it can lead to poorly discriminant correlation and its performance deteriorate rapidly when the input image is subject to distortions (rotations, scale, view angle, structured background ...) or when the input noise is not additive with a known spectral density.

These last twenty years, many different solutions have been proposed for filter design [1–13]. For example, when invariant or tolerant pattern recognition is required, specific methods for improving recognition capabilities have been introduced. For in-plane distortions, several methods have been proposed [2, 14]. When distortions cannot be described analytically, a supervised learning method can be used for which the filter is determined using training images that are sufficiently descriptive of the expected distortions. A well known example of these filters are the synthetic discriminant function (SDF) filters and their different variations [15–17].

However, the different heuristic criteria generally introduced in filter synthesis are not sufficient to describe all needed characteristics of the filter. In general, an optimal filter is obtained by minimizing some criterion, or a combination of criteria with the optimal trade-off (OT) approach [17], under constraints for optical implementation and/or for obtaining specific values on a training set. The criteria are usually dependent on some expected distortions. For example, the optimization of noise robustness or of signal to noise ratio (SNR) is dependent on a noise model. A filter optimal for this noise model can be greatly sub-optimal for other type of distortions or noise models [18]. Since this problem results from the unstability of the filter [19], we will analyze precisely in this paper different solutions to regularize filtering methods.

The detection and location of a target in a scene is a classical problem, pervasive to many image processing applications. However, the improved linear filtering techniques, as well as the matched filter [20], have been shown to perform poorly on many real-world images [21, 22]. This is because such images often do not belong to the class for which linear filtering is optimal. Indeed, in general, in real images, the main source of noise is not the additive detector noise, but the whole background of the scene (clutter), which is nonoverlapping [21]. Secondly, the gray levels of the reference object can be unknown *a priori*. Thirdly, the power spectral density (psd) of the noise is also often unknown *a priori*. In each of these three cases, an important assumption which is necessary to demonstrate the optimality of linear filters is not fulfilled. As a practical consequence, it has been frequently observed that linear filters yield poor performance in real applications of optical correlation.

On the other hand, the statistical approach is a very powerful method in order to describe specific image models and to design the corresponding optimal filtering techniques.

In this paper we propose to review two main approaches used in filter design theory. The first approach is based on filtering techniques with an imposed structure (mainly linear) and the filter is designed by optimizing some heuristic criteria. With the second approach, the images are described by a statistical model and the optimal filtering technique is the optimal statistical method for that model. In that case the structure of the filtering techniques is not imposed.

1.2 The different approaches to filter design

Let us first discuss more precisely the different approaches that can be used to design optical pattern recognition filters.

With **the empirical approach**, the mathematical expression of the location processor is chosen *a priori* without optimality consideration. Of course, this approach is far from being satisfactory. However, it can provide powerful techniques capable of solving problems unsolved with the known optimal methods. This has been the case for example with Nonlinear Joint-Transform Correlators (NLJTCs) proposed in [23, 24]. These processors were not first defined by the optimization of some criteria, but they have been shown to be attractive in many difficult instances because of their high discriminating performance [24–26]. Their basic properties in terms of signal processing and pattern recognition are still a subject of intensive investigations.

The empirical approach is far from being satisfactory since it does not elucidate in which context the considered method is superior to another one. Systematic simulations and experiments are necessary to answer this question. However, in the field of optical pattern recognition, the proof from experiments is very hard to obtain, due to the very different input scenes that can occur and to the large number of available techniques which all have their own parameters. This approach is thus not of particular interest for efficient filter design.

With the second approach (**the heuristic approach**), the filtering method is obtained by optimizing heuristic criteria using an *a priori* chosen class of processors (in general, linear filters). This is a very powerful and versatile technique which has been the main subject of investigation for optical pattern recognition filters these last twenty years [3, 27, 20, 16, 17]. Furthermore, with this approach, it is quite easy to determine filters that are invariant to some expected distortions of the object in the input scene compared to the reference object [14–17, 28]. It is also possible to determine filters which perform optimal tradeoff between several antagonist criteria [17, 29].

Recently, the optimal tradeoff approach has been extended to nonlinear joint transform correlation (NLJTC) [30,31], and it has been shown that some NLJTC techniques are optimal tradeoff solutions when some specific criteria are considered. Nonlinear global filtering (NGF) techniques, which can be implemented with NLJTCs, can be considered as the natural extension of the basic linear filtering techniques. However, unlike the linear matched filter, which was designed on the basis of the statistical decision theory before its implementation in optical correlators, investigations on nonlinear filters still suffer, in general, from insufficient

theoretical analyses.

In particular, the main limitation of the heuristic approach is that the criteria are in general chosen *a priori* by the filter designer. Of course, there have been many discussions in the scientific community about the different criteria [29], but none of these criteria has been shown to be adapted to most of the different situations which can occur in pattern recognition. In particular, as far as noise robustness is concerned, all the classical criteria are adapted to an additive noise with known psd and to a target with known gray levels.

The analyses of some properties of the heuristic approach for filter design will be reviewed in the first part of this article.

The third approach we will discuss here is based on **statistical decision theory**. In that case, the probability of error is minimized while the specific sources of randomness are clearly specified. It is thus necessary to know the probability density function (pdf) of the noise, although it is not necessary to specify all the parameters of this pdf. For example, as already mentioned above, it is well known that the classical matched filtering optimizes the probability of correct location of a target with known gray levels if the input noise is additive, Gaussian, stationary and ergodic with a known psd [32]. For this type of noise, any other filter, either linear or nonlinear, will result in a lower probability of correct location than the matched filter. However, if one or more of the above assumptions does not hold, there may exist a better solution than the matched filter. We will discuss some examples of this situation in section 2.

1.3 Notations

The notations used in this paper are:

x : for a scalar,
$\mathbf{x}$: for a vector,
X : for an operator or a matrix,
x^* : complex conjugate value of x,
$\mathbf{x}^\dagger$: complex conjugate transpose vector of $\mathbf{x}$,
$|x|$: modulus of x,
$\mathbf{x}^\dagger.\mathbf{y}$: scalar product of $\mathbf{x}$ and $\mathbf{y}$,
$\mathbf{x} \star \mathbf{y}$: correlation function of $\mathbf{x}$ and $\mathbf{y}$,
θ : an unknown parameter which can be a real or a complex vector,

$\underset{\delta}{max}\ F(\delta)$: maximum value of $F(\delta)$ for the different possible values of δ,

$\underset{\delta}{argmax}\ F(\delta)$: value of δ which maximizes $F(\delta)$.

In cartesian coordinates, the value of the image $\mathbf{x}$ at pixel i (which is indeed two dimensional, say $i = (i_u, i_v)$) will be denoted x_i, and $\hat{x}_k$ will denote a value in the Fourier domain (where k is also two dimensional, say $k = (k_u, k_v)$). The total number of pixels will be denoted N. C denotes the set of complex numbers.

So one can also written:

$$\mathbf{x}^\dagger.\mathbf{y} := \sum_{i=1}^{N} x_i^* y_i,$$
$$||\mathbf{x}||^2 := \sum_{i=1}^{N} |x_i|^2,$$
$$\mathbf{x}^\dagger A\mathbf{y} := \sum_{i=1}^{N} x_i^* A_{i,j} y_i,$$
$$\mathbf{x} \star \mathbf{y}(j) := \sum_{i=1}^{N} x_i^* y_{i+j},$$

If x is a random variable with porbability density function $P_X(x)$, we also define the mean value $< x >= \int x P_X(x) dx$. The mean value of a function $g(x)$ of a random variable is thus: $< g(x) >= \int g(x) P_X(x) dx$.

A lexicon can be found at the end of this paper.

1.4 Structure of the paper

This paper is organized in two main parts. In the first one we review some well known heuristic approaches with a particular attention to the stability of the methods. In the second part, we present some recent results obtained with the statistical theory for nonoverlapping noise, fluctuating gray level targets or for additive noise with unknown spectral density. We also discuss how to increase the stability of the statistical methods.

Part I
Heuristic theory and stability

2 Heuristic criteria and optimal filters

In this section, we present different criteria which classically have been considered in filter design (a more complete review can be found in [29]). Optimal filters for these criteria are discussed in the context of the heuristic approach for which the structure of the filtering technique is imposed. The structure is in general the linear filtering but we also describe the design of an optimal nonlinear processor by optimizing the discrimination capabilities. Finally, at the end of this section, SDF filters are introduced and discussed.

2.1 Noise robustness characterization and matched filter

Noise robustness can be characterized by the output variance (or mean square error - MSE) of the correlation peak when the input image is corrupted by noise:

$$MSE =< |c_0 - (\mathbf{r}+\mathbf{n})^\dagger.\mathbf{h}|^2 >= \mathbf{h}^\dagger S\mathbf{h}, \tag{1}$$

where $< . >$ represents the mean over the realizations of noise $\mathbf{n}$, which has covariance matrix S, zero mean and where c_0 is the mean of the central values of the correlation peaks. With our assumption of spatially stationary noise, we also have [33]:

$$MSE = \hat{\mathbf{h}}^\dagger \hat{S}\hat{\mathbf{h}} \tag{2}$$

where the diagonal element $\hat{S}_k$ of $\hat{S}$ is the noise spectral density at frequency k. The MSE is dependent on the modulus of $\mathbf{h}$, so in general the SNR is preferred in order to characterize noise robustness:

$$SNR = |c_0|^2/MSE \quad \text{with} \quad c_0 = \mathbf{h}^\dagger.\mathbf{r}. \tag{3}$$

Optimization of the SNR in the Fourier domain leads to the well known matched filter: $\hat{h}_k = \hat{r}_k/\hat{S}_k$, a proof of this result is given in appendix A. This filter is equivalent to the optimal solution in the bayesian approach with gaussian additive noise with known psd for a detection problem (see part 2).

This is an important theoretical approach to justify the use of filtering techniques (or correlation techniques) for pattern recognition.

Of course, the first limit of this approach with images is the assumption of additive gaussian noise. In particular, the assumption of additive noise is not completely realistic in image processing as discussed in [21]. With non-overlapping noise, different solutions have been proposed in [34] and [35] (see also part 2). The second important point is that this method assumes that the spectral density of the noise $\hat{S}$ is known. This is not the case in general in image processing, where in contrast to radar processing, it is very difficult to estimate $\hat{S}$. Then, an important question in the context of pattern recognition is the determination of an appropriate model for $\hat{S}$. Furthermore, there is no reason to consider

that the realizations of noise are obtained with a temporal stationary density probability law. For example, this is clear if images of the ground obtained from an airplane are considered. If the background is modeled by a noise, its spectral density can be very different for the sea [36], mountains or fields. Furthermore, even for application for which the noise is additive with a known spectral density, an important limitation of the matched filter is its low discrimination capabilities and the presence of sidelobes in the correlation plane which can result in false detection.

2.2 Sharpness of the correlation peak and inverse filter

In order to control the whole correlation plane and to minimize the possibility of large sidelobes which could result in false detection, the Peak to Correlation Energy (PCE) [29] is classically optimized:

$$PCE = |c_0|^2/CPE \tag{4}$$

where the Correlation Plane Energy (CPE) is defined as: $CPE = \mathbf{c}^\dagger.\mathbf{c} = \hat{\mathbf{c}}^\dagger.\hat{\mathbf{c}}$ (i.e the integral of the square modulus of the correlation function) :

$$CPE = \sum_{k=0}^{N-1} |\hat{c}_k|^2 = \sum_{k=0}^{N-1} |\hat{h}_k^* \hat{r}_k|^2.$$

It is easy to verify that the inverse filter ($\hat{h}_k = \hat{r}_k/|\hat{r}_k|^2$) optimizes this criterion (the proof is analogous to the one proposed in appendix A). This filter is equivalent to a matched filter designed for noise with spectral density matrix equal to $|\hat{r}_k|^2$. In that case, noise and target would have the same spectral density. At first sight, this situation may appear to be the most unfavorable one. We will see in the following section that this is not the case.

Other criteria have been proposed in order to characterize the sharpness of the correlation peak and to limit the existence of sidelobes (see [29] for a critical discussion about these criteria). However, they do not lead, up to now, to explicit solutions (i.e. for which the optimal filter has a known analytical expression).

2.3 Optical efficiency and phase only filter

For optical correlators, it is important to consider a third criterion [3] which characterizes the amount of light which will be detected for the determination of the correlation function. Indeed, if a linear filter is implemented with an optical Vander Lugt correlator, for each spatial frequency, one necessary has: $|\hat{h}_k| \leq 1$. The optical efficiency η_0 can thus be quantitatively characterized by:

$$\eta_0 = |c_0|^2, \tag{5}$$

The optical efficiency is therefore optimized with Phase-Only-Filters (POF) among which the choice $\hat{h}_k = \hat{r}_k/|\hat{r}_k|$ leads to high correlation peaks.

It is possible to obtain an interesting interpretation of this result in comparison with the matched filter. As mentioned above, an important question with the matched filter is to

infer a spectral density for the noise model. A possible approach may be to consider the worst case for the SNR for a given total noise power:

$$S_{total} = \sum_k \hat{S}_k. \tag{6}$$

In this approach, under the constraints of Eq. (6), by minimization of the SNR defined by Eq. (3) using the Lagrange multipliers technique, it can be shown that the worst density is: $\hat{S}_k = \beta\ |\hat{r}_k|$. The matched filter to this noise spectral density is then: $\hat{h}_k = \hat{r}_k/|\hat{r}_k|$, which is the POF. This property has been illustrated with numerical simulations in [18].

It is interesting to note that in that case the spectral density of the noise is not the spectral density of the reference ($|\hat{r}_k|^2$) but its square root. (We have seen that the optimization of the CPE leads to a filter optimal to noise with a spectral density matrix equal to $|\hat{r}_k|^2$).

2.4 Discrimination capabilities and a new approach

It can be important to improve the discrimination capabilities of the filter between the object to be recognized and objects to be rejected. For this purpose, a method discussed in [37] can be generalized. Let $\hat{\mathbf{y}}^\ell$ denote (with $\ell = 1, ..., P$) some objects to be rejected. For improving the discrimination capabilities of the filter, a possible method consists of minimizing the energy of the correlation with the images $\hat{\mathbf{y}}^\ell$, with the constraint that the correlation with the object $\hat{\mathbf{r}}$ to be recognized is equal to a given value.

In this case, the criterion to minimize is:

$$D_E = \sum_{\ell=1}^{P} \sum_k |\hat{h}_k^*.\hat{y}_k^\ell|^2. \tag{7}$$

Let $\hat{S}_k^{(rj)} = \sum_{\ell=1}^{P} |\hat{y}_k^\ell|^2$ be the average spectral density of patterns to be rejected. An optimal filter for this purpose is given by [37]: $\hat{h}_k = \hat{r}_k/\hat{S}_k^{(rj)}$.

The interpretation is straightforward, optimizing the discrimination capabilities leads to the consideration of a matched filter with a spectral density of noise equal to $\hat{S}_k^{(rj)}$.

Optimizing discrimination is a very attractive approach to the optimization of filters. Indeed, realizations of noise can be considered as images to be rejected. However, the general question of inferring the appropriate images $\hat{\mathbf{y}}^\ell$ which correspond to objects to be rejected is still not obvious.

It is interesting to note that this approach allows us to obtain a new interpretation of the CPE. The most difficult situation for discrimination can be expected when $\hat{\mathbf{y}}^\ell \simeq \hat{\mathbf{r}}$ for all ℓ. In that case, D_E is equivalent to the CPE and the inverse filter is optimal. However, as it will be seen below, this filter is in general unstable.

A new approach to overcome this drawback was proposed in [30]. Let $\hat{\mathbf{s}}$ denote the Fourier transform of the input image (i.e. the analyzed image). A discriminant filter for this image is obtained by minimizing the energy of the correlation function:

$$||\mathbf{c}||^2 = \sum_k |\hat{h}_k|^2 |\hat{s}_k|^2 \tag{8}$$

Of course, minimization of (8) leads to the null filter. However, if this minimization is performed under the constraint that there still is a correlation peak with the target (i.e. $\sum_k \hat{h}_k^* \hat{r}_k = c_0$), we will show below that an optimal nonlinear filtering technique can be obtained. We will also show that this approach can be very attractive if it is correctly regularized. Right now, we can observe that this criterion (8) is dependent of the input image. For each input image, a different filter will be obtained, the filter is thus adaptive with the input image.

2.5 Optimal SDF filters

In some practical situations of pattern recognition, the object to recognize may appear with different attitudes. For example the object have to be localized or recognized for different view angles and/or at different scales. The linear filtering technique with a filter designed for a single attitude (at given in-plane and out-of-plane rotation angles and at a specific scale) is in general unable to recognize the object with different attitudes.

In that case, two main approaches can be considered. The first one is brute force: a filter is designed for each attitude which can appear in the input image. The input image is then correlated with all the different filters. One can show that this approach is in general the best in terms of probability of good recognition, detection or location (see part 2 section 8.3.1). However, if the number of different attitudes is large, this approach can be very tedious and incompatible with the requirement of fast processing for most of the applications which may require optical correlators. Indeed, for a large number of attitudes (for invariant recognition for example) this optimal approach can lead to the computation of a very large number of correlations.

In order to overcome this bottleneck, suboptimal strategies have been proposed. The second approach belongs to that category and consists in designing filters which have some invariance or tolerance properties. A possible solution is to define a learning set of images which contains reference images with sufficient examples of the different possible attitudes. This is the basic concept of the Synthetic Discriminant Funcition filter (SDF) we now describe. Let $\mathbf{r}^\ell$ with $\ell = 1, ..., P$ denote the P reference images. SDF filters are defined in order to obtain some specified values at the center of the correlation function for the patterns belonging to the training set. Then, the SDF filter must satisfy the following constraints:

$$\hat{\mathbf{h}}^\dagger . \hat{\mathbf{r}}^\ell = d_\ell \qquad ; \forall \ell = 1, ..., P \tag{9}$$

where the d_ℓ are in general equal to one if the pattern $\mathbf{r}^\ell$ has to be recognized and is equal to zero otherwise. There exist in general an infinite number of solutions for the filter $\mathbf{h}$ (since in general $N >> P$) and the optimization of some relevant criteria is thus possible.

We thus now propose to generalize the definitions of the preceding criteria to the SDF problem. For the sake of simplicity, we will not consider the optical efficiency (an analysis of this point can be found in [38]).

With the constraints of Eq. (9), minimization of the MSE leads to the minimum variance SDF filter (MVSDF) [15]:

$$\hat{\mathbf{h}} = \hat{S}^{-1} \hat{R} \left[\hat{R}^\dagger \hat{S}^{-1} \hat{R} \right]^{-1} \mathbf{d} \tag{10}$$

where $\hat{R} = [\hat{\mathbf{r}}^1, \hat{\mathbf{r}}^2, ..., \hat{\mathbf{r}}^P]$.

A genereral analysis of determining optimal SDF filters for quadratic criteria is proposed in appendix B.

This filter shows the same limitation as the matched filter to lead to correlation functions with large sidelobes. It has thus been proposed to generalize the Correlation Peak Energy criterion for SDF filters. The average correlation peak energy (ACPE) for all the training patterns is:

$$ACPE = \frac{1}{P} \sum_{\ell=1}^{P} \sum_{k=1}^{N} |\hat{r}_k^\ell \hat{h}_k|^2 = \hat{\mathbf{h}}^\dagger \hat{D} \hat{\mathbf{h}} \tag{11}$$

with: $\hat{D}_k = 1/P \sum_{\ell=1}^{P} |\hat{\mathbf{r}}_k^\ell|^2$. The minimization of the $ACPE$ leads to the minimum average correlation energy (MACE) filter [16] :

$$\hat{\mathbf{h}} = \hat{D}^{-1} \hat{R} \left[\hat{R}^\dagger \hat{D}^{-1} \hat{R} \right]^{-1} \mathbf{d} \tag{12}$$

The MACE filter shows the same drawback as the inverse filter: it is very sensitive to input noise on the analyzed image [39].

2.6 Optimal Trade-off filters

Here, the general concept of Optimal Trade-off (OT) filters is recalled. In general the filter is subect to some constraints. For example, for invariant filtering, the central values of the correlation functions are imposed on a set of training patterns. For optical implementation reasons, $\hat{\mathbf{h}}$ can be constrained to belong to a particular subset of C ($\hat{h}_k \in \mathcal{S}$), for each frequency k [40]. These constraints impose that the filter $\hat{\mathbf{h}}$ can only be in a subset $\mathcal{D}$ of $\mathbf{C}^N$. For example, for optical implementation constraints we thus have $\mathcal{D} = \mathcal{S}^N$ and for SDF filters $\mathcal{D}$ is the hyperplane defined by the constraint of Eq. 9.

Let $E_1(\hat{\mathbf{h}})$, $E_2(\hat{\mathbf{h}})$, ...,$E_M(\hat{\mathbf{h}})$ denote the values of the M considered criteria for the filter $\hat{\mathbf{h}}$ (for example MSE and $ACPE$). Each criterion is defined in order to correspond to a minimization problem.

OT filters are defined in order to lead to the most interesting trade-off between the considered criteria with the constraint that the filter $\hat{\mathbf{h}}$ belongs to $\mathcal{D}$. Consider a filter $\hat{\mathbf{h}}^{(1)}$ in $\mathcal{D}$. The values of its criteria are: $E_1(\hat{\mathbf{h}}^{(1)})$, $E_2(\hat{\mathbf{h}}^{(1)})$, ..., $E_M(\hat{\mathbf{h}}^{(1)})$. If there exist a filter $\hat{\mathbf{h}}^{(2)}$ in $\mathcal{D}$ such that:

$$E_j(\hat{\mathbf{h}}^{(2)}) \; < \; E_j(\hat{\mathbf{h}}^{(1)}) \quad \forall j = 1, ..., M.$$

then it is clear that $\hat{\mathbf{h}}^{(2)}$ is a more interesting filter than $\hat{\mathbf{h}}^{(1)}$. OT filters are the filters $\hat{\mathbf{h}}^{(OT)}$ for which there is no filter $\hat{\mathbf{h}}$ in $\mathcal{D}$ such that:

$$E_j(\hat{\mathbf{h}}) \leq E_j(\hat{\mathbf{h}}^{(OT)}) \quad \forall j = 1, ..., M. \tag{13}$$

with at most one strict inequality. In order to obtain optimal trade-off (OT) filters it is necessary to minimize in $\mathcal{D}$ [40]:

$$E(\hat{\mathbf{h}}) = \sum_{j=1}^{M} \lambda_j E_j(\hat{\mathbf{h}}) \tag{14}$$

where the λ_j are positive numbers that allow to balance the optimization between the different criteria.

It is important to stress that the minimization of a linear combination of criteria in Eq. (14) does not arise from empirical considerations, but is a strategy in order to find OT filters.

2.7 Optimal trade-off SDF filters

The MACE filter is very sensitive to input noise on the analyzed image [39], while the MVSDF filter is not enough discriminant in general. It is the reason why Optimal trade-off SDF (OTSDF) filters were introduced [17].

When criteria are quadratic forms of the filter $\mathbf{h}$, the OT approach also leads to the minimization in $\mathcal{D}$ of a quadratic form which will be written:

$$E(\mathbf{h}) = \mathbf{h}^\dagger B \mathbf{h} = \hat{\mathbf{h}}^\dagger \hat{B} \hat{\mathbf{h}}$$

Let us, for example, consider the filter which corresponds to the optimal trade-off among the two previous criteria (MSE and $ACPE$). The quadratic form to optimize is a linear combination of the MSE and $ACPE$ which leads to the OT SDF filter:

$$\hat{\mathbf{h}} = \hat{B}^{-1}\hat{R}\left[\hat{R}^\dagger \hat{B}^{-1}\hat{R}\right]^{-1}\mathbf{d} \qquad \text{where:} \qquad \hat{B} = (1-\mu)\hat{S} + \mu\hat{D} \tag{15}$$

where the parameter μ allows us to balance optimally between the different criteria (μ is positive and smaller than 1). If $\mu = 0$, the MVSDF filter [15] is obtained (see Eq. 10). If $\mu = 1$ the MACE filter [16] is obtained (see Eq. 12).

Since this result is derived in [17], let us only comment this equation. $\hat{B}$ is a diagonal matrix of size $N \times N$ and R is a rectangular matrix of size $N \times P$, so that $\hat{R}^\dagger \hat{B}^{-1}\hat{R}$ is a matrix of size $P \times P$. The inverse matrices $\hat{B}^{-1}$ and $\left[\hat{R}^\dagger \hat{B}^{-1}\hat{R}\right]^{-1}$ must be understood as inverse matrices in the space spanned by the eigenvectors associated to non null eigenvalues. Let us denote:

$$b_\ell = \sum_{n=1}^{P}\left(\left[\hat{R}^\dagger \hat{B}^{-1}\hat{R}\right]^{-1}\right)_{\ell,n} d_n$$

we can then write:

$$\hat{h}_k = \sum_{\ell=1}^{P} b_\ell \frac{\hat{r}_k^\ell}{B_k} \tag{16}$$

The Minace filter [28] was more recently introduced in order to obtain trade-offs between $ACPE$ and MSE. This filter can be written:

$$\hat{\mathbf{h}} = \hat{N}^{-1}\hat{R}\left[\hat{R}^\dagger \hat{N}^{-1}\hat{R}\right]^{-1}\mathbf{d} \quad \text{with:} \quad \hat{N}_k = Max[(1-\mu)\hat{S}_k, \mu\hat{D}_k] \tag{17}$$

and where $Max[x, y]$ means maximum value between x and y.

Minace and OTSDF filters were compared in [41]. The stability of these filters will be analyzed in the following. The choice of (17) is still an empirical approach.

In conclusion, it can be seen that optimal SDF filters have a mathematical expression which can be written:

$$\hat{\mathbf{h}} = \hat{A}^{-1}\hat{R}\left[\hat{R}^{\dagger}\hat{A}^{-1}\hat{R}\right]^{-1}\mathbf{d} \tag{18}$$

where $\hat{A}$ is a diagonal matrix which corresponds to the optimization of the criterion $E = \sum_k |\hat{h}_k|^2 \hat{A}_k$. The choice of a matrix $\hat{A}$, and therefore of the criterion which is optimized, can also be considered as an inference of the spectral density of the noise model (i.e. $\hat{S}_k = \hat{A}_k$). In other words, OTSDF and Minace filters can be considered as MVSDF filters with respectively noise spectral density equal to $\hat{B}_k$ or to $\hat{N}_k$.

3 Stability of the solutions

The solution of the general filter synthesis problem formulated above is not always stable with respect to the reference model.

We have seen that the filter synthesis problem can be written:

$$\mathbf{h}^{opt} = \underset{\mathbf{h} \in \mathcal{D}}{argmin}\ E(\mathbf{h}) \tag{19}$$

where $\underset{\mathbf{h} \in \mathcal{D}}{argmin}\ E(\mathbf{h})$ means: the value of $\mathbf{h}$ which minimizes $E(\mathbf{h})$ in $\mathcal{D}$.

Let us first consider the case for which there is a single pattern in the training set. The optimal filter of Eq. (19) is a function of the reference image. Let us emphasize this point by writing: $\mathbf{h}^{opt} = \mathbf{h}^{opt}[\mathbf{r}]$.

A filter is stable if a small variation in the reference pattern ($\mathbf{r} \longrightarrow \mathbf{r}+\delta\mathbf{r}$ with $||\delta\mathbf{r}|| << 1$) does not induce a large variation in the filter ($\delta\mathbf{h}^{opt}[\mathbf{r}] = \mathbf{h}^{opt}[\mathbf{r}+\delta\mathbf{r}] - \mathbf{h}^{opt}[\mathbf{r}]$). It is clear that if the filter is not stable, a small deviation $\delta\mathbf{r}$ of the reference will result in a large mismatch between the filter and the reference $\mathbf{r}+\delta\mathbf{r}$. In other words, a filter which is not stable against small perturbations in the pattern $\mathbf{r}$ will be very discriminant but will have very poor tolerance capabilities against distortions of the input image.

A possible distortion can be the addition of a noise which is very different from the one expected [18]. The possible distortion can also be small deformations, due for example to in-plane or out-of-plane rotation.

This concept of stability can be generalized when there are several patterns in the training set. In that case, the filter will be considered stable if it is stable for each possible modifications: $\mathbf{r}^{\ell} \longrightarrow \mathbf{r}^{\ell} + \delta\mathbf{r}^{\ell}$ $(\forall \ell = 1, ..., M)$. Since in general all the possible distortions cannot be included in the learning set, robustness to small in-plane, out-of-plane rotation or scale variations, not included in the training set is still important for SDF filters.

The importance of stability is thus clear in order to design filters which are robust to input images with properties slightly different to the ones expected during the filter design.

3.1 Analysis of the stability of optimal filters

Let us first consider filters with a single pattern in the training set.

Since OT filters include the matched and the inverse filters we now consider that case. The mathematical expression of the OT filter is (Eq. 15): $\hat{h}_k = \beta \hat{r}_k / \hat{B}_k$, where β is a multiplicative factor and $\hat{B}_k = (1-\mu)\hat{S}_k + \mu \hat{D}_k$.

If there is a frequency k for which $\hat{B}_k$ is small, then a small perturbation in $\mathbf{r}$ may result in a large variation of $\mathbf{h}$ and of $\mathbf{h} \star \mathbf{r}$. More precisely, if $\delta\hat{\mathbf{r}}$ is a small perturbation in $\hat{\mathbf{r}}$, the induced perturbation $\delta\hat{\mathbf{h}}$ in $\hat{\mathbf{h}}$ is:

$$\delta\hat{h}_k \simeq \beta \left((1-\mu)\frac{\hat{S}_k}{[\hat{B}_k]^2}\delta\hat{r}_k \; - \; \mu \frac{\hat{r}_k^2}{[\hat{B}_k]^2}\delta\hat{r}_k^* \right).$$

The norm of $\delta\hat{h}_k$ can be large if $\hat{B}_k$ is small.

If $\mathbf{h}$ is fixed ($\hat{h}_k = \beta \hat{r}_k / \hat{B}_k$), the variation of the center of the correlation function induced by $\delta\mathbf{r}$ is:

$$\delta c_0 = \beta \sum_k \frac{\hat{r}_k^*}{\hat{B}_k} \delta\hat{r}_k.$$

Thus, it is clearly seen that, in that case, neither the correlation function nor the filter are stable against small perturbations in the pattern $\mathbf{r}$.

However with a white noise model and μ not equal to 1, it can be expected to obtain a stable filter. More precisely the condition of stability is that, for all k, $(1-\mu)\hat{S}_k$ is not to small. We will precise this point in the following. On the other hand, the inverse filter is clearly unstable.

An equivalent approach (although the calculations are more tedious) would show that the Minace filter shows the same behavior.

Let us analyze the stability of the POF. We have seen that this filter can be written: $\hat{h}_k = \hat{r}_k / |\hat{r}_k|$. If $\delta\hat{\mathbf{r}}$ is a small perturbation in $\hat{\mathbf{r}}$, the induced perturbation $\delta\hat{\mathbf{h}}$ in $\hat{\mathbf{h}}$ is:

$$\delta\hat{h}_k \simeq \frac{1}{2|\hat{r}_k|}\delta\hat{r}_k - \frac{[\hat{r}_k]^2}{2|\hat{r}_k|^3}\delta\hat{r}_k^*.$$

The norm of $\delta\hat{h}_k$ can be large if $|\hat{r}_k|$ is small which is a typical situation for high frequencies with images. It is then also clear that the POF will not be stable in general. This is the main reason for its high level of discrimination and low level of tolerance (such as noise or rotation robustness for example). We proposed in section 2.3 an interpretation of the POF as a matched filter designed with the spectral density noise model which corresponds to the worst situation. We see that this approach leads to an unstable filter, i.e. which is too specialized to this particular noise model. We will discuss in the following a simple method in order to regularize the POF.

3.2 Stability of optimal SDF filters

Let us consider the case of the OTSDF filter (which contains as limit cases the MVSDF and the MACE filter). The expression of Eq. (16) ($\hat{h}_k = \sum_{\ell=1}^{P} b_\ell \hat{x}_k^\ell / B_k$) of the OTSDF filter leads

to a simple analysis. Indeed, the stability of SDF filters can be deduced from the stability of the filters of the previous section.

Let us denote $\delta\hat{\mathbf{r}}^\ell$ a small perturbation in $\hat{\mathbf{r}}^\ell$, the induced perturbation $\delta\hat{\mathbf{h}}$ in $\hat{\mathbf{h}}$ is:

$$\delta\hat{h}_k = \sum_{\ell=1}^{P} b_\ell \left[\frac{\hat{B}_k - \mu|\hat{r}_k^\ell|^2}{[\hat{B}_k]^2}\ \delta\hat{r}_k^\ell - \mu\ \frac{(\hat{r}_k^\ell)^2}{[\hat{B}_k]^2}(\delta\hat{r}_k^\ell)^* \right]. \tag{20}$$

Here again, the norm of $\delta\hat{h}_k$ can be large if $\hat{B}_k$ is small.

The MACE filter is unstable because the spectral density of images is generally very small at high spatial frequencies.

The MVSDF filter is stable for a white noise model, but is unstable for low band frequency noise models. The OTSDF filter, as well as the Minace filter, shows a similar behavior. It is stable if it has been designed with a white noise model and unstable if it has been designed for a low band frequency noise model.

4 Regularization of filters

It has been shown in the previous section that the general problem of filter synthesis can be formulated as the minimization of some criterion, in a definition set $\hat{\mathcal{D}}$. The optimal filter $\hat{\mathbf{h}}^{opt}$ can thus be written:

$$\hat{\mathbf{h}}^{opt} = \underset{\hat{\mathbf{h}} \in \hat{\mathcal{D}}}{argmin}\ E(\hat{\mathbf{h}}) \tag{21}$$

In order to regularize this inverse problem, different solutions can be employed. We will first analyze the simple method of regularization by truncation and then review some well known heuristic approaches. Then we will consider the regularization of the inverse problem with the use of a stabilizing functional.

4.1 Truncature method for regularization

A classical approach to the regularization of inverse problems is to forbid the inversion of small values. Let us first consider the case of filters designed for a single reference object. Their general expression can be written: $\hat{h}_k = \hat{r}_k/\hat{A}_k$ (which optimizes the criterion $\sum_k \hat{h}_k^* \hat{A}_k \hat{h}_k$). The truncature method of regularization replaces this filter by:

$$\hat{h}_k = \begin{cases} \hat{r}_k/\hat{A}_k & \text{if} \quad \hat{A}_k > \epsilon \\ \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Of course, it is easily verified that the larger ϵ, the more regularized the solution, but the smaller the correlation peak and the larger the optimized criterion $\sum_k \hat{h}_k^* \hat{A}_k \hat{h}_k$.

Because of the all-pass nature of the POF and of the Binary POF [42], the resulting SNR can be very low [10, 43]. This remark has led to the concept of region of support [43]

of the POF in order to improve its noise robustness. This concept was also generalized in order to find optimal trade-offs between different criteria [44].

Let us consider the POF: $\hat{h}_k = \hat{r}_k/|\hat{r}_k|$. The regularization technique by truncature leads to:

$$\hat{h}_k = \begin{cases} \hat{r}_k/|\hat{r}_k| & \text{if} \quad |\hat{r}_k| > \epsilon \\ \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

which allows us to analyze the binary amplitude POF [45, 46, 44] as a regularized version of the POF. The same approach would allow us to analyze ternary filters as a regularized version by truncature of the binary POF. The truncature approach thus provides a new interpretation of binary amplitude and ternary filters [47, 48]. These filters are attractive since they can be easily implemented on available spatial light modulators.

This technique generalizes previous approaches of optimization of the performance and tolerance of POF [45, 46, 44]. In particular, with a non white noise model in the previous methods, the filter may be unstable. The truncature regularization will guarantee its stability.

This approach is easily generalized to optimal SDF filters. In that case, the truncature regularization leads to:

$$\hat{\mathbf{h}} = \hat{F}^{-1}\hat{R}\left[\hat{R}^{\dagger}\hat{F}^{-1}\hat{R}\right]^{-1}\mathbf{d} \tag{24}$$

where:

$$\hat{F}_k = \begin{cases} \hat{B}_k & \text{if} \quad \hat{B}_k > \epsilon \\ \\ 0 & \text{otherwise} \end{cases}$$

remembering that B, and then F, is diagonal in the Fourier domain, and that the inverse matrices must be understood as the pseudo–inverse.

It is also interesting to remark that truncature regularization applied to the Minace filter leads to a filter analog to Eq. (24) but with kernel:

$$\hat{F}_k = \begin{cases} \hat{N}_k & \text{if} \quad \hat{N}_k > \epsilon \\ \\ 0 & \text{otherwise} \end{cases}$$

4.2 Stabilizing functional

In order to regularize filters, a stabilizing functional $\Omega(\mathbf{h})$ can be introduced to stabilize the solution to small perturbations in the data (i.e. training patterns).

The regularized solution is simply defined as the solution of the following problem:

$$\hat{\mathbf{h}}^{reg} = \underset{\hat{\mathbf{h}} \in \hat{\mathcal{D}}}{argmin} \left[E(\hat{\mathbf{h}}) + \alpha\Omega(\hat{\mathbf{h}})\right] \tag{25}$$

The parameter α can be considered as a Lagrange parameter introduced in order to satisfy the inequality:

$$\Omega(\hat{\mathbf{h}}) \leq \epsilon. \tag{26}$$

where $\epsilon > 0$. In general, the Lagrange parameter α is not identified, but it is clear that the larger α the smaller $\Omega(\hat{\mathbf{h}}^{reg})$ [19].

When the stabilizing functional is quadratic, this is a very convenient approach since an explicit mathematical equation for the filter can still be found (this would not be the case, for example, with maximum entropy methods [49–51]).

Furthermore, as mentioned above, an interesting aspect of the stabilizing functional approach is to introduce *a priori* knowledge in the filter synthesis. Indeed, there are some *a priori* properties of the filter which are not easily defined as precise criteria. This is in particular the case for insensitiveness to distortions of the input pattern when all the possible distortions cannot be simply described, or included in a training set.

In the following, some examples of stabilizing functionals which improve the robustness to input distortions are proposed.

4.3 The minimum norm stability functional

Let us consider a filter $\mathbf{h}$, and the perturbated version $\mathbf{r} + \delta\mathbf{r}$ of an input pattern $\mathbf{r}$. The induced variation of the correlation function is:

$$\delta\mathbf{c} = \mathbf{h} \star [\mathbf{r} + \delta\mathbf{r}] - \mathbf{h} \star \mathbf{r} = \mathbf{h} \star \delta\mathbf{r}. \tag{27}$$

This variation of the correlation function can be bounded using the Cauchy-Schwarz inequality:

$$\| \delta\mathbf{c} \| \leq \| \mathbf{h} \| \| \delta\mathbf{r} \| \tag{28}$$

where $\| \ . \ \|$ represents the euclidian norm.

Thus, imposing a maximum value for $\| \mathbf{h} \|$ in turn imposes a bounding value for the norm $\| \delta\mathbf{c} \|$ of the variation of the correlation function for a given norm $\| \delta\mathbf{r} \|$ of the perturbation $\delta\mathbf{r}$.

In order to obtain a quadratic criterion, it is appropriate to choose the stabilizing functional:

$$\Omega(\mathbf{h}) = \| \mathbf{h} \|^2 \tag{29}$$

which corresponds to a classical regularization method [52]. When the criterion $E(\hat{\mathbf{h}})$ is also quadratic (i.e. $E(\hat{\mathbf{h}}) = \hat{\mathbf{h}}^{\dagger} \hat{B} \hat{\mathbf{h}}$), the regularized filter $\hat{\mathbf{h}}^{reg}$ is given by:

$$\hat{\mathbf{h}}^{reg} = \underset{\hat{\mathbf{h}} \in \hat{\mathcal{D}}}{argmin} \left[\hat{\mathbf{h}}^{\dagger} \hat{B} \hat{\mathbf{h}} + \alpha \hat{\mathbf{h}}^{\dagger} . \hat{\mathbf{h}} \right] \tag{30}$$

It is easy to check that the regularized filter is thus given by:

$$\hat{\mathbf{h}}^{reg} = \left[\hat{B} + \alpha \ I_d \right]^{-1} \hat{R} \left[\hat{R}^{\dagger} \left[\hat{B} + \alpha \ I_d \right]^{-1} \hat{R} \right]^{-1} \mathbf{d} \tag{31}$$

where I_d is the identity matrix in C^N (i.e. diagonal with unitary elements).

With a single pattern in the training set, the regularized filter is given by: $\hat{h}_k = \beta \hat{r}_k / [\hat{B}_k + \alpha]$, where β is a multiplicative factor and where $\hat{B}_k$ is defined by Eq. (15).

Figure 1: Image of an airplane.

Then, if there exist some frequencies k for which $\hat{B}_k$ is small, the filter will be approximately equal to $\hat{r}_k/\alpha$ at these frequencies, which limits the effect of a small perturbation of $\hat{\mathbf{r}}$ on the modulus of the correlation function.

For a small perturbation $\delta\hat{\mathbf{r}}$, we now obtain:

$$\delta\hat{h}_k \simeq \beta \frac{\hat{B}_k + \alpha - \mu |\hat{r}_k|^2}{[\hat{B}_k + \alpha]^2} \delta\hat{r}_k - \mu \frac{\hat{r}_k^2}{[\hat{B}_k + \alpha]^2} \delta\hat{r}_k^*.$$

which modulus is now bounded. It is easy to see that for an OT filter with a white noise model, this kind of regularization is naturally obtained.

4.4 The minimum variation stability functional

In the previous section, no particular form for the perturbation $\delta\mathbf{r}$ was assumed. However, in general, it is needed that the filter be robust to small distortions of the input image, which can be introduced for example by small variations of the attitude of the object (scale, view angle...). It is then natural to impose that the filter variation be bounded for perturbations $\delta\mathbf{r}$ which look like the ones induced by small attitude variations. Since such small attitude variations emphasize mainly the edges of the object, $\delta\mathbf{r}$ is mostly composed of high frequencies. In order to support this conjecture, we show the difference between an image and a small attitude variation of itself. More precisely, Fig. 1 shows an image of an airplane, and Fig. 2 the modulus of the subtraction of this image with its 5 degrees rotated version. It is clearly seen that the difference image is essentially an edge enhanced image of the airplane.

Thus, a first approximation of $\delta\mathbf{r}$ might be:

$$\delta\mathbf{r} = \mathbf{g}_{hp} \star \mathbf{r} \tag{32}$$

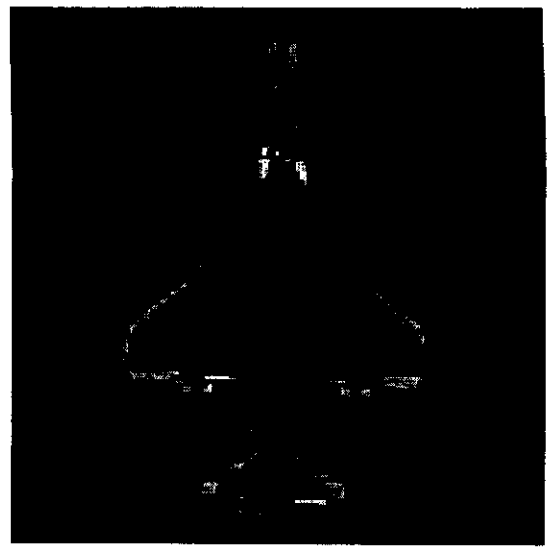


Figure 2: Modulus of the subtraction of the previous image with its rotated version by an angle of 5 degrees.

where $\mathbf{g}_{hp}$ is a high pass filter, enhancing the edges of $\mathbf{r}$. Using the approach of the previous section, the Cauchy-Schwarz inequality leads to the minimization of the following stabilizing functional:

$$\Omega(\mathbf{h}) = \| \mathbf{h} \star \mathbf{g}_{hp} \|^2 \tag{33}$$

When the criterion $E(\hat{\mathbf{h}})$ is quadratic, the regularized filter $\hat{\mathbf{h}}^{reg}$ is given by:

$$\hat{\mathbf{h}}^{reg} = \left[\hat{B} + \alpha\, \hat{G}\right]^{-1} \hat{R} \left[\hat{R}^{\dagger} \left[\hat{B} + \alpha\, \hat{G}\right]^{-1} \hat{R}\right]^{-1} \mathbf{d} \tag{34}$$

where $\hat{G}_k = \left|[\hat{g}_{hp}]_k\right|^2$. Then, if there are some high frequencies k for which $\hat{B}_k$ is small, the filter will be approximately equal to $\hat{r}_k / \alpha \hat{G}_k$ at these frequencies, which bounds the modulus of the effect of a small perturbation of $\mathbf{r}$ due to a small edge shifting.

4.5 The minimum elastic energy stability functional

The previous considerations lead to analyze the problem with the following approach. The filter needs to be robust to small distortions of the input image which can be induced by small variations of the object (scale, view angle...). This can be achieved if the filter values do not vary rapidly in the image domain. In other words, it is required that the filter be a smooth function of the spatial coordinate (u, v). For the sake of simplicity, let us first consider a non sampled case. The filter will be denoted $h(u, v)$. It can be imposed to the filter to be a smooth function by minimization of the elastic energy:

$$\int\int \left| \frac{\partial^2 h(u,v)}{\partial u^2} + \frac{\partial^2 h(u,v)}{\partial v^2} \right|^2 du dv \tag{35}$$

Using the Fourier transform of continuous functions, and after sampling, it is easily shown [19] that this approach leads to the stabilizing functional:

$$\Omega(\hat{\mathbf{h}}) = \sum_{k_u} \sum_{k_v} \left[{k_u}^2 + {k_v}^2 \right]^2 \left| \hat{h}(k_u, k_v) \right|^2 . \tag{36}$$

When the criterion $E(\hat{\mathbf{h}})$ is quadratic, the regularized filter $\hat{\mathbf{h}}^{reg}$ is given by:

$$\hat{\mathbf{h}}^{reg} = \left[\hat{B} + \alpha \, \hat{L} \right]^{-1} \hat{R} \left[\hat{R}^{\dagger} \left[\hat{B} + \alpha \, \hat{L} \right]^{-1} \hat{R} \right]^{-1} \mathbf{d} \tag{37}$$

where:

$$\hat{L}_{(k_u,k_v)} = \left[{k_u}^2 + {k_v}^2 \right]^2 .$$

It is then seen again that if the frequencies for which $\hat{B}_k$ is small are high frequencies, $\hat{L}$ will stabilize the filter.

The minimum elastic energy stability functional presents a strong analogy with the minimum variation stability functional since both measure the energy of high frequencies.

4.6 Application to optimal discriminant processor

We saw in section 2.4 that the discrimination capabilities of filters is generally optimized indirectly by minimizing either the sharpness of the correlation function [17], or the energy of the correlation function with objects to be rejected or background models to be discriminated against [37]. We also discussed the limitations of this approach and introduced a new criterion.

With the use of the minimum norm stabilizing functional, we are now able to determine a method for optimizing the discrimination capabilities of the processor which does not need *a priori* knowledge of objects to be rejected or of the background [30].

Let $\mathbf{r}$ and $\mathbf{s}$ denote respectively the reference and input image. The output of the processor is still written $c_i = \sum_{j=1}^{N} h^*_{i+j} s_j$, but $\mathbf{h}$ may now be a function of $\mathbf{s}$.

When the input image is the reference object, it is imposed that the filter produces the correlation peak $c_0 = d$. This leads to the constraint:

$$\sum_{j=1}^{N} h^*_j r_j = d. \tag{38}$$

As mentioned in section 2.4 and in [30], in order to optimize the discrimination capabilities of the processor, the energy of the correlation function with the input image $\mathbf{s}$ is minimized. Since we consider Eq. (38) as a constraint, this is equivalent to minimizing:

$$E_s[h] = \sum_{k} |\hat{h}_k|^2 \, |\hat{s}_k|^2. \tag{39}$$

which can also be written: $E_s[h] = \| \mathbf{c} \|^2$ with $\hat{c}_k = \hat{h}^*_k \, \hat{s}_k$.

In order to be robust to modifications or distortions of the reference image $\mathbf{r}$, and with no *a priori* knowledge of the perturbations, we consider the minimum norm stabilizing functional:

$\Omega(\mathbf{h}) = \| \mathbf{h} \|^2$. The problem is now clearly defined, $\hat{\mathbf{h}}^{reg} = \underset{\hat{\mathbf{h}} \in \hat{\mathcal{D}}}{argmin} \left[\| \mathbf{c} \|^2 + \alpha \| \mathbf{h} \|^2 \right]$ where $\mathcal{D}$ is defined by the constraint: $\sum_k \hat{h}_k^* \hat{r}_k = d$.

After appropriate substitution and taking into account that d is arbitrary, it can be shown [30] that the optimum processor is:

$$\hat{c}_k = \frac{\hat{r}(k)^* \, \hat{s}(k)}{\sigma^2 + |\hat{s}(k)|^2}. \tag{40}$$

where $\hat{\mathbf{c}}$ is the Fourier transform of the correlation output and where σ^2 is a given positive constant. It is obvious from Eq. (40) that this optimal processor is nonlinear since it requires a nonlinear transformation of the input image. This nonlinear processor is adaptive since the filter function is dependent on the input image energy spectrum. The optimum nonlinear processor can be generalized with any of the previous stabilizing methods or functionals. Furthermore, if $|\hat{s}(k)|^2$ is replaced by $|\hat{r}(k)|^2$ in Eq. (40), the correlation with an OT filter for peak sharpness and noise robustness with a white noise model is obtained [17]. It can be shown [30] that the regularized solution with the minimum norm functional of the optimal trade-off filter between discrimination and correlation peak sharpness (PCE) leads to the optimum processor: $\hat{c}_k = \hat{r}_k^* \hat{s}_k / [\sigma^2 + \mu \, |\hat{r}_k|^2 + (1 - \mu) \, |\hat{s}_k|^2]$ (with $\mu \in [0; 1]$). This processor is very similar to the Fourier transform of a nonlinear JTC output [24] which can be implemented optically. Both have the same Fourier phase and the amplitude modulation requires a nonlinear transformation of both Fourier magnitudes of the reference function and the input function.

In summary, we have seen that regularization with stabilizing functionals can be applied to optimal nonlinear processor design.

4.7 Application to optimal optical implementation

If the filter is to be implemented in an optical correlator with a SLM in the Fourier plane, all complex values for each spatial frequency of the filter cannot be reached. Indeed, in general a voltage is applied to each pixel of the SLM and only a bounded curve in the complex plane can be obtained. More precisely, this constraint results in a coupled modulation between amplitude and phase with bounded values. Well known examples are phase only modulation, pure amplitude modulation, binary phase only modulation or ternary modulation. In that latter case, for each spatial frequency, the ternary filter can only take the values (-1, 0 ,1). POF, binary POF and ternary filters have already been analyzed in the context of truncature regularization. However, it was shown more generally (see [53] and references therein) that other complex codings can be obtained, such as spirals in the complex plane.

We will take the example of Fourier plane optical filters. Let $\mathcal{S}$ denote the admissible domain of coding for each spatial frequency (for example $\mathcal{S} = (-1, 0, 1)$ for ternary filters). This constraint imposes that the filter $\hat{\mathbf{h}}$ can only be in a subset $\mathcal{D}$ of $\mathbf{C}^N$ which is defined by: $\mathcal{D} = \mathcal{S}^N$. If $\hat{h}_k = 0$ can be obtained with a good approximation (i.e. with a good contrast), the truncation regularization can be applied.

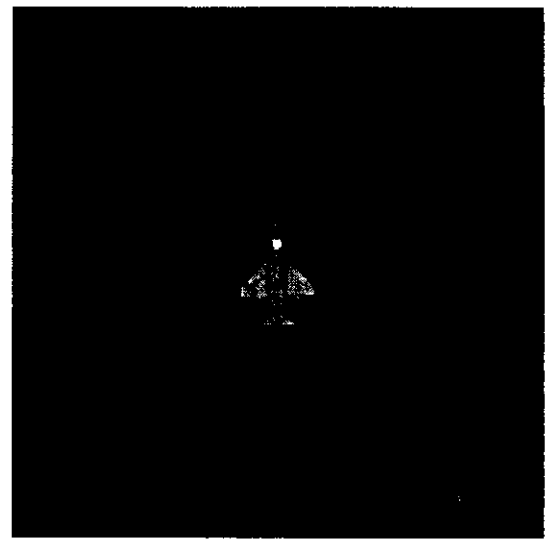


Figure 3: Target used in numerical simulations.

However, optimal realizable filters are obtained by optimizing a criterion [27]. Then, if this criterion is written as a function $E(\hat{\mathbf{h}})$ to be minimized, the stabilizing approach can also be applied. If we consider the minimum norm stabilizing functional, the optimal filter is given by:

$$\hat{\mathbf{h}}^{reg} = \underset{\hat{\mathbf{h}} \in \hat{\mathcal{D}}}{argmin} \left[E(\hat{\mathbf{h}}) - \lambda \hat{\mathbf{h}}^{\dagger}.\hat{\mathbf{r}} + \alpha \, ||\hat{\mathbf{h}}||^2 \right] \tag{41}$$

where $\mathcal{D} = \mathcal{S}^N$, and λ is a parameter which is introduced in order to balance between the optimization of the criteria and the optical efficiency. Indeed, since $\mathcal{S}$ is a bounded region, a trade-off is necessary in order to obtain a detectable correlation peak ($|\hat{\mathbf{h}}^{\dagger}\hat{\mathbf{r}}|^2$) in the correlation plane [54, 38].

In general this optimization does not lead to an explicit expression for the filter. The investigation of the iterative procedure is beyond the scope of this paper (see for example [40]). However, it is clearly seen that the general formalism of stabilizing functionals is well adapted to the regularization of optimal realizable filters which are obtained by optimizing a criterion.

5 Illustration with simple examples

Numerical simulations are now shown in order to illustrate in simple examples the effect of stabilization of filtering techniques for pattern recognition. The minimum norm stabilizing functional is first illustrated with the optimal nonlinear discriminant processor. We will also illustrate the effect of the elastic energy stabilizing functional in colored noise. For these different numerical simulations, the image of the target used is shown in Fig. 3.

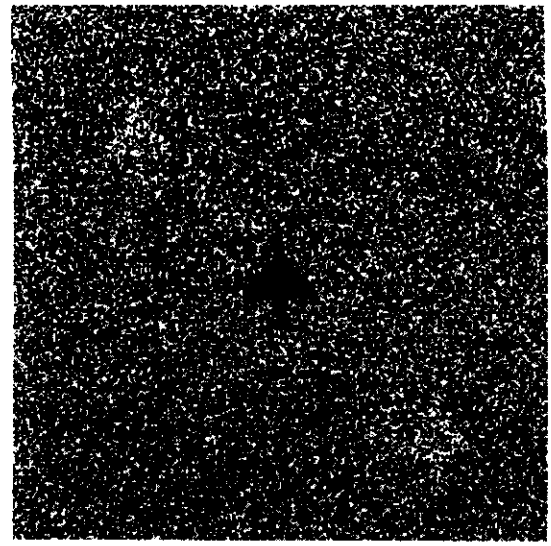


Figure 4: Input image used for numerical simulations with the optimal discriminant nonlinear processor.

5.1 Nonlinear processor

The input image is shown in Fig. 4. In Fig. 5, we show our conventions for the different objects in the input image, and the correspondence with the different correlation peaks. The airplanes appear buried in white noise. At the center (position B), the noise is nonoverlapping with the target. However, at positions A and C, the noise is additive (overlapping) with the target.

In Fig. 6, we show the result of the nonlinear processor of Eq. (40) with $\sigma^2 = 0$ and $\mu = 1/2$, i.e. without regularization. It is clearly seen that in this case, the correlation is not regular enough to identify correctly the different correlation peaks. In Fig. 7, a minimum norm regularization is considered (i.e. $\sigma^2 \neq 0$ and $\mu = 1/2$). The efficiency of the minimum

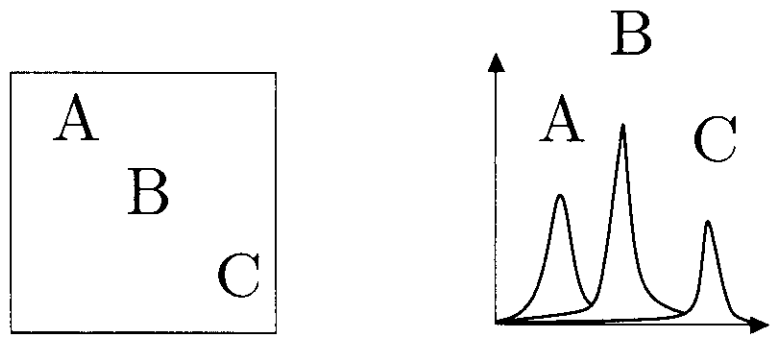


Figure 5: Conventions for the different objects present in the input image, and correspondence with the correlation peaks.

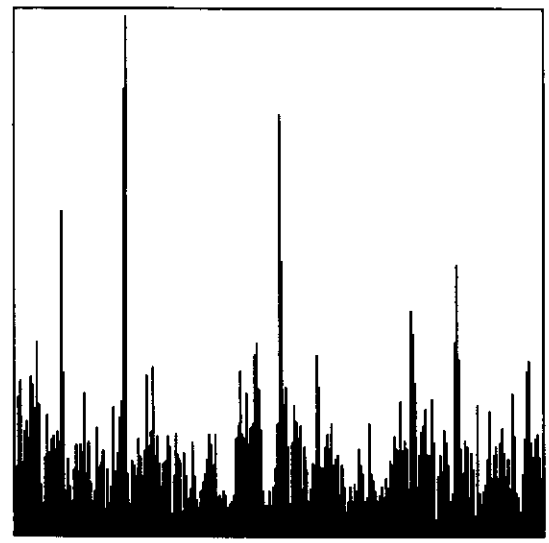


Figure 6: Correlation with the optimal discriminant nonlinear processor without regularization.

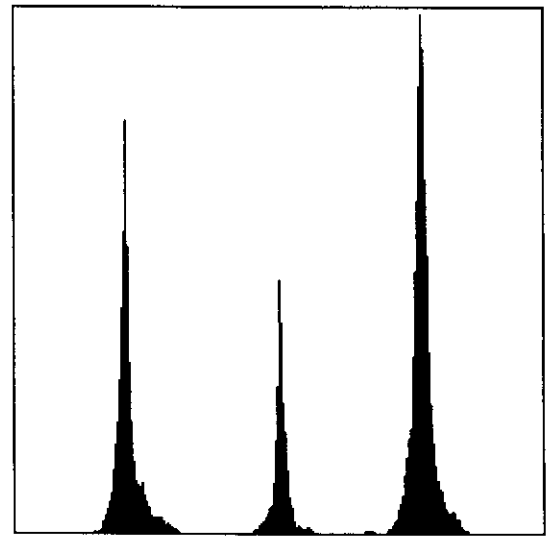


Figure 7: Correlation with the optimal discriminant nonlinear processor with regularization.

Figure 8: Input image used for numerical simulations with OT filters.

norm regularization is clear, since now the correlation peaks can easily be detected. Other numerical simulations comparing the nonlinear processor with OT filters were published in [30].

5.2 Elastic energy stabilizing functional

We now illustrate the effect of the minimum elastic energy stabilization of the simple OT filter (i.e. without learning capabilities).

The target is still the airplane of Fig. 3, however the input image is now that of Fig. 8.

The airplanes now appear on colored noise, with a spectral density equal to $1/[f_x^2 + f_y^2]$ (which will be denoted $1/f^2$ in the following). The noise is overlapping with the target. At the center (position B), the target has the same orientation as the one of the reference target. At location A, the target is rotated by an angle of 10 degrees and at location C, the target is rotated by an angle of 5 degrees.

In Fig. 9, the results of the OT filter (Eq. (37)) without regularization are shown. We clearly see that in this case, the correlation is not regular enough to identify the different correlation peaks for rotated targets. This is a consequence of the $1/f^2$ nature of the noise. In Fig. 10, an elastic energy stabilizing functional regularization is considered. The efficiency of this regularization, which is defined in order to stabilize the filter to small variations of the reference object, is clearly seen. Other numerical simulations analyzing the performance of regularized OTSDF filters were published in [19]. It is important to notice that MVSDF and MACE filters are limiting cases of the OTSDF filter and they thus show very unstable behavior in this example.

These two examples clearly illustrate the advantage of using stabilizing functionals when the input image is perturbated.

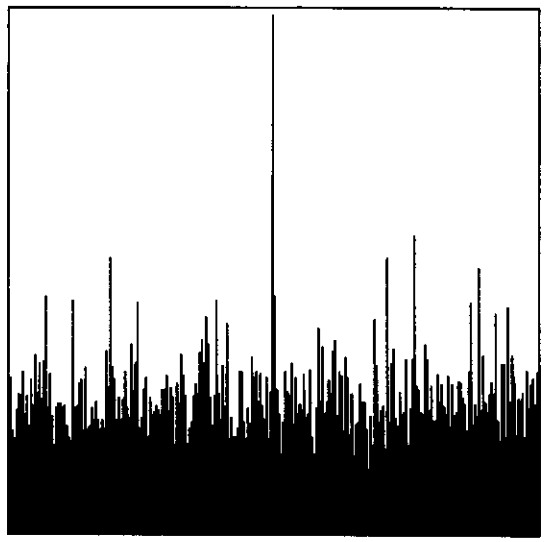


Figure 9: Correlation function with an OT filter, optimized for $1/f^2$ noise and without regularization.

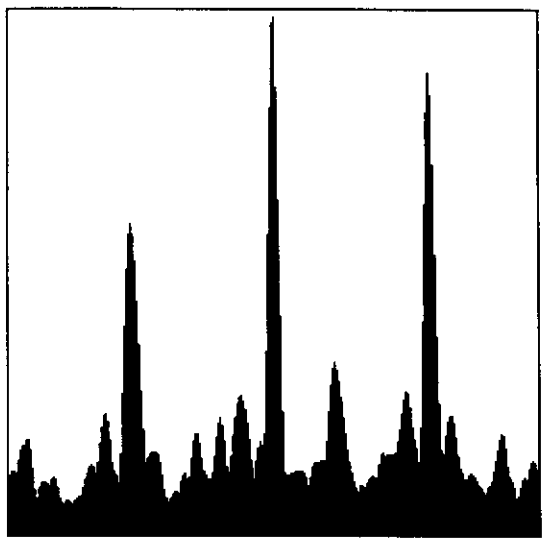


Figure 10: Correlation function with an OT filter, optimized for $1/f^2$ noise and regularized with the minimum elastic energy stabilizing functional.

6 Conclusion of part 1

In this part, we have reviewed heuristic filtering methods for pattern recognition. We have analyzed different criteria which are classically considered for filter design. We have shown that optimizing these criteria (or trade-offs between these criteria) leads to well known filters.

In this context, the importance of the stability of the resulting filters has been analyzed. We have seen that stability is not in general obtained with classical heuristic approaches. We have then proposed two methods of regularization. The first one is the truncature approach, which allowed us to give a new insight in the use of the region of support of phase–only filters. The second method is the stabilizing functional approach, which allows one to introduce more knowledge about the *a priori* expected distortions.

These concepts were illustrated with numerical experiments.

Part II
Statistical theory

7 Introduction

7.1 Filtering techniques and image model

As mentioned in the introduction, the detection and location of a target in a scene is a classical problem, pervasive to many image processing applications. However, the matched filter, as well as the improved linear filtering techniques [20], have been shown to perform poorly on many real-world images [21, 22]. This is because such images often do not belong to the class for which linear filtering is optimal.

In general, in real images, the main source of noise is not the additive detector noise, but the whole background of the scene (clutter), which is nonoverlapping [21] with the target. Secondly, the gray levels of the reference object can be unknown *a priori*. Thirdly, the power spectral density (psd) of the noise is also often unknown *a priori*. In each of these three cases, an important assumption which is necessary to demonstrate the optimality of the matched filter is not fulfilled. As a practical consequence, it has been frequently observed that the matched filter yields poor performance in real applications of optical correlation.

7.2 Examples of noise actually present in real images

Let us discuss more precisely the three cases above, in which one of the assumptions necessary for the optimality of the matched filter is not fulfilled. In particular, let us illustrate the behavior of the linear filters in these cases.

7.2.1 Nonoverlapping noise

If the main source of noise in the input image is not the detector noise, which can be considered as additive, but the whole background of the scene, the noise is non-additive since it does not affect the pixels of the target. The importance of this point has been clearly emphasized by B. Javidi et al. in [21]: They have named this type of noise "nonoverlapping". The difference between additive and nonoverlapping noise is illustrated in Figure 11. In this figure, it is also pointed out that many real-life scenes can be more accurately represented by a nonoverlapping noise model than by an additive noise model.

In presence of nonoverlapping noise, the important task is to discriminate the target with the background clutter, and classical linear filters often fail to correctly locate or detect the target in that case. We show in Figure 12 (bottom line) an example of such situation, in which we can clearly observe that the matched filter, the Optimal Tradeoff [17] and the POF [6] filters, which are all linear filters, are unable to locate the target although it can be seen very easily on the scene image. The same linear filters are efficient (when slightly regularized) in presence of additive noise (Figure 12, top row). We will discuss in the following

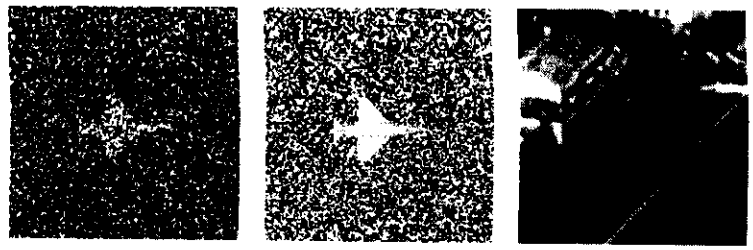


Figure 11: *Left: additive noise. Center : nonoverlapping noise. Right: realistic image ; it can be noticed that the main source of noise is nonoverlapping.*

recent approaches which can solve this problem and in particular the maximum likelihood solution [34].

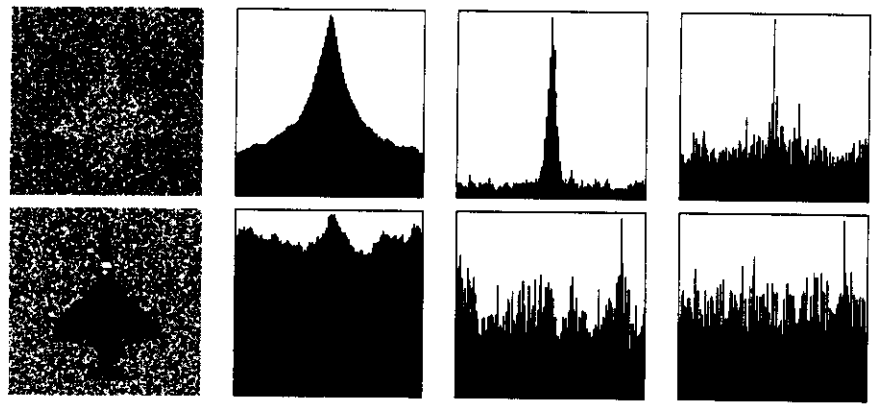


Figure 12: *Top row: scene with additive noise and result of the processing of this scene with the classical matched filter (CMF), an Optimal Tradeoff filter (OT) and a Phase-Only filter (POF) slightly regularized. Bottom row: scene containing the same object (airplane) than above, but corrupted with nonoverlapping noise, and result of the processing of this scene with the same linear filters as above.*
NB: the curves represent the maximum value of each line of the modulus square of the correlation plane obtained with the corresponding filter.

7.2.2 Fluctuations of the target's gray levels

Until now [20], most optimal filtering techniques have assumed that the internal structure of the target, that is, its gray level distribution, is deterministic and known. This is in particular the case of the matched filter, the MACE [16], the Optimal Tradeoff (OT) and the MINACE [28] filters. However, in many applications, this assumption is not realistic. For example, the target may be subject to sun reflections in optical images or temperature

changes in infrared images: the gray levels of the target then fluctuate in an unpredictable way.

In order to illustrate this problem, let us look at Figure 13. The images in the first column are two scenes representing an object with the same shape in presence of nonoverlapping noise. Both scenes are processed with a linear filter (second column) and the maximum likelihood processor of Ref. [34]. The images in the first line are the reference objects utilized to design the corresponding processors. It can be seen that the linear filter fails to locate the target in the scene of the second line, since it is not adapted to nonverlapping noise. However, the processor of Ref. [34] correctly locates the target. In the third line, the scene also contains nonoverlapping noise, but its gray level internal structure is different of that of the reference object utilized to build the processors. It can be seen that the optimal processor for nonoverlapping noise also fails on this image. An optimal maximum likelihood processor adapted to such an image model has been designed in Ref. [55]. Its behavior is represented in the fourth column of Figure 13: it is robust to nonoverlapping noise and to fluctuations of the target's gray levels. The principle of this processor is summarized in Section 8.3.3.

7.2.3 Additive noise with unkown psd

When the noise is additive with a known power spectral density (psd) we will show below that the optimal processor in the sense of decision theory (maximum likelihood) is the well known matched filter. However, in many applications, the psd of the noise is unknown *a priori* and can vary rapidly from one image to another. The consequences of a difference between the actual psd of the noise and the psd used in the filter synthesis have been precisely studied in [18, 36] for Optimal Tradeoff filters and nonlinear filters. It has been shown that the matched filer is in general very sensitive to such difference. It has been demonstrated in [18] that Optimal Tradeoff filters designed with a white noise model are less sensitive to that problem than the matched filter, since they are stable [56] while the matched filter can be unstable. However, Optimal Tradeoff filters are not optimal when the psd of the noise is unknown. It would thus be useful to find out the optimal processor in that case. Section 9 will be devoted to this topic.

8 Background on the statistical decision theory approach

We propose in this section to review the statistical decision theory and to apply it to the target location problem.

8.1 Statistical decision theory without nuisance parameters

Let δ denote the parameter to be estimated. In the particular case of location problems, δ is the coordinates of the target's position in the scene, which will be denoted j in Section 9. In the current section, however, we use a different notation since, at this level, we do not need to particularize the discussion to location estimation. However, in order to make the

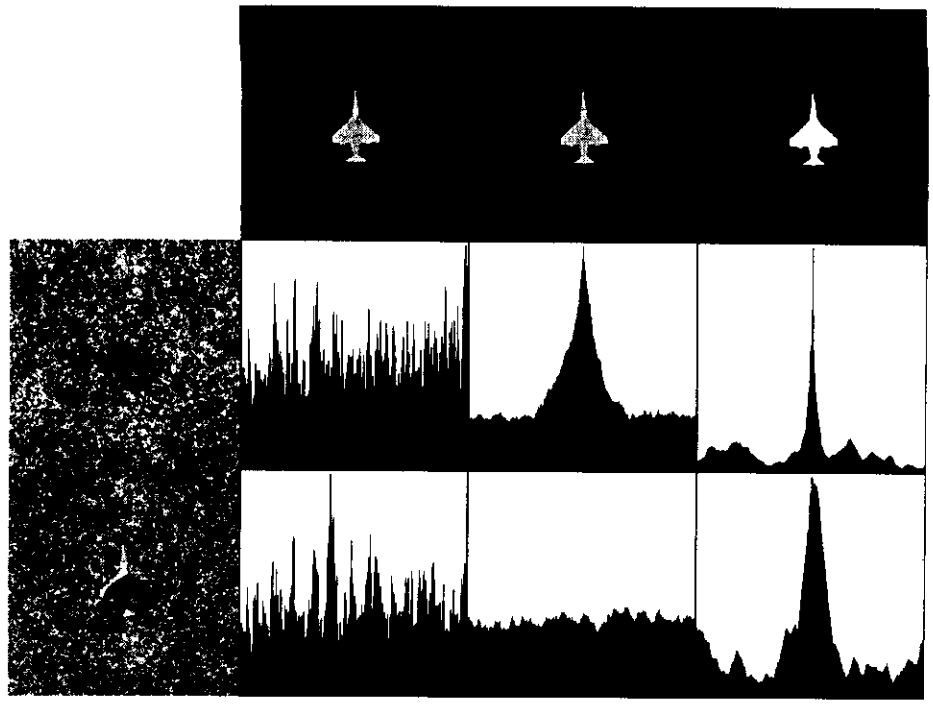


Figure 13: *Behavior of a linear filter, of the maximum likelihood optimal processor for nonverlapping noise of Ref. [34] (cf. Eq. 67), and of the maximum likelihood optimal processor for fluctuating targets of Ref. [55] (cf. Eq. 73).*

discussion easier, we will consider that the parameters δ or j take discrete values. $\mathbf{s}$ will still denote the input image.

The first point to specify in order to apply the decision theory approach is the image formation model. This is indeed an important property of the decision theory to allow one to design an optimal method particularly adapted to each image formation model. It is in fact because of this adaptation that better results can be obtained with decision theory approaches than with heuristic techniques. From a methodological point of view, it is also an important property of the decision theory approach to clearly specify the problem which will be solved.

In order to specify the image formation model, we have to clearly state the relation between the image $\mathbf{s}$, the reference object $\mathbf{r}$ and the noise $\mathbf{n}$ for a given hypothesis on the value of the parameter δ:

$$\mathbf{s} = F_\delta(\mathbf{r}, \mathbf{n}) \tag{42}$$

For example, in the presence of additive noise and for target location, this model is:

$$s_i = r_{i-\delta} + n_i \quad \forall i \in [0; N-1] \tag{43}$$

while with multiplicative noise, it is:

$$s_i = n_i \; r_{i-\delta} \quad \forall i \in [0; N-1] \tag{44}$$

In the general case, the noise is not independent of the target and the general image formation model can be deduced from the conditional probability $P[\mathbf{s}|\mathbf{r}, \delta]$, which represents the probability density of observing $\mathbf{s}$ assuming that the value of the unknown parameter is δ and the gray levels of the reference are $\mathbf{r}$. $P[\mathbf{s}|\mathbf{r}, \delta]$ has a central position in the decision theory and is called *likelihood*. Indeed, it is the likelihood of the hypothesis that the value of the parameter we want to determine is equal to δ, assuming that the target is known and equal to $\mathbf{r}$. For that reason, it will be also denoted $L[\mathbf{s}|\mathbf{r}, \delta]$ in the following.

When the noise is independent of the target's gray level values, the likelihood can be determined very easily from the image model itself. For that purpose, let $P_{\mathbf{n}}(\mathbf{n})$ denote the probability density function (pdf) of the noise and assume that Eq. 42 can be rewritten in the form:

$$\mathbf{n} = F_{\delta}^{-1}(\mathbf{r}, \mathbf{s}) \tag{45}$$

The likelihood is thus:

$$L[\mathbf{s}|\mathbf{r}, \delta] = P_{\mathbf{n}}\left(F_{\delta}^{-1}(\mathbf{r}, \mathbf{s})\right) \tag{46}$$

In the following, we will consider a general image formation model for which only the shape $\mathbf{w}$ of the target is assumed to be known, but not its gray levels $\mathbf{r}$. In that case, the likelihood will be written: $L[\mathbf{s}|\mathbf{w}, \delta]$. We can now describe the Maximum Likelihood (ML) principle which is frequently used in statistical inference problems.

8.1.1 Maximum Likelihood principle

The maximum likelihood estimate of the unknown parameter δ is the value which maximizes $L[\mathbf{s}|\mathbf{p}, \delta]$ where $\mathbf{p}$ is any assumed known parameter. The parameter $\mathbf{p}$ can be for example $\mathbf{r}$ or $\mathbf{w}$. The maximum likelihood estimate of δ can thus be written:

$$\delta^{ML} = \underset{\delta}{argmax} \; L[\mathbf{s}|\mathbf{p}, \delta] \tag{47}$$

Since the likelihood $L[\mathbf{s}|\mathbf{p}, \delta]$ is the probability of observing $\mathbf{s}$ under the assumption that the value of the unknown parameter is δ, the basic physical idea of the ML principle is to choose the value of δ which makes very probable (or likely) the data we have observed. This is a very reasonable approach but it can also appear arbitrary at this level.

8.1.2 Maximum a Posteriori principle

In fact, it is well known [57] that if we want to minimize the probability of wrong estimation of the value of δ and maximize the probability of true estimation, we have to consider the Maximum a Posteriori (MAP) estimation obtained from $P[\delta|\mathbf{s}, \mathbf{p}]$. We thus obtain the MAP estimation principle which is also frequently used in statistical inference problems. The MAP

estimation of the unknown parameter δ is the value which maximizes $P[\delta|\mathbf{s},\mathbf{p}]$ where $\mathbf{p}$ is any assumed known parameter. The MAP estimation of δ can thus be written:

$$\delta^{MAP} = \underset{\delta}{argmax}\ P[\delta|\mathbf{s},\mathbf{p}] \tag{48}$$

The probability density used for the MAP estimate is directly related to the likelihood through the Bayes relation. Indeed, we have:

$$P[\delta|\mathbf{s},\mathbf{p}] = \frac{L[\mathbf{s}|\mathbf{p},\delta]P[\delta]}{P[\mathbf{s}]} \tag{49}$$

$P[\delta]$ is a prior probability density for the values of the unknown parameter δ, it is thus frequently simply denominated *prior*. For example, when tracking a target in an image sequence, this prior can be obtained from the previous estimated location using a Kalman filter [58].

The MAP estimate is easily obtained from the likelihood and the prior:

$$\delta^{MAP} = \underset{\delta}{argmax}\ [L[\mathbf{s}|\mathbf{p},\delta]P[\delta]] \tag{50}$$

If we introduce the loglikelihood:

$$\ell[\mathbf{s}|\mathbf{p},\delta] = \log\left[L[\mathbf{s}|\mathbf{p},\delta]\right] \tag{51}$$

the MAP estimate becomes:

$$\delta^{MAP} = \underset{\delta}{argmax}\ [\ell[\mathbf{s}|\mathbf{p},\delta] + \log\left[P[\delta]\right]] \tag{52}$$

which clearly shows that when the loglikelihood is determined, it is in general simple to obtain the MAP estimate. We will thus consider in the following ML estimation of the parameter δ.

8.2 Decision theory in presence of nuisance parameters

In the general case, the conditional probability $P[\mathbf{s}|\mathbf{r},\delta]$ is not precisely known. It is possible to consider a parametric model for this function, which depends on unknown parameters (denoted μ in the following). In that case, it must be written: $P[\mathbf{s}|\mu,\mathbf{p},\delta]$ since it can be evaluated only for some *a priori* assumed values of μ. The parameter μ is generally a nuisance parameter since we are not interested in its value. A classical example of such situation is the location of a target with unknown illumination β and in the presence of additive noise:

$$s_i = \beta\ r_{i-\delta} + n_i \quad \forall i \in [0; N-1] \tag{53}$$

In that example, the nuisance parameter is the scalar value β. We will consider more complex examples in the following sections.

There exist several methods to deal with nuisance parameters. We propose to briefly describe three of them, which are frequently used and which will be useful in the following developments.

8.2.1 ML estimation of the nuisance parameter

In the first approach, we determine the ML estimate of the nuisance parameter and then insert the value of this estimate in the expression of the likelihood.

The ML estimation of the nuisance parameter μ is determined by:

$$\mu^{ML}(\delta) = \underset{\mu}{argmax}\ L[\mathbf{s}|\mu, \mathbf{p}, \delta] \tag{54}$$

The ML estimate of δ can thus be written:

$$\delta_{ML}^{ML} = \underset{\delta}{argmax}\ L[\mathbf{s}|\mu^{ML}(\delta), \mathbf{p}, \delta] \tag{55}$$

In practice, the maximum likelihood estimate often appears to be unstable. The likelihood is said to be unstable [56] if a small variation of the initial conditions (here $\mathbf{s} \rightarrow \mathbf{s} + d\mathbf{s}$) can lead to a large variation of $L[\mathbf{s}|\mu^{ML}(\delta), \mathbf{p}, \delta]$ [1]. The MAP approach can be used to overcome this problem.

8.2.2 MAP estimation of the nuisance parameter

The MAP approach is interesting when some values of the nuisance parameter are very unlikely or lead to unstable estimate of the likelihood and thus of the parameter δ. The MAP estimate of the nuisance parameter μ is determined by:

$$\mu^{MAP}(\delta) = \underset{\mu}{argmax}\ [L[\mu|\mathbf{s}, \mathbf{p}, \delta]P[\mu]] \tag{56}$$

The ML estimate of δ can thus be written:

$$\delta_{MAP}^{ML} = \underset{\delta}{argmax}\ \left[L\left[\mathbf{s}|\mu^{MAP}(\delta), \mathbf{p}, \delta\right] P\left[\mu^{MAP}(\delta)\right]\right] \tag{57}$$

The prior $P[\mu]$ is thus useful to penalize undesirable values of μ. It is very easy to apply this approach when computing the MAP estimate of δ, which is:

$$\delta_{MAP}^{MAP} = \arg\max_{\delta} \left[L\left[\mathbf{s}|\mu^{MAP}(\delta), \mathbf{p}, \delta\right]\ P\left[\mu^{MAP}(\delta)\right]\ P[\delta]\right] \tag{58}$$

It is clear that if the prior $P[\mu]$ is uniform, the MAP approach is equivalent to the ML method.

[1] One can also consider stability against reference variation $\mathbf{r} \rightarrow \mathbf{r} + d\mathbf{r}$ as in part 1. Both approaches should be analyzed but for statistical models considered here stability requirement against input image variation implies stability against reference variation.

8.2.3 Bayesian approach

The Bayesian method has been the subject of many investigations and theoretical studies [59]. Here, we will not discuss this approach in detail since critical analyses can be found in references [57, 59] for example. The motivation for this approach for the applications considered here are the same as for the MAP [2].

In the Bayesian approach, the ML estimate of δ is obtained with:

$$\delta_{Bay}^{ML} = \underset{\delta}{argmax} \int L[\mathbf{s}|\mu, \mathbf{p}, \delta] P[\mu] d\mu \tag{59}$$

where, when μ is a real vector of dimension n. $\int F[\mu]d\mu$ is a formal notation for: $\int_{\mu_1} \int_{\mu_2} \cdots \int_{\mu_n} F[\mu]d\mu_1 d\mu_2 ... d\mu_n$.

8.2.4 Conclusion

With both the MAP or the Bayesian approaches, the value of the estimate depends on the expression of the prior $P[\mu]$. This point is sometimes considered as a limitation of these methods [59]. However, it should be noted that our goal is mainly to suppress the cases where the likelihood is unstable. Thus, from a practical point of view, the choice of the expression of $P[\mu]$ generally results from the following considerations:

1. to penalize the undesirable values of μ,
2. to lead to MAP or Bayesian estimates that are close to the ML estimate for the values of the nuisance parameter for which the likelihood is stable,
3. to lead to tractable mathematical equations.

In Section 9, we propose to illustrate these approaches in a particular case for which the psd of the noise is unknown *a priori*, and can thus be considered as a nuisance parameter. However, let us first rapidly discuss some classical and more recent results obtained with the ML method in the field of optical correlation.

8.3 Examples of location problems and statistical decision theory

In this section, we first review the well known case of the matched filter. We then describe two more recent approaches that utilize decision theory in order to solve original target location problems.

[2] One can show with the risk theory [57, 59] that with the bayesian method the nuisance parameter are really considered of no interest while in the MAP approach they are in fact considered as parameter of interest.

8.3.1 Matched filter

The correlation operation allows one to compare a reference pattern with an input scene. This comparison can be useful not only for detection or classification of a target, but also to estimate a parameter (for example the attitude) or to locate this target in an input image. In the following, we will speak about classification for these general tasks.

We propose to review this classical problem of signal processing when the input noise on the image (which represents uncertainty) is additive, gaussian, stationary and with a kown spectral density. We consider in this classical problem the ML estimations of the nuisance parameters.

Let $\mathbf{s}$ denote the input image and $\mathbf{r}^\delta$ the references. The references are dependent on a parameter δ. Let us consider some examples. If we want to locate a reference $\mathbf{r}$ in the input image, δ denotes the different possible locations of the target. In that case $r_i^\delta = r_{i-\delta}$. If we try to estimate the angle of a rotation around a given axis, $\mathbf{r}^\delta = R_\delta \mathbf{r}$ where R_δ is the rotation operator with an angle δ. If the problem is a detection problem, there are two possibilities: $\delta = 0$ or $\delta = 1$. $\delta = 1$ corresponds to the hypothesis that the target is present and $\delta = 0$ that it is not present. More generally, if P types of target are possible (for example P different objects or P different attitudes of the object) the problem is a classification one and δ can be a label for the type of the targets ($\delta = 1, 2, ..., P$).

For invariant recognition, we are not interested in the particular value of δ but we try to know if δ belongs to the set of objects to be recognized. This last problem clearly shows strong analogies with the problem of invariance analyzed in part 1 and for which we introduced the SDF approach.

In realistic applications, the problem is a mixture of these different problematics and δ is a vector. We consider for the matched filter design that the input image $\mathbf{s}$ is the addition of a noise and of the target with an unknown value of δ and, in general, with an unknown illumination β [3]. Let $\gamma[\delta, \beta]$ denote the hypothesis that the input target is in the input image $\mathbf{s}$ with parameter δ and illumination β. Under hypotheses $\gamma[\delta, \beta]$, we can write:

$$\mathbf{s} = \beta\ \mathbf{r}^\delta + \mathbf{n} \tag{60}$$

where $\mathbf{n}$ is a gaussian additive noise with a known spectral density $\hat{\mathbf{S}}$.

The probability density function $P_n(\mathbf{n})$ of the noise can be written:

$$P_n(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}}\ exp[-\frac{1}{2}\mathbf{n}^T S^{-1}\mathbf{n}], \tag{61}$$

where S is the noise covariance matrix (which is a Toeplitz matrix [60] since the noise is assumed stationary), and $|S|$ is its determinant. Furthermore, cyclic boundary conditions are assumed. In other words, $S_{|i-j|} = S_{|i-j+pN|}$ for any integer p and S is a circulant Toeplitz matrix [60]. In that case, S is diagonal in the Fourier domain with value equal to the spectral density $\hat{\mathbf{S}}$.

We can write:

$$P(\mathbf{s}|\gamma[\delta, \beta]) = \frac{1}{\sqrt{(2\pi)^N |S|}}\ exp[-\frac{1}{2}(\mathbf{s} - \beta\ \mathbf{r}^\delta)^T S^{-1}(\mathbf{s} - \beta\ \mathbf{r}^\delta)]. \tag{62}$$

[3] The parameter β is introduced with its specific notation since its ML estimation is very easy.

The ML estimation of δ and β is obtained by maximizing the likelihood 62. With our assumption of additive gaussian noise, we obtain:

$$\beta^{ML} = \frac{[\mathbf{r}^\delta]^T S^{-1}\mathbf{s}}{[\mathbf{r}^\delta]^T S^{-1}\mathbf{r}^\delta} \qquad \text{and:} \qquad \delta^{ML} = \underset{\delta}{argmax}\ \frac{|[\mathbf{r}^\delta]^T S^{-1}\mathbf{s}|^2}{[\mathbf{r}^\delta]^T S^{-1}\mathbf{r}^\delta} \tag{63}$$

In order to be somewhat more concrete, let us assume that the target can belong to one of P possible classes. Furthermore, let us consider that this recognition task has to be performed with translation invariance.

The parameter δ is now composed of a value for the class ℓ and a value p for the location, $\delta = (\ell, p)$. The optimal estimation of ℓ and p is given by:

$$(p^{ML},\ \ell^{ML}) = \underset{\ell,\ p}{argmax}\ \frac{|\sum_{i,j=1}^{N} r_{i-p}^{\ell}[S^{-1}]_{i,j}s_j|^2}{\sum_{i,j=1}^{N} r_{i-p}^{\ell}[S^{-1}]_{i,j}r_{j-p}^{\ell}} \tag{64}$$

Since the noise is assumed stationary with cyclic boundary conditions, it holds $[S]_{i,j}^{-1} = [S]_{i-p,j-p}^{-1}$, and thus:

$$\sum_{i,j=1}^{N} r_{i-p}^{\ell}[S^{-1}]_{i,j}r_{j-p}^{\ell} = \sum_{i,j=1}^{N} r_i^{\ell}[S^{-1}]_{i,j}r_j^{\ell}$$

Furthermore, the expression of Eq. (64) can be simplified if the matched filter is introduced:

$$h_j^{\ell} = \sum_{i=1}^{N} r_i^{\ell}[S^{-1}]_{i,j}$$

Thus, Eq. (64) becomes:

$$\ell^{ML} = \underset{\ell}{argmax}\ C_{max}[\ell]\ /\ A[\ell] \tag{65}$$

with:

$$C_{max}[\ell] = \underset{p}{max}\ |\sum_{j=1}^{N} h_{j-p}^{\ell}s_j|^2$$

and:

$$A[\ell] = \sum_{i,j=1}^{N} r_i^{\ell}[S^{-1}]_{i,j}r_j^{\ell}$$

$C_{max}[\ell]$ is the maximum value in the correlation plane of the modulus square of the correlation function between $\mathbf{s}$ and $\mathbf{h}^{\ell}$. The ML classification is thus performed in two steps. In the first step the maximum value of the modulus square of the correlation function between $\mathbf{s}$ and $\mathbf{h}^{\ell}$ is determined for each ℓ. In the second step, the class ℓ which maximizes over all the hypotheses this value of the modulus square of the correlation correctly normalized (i.e. $C_{max}[\ell]/A[\ell]$) is chosen.

This is an important theoretical approach to justify the use of linear filtering techniques (or correlation techniques) for pattern recognition.

Of course, the first limit of this approach with images is the assumption of additive gaussian noise. In particular, the assumption of additive noise is not completely realistic in image processing as discussed in [21]. With non–overlapping noise, different solutions have been proposed in [34] and [35]. The second important point is that this method assumes that the spectral density of the noise $\hat{S}$ is known. This is not the case in general in image processing, where in contrast to radar processing, it is very difficult to estimate it. Then, an important question in the context of pattern recognition is the determination of an appropriate model for $\hat{S}$. Furthermore, as mentioned in part 1, there is no reason to consider that the realizations of noise are obtained with a temporal stationary density probability law.

It appears also clearly that if the number of classes P is large, determining the correlation function with all the $\mathbf{h}^\ell$ can be prohibitive in terms of memory in order to store all the $\mathbf{h}^\ell$, and in terms of computational power to determine the correlations. This problem is an important motivation for using SDF filters as noted in part 1, even though they are sub–optimal in the statistical decision theory framework.

8.3.2 Deterministic target and nonoverlapping noise

This model has been developed in order to describe situations in which the main source of noise is the background clutter. In that case we introduce the support function $\mathbf{w}$ of the target $\mathbf{r}$ which is defined by setting w_i is equal to 1 within the support of the target (i.e. where $r_i \neq 0$) and is equal to 0 elsewhere. The nonoverlapping noise is equal to zero in the support function of the target while it can have a positive mean value outside this support function. This background noise is thus represented by $(1 - w_{i-\delta})\ b_i$ where b_i are the gray levels of the background. The image formation model is:

$$s_i = \beta\ r_{i-\delta} + (1 - w_{i-\delta})\ b_i + n_i \quad \forall i \in [0; N-1] \tag{66}$$

where the target is assumed to have known internal gray level structure r_i, the noise n_i is independent of the target and is assumed to be white, Gaussian and additive, and β is a nuisance parameter. It has been shown in [34] that when the power of n_i tends to 0, the ML estimate of the location is obtained by maximizing

$$\delta_{ML}^{ML} = \arg\max_{\delta} \left[-\sum_i w_{i-\delta}\ s_i^2 + \frac{\left[\sum_i r_{i-\delta} s_i\right]^2}{\sum_i r_i^2} \right] \tag{67}$$

In that case, the nuisance parameter β has also been estimated with the ML approach.

In fact, we show in appendix C that the processor of Eq. 67 is optimal in the maximum likelihood sense for any value of the variance of n_i within the target support. More precisely, we demonstrate that if the values b_i are considered as nuisance parameters, the ML estimation method described in Section 8.2.1 for the b_i leads to the processor of Eq. 67. This result is important, since it precisely defines the domain of optimality of this processor.

8.3.3 White random target and nonoverlapping noise

The previous model can lead to very interesting results. However, only a global fluctuation of the target's gray levels is taken into account with the model of Eq.66, through the parameter β: the gray level structure r_i of the object is assumed to be precisely known. In some practical situations, this assumption is not realistic. This can happen, for example, when the gray levels of the reference object are not known at the moment where the reference object is defined. Moreover, these gray levels can fluctuate rapidly from one image to another (due to sun reflections for example). In order to describe such situations, one can consider [61] a new image formation model:

$$s_i = (w_{i-\delta})\ a_i + (1 - w_{i-\delta})\ b_i \quad \forall i \in [0; N-1] \tag{68}$$

where a_i and b_i are two white Gaussian random fields. They are considered as statistically independent, so that this model has been named *Statistically Independent Region* (SIR) image model. The corresponding ML or MAP optimal location processors are thus denoted SIR *processors.*

Since a white Gaussian random field is described by its mean and its variance, there are four nuisance parameters in the model (the means and the variances of each random field, i.e. m_a, m_b, σ_a and σ_b). It can be found that the ML estimates of these parameters are the well known empirical estimates of the mean and the variance of a data set [61]:

$$m_a^{ML}(\delta) = \sum_i (w_{i-\delta}) s_i \tag{69}$$

$$m_b^{ML}(\delta) = \sum_i (1 - w_{i-\delta}) s_i \tag{70}$$

$$\sigma_a^{ML}(\delta) = \sum_i (w_{i-\delta}) \left[s_i - m_a^{ML}(\delta)\right]^2 \tag{71}$$

$$\sigma_b^{ML}(\delta) = \sum_i (1 - w_{i-\delta}) \left[s_i - m_b^{ML}(\delta)\right]^2 \tag{72}$$

The ML estimation of the location is thus obtained by:

$$\delta_{ML}^{ML} = \arg\max_{\delta} \left[N_w\ \log\left[\sigma_a^{ML}(\delta)\right] + (N - N_w)\ \log\left[\sigma_b^{ML}(\delta)\right]\right] \tag{73}$$

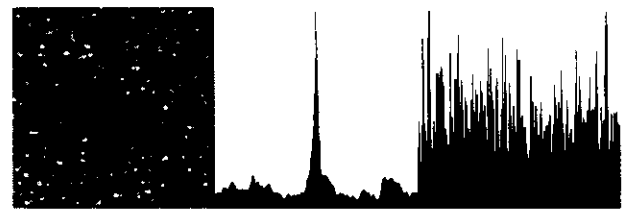
where N_w is the number of pixels in the support function ($\mathbf{w}$) of the target.

It has been shown that this processor can solve problems that cannot be solved by linear filters. Figures 14 and 15 show examples of such situations. Moreover, a similar algorithm can be designed for images with χ^2 statistics [62].

However, the processor of Eq. 73 is designed to be optimal for white (uncorrelated) statistics of the target and of the background. Experience shows that its performance decreases when the statisitics become correlated, and when their correlation lengths increase. An example of such situation can be found in Figure 16.

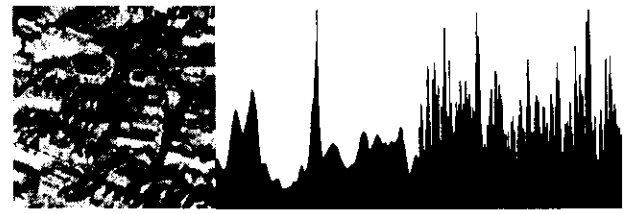
8.3.4 Correlated random target and nonoverlapping noise

In order to overcome the problem of correlated statistics, several strategies may be envisaged. One of them is to model the vectors $\mathbf{a}$ and $\mathbf{b}$ of the SIR image model (cf. Eq. 68) with a



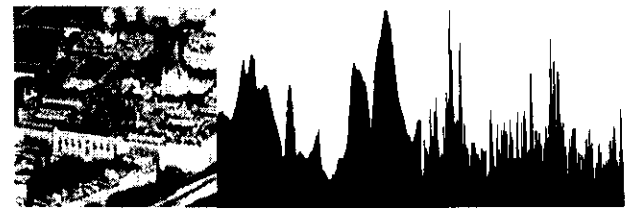
(a) (b) (c)

Figure 14: **(a)** *Scene image.* **(b)** *Result of the processing of (a) with the optimal SIR processor of Eq. 73.* **(c)** *Result of the processing of (a) with an Optimal Tradeoff filter. The matched filter [32], the POF [6] and other linear filters also fail to detect the target. NB:* **(b)** *and* **(c)** *are plots of the maximum of each line of the correlation plane.*



(a) (b) (c)

Figure 15: *Same legend as Figure 14.*



(a) (b) (c)

Figure 16: *Same legend as Figure 14.*

correlated random field. Markov Random Fields (MRF) are well adapted to this aim [63]. Let us assume that the statistics of both the target and the background are represented by the same MRF, which has the following expression:

$$P(\mathbf{x}) = A\ exp\left[-\left(\sum_{i=1}^{N}\sum_{j\in V_i}(x_i - x_j)^2 + g(x_i)\right)/T\right] \tag{74}$$

where A is a normalization constant, T is a parameter and $g(x_i)$ is the realization of a random function which depends only on x_i. It can be shown [63] that the ML optimal processor is:

$$\delta^{ML} = \arg\max_{\delta}\left[\sum_{i=0}^{N-1}\Delta w^h_{i-\delta}\ s^h_i + \sum_{i=0}^{N-1}\Delta w^v_{i-\delta}\ s^v_i\right] \tag{75}$$

For each pixel i, i^h denotes its horizontal right neighbor and i^v its vertical top neighbor. The two images $\mathbf{s^h}$ and $\mathbf{s^v}$ are defined in the following way

$$s^h_i = (s_i - s_{i^h})^2 \quad \text{and} \quad s^v_i = (s_i - s_{i^v})^2 \tag{76}$$

These images contain the modulus square of the gradient of $\mathbf{s}$ respectively in the horizontal right and in the vertical top directions. $\Delta\mathbf{w^h}$ and $\Delta\mathbf{w^v}$ are two edge images of the support of the reference object:

$$\Delta w^h_{i+k} = [(w_k)_i - (w_k)_{i^h}]^2 \quad \text{and} \quad \Delta w^v_{i+k} = [(w_k)_i - (w_k)_{i^v}]^2 \tag{77}$$

An illustration of the performance of this processor on a realistic image is represented in Figure 17. It can be shown that the processor of Eq. 75 is not optimal in presence of white statistics, and that its performance increases as the image statistic gets more correlated. Its behaviour is thus inverse of that of the ML SIR processor for white statistics (cf. Eq. 73). This can be seen very clearly on Figure 18: for small correlation lengths of the scene, the optimal SIR processor for white statistics is more efficient, whereas for large correlation lengths, the optimal SIR processor for Random Markov Fields is more adapted. Consequently, these two processors are complementary, so that it would be interesting to utilize them at the same time on the same image, and to perform a *fusion* of their results. A global processor based on this principle would be efficient for a wider class of images than the two sub-processors used separately.

8.3.5 Whitening preprocessing and Statistically Independent Region processor

In some real images, as mentioned in setion 8.3.4, the statistics of both the target and the background cannot be approximated with good precision by uncorrelated random fields. In these situations, the SIR processor is then suboptimal and can fail. In Fig.19, two scenes and their respective output planes obtained with the SIR processor are shown. The maximum of the output plane represents the estimated location of the target. On scene (b), the probability density functions (pdf's) of both the target and the background are white and Gaussian whereas those of scene (c) are also Gaussian but correlated. One can note that

(a) (b) (c)

Figure 17: **(a)** *Reference object utilized to design the location processor.* **(b)** *Scene.* **(c)** *Result of the processing of (b) with the ML processor for RMF correlated statistics of Eq. 75.*

even if the target is easier to see on scene (c) than on scene (b), the SIR processor fails on scene (c) whereas it is able to locate the target on scene (b).

Up to now, it seems difficult to design a general optimal processor for the location of a random correlated target appearing on a random correlated background. The main problem consists in finding simple models for the target and for the background that characterize the different possible situations and that allow one to design an optimal processor in the maximum likelihood sense for these models. The processor designed in section 8.3.4 assumes that the textures are described by the same random Markov field model for both the target and the background. However, although that model represents a difficult case, it corresponds to a particular situation.

We analyze here another practical approach which allows one to satisfy the optimality conditions in order to apply the SIR processor. We define the nonlinear whitening filter in the Fourier domain by :

$$\hat{h}_k = \frac{1}{|\hat{s}_k|} = \frac{1}{\sqrt{\hat{s}_k^* \hat{s}_k} + \mu} \tag{78}$$

where μ is a small positive parameter which allows one to obtain stable filters (see part I). The Fourier Transform $\hat{\mathbf{z}}$ of the preprocessed image $\mathbf{z}$ is thus defined by :

$$\hat{z}_k = \hat{h}_k \hat{s}_k. \tag{79}$$

One can note that since $\mathbf{s}$ and $\hat{\mathbf{h}}$ are real, $\mathbf{z}$ is also real. It is easy to show that the square modulus of $\hat{\mathbf{z}}$ is approximtely constant. Thus we can conjecture that the pixel values of the preprocessed image $\mathbf{z}$ are uncorrelated variables. In Fig.20, we show a target with a correlated texture which appears on a random correlated background, and the obtained preprocessed image. The impulse response of the preprocessing filter ($\hat{h}_k$ in the Fourier domain) and the histogram of the preprocessed background (which is similar to the histogram of the preprocessed target) are also represented. One can note that describing the pixel values of the preprocessed image as Gaussian random variables is a good approximation. If we model the preprocessed image with two independent regions, according to [61], the 2-SIR processor can be defined by :

$$F_j^{(1)} = -N_w \log[\hat{\sigma}^2(j)] - (N - N_w) \log[\hat{\bar{\sigma}}^2(j)] \tag{80}$$

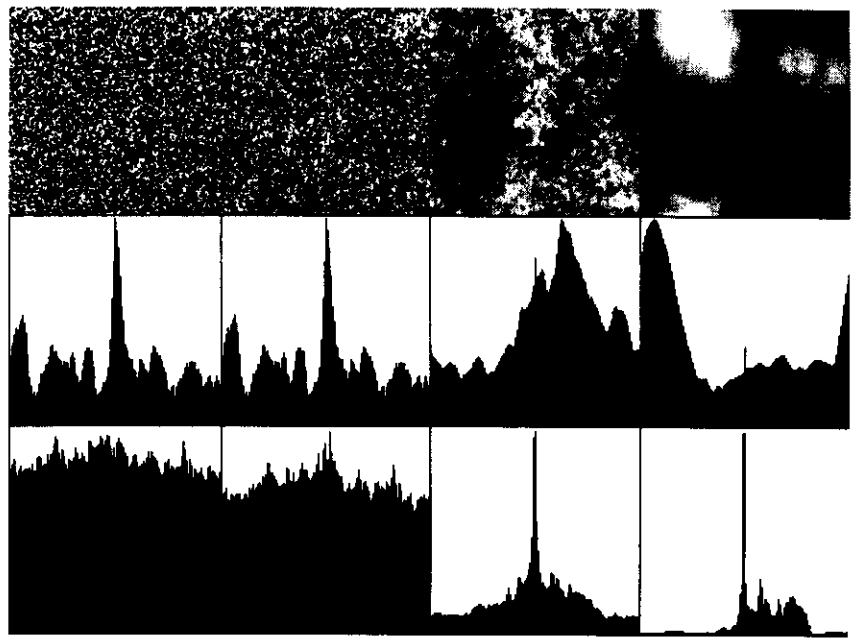


Figure 18: *First line: scene where the target's and the background's gray levels are noises with increasing correlation lengths. Second line: Result of the application of the SIR processor optimal for white statistics (cf. Eq. 73) to the corresponding scene. Third line: Result of the application of the SIR processor optimal for Random Markov Fields (cf. Eq. 75) to the corresponding scene.*

where $\hat{\sigma}^2(j)$ and $\hat{\tilde{\sigma}}^2(j)$ are computed from the preprocessed image $\mathbf{z}$ as follow:

$$\hat{\sigma}^2(j) = \frac{1}{N_w}[\mathbf{z}^2 \star \mathbf{w}]_j - \frac{1}{N_w^2}[\mathbf{z} \star \mathbf{w}]_j^2 \tag{81}$$

$$\hat{\tilde{\sigma}}^2(j) = \frac{1}{N - N_w}\left(\sum_{i=1}^{N} {z_i}^2 - [\mathbf{z}^2 \star \mathbf{w}]_j\right) - \frac{1}{(N - N_w)^2}\left(\sum_{i=1}^{N} z_i - [\mathbf{z} \star \mathbf{w}]_j\right)^2 \tag{82}$$

and

$$\mathbf{z}^2 = \left\{ {z_i}^2 | i \in [1, N] \right\} \tag{83}$$

However, as one can remark in Fig. 20, the preprocessing can introduce three regions in the preprocessed image. Indeed, when the textures are strongly correlated, a boundary region appears in the preprocessed image between the target and the background.

We have determined the width of that boundary from the width of the impulse response of the preprocessing filter $\mathbf{h}$ (see Fig. 20). Let $\mathbf{f}$, $\mathbf{t}$ and $\mathbf{b}$ denote respectively the boundary, the target and the background composed of N_f, N_t and N_b pixels, and let $\mathbf{w^f}$, $\mathbf{w^t}$ and $\mathbf{w^b}$ define respectively the shapes of the boundary, the target and the background reference centered on the pixel 0 so that ${w_i}^f$ (respectively ${w_i}^t$ and ${w_i}^b$) is equal to one within the boundary (respectively the target and the background) reference and to zero elsewhere. In

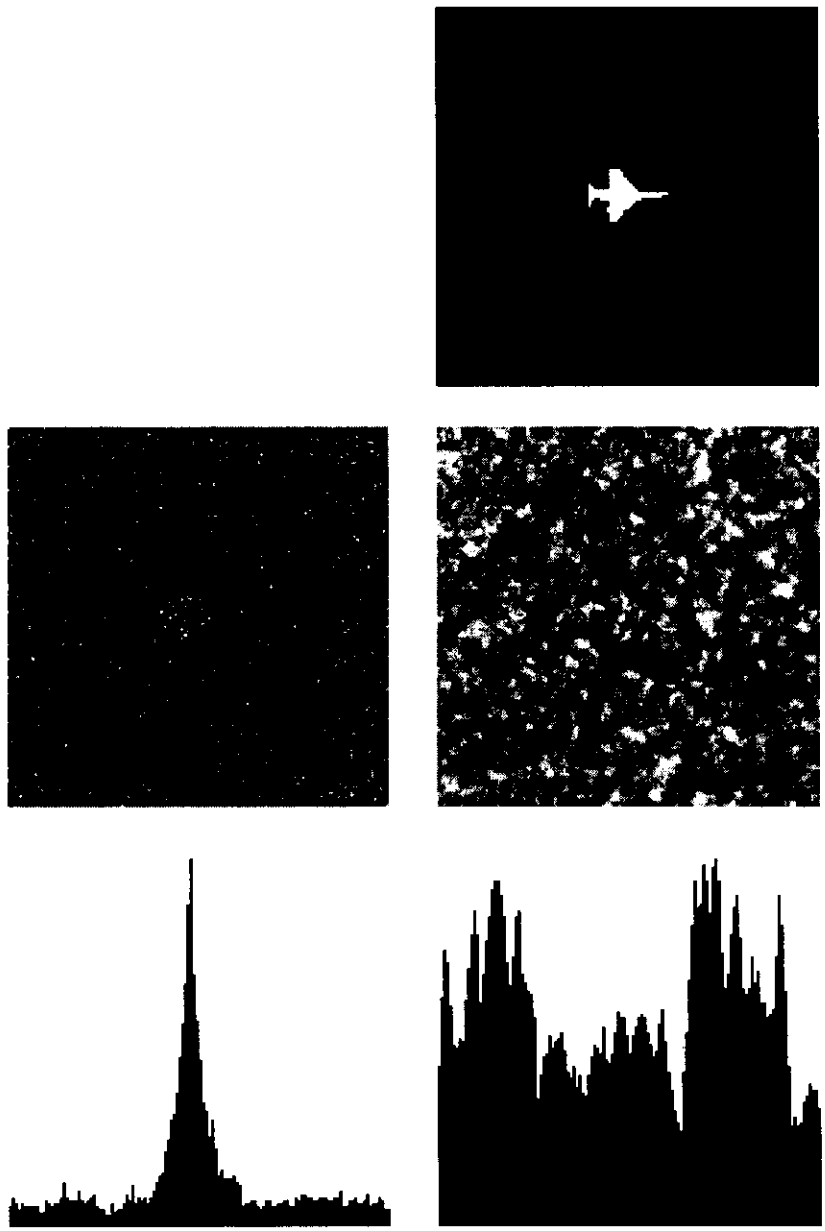


Figure 19: *(a) Binary support of the reference. (b) Example of a scene with white gaussian textures for both the target and the background. The texture parameters are* $m_v = 0$ *and* $\sigma_v = 1$ *for the background and* $m_u = 0.5$ *and* $\sigma_u = 1.5$ *for the target. (c) Example of a scene with correlated gaussian textures for both the target and the background. The texture parameters are* $m_v = 0$, $\sigma_v = 1$ *and* $l_v = 1$ *for the background and* $m_u = 0.5$, $\sigma_u = 1.5$ *and* $l_u = 1$ *for the target. (d) Result of the processing of (b) with the 2-SIR processor. (e) Result of the processing of (c) with the 2-SIR processor.*
Note : (d) and (e) are plots of the maximum of each line of the output plane of the considered method.

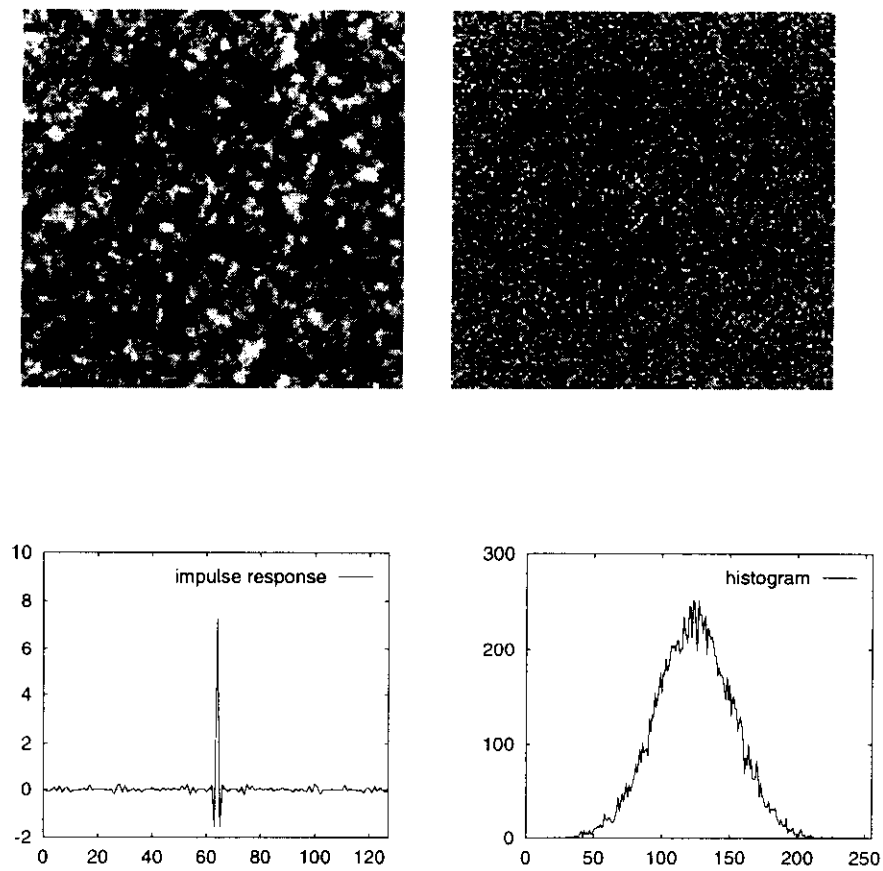


Figure 20: *(a) Example of a scene with correlated gaussian textures for both the target and the background. The pdf parameters are* $m_v = 0$, $\sigma_v = 1$ *and* $l_v = 1$ *for the background and* $m_u = 0.5$, $\sigma_u = 1.5$ *and* $l_u = 1$ *for the target. (b) Preprocessed image of (a) with the whitening filter. (c) Impulse response of the whitening preprocessing obtained with (a). (d) Histogram of the background of (b).*

that situation, we thus propose to describe the preprocessed image $\mathbf{z}$ using a 3-SIR model as follow:

$$z_i = t_i w^t_{i-j} + f_i w^f_{i-j} + b_i w^b_{i-j} \tag{84}$$

when the target is supposed to be centered on the j^{th} pixel of the image.

With an approach analogous to the one in 8.3.3 and [61], we can design a 3-SIR processor that takes into account three regions. This leads to :

$$F^{(2)}_j = -N_t \log[\hat{\sigma}^2_t(j)] - N_f \log[\hat{\sigma}^2_f(j)] - N_b \log[\hat{\sigma}^2_b(j)] \tag{85}$$

with :

$$\hat{\sigma}^2_t(j) = \frac{1}{N_t}[\mathbf{z}^2 \star \mathbf{w}^t]_j - \frac{1}{N_t^2}[\mathbf{z} \star \mathbf{w}^t]^2_j \tag{86}$$

$$\hat{\sigma}^2_f(j) = \frac{1}{N_f}[\mathbf{z}^2 \star \mathbf{w}^f]_j - \frac{1}{N_f^2}[\mathbf{z} \star \mathbf{w}^f]^2_j \tag{87}$$

$$\hat{\sigma}^2_b(j) = \frac{1}{N_b}[\mathbf{z}^2 \star \mathbf{w}^b]_j - \frac{1}{N_b^2}[\mathbf{z} \star \mathbf{w}^b]^2_j \tag{88}$$

$\hat{\sigma}^2_t(j)$, $\hat{\sigma}^2_f(j)$ and $\hat{\sigma}^2_b(j)$ are respectively the estimated variances of the target, the boundary and the background of the whitened image when the target is supposed to be centered on the j^{th} pixel of the image. These quantities can be determined by correlating the images $\mathbf{z}$ and $\mathbf{z}^2$ with binary masks. They can be obtained with a simple optoelectronical architecture or using FFT algorithm applied to the images $\mathbf{z}$ and $\mathbf{z}^2$.

To illustrate the domain of application of the 3-SIR processor, we finally show in Fig. 21, a realistic scene where the target appears on a structured background. The shape of the target is the same as Fig. 1 and the three windows are defined with a structured element of size 3×3 pixels. We can observe in the figure that the 3-SIR processor and the proposed preprocessing are efficient in the considered case. It can also be shown that the 2-SIR processor succeeds in locating the target on the whitened image whereas it fails when it is applied directly to the input image.

8.4 Conclusion

In conclusion of these examples, one can see that the statistical decision theory approach is a very powerful tool in order to design optimal processors adapted to different practical situations. There exist many other examples, but it is not our purpose to detail them here. One can however remark with all previous examples that the optimal solutions are rarely linear correlations. Only the matched filtering technique is a linear correlation followed by the selection of the maximum value of the modulus square of the correlation plane. However, it is easy to show that with the three above optimal solutions, the most intensive computations which are needed are linear correlations. All the other computations can be considered as low-complexity preprocessings of the input image and postprocessings of the correlation planes.

In the following section, we propose to discuss in more detail another image formation model, which leads to nonlinear correlations that are analogous to the ones which can be performed by a NLJTC. Moreover, this example will provide a more detailed illustration of the different approaches for the estimation of the nuisance parameters.

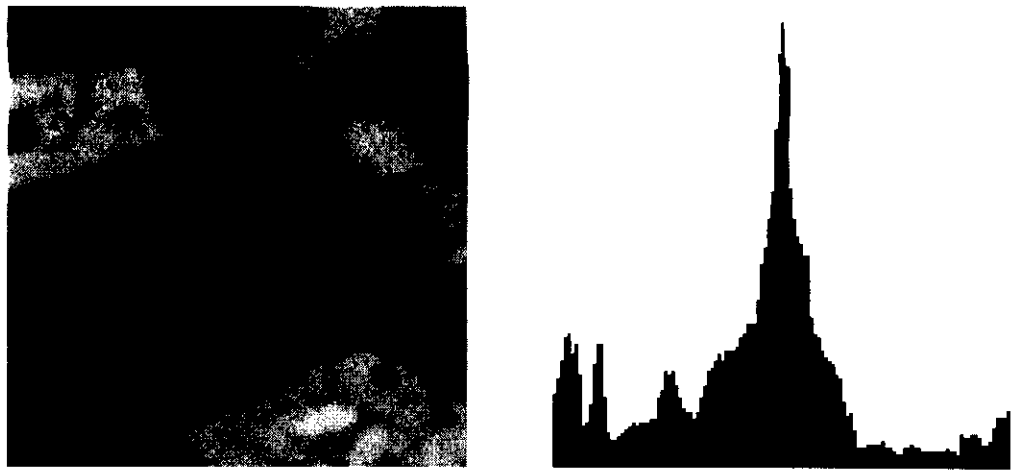


Figure 21: *(a) Scene with a structured background. (b) Result of the processing of (a) with the 3-SIR processor. The three references* $\mathbf{w^f}$, $\mathbf{w^t}$ *and* $\mathbf{w^b}$ *used in the processing are determined from the reference* $\mathbf{w}$ *shown on figure 1(a).*
Note : (b) are plots of the maximum of each line of the output plane of the considered method.

9 Theoretical approaches to Nonlinear Joint Transform Correlations

In this section, we assume that the observed scene is corrupted by a cyclostationary [31], additive, Gaussian noise. The image model can thus be described by:

$$s_i = r_i^j + n_i \tag{89}$$

However, contrary to the usual model which leads to the classical matched filter (cf. section 2.1), we suppose that the psd $\mathbf{\Gamma}$ of the noise n_i is unknown *a priori*. For this model, let us determine the optimal location processor with the different statistical approaches described above.

9.1 Mathematical expression of the likelihood

Since we consider here sampled random signals with a finite number of pixels, it is possible to define their Fourier transforms:

$$\hat{s}_k = \hat{r}_k^j + \hat{n}_k \tag{90}$$

Let us now compute the likelihood of $\hat{\mathbf{s}}$, that is, the probability of observing the input image $\hat{\mathbf{s}}$ with the hypothesis that the spectral density is $\mathbf{\Gamma}$ and the object location is j. Since $\hat{\mathbf{n}}$ is a cyclostationary Gaussian random vector, one has:

$$P(\hat{\mathbf{n}}|\mathbf{\Gamma}) = \prod_{k=0}^{N-1} \frac{1}{\sqrt{2\pi\Gamma_k}} \exp\left[-\frac{|\hat{n}_k|^2}{2\Gamma_k}\right] \tag{91}$$

where Γ_k denotes the component of $\mathbf{\Gamma}$ at frequency k. It will be interesting in the following mathematical developments to write the pdf of Eq. 91 in the canonical form of the exponential family [59] by introducing $\alpha_k = 1/(\Gamma_k)$. We thus define a vector **alpha** with components α_k. We thus obtain the following expression for the probability of observing the input image $\hat{\mathbf{s}}$ (i.e. the likelihood):

$$L(\hat{\mathbf{s}}|\alpha, j) = \prod_{k=0}^{N-1} \frac{\sqrt{\alpha_k}}{\sqrt{2\pi}} \exp\left[-\frac{\alpha_k}{2}|\hat{s}_k - \hat{r}_k^j|^2\right] \tag{92}$$

9.2 ML estimation of the spectral density

We propose to estimate the spectral density $\mathbf{\Gamma}$ of the noise and the target position which maximize the likelihood $L(\hat{\mathbf{s}}|\mathbf{\Gamma}, j)$. In order to obtain this result, one has to solve the following equation:

$$\frac{\partial P(\hat{\mathbf{s}}|\mathbf{\Gamma}, j)}{\partial \Gamma_k} = 0 \tag{93}$$

The solution of this equation leads to the following estimate:

$$\Gamma_k^{ML}(j) = |\hat{s}_k - \hat{r}_k^j|^2 \tag{94}$$

After injecting this estimate into Eq. 92 and taking its logarithm, we obtain the following expression of the loglikelihood:

$$\ell_{ML}(j) = -\sum_{k=0}^{N-1} \frac{1}{2} \log\left[|\hat{s}_k - \hat{r}_k^j|^2\right] + K \tag{95}$$

where K is a constant independent of j.

One can remark that $\ell_{ML}(j)$ diverges if $|\hat{s}_k - \hat{r}_k^j| = 0$ for some frequency k: this estimate can thus be unstable. In order to have a stable estimate, it is thus necessary to assume an *a priori* knowledge with a prior pdf $P(\mathbf{\Gamma})$ which penalizes null values for Γ_k: this will be the subject of the next sections.

However, let us now consider Eq. 95, which represents the optimal processor in the maximum likelihood sense for the considered image model. It is very computation intensive. Thus in the following, we will study how a first order development of this expression leads to a processor which has the same computational complexity as a linear filter, and which is identical to some previously proposed NLJTC processors.

We first assume that for every frequency k, we always have $|\hat{s}_k - \hat{r}_k^j| > \sigma^2$ with $\sigma^2 > 0$. In order to simplify the following analysis, we propose to introduce the notations:

$$\Delta_k^j = |\hat{s}_k - \hat{r}_k^j|^2 \tag{96}$$

$$U_k^j = [\hat{s}_k]^* \, \hat{r}_k^j + [\hat{r}_k^j]^* \, \hat{s}_k \tag{97}$$

$$D_k = |\hat{s}_k|^2 + |\hat{r}_k|^2. \tag{98}$$

We have used the fact that $\hat{r}_k^j = \hat{r}_k \exp(-i2\pi \; j \; k)$, where $\hat{r}_k$ is the Fourier transform of the target when it is centered at the origin. With these definitions, one has: $\Delta_k^j = D_k - U_k^j$ (U_k^j

is real) and the loglikelihood can be written:

$$\ell_{ML}(j) = K \; - \; \frac{1}{2} \sum_{k=0}^{N-1} \left(\log[D_k] + \log[1 \; - \; \frac{U_k^j}{D_k}] \right) \tag{99}$$

Since D_k is independent of j and since the assumption $\Delta_k^j > \sigma^2$ leads to $\frac{U_k^j}{D_k} < 1 - \frac{\sigma^2}{D_k}$, one can consider the first order development of the loglikelihood:

$$\ell_{ML}(j) = K' \; + \; \frac{1}{2} \sum_{k=0}^{N-1} \frac{U_k^j}{D_k} \tag{100}$$

where K' is a constant independent of j.

It can be shown [64] that maximizing $\ell_{ML}(j)$ is equivalent to maximizing the inverse Fourier transform of $\hat{F}_k$:

$$\hat{F}_k = \frac{\hat{s}_k(\hat{r}_k)^*}{|\hat{s}_k|^2 + |\hat{r}_k|^2}. \tag{101}$$

We have utilized the fact that, when **r** and **s** are real, one has:

$$\sum_{k=0}^{N-1} \hat{F}_k \; \exp(-i2\pi \; j \; k) = \sum_{k=0}^{N-1} (\hat{F}_k)^* \; \exp(i2\pi \; j \; k) \tag{102}$$

The Fourier transform of the nonlinear JTC introduced in section 4.6(see also [30]) is described in the general case by:

$$\hat{C}_k = \frac{\hat{s}_k(\hat{r}_k)^*}{a + \mu \; |\hat{s}_k|^2 + (1 - \mu) \; |\hat{r}_k|^2}. \tag{103}$$

One can thus remark that $\hat{F}_k$ is a particular case of $\hat{C}_k$ with the regularization term a equal to 0 and $\mu = 1 - \mu = 1/2$.

If the assumption $\Delta_k^j > \sigma^2$ is not fulfilled for σ^2 sufficiently large, the nonlinear JTC defined by $\hat{F}_k$ can be unstable. We thus propose in the following sections to analyze respectively the MAP and the Bayesian approaches which allow one to regularize the ML solution.

9.3 MAP estimation of the spectral density

We now propose to determine the estimates of the spectral density of the noise and of the target location which maximize the posterior probability $P(\alpha, j|\hat{\mathbf{s}})$ (MAP estimate). We have shown in Section 2 that according to Bayes law, maximizing $P(\alpha, j|\hat{\mathbf{s}})$ is equivalent to maximizing $P(\hat{\mathbf{s}}|\alpha, j) \; P(\alpha)$. We have assumed that the variables α and j are statistically independent (i.e. that $P(\alpha, j) = P_p(\alpha) \; P(j)$). We have seen in Eq. 92 that, in order to obtain an exponential mathematical form (the canonical expression of the exponential family), the likelihood $L(\hat{\mathbf{s}}|\alpha, j)$ can be written:

$$L(\hat{\mathbf{s}}|\alpha, j) = \prod_{k=0}^{N-1} \frac{\sqrt{\alpha_k}}{\sqrt{2\pi}} \; \exp\left[-\frac{\alpha_k}{2} \Delta_k^j \right] \tag{104}$$

The likelihood becomes unstable for null values of $\Gamma_k^{ML}(j)$, so from a practical point of view, the main goal of the choice $P[\alpha]$ is to penalize large values of α_k. There are many types of such functions. In the following, we will analyze two of them: the uniform prior and the exponential prior (cf. Figure 22).

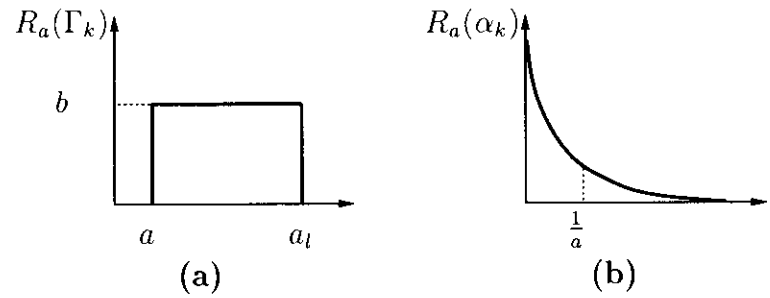


Figure 22: *Examples of possible priors.* **(a)**: *Uniform prior (on Γ_k).* **(b)**: *Exponential prior (on $\alpha_k = 1/\Gamma_k$).*

9.3.1 Uniform prior

The expression of the uniform prior on Γ_k is $P[\mathbf{\Gamma}] = \prod_{k=0}^{N-1} R_a[\Gamma_k]$ where:

$$\begin{cases} R_a[\Gamma_k] = b & \text{if } a \leq \Gamma_k \leq a_l \\ R_a[\Gamma_k] = 0 & \text{otherwise} \end{cases} \tag{105}$$

a_l is determined so that on the current scene $\Gamma_k < a_l$ for all k. Moreover $b = 1/(a_l - a)$. This prior fits the second point of Section 8.2.4: it is obvious that it will lead to psd estimate values close – in fact equal — to the ML estimate values when the likelihood is stable.

The optimal processor in the MAP sense corresponding to this prior is derived in appendix D. The computations are complex, and so is its final expression. However, after one approximation, it can be shown that its first order term is:

$$\hat{F}_k = \frac{\hat{s}_k(\hat{r}_k)^*}{\max\left[|\hat{s}_k|^2 + |\hat{r}_k|^2\ ,\ a\right]} \tag{106}$$

This expression is quite satisfying for the intuition, since it penalizes the very little values of $|\hat{s}_k|^2 + |\hat{r}_k|^2$, which could make the expression of the filter unstable. This prior leads to solutions which present strong analogies with the truncature method discussed in part1.

9.3.2 Exponential prior

As we discussed in Section 8.2.4, another interesting property for the choice of the prior is to lead to easily tractable mathematical equations. The exponential prior:

$$P[\alpha] = \prod_{k=0}^{N-1} a\ exp(-a\ \alpha_k) \tag{107}$$

fulfills this requirement. If this prior is used, it is easy to show that the MAP estimate of the spectral density has the following expression [64]:

$$\Gamma_k^{MAP}(j) = 2\ a\ +\ \Delta_k^j \tag{108}$$

After injecting this estimate into Eq. 92 and taking its logarithm, we obtain the following expression of the loglikelihood:

$$\ell_{MAP}(j) = -\sum_{k=0}^{N-1} \frac{1}{2} \log[2\ a\ +\ |\hat{s}_k - \hat{r}_k^j|^2]\ +\ K' \tag{109}$$

where K' is a constant independent of j.

The first order of the Taylor expansion of the logarithm shows that the MAP estimate of j is thus obtained by maximizing the inverse Fourier transform of $F(j)$:

$$\hat{F}_k = \frac{[\hat{s}_k]^*\ \hat{r}_k}{2\ a\ +\ |\hat{s}_k|^2 + |\hat{r}_k|^2} \tag{110}$$

Within this approximation, the optimal estimate of the location j is the one which maximizes $F(j)$. This expression is identical to that of the nonlinear JTC introduced in section 4.6 (see also [30]) with $\mu = 1 - \mu = 1/2$. The MAP approach has thus allowed us to determine the optimal value of a parameter which remained free in Ref. [30]. This is an example of practical advantage of the decision theoretical approaches versus the heuristic approaches.

In order to illustrate the previous results, let us look at Figure 23. We have constructed two scenes containing the same object and an additive white noise. The noise statistic is uniform. The Signal to Noise Ratio (SNR) is respectively 10 (very low noise) and 0. In the second line appears the correlation plane obtained by applying Eq. 110 to the corresponding scene. In the third line appears the result of the MAP processor of Eq. 109. We can see that the correlation planes corresponding to both filters are almost identical. This shows that the higher-order terms of Eq. 109 are negligible with respect to the first-order approximation. The Figure 24, which represents the same data computed from noisier images (SNR of -10 and -20 dB) leads to the same conclusions. Whatever the level of noise present in the image, the nonlinear JTC of [30] is thus an accurate approximation of the optimal alternative to the matched filter when the spectral density of the noise is unknown.

9.4 Bayesian approach

Let us consider here again the exponential prior of Eq. 107. In the Bayesian approach, one has to determine:

$$P(\hat{\mathbf{s}}|j) = \int L(\hat{\mathbf{s}}|\alpha, j) P[\alpha] d\alpha \tag{111}$$

where the integral has to be interpreted with same notations as in Section 8.2.3.

One can show that

$$P(\hat{\mathbf{s}}|j) = \prod_{k=0}^{N-1} \frac{a}{[2\ a\ + \Delta_k^j]^{(\frac{3}{2})}}. \tag{112}$$

Its logarithm is:

$$\ell_{Bay}(j) = -\sum_{k=0}^{N-1} \frac{3}{2} \log[2\ a\ +\ |\hat{s}_k - \hat{r}_k^j|^2]\ +\ K'' \tag{113}$$

where K'' is a constant independent of j. One can remark that the j-dependent term of $\ell_{Bay}(j)$ is proportional to the j-dependent term of $\ell_{MAP}(j)$ (cf. Eq. 109). The maximization

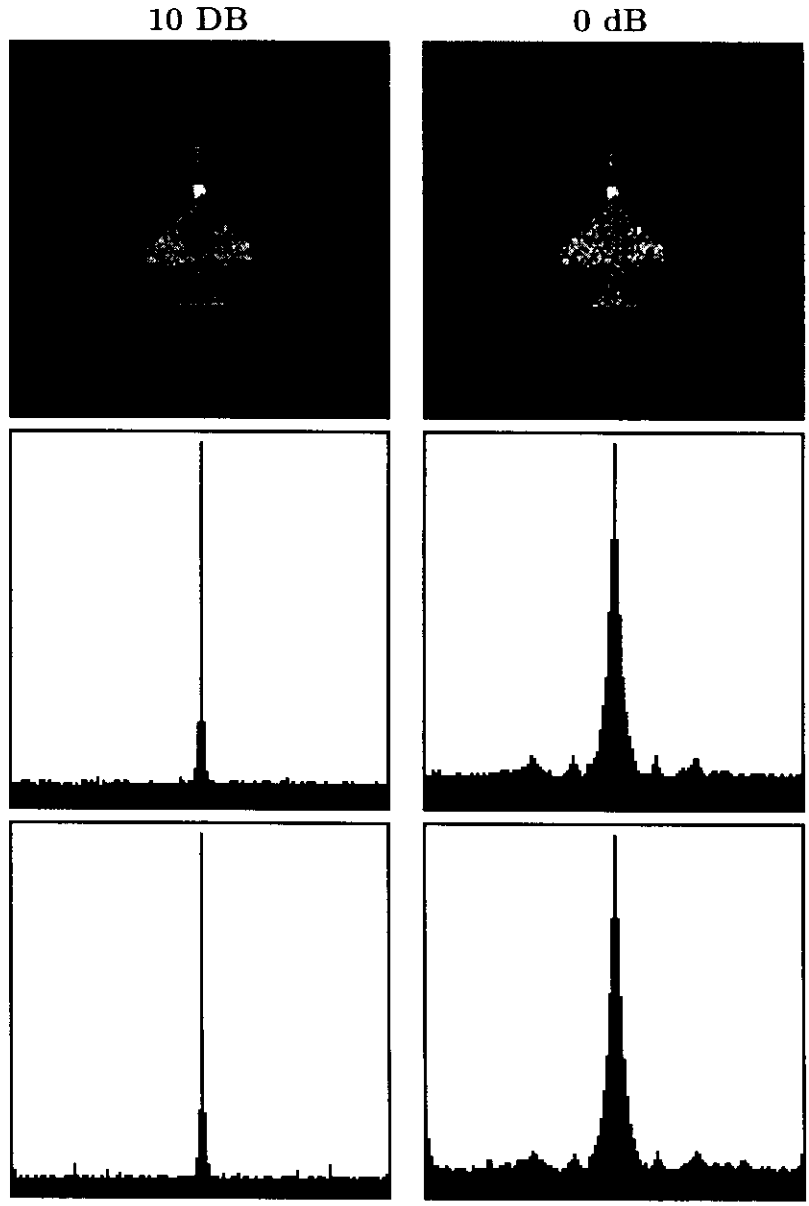


Figure 23: *First line: scenes with additive white uniform noise (low levels of noise). Second line: result of the application of the NLJTC of Eq. 110 to the corresponding scene. Third line: result of the application of the MAP optimal processor of Eq. 109.*
NB: the graphs of second and third lines are plots of the maximum of each line of the output plane obtained with the corresponding algorithm.

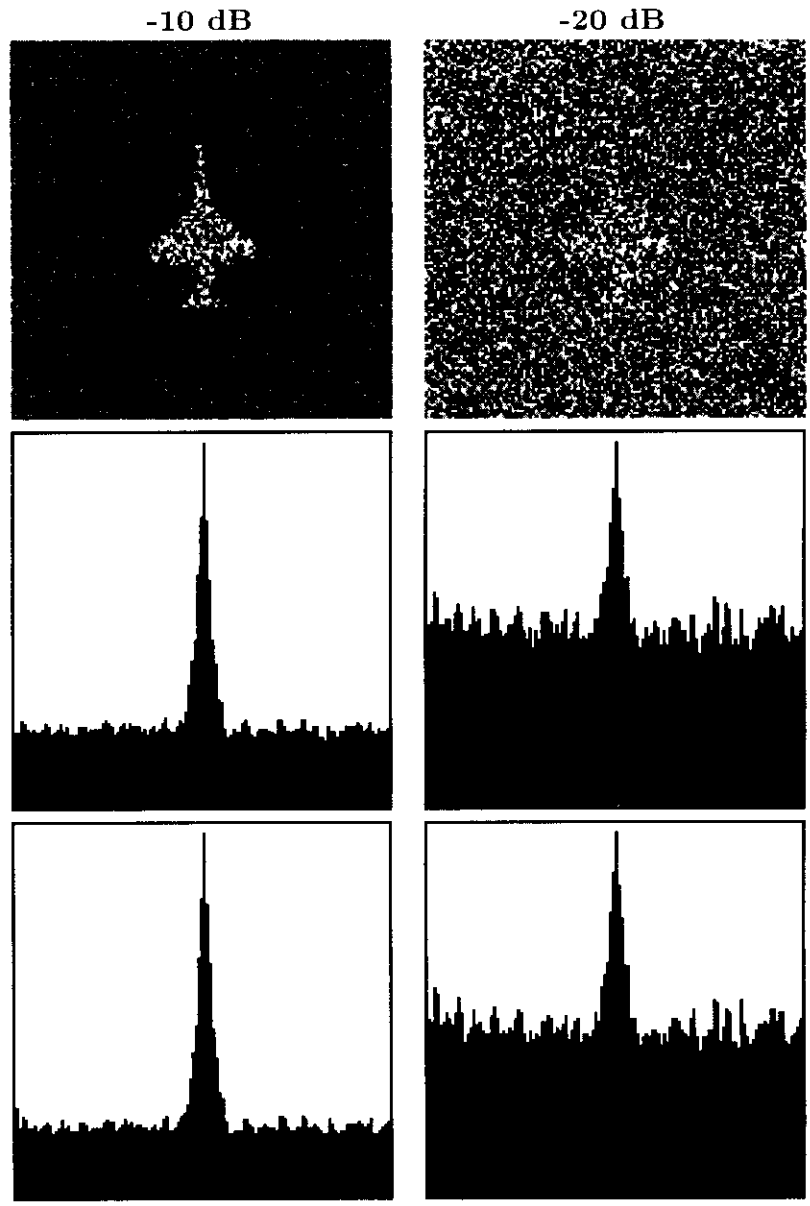


Figure 24: *First line: scenes with additive white uniform noise (higher levels of noise). Second line: result of the application of the NLJTC of Eq. 110 to the corresponding scene. Third line: result of the application of the MAP optimal processor of Eq. 109.*
NB: the graphs of second and third lines are plots of the maximum of each line of the output plane obtained with the corresponding algorithm.

of $\ell_{Bay}(j)$ with respect to j is thus equivalent to the maximization of $\ell_{MAP}(j)$. In other terms, for the exponential prior, the optimal Bayesian solution is equivalent to the MAP approach. Of course, this is also true for the first order developments, and thus the NLJTC of Eq. 110 is also a good approximation of the Bayesian solution. The NLJTC is thus close to the optimal decision theoretical solution for an exponential prior, whatever the derivation technique that is used.

9.5 Conclusion

In this section, we have analyzed statistical approaches adapted to practical target detection and location tasks when the psd of a Gaussian additive noise is unknown *a priori*. In that case, the psd can be considered as a nuisance parameter. We have discussed the ML, the MAP and the Bayesian solutions to that problem. It has clearly appeared that the MAP and the Bayesian solutions can lead to regularized versions of the ML solution.

Moreover, we have determined the precise prior pdf of the spectral density which leads to the optimal nonlinear JTC of [30] and we have shown that this approximation is a particular case of the nonlinear JTC introduced in [30]. More precisely, with the heuristic approach, there were two independent parameters while with the statistical decision theory approach there is only one free parameter. Using numerical simulations, we have shown that the approximation by the nonlinear JTC is quite accurate.

10 Conclusion of part 2

In this part, we have analyzed the statistical decision theory in the context of optical correlation applied to object location. We have first presented the general theory, with the ML and the MAP approaches. We have also discussed the different way of handling nuisance parameters, i.e. the ML, the MAP and the Bayesian methods. We have then illustrated these different approaches with recent results obtained for object location. These results show that the statistical decision theory approach is very efficient and enables to solve problems that could not be solved by usual heuristic approaches. The main reason for this fact is that with decision theoretical approaches, the starting point is the image model itself, which represents – as accurately as possible – the characterisitcs of real scenes.

As a further application of the decision theoretical techniques, we have studied the case of a scene corrupted by a Gaussian additive noise with unknow psd. We have determined the optimal processors in the ML, MAP and Bayesian approaches. We have demonstrated that the nonlinear joint-transform correlation, which is frequently used in optical correlators, can be considered as an approximation of these optimal processors. This result constitutes a theoretical support in the context of statistical detection theory for the use of nonlinear joint transform correlators. In particular, its practical consequence is to accurately specify the domain of applications for which one can expect that nonlinear joint transform correlators will provide better performance than linear Vander Lugt correlators.

In this part, we have illustrated the efficiency and the versatility of the statistical decision theoretical approach for object location applications. In the context of optical correlation, it is important to notice that all the optimal algorithms presented in this paper are mainly

based on correlations. Some of them consist of nonlinear preprocessings of the input scene followed by correlations. Others are based on nonlinearities in the Fourier plane (NLJTC). Their computational complexity is thus equivalent to that of linear filters. It is important to notice that this result was not postulated *a priori*, as in the heuristic approaches, but obtained *a posteriori* through optimal decision theoretical filter design techniques.

Acknowledgments

The author gratefuly acknowledge P. Chavel, J. Figue, C. Chesnaud, F. Goudail, F. Guérault, J.-P. Huignard, B. Javidi and V. Laude for fruitful discussions.

Lexique

ACPE : average CPE,
CPE : Correlation Plane Energy,
MACE : minimum average correlation energy,
MAP : Maximum a Posteriori,
ML : maximum likelihood,
MVSDF : minimum variance SDF,
MSE : mean square error,
NLJTC : nonlinear joint transform correlation,
OT : optimal trade-off,
OTSDF : optimal trade-off SDF,
PCE : Peak to Correlation Energy,
pdf : probability density function,
POF : Phase-Only-Filters,
psd : power spectrum density,
SDF : synthetic discriminant function,
SNR : signal to noise ratio.

A Appendix

The criterion to optimize is

$$SNR = \frac{|\mathbf{h}^{\dagger}.\mathbf{r}|^2}{\hat{\mathbf{h}}^{\dagger}\hat{S}\hat{\mathbf{h}}} \tag{A1}$$

In the Fourier domain, one obtains:

$$SNR = \frac{|\sum_k \hat{h}_k^* \hat{r}_k|^2}{\sum_k \hat{h}_k^* \hat{S}_k \hat{h}_k} \tag{A2}$$

Let us introduce:

$$\hat{g}_k = \sqrt{\hat{S}_k}\hat{h}_k \tag{A3}$$

and:

$$\hat{f}_k = \frac{\hat{r}_k}{\sqrt{\hat{S}_k}} \tag{A4}$$

one thus has:

$$SNR = \frac{|\sum_k \hat{g}_k^* \hat{f}_k|^2}{\sum_k \hat{g}_k^* \hat{g}_k}. \tag{A5}$$

The Cauchy Scwartz inequality let us know that:

$$||\hat{\mathbf{g}}^\dagger.\hat{\mathbf{f}}||^2 \leq ||\hat{\mathbf{g}}||^2 ||\hat{\mathbf{f}}||^2 \tag{A6}$$

Furhtermore, one knows that there is an equality if there exist a scalar number λ such as:

$$\hat{\mathbf{g}} = \lambda \hat{\mathbf{f}} \tag{A7}$$

So the ratio defined by the SNR is optimized if

$$\hat{g}_k = \lambda \hat{f}_k \tag{A8}$$

which leads to:

$$\hat{h}_k \propto \frac{\hat{r}_k}{\hat{S}_k}. \tag{A9}$$

The multiplicative constant is often choosen equal to 1 since it does not modify the SNR.

B Appendix

Let $\mathbf{r}^\ell$ with $\ell = 1, ..., P$ denote the P reference images of dimension N. SDF filters are defined in order to obtain some specified values at the center of the correlation function for the patterns belonging to the training set.

The mathematical problem of optimal SDF filter for quadratic criteria is to optimaize a quadratic function of the form:

$$\hat{\mathbf{h}}^\dagger A \hat{\mathbf{h}} \tag{B1}$$

with the following constraints:

$$\hat{\mathbf{h}}^\dagger.\hat{\mathbf{r}}^\ell = d_\ell \qquad ;\forall \ell = 1, ..., P \tag{B2}$$

We consider positive circulant Toeplitz matrices A, i.e. matrices which are diagonal in the Fourier domain with positive values. The optimization problem is thus to minimize:

$$\sum_k \hat{h}_k^* A_{k,k} \hat{h}_k \tag{B3}$$

with the constraints:

$$\sum_k \hat{h}_k^*.\hat{r}_k^\ell = d_\ell \qquad ;\forall \ell = 1, ..., P \tag{B4}$$

Let us introduce the matrix $\hat{R}$ such that its column numbet ℓ is the vector $\hat{\mathbf{r}}^\ell$. In other words: $\hat{R} = [\hat{\mathbf{r}}^1, \hat{\mathbf{r}}^2, ..., \hat{\mathbf{r}}^P]$

As we have done in appendix A, let us introduce:

$$\hat{g}_k = \hat{h}_k \sqrt{\hat{A}_{k,k}} \tag{B5}$$

and:

$$\hat{x}_k^\ell = \frac{\hat{r}_k^\ell}{\sqrt{\hat{A}_{k,k}}} \tag{B6}$$

We also define the matrix $\hat{X}$ such that its column numbet ℓ is the vector $\hat{\mathbf{x}}^\ell$ and the vector $\mathbf{d}$ with components d_ℓ. It is easy to verify that:

$$\hat{X} = \hat{A}^{-1/2} \hat{R} \tag{B7}$$

and:

$$\hat{\mathbf{g}} = \hat{A}^{1/2} \hat{\mathbf{h}} \tag{B8}$$

where $\hat{A}^{1/2}$ and $\hat{A}^{-1/2}$ are the diagonal matrices with elements:

$$(\hat{A}^{1/2})_{k,k} = \sqrt{\hat{A}_{k,k}} \tag{B9}$$

$$(\hat{A}^{-1/2})_{k,k} = 1/\sqrt{\hat{A}_{k,k}} \tag{B10}$$

The optimization problem is thus to find the solution of:

$$\hat{X}^\dagger \hat{\mathbf{g}} = \mathbf{d} \tag{B11}$$

which mininizes $||\hat{\mathbf{g}}||$ i.e. the norm of $\hat{\mathbf{g}}$.

Let us write:

$$\hat{\mathbf{g}} = \hat{X}\mathbf{a} + \hat{\mathbf{p}} \tag{B12}$$

where $\mathbf{a}$ is a parameter of dimension P and $\hat{\mathbf{p}}$ is a vector of dimension N and which is orthogonal to the P vectors $\hat{x}_k^\ell$. With Eq.B11 one obtains:

$$\hat{X}^\dagger \hat{X} \mathbf{a} = \mathbf{d} \tag{B13}$$

and thus:

$$\mathbf{a} = [\hat{X}^\dagger \hat{X}]^{-1} \mathbf{d} \tag{B14}$$

where $[\hat{X}^\dagger \hat{X}]^{-1}$ is the pseudo inverse of $\hat{X}^\dagger \hat{X}$ (which means that only the non null eigenvalues of the symetrical matrix $\hat{X}^\dagger \hat{X}$ are inverterd). We thus obtain:

$$\hat{X}\mathbf{a} = \hat{X}[\hat{X}^\dagger \hat{X}]^{-1} \mathbf{d} \tag{B15}$$

It is thus clear that the orthogonal vector $\hat{\mathbf{p}}$ has on influence on the constraints and the vector which minimizes the norm of $\hat{\mathbf{g}}$ is:

$$\hat{\mathbf{g}} = \hat{X}[\hat{X}^\dagger \hat{X}]^{-1} \mathbf{d} \tag{B16}$$

which leads to:

$$\hat{\mathbf{g}} = A^{-1} \hat{R} [\hat{R}^\dagger A^{-1} \hat{R}]^{-1} \mathbf{d} \tag{B17}$$

C Appendix

In this appendix, we demonstrate that the processor proposed in Ref. [34] is optimal in the ML sense even if the target is noisy. Let us first briefly review the derivation of this processor which is performed in Ref. [34]. The background pixels values b_i are considered as random gaussain variables, so that the pixel values in the background of the noisy scene $\bar{\mathbf{s}}^\delta$ are distributed with a probability density function $P_{\mathbf{t}}\left(\bar{\mathbf{s}}^\delta\right)$ where $t_i = b_i + n_i$. The loglikelihood corresponding to the model (after having estimated β in the ML sense) is thus:

$$\ell[\mathbf{s}|\mathbf{r},\mathbf{w},\delta] = -\frac{1}{2\sigma_n^2}\left[\sum_i w_{i-\delta}\ s_i^2 - \frac{\sum_i r_{i-\delta}\ s_i}{\sum_i r_i^2}\right] + \log P_{\mathbf{t}}\left(\bar{\mathbf{s}}^\delta\right) \tag{C1}$$

When the variance σ_n^2 of the additive noise tends to 0, the second term of this equation becomes negligible with respect to the first one. The authors conclude that the first term (which is equivalent to Eq. 67) is the optimal location processor when the variance of the additive noise is zero. This result is satisfying for the intuition since when there is no additive noise, it can be shown that this processor exactly reaches its minimum value when δ is the true position of the target.

However, this processor is optimal in an other type of situation. Indeed, let us now assume that the pixel values of the background b_i are unknown parameters, that is, nuisance parameters. In that case, the term $\log P_{\mathbf{t}}\left(\bar{\mathbf{s}}^\delta\right)$ in the likelihood must be replaced with $\log P_{\mathbf{n}}\left(\bar{\mathbf{s}}^\delta\right) = -\frac{1}{2\sigma_n^2}\ \sum_i (1 - w_{i-\delta})\ (s_i - b_i)^2$ The expression of the likelihood thus becomes:

$$\begin{aligned}\ell[\mathbf{s}|\mathbf{r},\mathbf{w},\mathbf{b},\delta] \ &= \ -\frac{1}{2\sigma_n^2}\left[\sum_i w_{i-\delta}\ s_i^2 \frac{\sum_i r_{i-\delta}\ s_i}{\sum_i r_i^2}\right] \\ &\quad -\frac{1}{2\sigma_n^2}\ \sum_i (1 - w_{i-\delta})\ (s_i - b_i)^2\end{aligned} \tag{C2}$$

It is then possible to estimate the b_i in the ML sense:

$$\frac{\partial \ell}{\partial b_i} = 0 \iff b_i^{ML} = s_i \tag{C3}$$

After injecting this estimate into the expression of the likelihood, on obtains:

$$\ell[\mathbf{s}|\mathbf{r},\mathbf{w},\hat{\mathbf{b}},\delta] = \frac{1}{2\sigma_n^2}\left[-\sum_i w_{i-\delta}\ s_i^2 + \frac{\sum_i r_{i-\delta}\ s_i}{\sum_i r_i^2}\right] \tag{C4}$$

which is the processor introduced in Ref. [34]. This processor is thus optimal in the ML sense whatever the power of the noise on the target, when we have no *a priori* knowledge about the statistical distribution of the background's pixel values.

D Appendix

In this appendix, we determine the MAP processor for a uniform prior on Γ (cf. Section 9.3 and Figure 22).

Let us recall the expression of the log-posterior density:

$$\ell(j|\mathbf{\Gamma},\mathbf{s}) = -\frac{1}{2}\sum_k \log\Gamma_k - \frac{1}{2}\sum_k \frac{\Delta_k^j}{\Gamma_k} + \log P_{\mathbf{\Gamma}}(\mathbf{\Gamma}) \tag{D1}$$

In order to maximize this expression with respect to Γ_k, one first computes the ML estimate (without taking the term $\log P_{\mathbf{\Gamma}}(\mathbf{\Gamma})$ into account). According to Eq. 94, its expression is:

$$\Gamma_k^{ML} = \Delta_k^j \tag{D2}$$

Then two cases may appear:

$$\Delta_k^j \geq a \implies \Gamma_k^{MAP} = \Delta_k^j \tag{D3}$$
$$\Delta_k^j < a \implies \Gamma_k^{MAP} = a \tag{D4}$$

Let us define the following two sets of integer values:

$$\mathcal{D} = \left\{k \in [0, N-1] \mid \Delta_k^j \geq a\right\} \tag{D5}$$
$$\bar{\mathcal{D}} = \left\{k \in [0, N-1] \mid \Delta_k^j < a\right\} \tag{D6}$$

After injecting the estimates of Γ_k^{MAP} into the posterior density, one obtains:

$$\ell\left(j|\mathbf{\Gamma}^{MAP},\mathbf{s}\right) = -\frac{1}{2}\sum_{k\in\mathcal{D}} \log\Delta_k^j - \frac{1}{2}\sum_{k\in\bar{\mathcal{D}}}(\log a - 1) - \frac{1}{2}\sum_{k\in\bar{\mathcal{D}}}\frac{\Delta_k^j}{a} + K \tag{D7}$$

where K is independent of j. Let us now perform on $\ell\left(j|\mathbf{\Gamma}^{MAP},\mathbf{s}\right)$ the same first order development as in Section 9.2. One obtains:

$$\begin{aligned} d\ell\left(j|\mathbf{\Gamma}^{MAP},\mathbf{s}\right) &= \frac{1}{2}\sum_{k\in\mathcal{D}}\frac{U_k^j}{D_k} - \frac{1}{2}\sum_{k\in\mathcal{D}}\log D_k - \frac{1}{2}\sum_{k\in\bar{\mathcal{D}}}(\log a - 1) \\ &\quad -\frac{1}{2}\sum_{k\in\bar{\mathcal{D}}}\frac{D_k}{a} + \frac{1}{2}\sum_{k\in\bar{\mathcal{D}}}\frac{U_k^j}{a} + K \end{aligned} \tag{D8}$$

This expression is quite involved since $\mathcal{D}$ and $\bar{\mathcal{D}}$ depend on j itself. In order to obtain a more tractable equation, one has to do a further approximation, which is consistent with the hypothesis of validity of the first order development: Let us suppose that $D_k >> U_k^j$ so that $\Delta_k^j \simeq D_k$. We can then replace the set $\mathcal{D}$ and $\bar{\mathcal{D}}$ with:

$$\mathcal{D}' = \{k \in [0, N-1] \mid D_k \geq a\} \tag{D9}$$
$$\bar{\mathcal{D}}' = \{k \in [0, N-1] \mid D_k < a\} \tag{D10}$$

The advantage of this operation is that $\mathcal{D}'$ and $\bar{\mathcal{D}}'$ do not depend on j, so that the expression of $d\ell\left(j|\mathbf{\Gamma}^{MAP},\mathbf{s}\right)$ becomes:

$$d\ell'\left(j|\mathbf{\Gamma}^{MAP},\mathbf{s}\right) = \frac{1}{2}\sum_{k\in\mathcal{D}'}\frac{U_k^j}{D_k} + \frac{1}{2}\sum_{k\in\bar{\mathcal{D}}'}\frac{U_k^j}{a} + K' \tag{D11}$$

where K' is a constant which do not depend on j. Consequently, following the same reasoning as in Section 9.2, maximizing $d\ell'\left(j|\mathbf{\Gamma}^{MAP}, \mathbf{s}\right)$ with respect to j is equivalent to maximizing the inverse Fourier transform of:

$$\hat{F}_k = \frac{\hat{s}_k(\hat{r}_k)^*}{\max\left[D_k, a\right]} \tag{D12}$$

References

[1] H. J. Caulfield and W. T. Maloney, "Improved discrimination in optical character recognition," Appl. Opt. **8**, 2354–2355 (1969).

[2] D. Casasent and D. Psaltis, "Position, rotation and scale invariant optical correlation," Appl. Opt. **15**, 1795–1799 (1976).

[3] J. L. Horner, "Light utilization in optical correlators," Appl. Opt. **21**, 4511–4514 (1982).

[4] Y. N. Hsu and H. H. Arsenault, "Optical pattern recognition using circular harmonic expansion," Appl. Opt. **21**, 4016–4019 (1982).

[5] B. V. K. Vijaya Kumar, "Efficient approach for designing linear combination filters," Appl. Opt. **22**, 1445–1448 (1983).

[6] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23**, 812–816 (1984).

[7] D. Casasent, "Unified synthetic discriminant function computation formulation," Appl. Opt. **23**, 1620–1627 (1984).

[8] F. M. Dickey and L. A. Romero, "Normalized correlation for pattern recognition," Opt. Lett. **16**, 1186–1188 (1991).

[9] F. M. Dickey and K. T. Stalker, "Binary phase only filters: Implications of bandwith and uniqueness on performance," J. Opt. Soc. Am. A **4**, 69 (1987).

[10] F. M. Dickey, T. K. Stalker, and J. J. Mason, "Bandwith considerations for binary phase-only filter," Appl. Opt. **27**, 3811–3818 (1988).

[11] Z. Zouhir Bahri and B. V. K. Vijaya Kumar, "Fast algorithms for designing optical phase-only-filters (POFs) and binary phase-only-filters (BPOFs)," Appl. Opt. **29**, 2992–2996 (1990).

[12] F. M. Dickey, B. V. K. Vijaya Kumar, L. A. Romero, and J. M. Connelly, "Complex ternary matched filters yielding high signal-to-noise ratios," Opt. Eng. **29**, 994–1001 (1990).

[13] A. A. S. Awwal, M. A. Karim, and S. R. Jahan, "Improved correlation discrimination using an amplitude modulated phase-only-filter," Appl. Opt. **29**, 233–236 (1990).

[14] Y. N. Hsu, H. H. Arsenault, and G. April, "Rotation-invariant digital pattern recognition using circular harmonic expansion," Appl. Opt. **21**, 4012–4015 (1982).

[15] B. V. K. Vijaya Kumar, "Minimum variance synthetic discriminant functions," J. Opt. Soc. Am. A **3**, 1579–1584 (1986).

[16] A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimun average correlation energy filters," Appl. Opt. **26**, 3633–3640 (1987).

[17] Ph. Réfrégier, "Filter design for optical pattern recognition: Multi-criteria optimization approach," Opt. Lett. **15**, 854–856 (1990).

[18] J. Figue and Ph. Réfrégier, "Influence of the noise model on correlation filters: peak sharpness and noise robustness," Opt. Lett. **17**, 1476–1478 (1992).

[19] Ph. Réfrégier, "Application of the stabilizing functional approach to pattern recognition filters," J. Opt. Soc. Am. A **11**, 1243–1251 (1994).

[20] B. V. K. Vijaya Kumar, C. Hendrix, and D. W. Carlson, "Tradeoffs in the design of correlation filters," in *Optical Pattern Recognition*, J. L. Horner and B. Javidi, eds., SPIE Optical Engineering Press, 191–215 (1992).

[21] B. Javidi and J. Wang, "Limitation of the classic definition of the correlation signal-to-noise ratio in optical pattern recogition with disjoint signal and scene noise," Appl. Opt. **31**, 6826–6829 (1992).

[22] F. Goudail, V. Laude, and Ph. Réfrégier, "Influence of non-overlapping noise on regularized linear filters for pattern recognition," Opt. Lett. **20**, 2237–2239 (1995).

[23] C. S. Weaver and J. W. Goodman, "Technique for optically convolving two functions," Appl. Opt. **5**, 1248–1249 (1966).

[24] B. Javidi, "Nonlinear joint power spectrum based optical correlation," Appl. Opt. **28**, 2358–2367 (1989).

[25] L. Pichon and J.-P. Huignard, "Dynamic joint-fourier transform correlator by bragg diffraction in photorefractive *bso* crystal," Opt. Com. **36**, 277–280 (1981).

[26] F. Turon, E. Ahouzi, J. Campos, K. Chalasinska-Macukow, and M. J. Yzuel, "Nonlinearity effects in the pure phase correlation method in multiobject scenes," Appl. Opt. **33**, 2188–2191 (1994).

[27] R. D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle," Appl. Opt. **32**, 5100–5111 (1993).

[28] G. Ravichandran and D. P. Casasent, "Minimum noise and correlation energy optical correlation filter," Appl. Opt. **31**, 1823–1833 (1992).

[29] B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," Appl. Opt. **29**, 2997–3006 (1990).

[30] Ph. Réfrégier, B. Javidi, and V. Laude, "Nonlinear joint-transform correlation: an optimal solution for adaptive image discrimination and input noise robustness," Opt. Lett. **19**, 405–407 (1994).

[31] Ph. Réfrégier, V. Laude, and B. Javidi, "Basic properties of nonlinear global filtering techniques and optimal discriminant solutions," Appl. Opt. **34**, 3915–3923 (1995).

[32] A. Vander Lugt, "Signal detection by complex filtering," IEEE Trans. Inform. Theory, **IT-10**, 139–145 (1964).

[33] J. V. Candy, *Signal Processing* (McGraw-Hill - Electrical Engineering Series, New York, 1988).

[34] B. Javidi, Ph. Réfrégier, and P. Willet, "Optimum receiver design for pattern recognition with nonoverlapping target and scene noise," Opt. Lett. **18**, 1660–1662 (1993).

[35] Ph. Réfrégier, B. Javidi, and G. Zhang, "Minimum mean-square-error filter for pattern recognition with spatially disjoint signal and scene noise," Opt. Lett. **18**, 1453–1456 (1993).

[36] R. Kotynski and K. Chalasinska-Macukow, "Optical correlator with dual nonlinearity," J. Mod. Opt. **43**, 295–310 (1996).

[37] L. P. Yaroslavsky, "Is the phase-only filter and its modifications optimal in terms of the discrimination capability in pattern recognition ?," Appl. Opt. **31**, 1677–1679 (1992).

[38] Ph. Réfrégier, "Optimal introduction of optical efficiency for pattern recognition filters," in *Optical Information Processing Systems and Architectures IV*, B. Javidi, ed., Proc. Soc. Photo-Opt. Instrum. Eng., 104–115 (1992).

[39] Ph. Réfrégier and J. Figue, "Optimal trade-off filters for pattern recognition and comparison with Wiener approach," Optical Computing and Processing **1**, 245–266 (1991).

[40] V. Laude and Ph. Réfrégier, "Multicriteria characterization of coding domains with optimal Fourier spatial light modulator filters," Appl. Opt. **33**, 4465–4471 (1994).

[41] J. Figue and Ph. Réfrégier, "On the optimality of trade-off filters," Appl. Opt. **32**, 1933–1935 (1993).

[42] D. Psaltis, E. G. Paek, and S. S. Wenkatesh, "Optical image correlation with a binary spatial light modulator," Opt. Eng. **23**, 668–704 (1984).

[43] B. V. K. Vijaya Kumar and Zouhri Bahri., "Phase-only filters with improved signal to noise ratio," Appl. Opt. **28**, 250–257 (1989).

[44] Ph. Réfrégier, B. V. K. Vijaya Kumar, and C. Hendrix, "Multicriteria optimal binary amplitude phase-only filters," J. Opt. Soc. Am. A **9**, 2118–2125 (1992).

[45] C. Hendrix, B. V. K. Vijaya Kumar, K. T. Stalker, B. Kast, and R. Shori, "Design and testing of 3-level optimal correlation filters," in *Optical Information Processing Systems and Architectures III*, B. Javidi, ed., Proc. Soc. Photo-Opt. Instrum. Eng., 2–13 (1991).

[46] D. L. Flannery, W. E. Phillips, and R. L. Reel, "A case study of the design trade-offs for ternary phase-amplitude filters," in *Optical Information Processing Systems and Architectures III*, B. Javidi, ed., Proc. Soc. Photo-Opt. Instrum. Eng., 65–77 (1991).

[47] B. A. Kast, M. Giles, S. Lindell, and D. Flannery, "Implementation of ternary phase-amplitude filters for improved correlation discrimination," Appl. Opt. **28**, 1044–1046 (1989).

[48] B. V. K. Vijaya Kumar and Zouhri Bahri., "Efficient algorithm for designing a ternary valued filter yielding maximum signal to noise ratio," Appl. Opt. **28**, 1919–1925 (1989).

[49] B. R. Frieden, "restoring with maximum likelihood and maximum entropy," J. Opt. Soc. Am. A **26**, 511–518 (1972).

[50] S. J. Wernecke and L. R. d'Addario, "Maximum entropy image reconstruction," IEE Trans. Comput. **C-26**, 351–364 (1977).

[51] M. Fleisher, U. Mahalab, and J. Shamir, "Entropy optimized filter for pattern recognition," Appl. Opt. **29**, 2091–2098 (1990).

[52] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems* (John Wiley and Sons, Inc., New York, 1977).

[53] V. Laude, S. Mazé, P. Chavel, and Ph. Réfrégier, "Amplitude and phase coding measurements of a liquid crystal television," Opt. Com. **103**, 33–38 (1993).

[54] Ph. Réfrégier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak and Horner efficiency," Opt. Lett. **16**, 829–831 (1991).

[55] F. Goudail and Ph. Réfrégier, "Optimal and suboptimal detection of a target with random gray levels imbedded in non-overlapping noise," Opt. Com. **125**, 211–216 (1996).

[56] Ph. Réfrégier and V. Laude, "Critical analysis of filtering techniques for optical pattern recognition: Are the solutions of this inverse problem stable ?," in *Workshop on optical pattern recognition*, Ph. Réfrégier and B. Javidi, eds., pages 58–84 (Proc. Soc. Photo-Opt. Instrum. Eng., Bellingham, Washigton USA, 1994).

[57] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis* (John Wiley and sons, Inc., New York, 1973).

[58] A. H. Jazwinski, *Stochastic processes and filtering theory* (Academic Press, Inc., New York, 1970).

[59] C. P. Robert, *The Bayesian Choice - A decision-Theoretic Motivation* (Springer-Verlag, New York, USA, 1996).

[60] N. Ahmed, ," in *Handbook of digital signal processing - Adaptive Filtering*, D. F. Elliott, ed., pages 857–896 (Academic Press, Inc., San Diego, 1987).

[61] F. Goudail and Ph. Réfrégier, "Optimal detection of a target with random gray levels on a spatially disjoint noise," Opt. Lett. **21**, 495–497 (1996).

[62] F. Guérault and Ph. Réfrégier, "Optimal χ^2 filtering method and application to targets and backgrounds with random correlated gray levels," Opt. Lett. **22**, 630–632 (1997).

[63] Ph. Réfrégier, F. Goudail, and Th. Gaidon, "Optimal location of random targets in random background: random markov fields modelization," Opt. Com. **128**, 211–215 (1996).

[64] Ph. Réfrégier and F. Goudail, "Bayesian theoretical approach to nonlinear joint-transform correlations," European Symposium on Lasers and Optics in Manufacturing, Conference (Munich), june 1997 (1997).