Statistically Independent Region models applied to correlation and segmentation techniques

Philippe Réfrégier, François Goudail and Christophe Chesnaud

Laboratoire Signal et Image
Ecole Nationale Supérieure de Physique de Marseille,
Domaine Universitaire de Saint-Jérôme
13 397 Marseille cedex 20 France

Abstract

Recently new approaches for location and/or segmentation of objects with unknown gray levels embedded in non-overlapping noise have been proposed. These techniques are based on the Statistically Independent Region (SIR) model and are optimal in the maximum likelihood sense. In this paper, we review their theoretical bases and propose a unified approach which enlarges their field of application.

Keywords: object location and segmentation, parameter estimation, optimal algorithms

Contents

1 Introduction

The introduction of optical correlators [1] has stimulated the study of pattern recognition filters based on correlation [2] [3]. Linear filters, such as the matched filter, have been designed to detect a target in the presence of additive noise. However, there exists a different type of noise which is inherent to image processing. It appears as soon as the problem of locating a target appearing on a random

background is addressed. In such cases, the background itself must be regarded as non-overlapping noise [4] [5]. In the presence of such noise, it has been shown that linear filters can fail in locating the target [6].

Furthermore, the efficiency of correlation based techniques can drastically decrease if the object to be detected, located or recognized becomes different from the reference used in the correlation operation. This occurs for example in target tracking applications where the target's attitude in the scene varies or when recognition has to be performed with a large amount of invariance capabilities. A classical solution to this issue has been to design composite filters [7] which allow one to store different attitudes of the target in a single filter.

Recently, algorithms optimal in the Maximum Likelihood (ML) sense [8] for location of an object embedded in non-overlapping noise have been proposed [9, 10] and a unified method has also been designed [11]. In these approaches, the input image is considered to be composed of two independent random fields and the corresponding methods are thus denominated Statistically Independent Region (SIR) methods. A new technique, based on the SIR model, and which allows one to segment an object in an input image has also been recently proposed [12] [13]. This technique is complementary to the correlation methods and is analogous to recently proposed approaches of active contours (snakes) [14] [15] [16]. However, our proposed approach presents clear optimal properties in the context of statistical estimation theory.

In this paper we propose a unified approach for the SIR models which we have presented in the past. We will thus enlarge the field of applications of these techniques in two directions. Firstly, we will be able to consider a large number of input noise statistics which correspond to different physical situations. Secondly, we will analyze these models in the general context of the estimation theory. This general approach will enable us to include such applications as detection, recognition, location, tracking, estimation of unknown parameters (for example the orientation of the object) and shape estimation or, in other words, segmentation.

We will also show that this approach can enlarge the field of application of optoelectronical correlators. As a matter of fact, SIR-based techniques consist of a preprocessing of the analyzed image followed by correlations with binary masks. A simple optoelectronical architecture could thus perform detection, tracking, estimation and segmentation with the same hardware, thus achieving efficient target tracking or recognition.

This paper is organized as follows. In Section 2, we present the mathematical SIR model and study its general solutions in the framework of the statistical theory of estimation for probability density functions (pdf) which belongs to the exponential family. In Section 3, we analyze the optimal solutions when the unknown parameters are estimated in the Maximum Likelihood (ML) sense for location applications. In Section 4, we discuss estimation problems and more particularly segmentation applications. Finally, in Section 5, we illustrate on synthetic and real-world images the efficiency of the proposed algorithms for location, segmentation and tracking applications.

Application	nature of θ
Detection	0 or 1
Discrimination	discrete set
Location	(x, y)
Attitude estimation	angles
Segmentation	Node coordinates

Table 1: *Examples of application and nature of the parameter θ.*

2 The SIR model

2.1 Image model

The SIR model is a probabilistic framework to determine algorithms for detection, recognition, location, parameter estimation or segmentation of an object in an image. We assume that the observed scene is composed of two zones: the target and the background. Furthermore, the target's and the background's gray level are supposed to be unknown and we model their values as independent random variables.

In the following mathematical developments, one-dimensional notations are used for simplicity, and bold font symbols will denote N-dimensional vectors. For example $s = \{s_i | i \in [1, N]\}$ denotes the input image composed of N pixels. For each considered case, the purpose of the image processing algorithm is to estimate an unknown parameter which will be denoted symbolically θ. For example, in detection applications, θ is a binary value, for recognition (or more precisely discrimination), it is a value belonging to a discrete set, and for location, it is the position of some characteristic points of the object (for example the center of gravity). For orientation estimation, θ is a set of possible angles. For segmentation purpose, θ is the shape of the object. In the latter case, if the shape is approximated by a polygonal contour, θ is the set of coordinates of the nodes of the polygon (see table 1).

Let $\mathbf{w}^\theta = \{w_i^\theta | i \in [1, N]\}$ denote a binary window function that defines a certain location, orientation or shape for the target, so that w_i^θ is equal to one within the target and to zero elsewhere. Note that in the following, we will use the same notation $\mathbf{w}^\theta$ for the previously defined binary function and for the set of pixels for which this function is 1. Let us consider the different hypotheses H_θ that consist in assessing a binary window $\mathbf{w}^\theta$ to the target in the input image s, so that we can write:

$$s_i = a_i w_i^\theta + b_i [1 - w_i^\theta] \tag{1}$$

where the target's gray levels **a** and the background noise **b** are random variables. These random variables are characterized by their respective pdfs $P^{\mu_a}[a_i]$ and $P^{\mu_b}[b_i]$ where μ_a and μ_b are the parameters of the pdfs which will be considered as a priori unknown. These parameters can be scalars or vectors if more than one scalar parameter is needed to determine the pdf.

Equation 1 with the pdfs $P^{\mu_a}[a_i]$ and $P^{\mu_b}[b_i]$ is the image model. The parameter of interest is θ while the parameters μ_a and μ_b are nuisance parameters. We will consider the maximum a posteriori (MAP) estimation of the parameter θ. The optimal estimate is thus obtained by maximizing the conditional probability $P[H_\theta|s]$. This conditional probability can be obtained by using Bayes law [8]:

$$P[H_\theta|s] = P[s|H_\theta]P[H_\theta]/P[s] \tag{2}$$

Considering that all hypotheses H_θ are equiprobable, the MAP estimation of θ is equivalent to the Maximum likelihood estimation obtained by maximizing $P[s|H_\theta]$. In the following, we will analyze the ML estimation since generalization to the MAP estimate is obvious using Eq. 2. With the image model of Eq. 1, the likelihood is:

$$P[s|H_\theta, \mu_a, \mu_b] = \prod_{i \in \mathbf{w}^\theta} P^{\mu_a}[s_i] \prod_{i \in \bar{\mathbf{w}}^\theta} P^{\mu_b}[s_i] \tag{3}$$

where we have explicitly denoted the dependence on the unknown parameters μ_a and μ_b. Please remember that we use the same notation $\mathbf{w}^\theta$ (or $\bar{\mathbf{w}}^\theta$) for binary support functions and for the set of pixels for which the value of these functions is 1.

The question is now how to deal with the nuisance parameters μ_a and μ_b in order to express the likelihood as a function of the only parameter of interest θ and of the input image s. There exists several methods to deal with nuisance parameters and the three most frequently used are the Maximum Likelihood (ML) estimation, the Maximum A Posteriori (MAP) estimation and the marginal Bayesian approach. With the marginal Bayesian and the MAP approaches, the nuisance parameters are considered as random variables and prior density probability functions have to be chosen [8, 17].

Let $\pi_a(\mu_a)$ and $\pi_b(\mu_b)$ denote these priors, the marginal Bayesian approach is simple from a theoretical point of view and is based on the Bayes relation:

$$P[s|H_\theta] = \int_{\mu_a} \int_{\mu_b} P[s|H_\theta, \mu_a, \mu_b]\pi_a(\mu_a)\pi_b(\mu_b)d\mu_a d\mu_b \tag{4}$$

where a symbolic notation has been used for the multidimensional integrals:

$$\int_{\mu_a} f(\mu_a)d\mu_a = \int_{\mu_{a1}} \int_{\mu_{a2}} \cdots \int_{\mu_{an}} f(\mu_{a1}, \mu_{a2}, ..., \mu_{an})d\mu_{a1}d\mu_{a2}...d\mu_{an}$$

and

$$\int_{\mu_b} f(\mu_b)d\mu_b = \int_{\mu_{b1}} \int_{\mu_{b2}} \cdots \int_{\mu_{bn}} f(\mu_{b1}, \mu_{b2}, ..., \mu_{bn})d\mu_{b1}d\mu_{b2}...d\mu_{bn}$$

if μ_a and μ_b are n dimensional parameters. With Eq. 4 the likelihood $P[s|H_\theta]$ is obtained and the problem is solved from a theoretical point of view.

With the MAP approach, instead of eliminating the nuisance parameters as with the marginal Bayesian approach, one considers estimates of their values. If we are not interested in the nuisance parameter's values, this approach is suboptimal (see [18] for a discussion in an analogous situation). Nevertheless, this method can be of interest from a practical point of view, since it can be easier to determine the MAP estimates of the nuisance parameters than integrating as in Eq.4. The MAP estimates of the nuisance parameters are the values which maximize $P[s, \mu_a, \mu_b | H_\theta]$. They are obtained by the following equation:

$$[\mu_a{}^{MAP}[s], \mu_b{}^{MAP}[s]] = \underset{(\mu_a, \mu_b)}{argmax} \ P[s|H_\theta, \mu_a, \mu_b]\pi_a(\mu_a)\pi_b(\mu_b) \qquad (5)$$

where $\underset{y}{argmax} \ (Z)$ is the value of the parameters y which maximizes Z. The estimate of θ can be obtained by maximizing the pseudo-likelihood:

$$\theta^{MAP} = \underset{\theta}{argmax} \ P\left[s, \mu_a{}^{MAP}[s], \mu_b{}^{MAP}[s] | H_\theta\right] \qquad (6)$$

It is worth noting that since $\mu_a{}^{MAP}[s]$ and $\mu_b{}^{MAP}[s]$ are functions of s, the considered criterion $P\left[s, \mu_a{}^{MAP}[s], \mu_b{}^{MAP}[s] | H_\theta\right]$ with the MAP approach is not a likelihood, as it is the case with the marginal Bayesian approach (see Eq.4).

The ML method is analogous to the MAP approach from a technical point of view but is not based on the modelization of the nuisance parameters as random variables. The important consequence is that no prior has to be introduced as for the marginal Bayesian and the MAP approaches. The ML estimates are given by:

$$[\mu_a{}^{ML}[s], \mu_b{}^{ML}[s]] = \underset{(\mu_a, \mu_b)}{argmax} \ P[s|H_\theta, \mu_a, \mu_b] \qquad (7)$$

The estimation of θ is obtained by maximizing the pseudo-likelihood:

$$\theta^{ML} = \underset{\theta}{argmax} \ P\left[s|H_\theta, \mu_a{}^{ML}[s], \mu_b{}^{ML}[s]\right] \qquad (8)$$

One can note that the ML approach is analogous to the MAP approach if a uniform (also denominated non informative) prior for the nuisance parameters is considered. The ML approach is simpler since no prior pdfs are needed (although the obtained solution can be unstable [19]) and in the following we will mainly discuss results for the ML solutions.

2.2 The exponential family

Members of the exponential family include the Bernoulli, Gamma, Gaussian, Poisson, Rayleigh and many other familiar statistical distributions [20]. These distributions can be used to describe realistic situations. The case of binary images is

Law	$pdf : P(x)$	$Parameters : \mu_u$
$Bernoulli$	$p\delta(x) + (1 - p)\delta(1 - x)$	p
$Gamma$	$\frac{x^{c-1}}{p^c \gamma(c)} exp[-x/p]$	p
$Gaussian$	$\frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{(x-m)^2}{2\sigma^2}]$	m, σ
$Poisson$	$\sum_{n \in \mathbf{N}} \delta(x - n)e^{-p}\frac{p^n}{n!}$	p
$Rayleigh$	$\frac{x}{p} exp[-\frac{x^2}{2p}]$	p

Table 2: *pdf of the considered laws of the exponential family and their corresponding parameters. $\delta(x)$ is the Dirac distribution, $\mathbf{N}$ is the set of integers and $n! = n(n - 1)..2.1$.*

simple to handle since the probability law for the gray levels can be described with a Bernoulli pdf. It is also well known that at low photon levels, the noise present in images is described by Poisson pdf (due to the discrete nature of the events, which are the arrivals of photons on the sensor). This situation occurs for example, in astronomical imagery when the exposure time of the sensor is short. Synthetic Aperture Radar (SAR) intensity images are corrupted by a multiplicative noise, also known as speckle [21], which can be described with Gamma pdf. This issue has been widely studied over the past years and it is now well known that in order to obtain efficient algorithms, the statistical properties of the speckle have to be taken into account in the design of image processing algorithms (see [22], [23] and references therein). Ultrasonic medical images correspond to amplitude detection of the incident acoustic field and the speckle noise can be described with a Rayleigh pdf [24]. Finally, we will also discuss the case of optronic images and the relevance of normal laws when a whitening preprocessing is used [25].

Probability density functions (pdf) which belong to the exponential family are defined by [20]:

$$P^\mu(x) = \kappa(x) \ exp\left[\alpha(\mu).\mathbf{f}(x) + \beta(\mu)\right] \tag{9}$$

where $\mu = [\mu_1, \mu_2, ..., \mu_n]^T$ is the vector of parameters of the pdf, $\kappa(x)$ is a scalar function of x while $\alpha(\mu)$ and $\mathbf{f}(x)$ are p-component vector functions of respectively μ and x. We summarize in table 2 the pdfs of the exponential family which will be discussed in the following as well as the parameters which will be considered unknown.

These pdf possess simple sufficient statistics [20]. Let us consider a sample χ_u

of n_u random variables distributed with a pdf $p^{\mu_u}(x)$: $\chi_u = x_1, x_2, ..., x_{n_u}$. A sufficient statistic $T[\chi_u]$ for μ_u is a function of the sample χ_u that contains all the information relevant to estimating the parameter μ_u in the ML sense. If the pdf belongs to the exponential family, the likelihood is:

$$P^\mu(\chi_u) = \prod_{i=1}^{n_u} \kappa(x_i) \ exp\left\{\alpha(\mu).\mathbf{f}(x_i) + \beta(\mu)\right\} \qquad (10)$$

The ML estimate of μ_u is thus:

$$\mu_u{}^{ML} = \underset{\mu}{argmax} \ \left[\prod_{i=1}^{n_u} \kappa(x_i) \ exp\left\{\alpha(\mu).\mathbf{f}(x_i) + \beta(\mu)\right\}\right] \qquad (11)$$

which can also be written:

$$\mu_u{}^{ML} = \underset{\mu}{argmax} \ \left[\exp\left\{\alpha(\mu).T[\chi_u] + n_u \ \beta(\mu)\right\}\right] \qquad (12)$$

with:

$$T[\chi_u] = \sum_{i=1}^{n_u} \mathbf{f}(x_i) \qquad (13)$$

which clearly defines the sufficient statistics of the exponential family. In table 3 we provide the sufficient statistics for the pdfs of the exponential family which will be discussed in the following. For that purpose, let us define S_u the set of n_u random variables from which the parameters are inferred (i.e. the set of pixels from which the unknown parameters are estimated). In particular n_a is the number of pixels in $\mathbf{w}^\theta$ and n_b is the pixel number in the background region $\bar{\mathbf{w}}^\theta$.

For the image processing problems we consider, the likelihood is a function of $\mathbf{w}^\theta$. Let us denote $L(\mathbf{s}, \mathbf{w}^\theta)$ the likelihood obtained with the marginal Bayesian approach or the pseudo likelihood obtained with the MAP or the ML approach, and $\ell(\mathbf{s}, \mathbf{w}^\theta)$ its logarithm. It is easy to show the following property.

Property:
Whatever the adopted approach to deal with the nuisance parameters, the loglikelihood of an hypothesis H_θ is:

$$\ell[\mathbf{s}|H_\theta] = -n_a \ F_a\left[T[\chi_a]\right] - n_b \ F_b\left[T[\chi_b]\right] + G(\mathbf{s}) \qquad (14)$$

with:

$$T[\chi_a] = \sum_{i \in \mathbf{w}^\theta} \mathbf{f}(s_i) \qquad (15)$$

$$T[\chi_b] = \sum_{i \in \bar{\mathbf{w}}^\theta} \mathbf{f}(s_i) \qquad (16)$$

$$G(\mathbf{s}) = \sum_{i \in \mathbf{M}^\theta} \ln[\kappa(s_i)] \qquad (17)$$

where $\bar{M}^\theta = \mathbf{w}^\theta \bigcup \bar{\mathbf{w}}^\theta$. Functions F_a and F_b depend on the considered pdf and on the prior on the nuisance parameters for the marginal Bayesian and MAP approaches. They are equal in the case of a ML estimation of the nuisance parameters.

Law	Parameters :	Sufficient statistics:$\mathbf{T}[\chi_u]$
Bernoulli	$p = T_1/n_u$	$T_1 = \sum_{i \in S_u} s_i$
Gamma	$p = T_1/n_u$	$T_1 = \sum_{i \in S_u} s_i$
Gaussian	$m = T_1/n_u$ $m^2 + \sigma^2 = T_2/n_u$	$T_1 = \sum_{i \in S_u} s_i$ $T_2 = \sum_{i \in S_u} (s_i)^2$
Poisson	$p = T_1/n_u$	$T_1 = \sum_{i \in S_u} s_i$
Rayleigh	$p = T_2/n_u$	$T_2 = \sum_{i \in S_u} (s_i)^2$

Table 3: *Mathematical expressions of the sufficient statistics for the parameters defined in Table 2.*

It is clear that the last term $G(s)$ is independent of the hypotheses H_θ if $\bar{\mathbf{M}}^\theta$ is the image to be analyzed or a subwindow in this image which is chosen independently of H_θ. We will see in the following that the estimation of the likelihood of H_θ is a function of the input image s through the determination of the sufficient statistics. In table 4, the expressions of the varying part of the loglikelihood defined in Eq.14 are provided when the nuisance parameters are estimated with the ML method. We propose in the next sections to illustrate these concepts with different kinds of applications.

3 Application to object location

3.1 Introduction and limitations of the ML approach

In order to perform the important task of detecting and locating a target appearing on a random background, a pattern recognition system must discriminate between the background and the target. The background can thus be considered as noise. This noise is not additive, since it does not affect the target: it is said to be non-overlapping. Classical linear filters have been shown to often fail in presence of such noise [4], and an explanation of this phenomenon has been presented in [6]. Different techniques [26] [9] have been proposed in the past for detection and location of a target with a known internal structure and an unknown uniform illumination in presence of non-overlapping background noise. When the target's gray levels

Law	$F_u(z)$	z
Bernoulli	$z \, ln[z] + (1 - z) \, ln[1 - z]$	$z = T_1/n_u$
Gamma	$ln[z]$	$z = T_1/n_u$
Gaussian	$ln[z]$	$z = T_2/n_u - [T_1/n_u]^2$
Poisson	$-z \, ln[z]$	$z = T_1/n_u$
Rayleigh	$ln[z]$	$z = T_2/n_u$

Table 4: *Mathematical expressions which define the varying part of the loglikelihod (see Eq. 14) in terms of the sufficient statistics defined in Table 3. ln(z) is the neperian logarithm.*

are unknown it is necessary to introduce different approaches [10, 27, 25, 28, 11]. Such a situation can happen when the target is subject to sun reflections in optical images, when temperature changes in infrared images or when only a shape model is available for the location of the target in the input image. In this case, with the proposed solutions [10, 27], the pixel values of both the target and the background have been modeled as random variables with Gaussian pdfs but with different parameters. The only *a priori* knowledge is thus the silhouette of the target, which defines the frontier between the target and the background. These models have been recently generalized to Gamma pdfs [27] and to binary images [29]. We discuss in the following the general solution for the exponential family which includes the previous cases as particular cases.

With the SIR approach the input image model is:

$$s_i = a_i \, w_{i-\theta} + b_i \, \bar{w}_{i-\theta} \tag{18}$$

a and **b** represent the gray levels of, respectively, the target and the background zone. The unknown parameter θ is now simply the position of the object in the scene. The ML solution for the estimation of the location θ can be written (see Eq. 14):

$$\theta^{ML} = \arg\max_\theta \ell[\mathbf{s}|H_\theta]$$

$$= \arg\max_\theta \left[-n_a \, F_a \left[\frac{1}{n_a} \sum_{i \in \mathbf{w}^\theta} \mathbf{f}(s_i) \right] - n_b \, F_b \left[\frac{1}{n_b} \sum_{i \, \bar{\mathbf{w}}^\theta} \mathbf{f}(s_i) \right] \right] \tag{19}$$
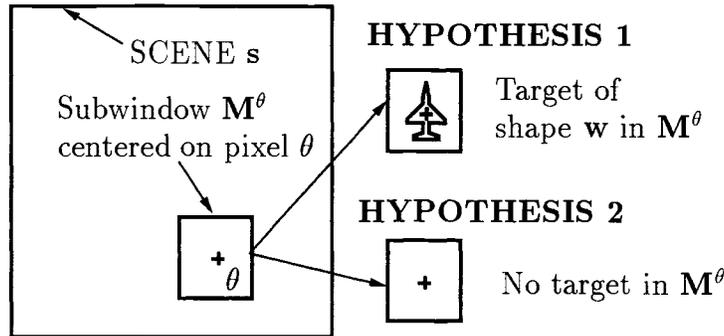
Figure 1: *Sketch of the two hypotheses considered in the MLRT scheme.*

Here again, the functions F_a and F_b depend on the considered pdf and on the prior on the nuisance parameters for the marginal Bayesian and MAP approaches but are equal in the case of the ML estimation of these parameters. The mathematical equations for Bernoulli, Gaussian, Gamma, Poisson and Rayleigh pdfs can be easily obtained from tables 2, 3 and 4.

The main practical problem with the SIR models is that the input image is assumed to be composed of two homogeneous random fields a_i and b_i. Furthermore, the random variables are assumed to be independently distributed. These conditions may not be fulfilled in real-world images. We discuss in the following two techniques in order to overcome these limitations.

3.2   The maximum likelihood ratio test (MLRT) approach

In the SIR image model, the background region $\bar{\mathbf{w}}^\theta$, that is, the whole image but the target, is considered to have homogeneous statistics. This is often a non-realistic assumption since real-world backgrounds are in general better modeled with several zones having different average values. In order to overcome this problem, we will estimate the statistics in a small subwindow $\mathbf{M}^\theta$ centered on the assumed target location θ (see Fig. 1). Indeed, if we consider a sufficiently small subwindow, the hypothesis that the background is homogeneous becomes a better approximation.

In order to better understand the method we propose, let us temporarily set aside the object location problem and let us consider the simpler problem of object detection. It consists in determining if there is an object of shape $\mathbf{w}$ in the center of the sub-image $\mathbf{M}^\theta$ or not. More precisely, we want to discriminate between the two following hypotheses :

- Hypothesis H_0 : the window $\mathbf{M}^\theta$ contains only background noise $\mathbf{b}$, so that

$$\forall i \in \mathbf{M}^\theta \quad s_i = b_i \tag{20}$$

- Hypothesis H_1 : the target is present in the center of the window $\mathbf{M}^\theta$, so that

$$\forall i \in \mathbf{M}^\theta \quad s_i = w_{i-\theta}\, a_i + \bar{w}_{i-\theta}\, b_i \tag{21}$$

Note that in this section, $\bar{\mathbf{w}}^\theta$ will denote the part of the complementary of $\mathbf{w}^\theta$ belonging to $\mathbf{M}^\theta$. In other words, $\mathbf{M}^\theta = \mathbf{w}^\theta + \bar{\mathbf{w}}^\theta$.

A very classical method for determining the best choice between these two hypotheses is the maximum-likelihood ratio test [30]. It consists in computing the likelihoods $L(H_0, \theta)$ and $L(H_1, \theta)$ of both hypotheses, and taking their ratio $\tau(\theta) = L(H_1, \theta)/L(H_0, \theta)$. Then select a threshold value τ_0, and perform the following test :

- if $\tau(\theta) > \tau_0$, there is a target in the center of $\mathbf{M}^\theta$,

- else there is no target.

The value of the threshold τ_0 sets a compromise between the probability of detection and the probability of false alarm. Using this method, we can determine if the target is present or not at each location θ. If there may be several targets in the image, it is thus possible to determine their locations.

Let us now return to the problem of object location, which is slightly different : We assume that we know that there is only one target in the image (which can be the case in tracking applications for example), and we want to determine its location. In order to do so, we can extend the previously described detection algorithm to a location algorithm by choosing as the estimate of the target location the position which maximizes $\tau(\theta)$. In other words, the estimated location will be :

$$\theta^{MLRT} = \arg\max_\theta \tau(\theta) \qquad (22)$$

In the following, we will call this estimation approach the *"maximum likelihood ratio test"* (MLRT). This procedure is a heuristic extension of the optimal detection algorithm. Note that similar procedures have been used for locating edges in optical images (with Gaussian grey level statistics) [31] and in Synthetic Aperture Radar images (with Gamma grey level statistics) [22].

We shall now specify the expression of the likelihood ratio $\tau(\theta)$ for a SIR image belonging to the exponential family:

$$ln[\tau(\theta)] \quad = \quad -n_a \, F_a[\frac{1}{n_a} \sum_{i \in \mathbf{w}^\theta} \mathbf{f}(s_i)] - n_b \, F_b[\frac{1}{n_b} \sum_{i \in \bar{\mathbf{w}}^\theta} \mathbf{f}(s_i)] \qquad (23)$$

$$+n_c \, F_b[\frac{1}{n_c} \sum_{i \in \mathbf{M}^\theta} \mathbf{f}(s_i)] \qquad (24)$$

where n_c is the number of pixels of the scanning subwindow $\mathbf{M}^\theta$ and thus $n_c = n_a + n_b$. Here again, the functions F_a and F_b are dependent of the considered pdf and the prior on the nuisance parameters for the marginal Bayesian and MAP approaches but are equal in the case of a ML estimation of these parameters.

We now specialize Eq. 24 to particular pdf's belonging to the exponential family when the ML estimation of the nuisance parameters is considered. For simplicity reasons, let us introduce the following notations:

$$\begin{array}{l} t_\ell^{(a)}(\theta) = 1/n_a \sum_{i \in \mathbf{w}^\theta} (s_i)^\ell \\ t_\ell^{(b)}(\theta) = 1/n_b \sum_{i \in \bar{\mathbf{w}}^\theta} (s_i)^\ell \\ t_\ell^{(c)}(\theta) = 1/n_c \sum_{i \in \mathbf{M}^\theta} (s_i)^\ell \end{array} \qquad (25)$$

where $\ell = 1$ or 2.

In the Bernoulli case, Eq. 24 becomes

$$
\begin{aligned}
ln[\tau(\theta)] = \quad & -n_a \left[t_1^{(a)}(\theta) \; ln[t_1^{(a)}(\theta)] + (1 - t_1^{(a)}(\theta)) \; ln[1 - t_1^{(a)}(\theta)] \right] \\
& -n_b \left[t_1^{(b)}(\theta) \; ln[t_1^{(b)}(\theta)] + (1 - t_1^{(b)}(\theta)) \; ln[1 - t_1^{(b)}(\theta)] \right] \\
& +n_c \left[t_1^{(c)}(\theta) \; ln[t_1^{(c)}(\theta)] + (1 - t_1^{(c)}(\theta)) \; ln[1 - t_1^{(c)}(\theta)] \right] \;,
\end{aligned}
\tag{26}
$$

in the Gaussian case,

$$
\begin{aligned}
ln[\tau(\theta)] \quad = \quad & -n_a \; ln \left[t_2^{(a)}(\theta) - (t_1^{(a)}(\theta))^2 \right] - n_b \; ln \left[t_2^{(b)}(\theta) - (t_1^{(b)}(\theta))^2 \right] \\
& +n_c \; ln \left[t_2^{(c)}(\theta) - (t_1^{(c)}(\theta))^2 \right] \;,
\end{aligned}
\tag{27}
$$

in the Gamma case,

$$
ln[\tau(\theta)] = -n_a ln \left[t_1^{(a)}(\theta) \right] - n_b ln \left[t_1^{(b)}(\theta) \right] + n_c ln \left[t_1^{(c)}(\theta) \right] \;,
\tag{28}
$$

in the Poisson case

$$
\begin{aligned}
ln[\tau(\theta)] \quad = \quad & n_a \; t_1^{(a)}(\theta) \; ln \left[t_1^{(a)}(\theta) \right] + n_b \; t_1^{(b)}(\theta) \; ln \left[t_1^{(b)}(\theta) \right] \\
& -n_c \; t_1^{(c)}(\theta) \; ln \left[t_1^{(c)}(\theta) \right] \;,
\end{aligned}
\tag{29}
$$

and in the Rayleigh case

$$
ln[\tau(\theta)] = -n_a ln \left[t_2^{(a)}(\theta) \right] - n_b ln \left[t_2^{(b)}(\theta) \right] + n_c ln \left[t_2^{(c)}(\theta) \right]
\tag{30}
$$

3.3 The whitening process

In some real-world images, the statistics of both the target and the background cannot be approximated with good precision with uncorrelated random fields. In these situations, the SIR filter is then suboptimal and can fail. In figure 2, two scenes and the maximum of each line of their respective generalized correlation planes obtained with the SIR technique (i.e. $\ell(\mathrm{s}|H_\theta)$) are shown. The maximum of the correlation plane represents the estimated location of the target. In scene (a), the pdf's of both the target and the background are white and Gaussian whereas those of scene (b) are also Gaussian but correlated. One can note that the SIR algorithm adapted to white Gaussian statistics fails on scene (b) whereas it is able to locate the target on scene (a).

We want to design an optimal algorithm for the location of a random correlated target appearing on a random correlated background. The main problem consists in finding texture models that characterize real situations and for which the optimal solution is mathematically simple. Such a method has been recently designed using the same random Markov field model for both the target and the background [32]. This case represents a difficult, but particular situation.

In this subsection, we propose to apply a preprocessing to the input image in order to obtain an image with white Gaussian textures and then to apply the SIR
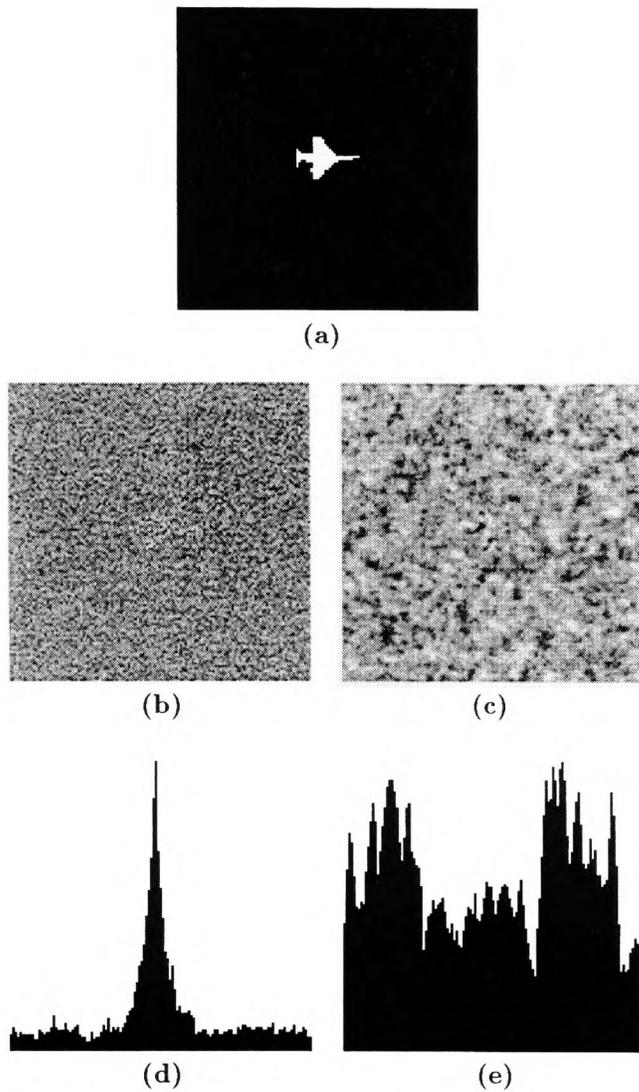
Figure 2: (a) : *Reference shape.* (b) : *Example of a scene with white Gaussian textures for the target and the background. The shape of the target is the reference shape in (a).* (c) : *Example of a scene with correlated Gaussian textures for the target and the background. The shape of the target is the reference shape in (a).* (d) : *Result of processing (b) with the SIR algorithm adapted to white Gaussian statistics.* (e) : *Result of processing (c) with the SIR algorithm adapted to white Gaussian statistics. Note : (d) and (e) are plots of the maximum of each line of the output plane of the considered method.*

method which is optimal in that case. However, as soon as the textures of both the target and the background are strongly correlated, the preprocessing introduces a third region in the preprocessed image, which characterizes the frontier between the target and the background. Following reference [25] we will thus discuss how to model this region with a white Gaussian random field and we design a SIR filter that takes into account the three regions (i.e. the background, the target and the frontier).

The Fourier Transform of s is denoted $\hat{s}$ (or $\hat{s}(\nu)$ at frequency ν), z^* is the complex conjugate value of z and $|z|$ its modulus. We define the whitening filter in the Fourier domain by :

$$\hat{h}(\nu) = \frac{1}{\epsilon + |\hat{s}(\nu)|} \tag{31}$$

where ϵ is a small positive constant introduced as a regularization parameter which avoids divergence when $|\hat{s}(\nu)|$ is close or equal to zero. The Fourier Transform $\hat{v}$ of the preprocessed image v is thus:

$$\hat{v}(\nu) = \hat{h}(\nu)\hat{s}(\nu). \tag{32}$$

One can note that since s and $\hat{h}$ are real, v is also real. It is easy to show that the square modulus of $\hat{v}$ is approximately constant. We can conjecture that the pixels of the preprocessed image z are approximately Gaussian uncorrelated variables. In figure 3, we show a target with a correlated texture which appears on a random correlated background and the obtained preprocessed image. One can show 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 and if the nuisance parameters are estimated in the ML sense, the SIR method leads to (see Eq. 14):

$$\ell[s|H_\theta] = -n_a \ln \left[\tilde{t}_2^{(a)}(\theta) - (\tilde{t}_1^{(a)}(\theta))^2 \right] - n_b \ln \left[\tilde{t}_2^{(b)}(\theta) - (\tilde{t}_1^{(b)}(\theta))^2 \right] \tag{33}$$

where

$$\tilde{t}_\ell^{(a)}(\theta) = \frac{1}{n_a} \sum_{i \in \mathbf{w}^\theta} (z_i)^\ell$$

$$\tilde{t}_\ell^{(b)}(\theta) = \frac{1}{n_b} \sum_{i \in \bar{\mathbf{w}}^\theta} (z_i)^\ell \tag{34}$$

where $\ell = 1$ or 2.

However, as one can remark in figure 3, the preprocessing can introduce three regions in the preprocessed image. Indeed, as soon as the textures are strongly correlated, a frontier appears between the target and the background.

Let f denote that frontier (and respectively a the target and b the background) and let $\mathbf{w}^f$ (resp. $\mathbf{w}^a$ and $\mathbf{w}^b$) define the new disjoint window functions composed of n_f (resp. n_a and n_b) pixels so that w_i^f (resp. w_i^a and w_i^b) is equal to one within the frontier (resp. the target and the background) and to zero elsewhere when the target is located at the center of the image. We thus propose to describe the preprocessed image z in the following way [33] :

$$z_i = a_i w_{i-\theta}^a + f_i w_{i-\theta}^f + b_i w_{i-\theta}^b \tag{35}$$
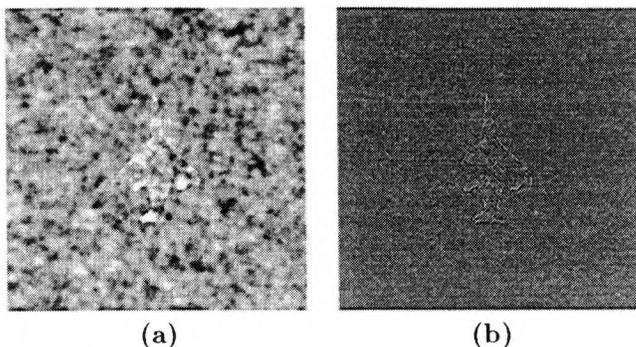
Figure 3: **(a)** : *Example of a scene with strongly correlated Gaussian textures for both the target and the background.* **(b)** : *Whitened version of (a).*

when the target is supposed to be centered on the θth pixel of the image.

Using an analogous approach as previously, we can design a SIR filter that takes into account three regions. This leads to :

$$
\begin{aligned}
\ell[\mathbf{s}|H_\theta] &= -n_a \ln\left[\tilde{t}_2^{(a)}(\theta) - (\tilde{t}_1^{(a)}(\theta))^2\right] - n_b \ln\left[\tilde{t}_2^{(b)}(\theta) - (\tilde{t}_1^{(b)}(\theta))^2\right] \\
&\quad -n_f \ln\left[\tilde{t}_2^{(f)}(\theta) - (\tilde{t}_1^{(f)}(\theta))^2\right]
\end{aligned}
\tag{36}
$$

where $\tilde{t}_\ell^{(a)}(\theta)$, $\tilde{t}_\ell^{(b)}(\theta)$ and $\tilde{t}_\ell^{(f)}(\theta)$ are defined as is Eq.34. All these quantities can be determined by correlating binary masks with images $\mathbf{z}$ and $\mathbf{z}^2$ [25]. They can be obtained with a simple optoelectronical architecture or using FFT algorithm applied to the images z_i and $z_i{}^2$.

3.4   The implementation issue

An interesting point is that $\ell[\mathbf{s}|H_\theta]$ in the standard SIR approach and $ln[\tau(\theta)]$ in the MLRT approach can easily be rewritten using correlation operations. Let us consider the MLRT approach, and let $[f \star g]_i$ denote the correlation between f and g:

$$
[f \star g]_i = \sum_j f_{i+j} g_j
\tag{37}
$$

and let $\mathbf{w} = \mathbf{w}^0$. Eq. 25 becomes :

$$
\begin{aligned}
t_\ell^{(a)}(\theta) &= \tfrac{1}{n_a}[\mathbf{w} \star (s)^\ell]_\theta \\
t_\ell^{(b)}(\theta) &= \tfrac{1}{n_b}[\bar{\mathbf{w}} \star (s)^\ell]_\theta \\
t_\ell^{(c)}(\theta) &= \tfrac{1}{n_c}[\mathbf{M} \star (s)^\ell]_\theta
\end{aligned}
\tag{38}
$$

where we remember that $\mathbf{M} = \mathbf{w} + \bar{\mathbf{w}}$. Since the most intensive computations are involved in $t_\ell^{(u)}(\theta)$ with $u = a, b, c$ or f, this new formulation is very attractive because it is closely connected to the detection architecture described in [10]. Indeed,

the detection and location steps require the same correlation functions. A simple optoelectronical architecture could thus perform the detection and/or location with this kind of hardware.

4 Application to segmentation

4.1 Introduction

An important goal of computational vision and image processing is to automatically recover the shape of objects from various types of images. Over the years, many approaches have been developed to reach this goal and in this section, we focus on the segmentation of a unique object in the scene. The unknown parameter θ is now the shape of the object in the scene.

A classical approach consists in detecting edges and linking them in order to determine the shape of the object presents in the image. However this approach does not use the knowledge that the object is simply connected. On the other hand, deformable models (also called "snakes") incorporate knowledge about the shape of the object from the start. Broadly speaking, a snake is a curve which has the ability to evolve (under the influence of a mathematical criterion) in order to match the contour of an object in the image. The first "snakes" [14] were driven by the minimization of a function in order to move them towards desired features, usually edges. This approach and its generalization [34] [35] [36] are edge-based in the sense the information used is strictly along the boundary. They are well adapted to a certain class of problems, but they can fail in presence of strong noise.

The SIR-based snake we will describe in the following belongs to the deformable template methods, which are parametric shape models with relatively few degrees of freedom. They constitute another interesting approach to recover the shape of an object [37] [38] [16]. The template is matched to an image, in a manner similar to the snake, by searching the value of a parameter vector θ (i.e. the node positions) that minimizes an appropriate mathematical criterion. One can cite for example strategies based on the consideration of the inner and the outer regions defined by the snake, which have been recently investigated [15] [16] [39] [40]. It is interesting to note that a statistical processing method can take full advantage of many suitable descriptions of the measured signals (see for example [41] [12] [24] [42] [43]).

The SIR approach allows one to determine this appropriate criterion. First, we generalize the approaches proposed in [12] [42] [44] to different statistical laws which belong to the exponential family and which are well adapted to describe physical situations. This technique is actually an extension of the optimal detection approach introduced in [10] and generalized in the previous sections.

4.2 The SIR snake model

The purpose of segmentation is therefore to estimate the most likely shape $\mathbf{w}^\theta$ for the target in the scene. Note the difference with the optimal location problem of section 3 where the silhouette of the target was known whereas its position had to be found. To achieve the shape estimation issue, we use a k-node polygonal active contour that defines the boundary of the shape. $\mathbf{w}^\theta$ is now a polygon-bounded

support function, one-valued on and within the snake and zero-valued elsewhere and θ is the set of the positions of each node of the contour. Let us consider the different hypotheses H_θ that consist in assessing a shape $\mathbf{w}^\theta$ to the target by assessing a position to each node of the contour, so that we can write:

$$s_i = a_i w_i^\theta + b_i [1 - w_i^\theta] \tag{39}$$

The optimal choice for $\mathbf{w}$ is the one which maximizes the conditional probability $P[H_\theta|\mathbf{s}]$. The ML estimation of the shape (i.e. θ) is obtained by maximizing the *a priori* probability $P[\mathbf{s}|H_\theta]$; which corresponds to the likelihood of the hypothesis.

Under the previous assumptions we can now specify the expression of the likelihood $\ell[\mathbf{s}|H_\theta]$ for a SIR image which belongs to the exponential family.

$$\ell[\mathbf{s}|H_\theta] = -n_a \, F_a \left[\frac{1}{n_a} \sum_{i \in \mathbf{w}^\theta} \mathbf{f}(s_i) \right] - n_b \, F_b \left[\frac{1}{n_b} \sum_{i \in \bar{\mathbf{w}}^\theta} \mathbf{f}(s_i) \right] \tag{40}$$

Here again, the functions F_a and F_b are dependent of the considered pdf and the prior on the nuisance parameters for the marginal Bayesian and MAP approaches but are equal in the case of a ML estimation of these parameters.

We now illustrate this result for some particular cases of the pdf family and we use the same notations as in Eq. 25.
In the Bernoulli case, Eq. 40 becomes :

$$\begin{aligned} \ell[\mathbf{s}|H_\theta] = \quad &-n_a \left[t_1^{(a)}(\theta) \, ln[t_1^{(a)}(\theta)] + (1 - t_1^{(a)}(\theta)) \, ln[1 - t_1^{(a)}(\theta)] \right] \\ &-n_b \left[t_1^{(b)}(\theta) \, ln[t_1^{(b)}(\theta)] + (1 - t_1^{(b)}(\theta)) \, ln[1 - t_1^{(b)}(\theta)] \right] \;, \end{aligned} \tag{41}$$

in the Gaussian case,

$$\ell[\mathbf{s}|H_\theta] = -n_a \, ln \left[t_2^{(a)}(\theta) - (t_1^{(a)}(\theta))^2 \right] - n_b \, ln \left[t_2^{(b)}(\theta) - (t_1^{(b)}(\theta))^2 \right] \;, \tag{42}$$

in the Gamma case,

$$\ell[\mathbf{s}|H_\theta] = -n_a ln \left[t_1^{(a)}(\theta) \right] - n_b ln \left[t_1^{(b)}(\theta) \right] \;, \tag{43}$$

in the Poisson case,

$$\ell[\mathbf{s}|H_\theta] = n_a \, t_1^{(a)}(\theta) \, ln \left[t_1^{(a)}(\theta) \right] + n_b \, t_1^{(b)}(\theta) \, ln \left[t_1^{(b)}(\theta) \right] \;, \tag{44}$$

and in the Rayleigh case,

$$\ell[\mathbf{s}|H_\theta] = -n_a \, ln \left[t_2^{(a)}(\theta) \right] - n_b \, ln \left[t_2^{(b)}(\theta) \right] \tag{45}$$

One can note that the whitening process introduced in the previous section can be also used in order to obtain white random fields well described with Gaussian pdf [13].

4.3 The implementation issue

The window function $\mathbf{w}^\theta$ that optimizes the criterion $\ell[\mathbf{s}|H_\theta]$ realizes the ML optimal segmentation of the target in the scene. The technical problem is thus to find the value of θ which maximizes $\ell[\mathbf{s}|H_\theta]$ (also denoted $l(\theta)$ in the following) :

$$\theta^{ML} = \underset{\theta}{argmax}\ \ell[\mathbf{s}|H_\theta] \tag{46}$$

We use a stochastic iterative algorithm to perform the optimization of $l(\theta)$ and thereby the segmentation. At each iteration m of the process, carry out the following steps:

- Consider a new shape θ^{m+1} by randomly moving a node of the polygon. This consequently defines a new window function $\mathbf{w}^{\theta^{m+1}}$.

- Accept this new shape if $l(\theta^{m+1}) > l(\theta^m)$ and refuse it otherwise.

This process is continued until $l(\theta^m)$ does not increase anymore.

An interesting point is that $l(\theta^m)$ can easily be rewritten using correlation operations. Let $[f \star g]_0$ denote the central value of the correlation between f and g:

$$[f \star g]_0 = \sum_j f_i\ g_j \tag{47}$$

One thus has:

$$\begin{aligned} t_\ell^{(a)}(\theta) &= \tfrac{1}{n_a}[\mathbf{w} \star (s)^\ell]_0 \\ t_\ell^{(b)}(\theta) &= \tfrac{1}{n_b}[\bar{\mathbf{w}} \star (s)^\ell]_0 \end{aligned} \tag{48}$$

Note the similarity between Eqs. 48 and 38. The detection, location and segmentation steps require the same correlation functions. A simple optoelectronical architecture could thus perform these tasks with the same hardware. As will be shown in Section 5.5, joint utilization of location and segmentation algorithms enables us to perform efficient target tracking.

4.4 Generalization of the SIR segmentation approach

Constrained deformation

In the previous section, the shape of the snake was not constrained, and could converge to any arbitrary polygon. In some applications, one may have some *a priori* knowledge about the object's shape, and thus constrain the evolution of the snake to a smaller class of possible shapes. This enables faster and more robust segmentation.

To formalize this approach, let θ be the node locations and let us consider a set of transformations $K_\alpha(\theta)$ with $\alpha \in \mathcal{S}$ where $\mathcal{S}$ is the set of possible values of α. In the case of the location task, the transformation is a translation of parameter α and thus:

$$K_\alpha(\theta) = \theta + \alpha \tag{49}$$

$$w_{i-K_\alpha(\theta)} = w_{i-\theta-\alpha} \tag{50}$$

More interesting can be the case of in-plane rotation R_α since the estimation of the target orientation can be performed with a more efficient technique than the general snake algorithm of the previous subsections. Indeed, instead of randomly moving the nodes one can select among all the rotated versions of $\mathbf{w}$ the angle α which maximizes the likelihood:

$$\alpha^{ML} = \underset{\alpha \in \mathcal{S}}{argmax} \; \ell[\mathbf{s}|H_{R_\alpha(\theta)}] \tag{51}$$

This concept can be generalized to other transformations such as isotropic or anisotropic scaling.

Recognition

Let us assume that the purpose is to recognize the object or, in other words, to discriminate between different classes. For example, one can imagine that the purpose is to determine whether the target is a car, a truck or a bus. One can now consider that θ belongs to the discrete set $\mathcal{D}$ of the possible classes of objects. For the above considered example one has $\mathcal{D} = \{car, \; truck, \; bus\}$. The recognition is thus obtained with:

$$\alpha^{ML} = \underset{\alpha \in \mathcal{D}}{argmax} \; \ell[\mathbf{s}|H_\alpha] \tag{52}$$

where $\ell[\mathbf{s}|H_\alpha]$ is a nonlinear function of the intercorrelation of s_i and of $(s_i)^2$ with the shapes w^α of the reference objects.

5 Simulation results

We propose in this section some numerical simulations to illustrate the performance of the location and segmentation algorithms described in this paper. We consider different noise statistics belonging to the exponential family and demonstrate the efficiency of the proposed algorithms on synthetic and real-world images. We also show how the location and the segmentation algorithms can be used together to efficiently track objects in images sequences.

5.1 Binary images

The image in figure 4.a represents an object (a bird) appearing against a complex background. Suppose that this image is to be processed with an optical correlator in which the input image is displayed on a binary spatial light modulator. We need to binarize the image before processing it. In many instances, it has been noticed that it was more efficient to edge-enhance an image before binarizing it ; this operation increases its contrast, making it easier to find a good threshold. The result of edge-enhancing and binarizing figure 4.a is represented in figure 4.b. Note that in binarized real-world images, the background noise is often non-homogeneous. For this type of images, the MLRT is thus more efficient than the ML algorithm [29].

The result of processing figure 4.b with the MLRT algorithm adapted to Bernoulli statistics is shown in Figure 4.d.

By looking more accurately at the binarized image in Figure 4.b, we can see that it is very noisy. This is because a low threshold has been chosen. It is better to have a low threshold since almost all the information-carrying edges are included in the images. Very little information is thus lost, but the drawback is that a lot of spurious edges remain after the binarization step. These edges are in general non-homogeneously distributed over the image, and this makes it important to use location algorithms robust to non-homogeneous background noise, such as the MLRT.

Figure 5 represents segmentation results on two binarized real-world images corrupted with additive Gaussian noise. The images in the left column display the initial shape of the snake. The images in the right column represent the snake after convergence. We can see that the searched shape has been correctly segmented. Note that the initial shape does not need to be very close to the true one for the snake to converge properly. This robustness to snake shape initialization is an important feature of the proposed algorithm in real-world applications.

5.2   Speckled images

Figure 6.a displays two tank-shaped small targets (78 pixels) appearing on a non-homogeneous background with exponential statistics. This background has been generated with the method described in Ref. [29], where we have replaced the Bernoulli variates with exponential variates. The tank in the upper right quadrant has been rotated by 10° with respect to the reference object. The result of processing the scene with the MLRT algorithm adapted to speckle statistics appears in Figure 6.b. We can see that the MLRT, as most correlation-based algorithms, is robust to small deformations of the target with respect to the reference object.

5.3   Low flux images

Figure 7.a displays a real-world image containing a boat appearing against a mountain background with atmospheric blurring. Figure 7.b represents the same image synthetically perturbed with Poisson noise, simulating for example photon-limited imaging. Figure 7.d shows the result of processing this image with the MLRT algorithm adapted to Poisson noise.

Figure 8 also represents a real image synthetically perturbed with some amount of Poisson noise. The car is segmented using the snake energy adapted to Poisson noise.

5.4   Optronic images

Figure 9.a is a synthetic image representing an airplane on a contrasted urban background. The whole scene is severely blurred. Note that the target gray levels are nonuniform. They are not known *a priori*, since the only information used by the algorithm is the binary shape displayed in figure 9.c. Fig 9.b represents a whitened version of the scene. It can be shown that the statistics of the gray levels in the whitened scene are approximately uncorrelated and Gaussian. The
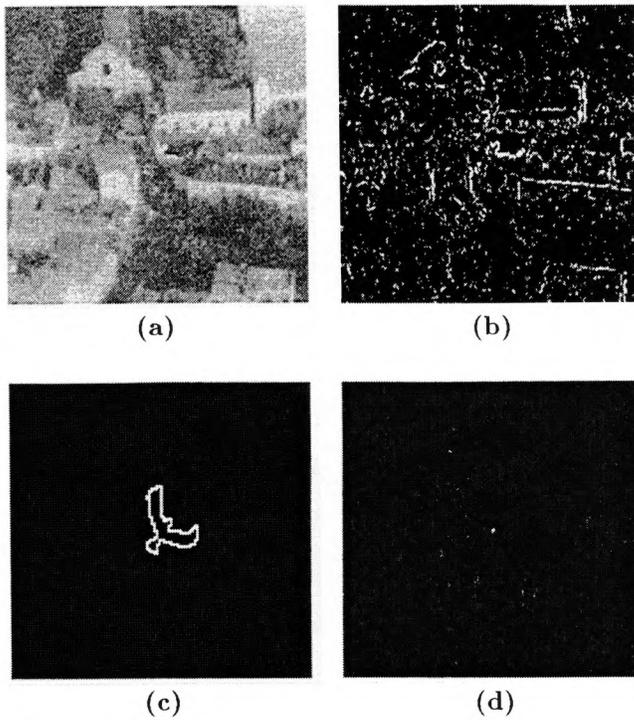
Figure 4: **(a)** : *Synthetic scene with an object (eagle) appearing on an urban background. The whole scene is corrupted with white Gaussian additive noise.* **(b)** : *Scene edge enhanced with Sobel operator and binarized.* **(c)** : *Reference object* **w**. **(d)** : *Result of processing (b) with the MLRT algorithm adapted to Bernoulli statistics. The reference object was (c), and the window* **M** *was constructed by dilating (c) two times with a* 3×3 *structuring element.*
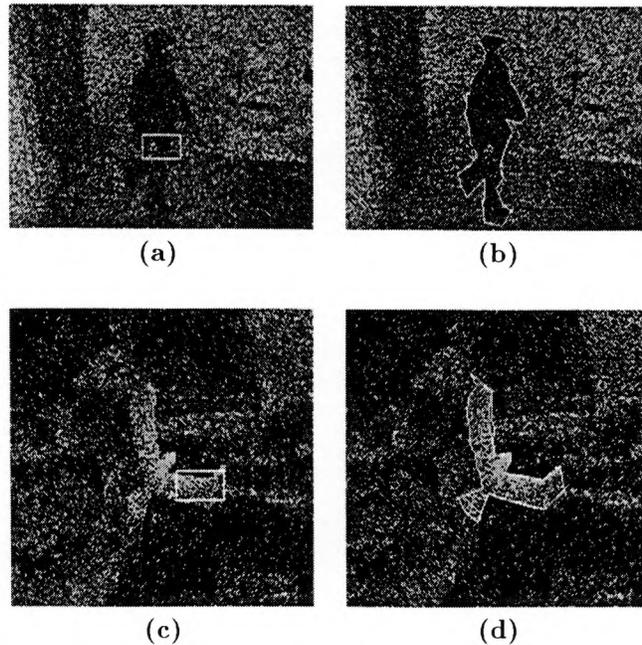
Figure 5: **(a)** and **(c)** : *Binarized versions of gray level real-world scenes. The white rectangle represents the initial shape of the snake.* **(b)** and **(d)** : *Snake after convergence.*
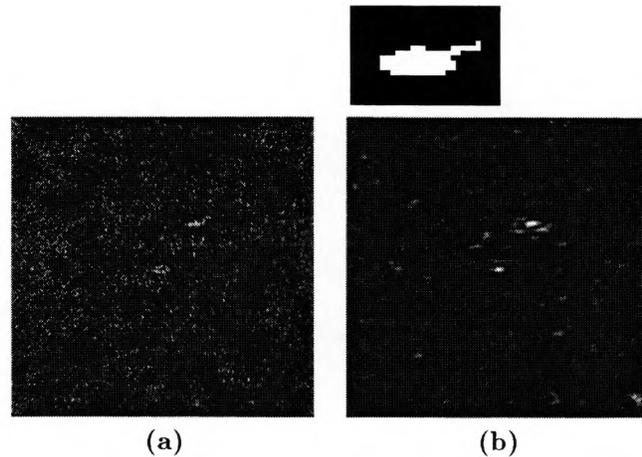
Figure 6: **(a)** : *Scene containing the searched object in the middle, and a rotated version (10°) of the searched object in the upper right quadrant. The background is nonhomogeneous with exponential statistics.* **(b)** : *Above : Reference object* **w**. *Below : Result of processing (a) with the MLRT algorithm adapted to Gamma statistics. The window* **M** *was constructed by dilating the reference object three times with a 3 × 3 structuring element.*
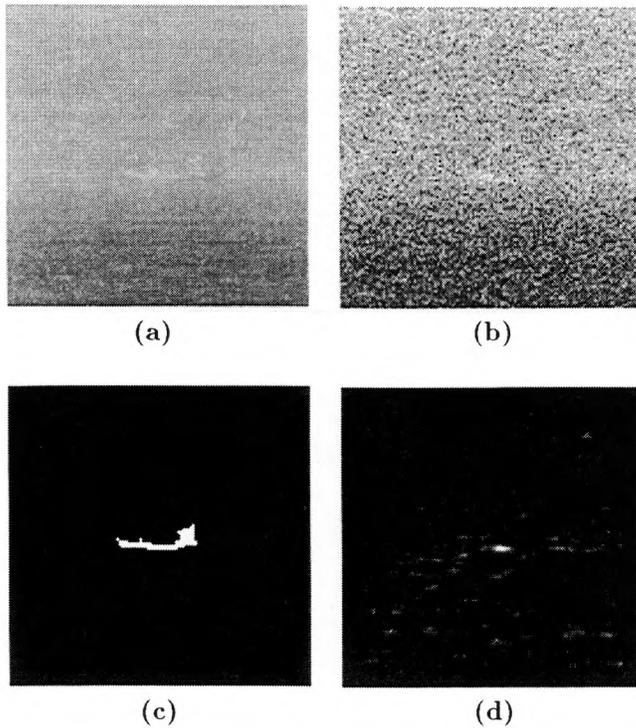
Figure 7: **(a)** : *Real-world gray level scene.* **(b)** : *Scene (a) perturbed with Poisson noise* **(c)** : *Reference object* **w**. **(d)** : *Result of processing (b) with the MLRT algorithm adapted to Poisson statistics. The reference object was (c), and the window* **M** *was constructed by dilating (c) three times with a* 3×3 *structuring element.*
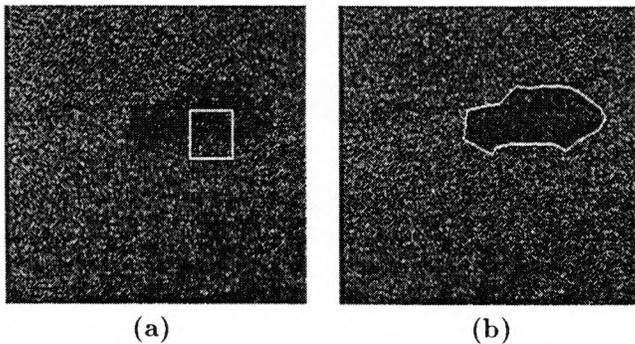
Figure 8: **(a)** : *Real-world scene synthetically corrupted with Poisson noise. The white rectangle represents the initial shape of the snake.* **(b)** : *Snake after convergence.*
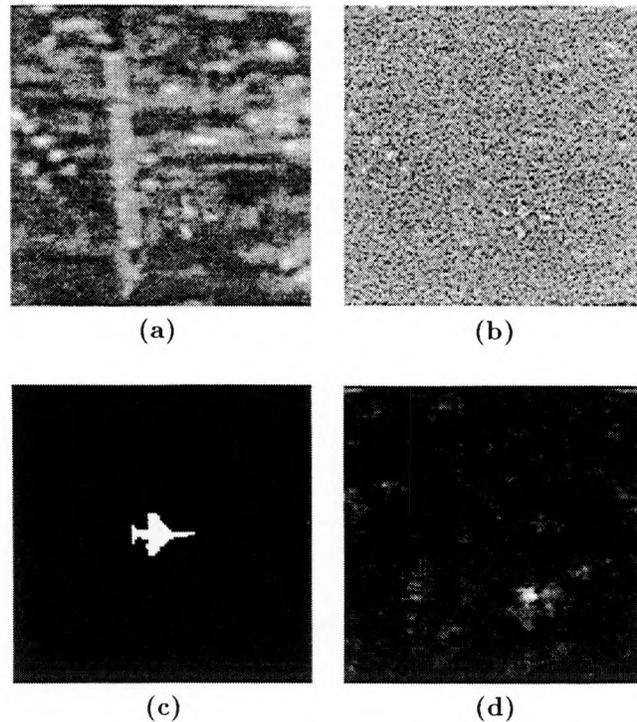
Figure 9: (a) : *Synthetic scene with an airplane appearing on an urban background. The whole scene is blurred and corrupted with white Gaussian additive noise.* (b) : *Whitened version of scene (a).* (c) : *Reference object* **w**. (d) : *Result of processing (b) with the ML algorithm adapted to Gaussian statistics. The reference object was (c).*

ML algorithm adapted to white Gaussian statistics is applied to the whitened image. The obtained result is displayed in figure 9.d. We can see that we are able to correclty locate the target despite its low contrast and the severely cluttered background.

Figure 10.a also represents an airplane on a contrasted urban background. The whole image is severely blurred. This means that edge-based snake techniques [14, 34, 35, 36] would not be efficient, since the edges between the target and the background are not sharper than the edges internal to the background. The proposed region-based snake method, which relies on all target and background pixels is able to segment the image, as can be seen in figure 10.b. Note that the snake has been applied to the whitened version of figure 10.a.

Figure 10.c represents a real-world image of a car on a road. Here again, the snake is applied to the whitened version of the scene. Note that the the snake has correctly converged although its initial shape (see figure 10.c) was very different from the true one. This is a further proof of the robustness of the proposed algorithm to the initial shape of the snake.
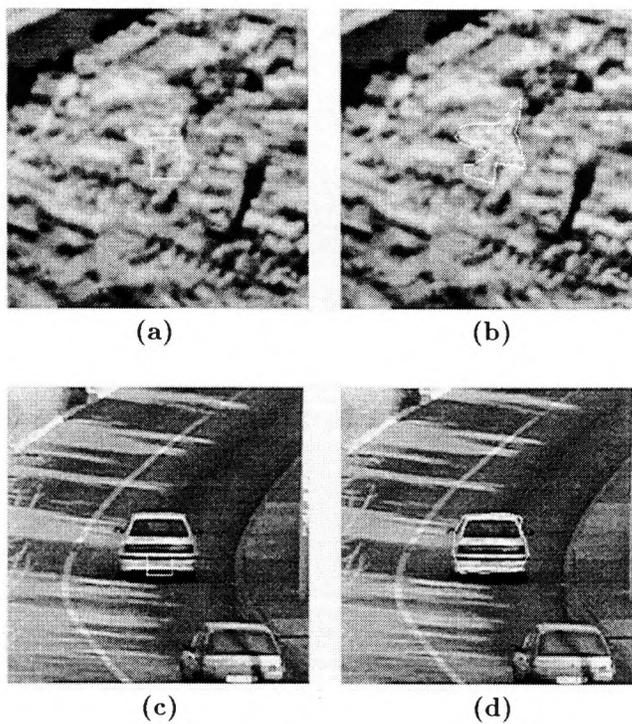
Figure 10: **(a)** and **(c)** : *Synthetic (a) and real-world (c) scenes. The white rectangle represents the initial shape of the snake.* **(b)** and **(d)** : *Snake after convergence (on the whitened version of the corresponding scene)*

5.5   Applications to tracking

We now illustrate the feasability of using cooperatively location and segmentation approaches to achieve efficient target tracking on image sequences, even if the shape of the target changes during the sequence. Assume that on the image acquired at time t, the target has been segmented using the proposed snake method. This segmentation produces a binary reference shape that enables us to locate the target in the image acquired at time $t+1$, e.g. with the MLRT algorithm. This is possible since the MLRT is robust to limited deformation of the target with respect to the reference object. It can thus locate the target even if its shape has slightly changed compared to the previous frame. We then use the obtained position estimate for centering the binary reference. This centered reference is used as the initial shape of a snake which has to converge to the new shape of the object. This process is repeated until the end of the sequence. In summary, using jointly MLRT and snake algorithms consists in first determining the object location (which corresponds to very constrained variation of the shape), and then in segmenting the shape whose position is approximately known. In many instances, the shape variations from one image to the next are small, and only few snake iterations are needed to converge to the new shape.

Let is first consider the problem of tracking walking persons. Typical images can be seen in figure 11. We can note that due to the walk, the apparent shape of the person changes during the sequence. In order to get rid of the influence of the structured background, we make an acquisition of the scene without people. We then substract each frame to this reference frame after having registered them. The MLRT algorithm for location and the snake are applied to this difference image. Such a procedure can be useful in surveillance applications for example. We can see in figure 11 that the object is correctly located and segmented in the image sequence.

The second example consists in tracking a car driving on a highway, and moving away from the camera. Due to this movement, the shape of the target varies during the sequence, and the binary reference used by the location algorithm must be periodically refreshed using the snake segmentation algorithm. We can see in figure 12 the result of applying the proposed tracking method to this sequence. Here again, each frame is substracted to an image of the highway without cars, and the segmentation algorithm is applied to a partially whitened version of this difference image. In order to show the robustness of the proposed algorithm, the snake has been initialized in each image to a square approximately centered on the object. We can see that the car is correctly tracked.

6   Conclusion and perspectives

We have presented a generic approach to parameter estimation in image processing using SIR models. Possible applications include object detection and location, attitude and scale estimation, segmentation and recognition. This approach is based on a simple statistical modeling of the image. This enables us to adapt the algorithms to the statistics of the noise actually present in the image, while keeping the same algorithmic architecture. When the considered model is not sufficient to describe the observed scene, we have described methods for adapting the image
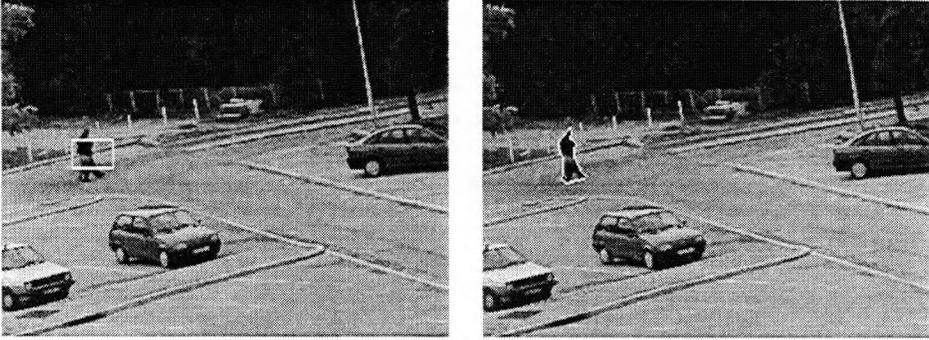
Image n°1

Image n°4

Image n°9

Figure 11: *Example of tracking on an image sequence. The sequence is composed of 9 frames, and frames n° 1, 4 and 9 are shown. Left column : image with initial snake. This initial snake is the shape segmented in the previous image translated to the estimated object location obtained with the MLRT algorithm. For image n°1, the initial shape is a square. Right column : image with snake after convergence.*

Image nᵒ1
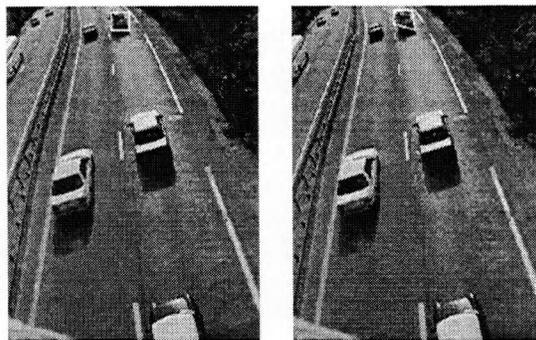
Image nᵒ4

Image nᵒ9

Figure 12: *Example of tracking on an image sequence. The sequence is composed of 9 frames, and frames nᵒ 1, 4 and 9 are shown. Left column : image with initial snake. This initial snake is a square approximately centered on the object. Right column : image with snake after convergence.*

(whitening preprocessing) or the algorithm (MLRT approach). The proposed technique is thus flexible, in the sense it can solve a variety of image processing tasks on a variety of images types, while keeping the same basic structure. This technique has been proven efficient on different types of synthetic and real-world images.

There are numbers of perspectives to this work. The unified algorithmic structure of SIR-based methods makes it possible to combine several tasks in a single application. An example has been given of the cooperation between the location and segmentation approaches for target tracking. The inclusion of attitude estimation (as a particular case of constrained-shape segmentation) and of recognition in such systems would be useful in many applications.

Another interesting development of this work is optical implementation of the described algorithms. We have shown that they are based on correlation operations with binary references. This makes it possible to benefit from the speed of binary SLM-based optical correlators. Note that in this case, the optical correlator would constitute the main building block of a system that would not only be able to perform location or recognition of a known target, but also segmentation.

References

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

[2] H. Arsenault, H., C. Ferreira, M. P. Levesque, and T. Szpolik, "Simple filter with limited rotation invariance," Appl. Opt. **25**, 3230–3234 (1986).

[3] B. V. K. Vijaya Kumar, D. P. Casasent, and A. Mahalanobis, "Correlation filters for target detection in a Markov model background clutter," Appl. Opt. **28**, 3112 (1989).

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

[5] V. Kober and J. Campos, "Accuracy of location measurements of a noisy target in nonoverlapping background," J. Opt. Soc. Am. A **13**, 1653–1666 (1996).

[6] 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).

[7] B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," Appl. Opt. **31**, 4773–4801 (1992).

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

[9] 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).

[10] 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).

[11] F. Guérault and Ph. Réfrégier, "Unified statistically independent region processor for deterministic and fluctuating target in non-overlapping background," Opt. Lett. **23**, 412–414 (1998).

[12] O. Germain and Ph. Réfrégier, "Optimal snake-based segmentation of a random luminance target on a spatially disjoint background," Opt. Lett. **21**, 1845–1847 (1996).

[13] C. Chesnaud, V. Pagé, and Ph. Réfrégier, "Robustness improvement of the statistically independent region snake-based segmentation method," Opt. Lett. **23**, 488–490 (1998).

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," International Journal of Computer Vision **1**, 321–331 (1988).

[15] R. Ronfard, "Region-based strategies for active contour models," International Journal of Computer Vision **2**, 229–251 (1994).

[16] C. Kervrann and F. Heitz, "A hierarchical statistical framework for the segmentation of deformable objects in image sequences," Proc. IEEE Conf. Comp. Vision Pattern Recognition, Seattle , 724–728 (1994).

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

[18] Ph. Réfrégier, "Bayesian theory for target location in noise with unknown spectral density," JOSA A **16**, 276–283 (1999).

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

[20] A. Azzalini, *Statistical Inference - Based on the likelihood* (Chapman and Hall, New York, 1996).

[21] J. W. Goodman, "Laser Speckle and Related Phenomena," in *Statistical Properties of laser Speckle Patterns*, pages 9–75 (Springer-Verlag (Topics in Applied Physics Vol. 9), Heidelberg, 1975).

[22] C. J. Oliver, I. Mc Connell, D. Blacknell, and R. G. White, "Optimum edge detection in SAR," in *Conf. on Satellite Remote Sensing, Paris, France*, **2584**, 152–163 (1995).

[23] O. Germain and Ph. Réfrégier, "Edge detection and localisation in SAR images : a comparative study of global filtering and active contour approaches," in *EurOpto Conf. on Image and Signal Processing for Remote Sensing, Barcelona, Spain*, **3500**, 111–121 (1998).

[24] José M. B. Dias and José M. N. Leitão, "Wall position and thickness estimation from sequences of echocardiographic images," IEEE Trans. on Medical Imaging **15**, 25–38 (1996).

[25] F. Guérault and Ph. Réfrégier, "Statistically independent region processor for target and background with random textures: whitening preprocessing approach," Optics Commun. **142**, 197–202 (1997).

[26] 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).

[27] 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).

[28] F. Guérault and Ph. Réfrégier, "Location of target in correlated background with the sir processor," Proceedings of the SPIE European Symposium on Lasers and Optics in Manufacturing (1997). Münich, Germany.

[29] H. Sjöberg, F. Goudail, and Ph. Réfrégier, "Optimal algorithms for target location in non-homogeneous binary images," Journal of the Optical Society of America A **15**, 2976–2985 (1998).

[30] P.H. Garthwaite, I.T. Jolliffe, and B. Jones, *Statistical Inference* (Printice Hall Europe, London, 1995).

[31] Y. Yakimovsky, "Boundary and object detection in real world images," Journal of the Association for Computing Machinery **23**, 599–618 (1976).

[32] 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).

[33] F. Guérault, L. Signac, F. Goudail, and Ph. Réfrégier, "Location of target with random gray levels in correlated background with optimal processors and preprocessings," Optical Engineering **36**, 2660–2670 (1997).

[34] L. D. Cohen, "On active contour models and balloons," CVGIP: Image Understanding **53**, 211–218 (1991).

[35] I. Chiou Greg and Jenq-Neng Hwang, "A neural network-based stochastic active contour model (nns-snake) for contour finding of distinct features," IEEE Transactions on Image Processing **4**, 1407–1416 (1995).

[36] Chenyang Xu and Jerry L. Prince, "Snakes, shapes, and gradient vector flow," IEEE Transactions on Image Processing **7**, 359–369 (1998).

[37] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," Journal of the Americain Statistical Association **86**, 376–387 (1991).

[38] T. Cootes, Hill A., C. J. Taylor, and Haslam J., "Use of active models for locating structure in medical images," Image and Vision computing **12**, 355–365 (1994).

[39] G. Storvik, "A bayesian approach to dynamic contours through stochastic sampling and simulated annealing," IEEE Trans. PAMI **16**, 976–986 (1994).

[40] Anil K. Jain, Yu Zhong, and Sridhar Lakshmanan, "Object matching using deformable template," IEEE Transactions on Pattern Analysis and machine intelligence **18**, 268–278 (1996).

[41] Mário Teles de Figueiredo and José M. N. Leitão, "Bayesian estimation of ventricular contours in angiographic images," IEEE Transactions on Medical Imaging **11** (1992).

[42] Ph. Réfrégier, O. Germain, and T. Gaidon, "Optimal snake segmentation of target and background with independent gamma density probabilities, application to speckled and preprocessed images," Optics Commun. **137**, 382–388 (1997).

[43] Mário A. T. Figueiredo, José M. N. Leitão, and Anil K. Jain, "Adaptative parametrically deformable contours," Energy Minimization Methods in Computer Vision and Pattern Recognition (1997).

[44] C. Chesnaud and Réfrégier, "Optimal snake region based segmentation for different physical noise model and fast algorithm implementation, france," in *First International symposium on Physics in Signal and Image Processing*, , 3–10 (1999).