UNSUPERVISED MARKED POINT PROCESS MODEL FOR BOAT EXTRACTION AND COUNTING IN HARBORS FROM HIGH RESOLUTION OPTICAL REMOTELY SENSED IMAGES

Paula Crăciun, Josiane Zerubia

INRIA, AYIN team
2004 Route des Lucioles, BP 93
06902 Sophia-Antipolis Cedex - France
firstname.lastname@inria.fr

Résumé

Les modèles de processus ponctuels marqués ont été appliqués avec succès pour l'extraction d'objets à partir d'images optiques de télédétection à haute résolution lors de la dernière décennie. Les modèles se composent typiquement de deux types de termes d'énergie : un terme d'attache aux données qui reflète la fidélité des configurations à l'image d'entrée et un terme qui intègre des connaissances a priori sur les objets à extraire. Dans cet article, nous traitons le problème de l'extraction des bateaux dans des ports. C'est un problème difficile en raison de la répartition particulière des objets dans ce cas. Nous décrivons un modèle de processus ponctuel marqué à base d'ellipses développé auparavant, pour lequel nous déterminons automatiquement l'un de ses principaux paramètres (la direction des bateaux). Nous présentons les inconvénients de ce modéle en raison des contraintes dures imposées, que nous relaxons ensuite pour proposer un nouveau modèle plus général.

Mots clés : Comptage de bateaux, extraction d'objet, télédétection, images optiques à haute résolution, géométrie stochastique

Abstract

Marked point process models have been successfully applied to object extraction in high resolution optical remotely sensed images during the last ten years. The models typically consist of two types of energy terms : a data term which reflects the fidelity of the configuration to the input image and a prior term which incorporates some knowledge about the objects to be extracted. In this paper we deal with the problem of extracting boats in harbors. This is a difficult problem due to the particular distribution of the objects in this case. We describe a previously developed marked point process model of ellipses for such a goal, for which we automatically determine one of its key parameters (the direction of the boats). We present the drawbacks of the model due to the hard constraints imposed, which we then relax and propose a new, more general model.

Keywords : *Boat counting, object extraction, remote sensing, high resolution optical images, stochastic geometry*

1. Introduction

Optical high resolution remotely sensed images are being increasingly used in different areas such as agriculture, urban development, disaster management, as well as by the army. Usual applications include object detection and tracking. A wide range of methods have been devised for solving the object extraction problem. These methods can be divided in two main categories : pixel-based and object-based approaches. While pixel-based approaches utilise only the spectral information of a pixel or its neighborhood, object-based approaches take additional information into account, like object texture, shape or relationships between adjacent objects.

Nevertheless, developing an object-based detection system is a difficult task. On the one hand, objects can be unevenly illuminated, partially occluded or blended with the background. On the other hand, the system has to be able to identify a specific class of objects. Thus, the system must implement a high inter-class and a low intra-

class variability. Practical grouping laws (i.e. vicinity or similarity), detailed by means of Gestalt Theory (see Desolneux et al. (2008)) can be added to the system to aid detection. Finally, the system must perform in reasonable time on large images.

Various boat extraction algorithms exist in literature. A model for detecting boats outside the harbor in high resolution optical images is provided in Proia and Pagé (2011), which uses Bayesian decision theory to determine the existence of boats, as well as in Máttyus (2013), which considers a binary classifier trained with both boat and background samples.

Stochastic geometry (Stoyan et al. (1987); Stoyan and Stoyan (1994)) provides a framework for handling object extraction problems. The idea is to model the desired objects by applying constraints on their shapes (see Descombes and Zerubia (2002)). In our case, boats can be modeled using ellipses (see Ben Hadj et al. (2010)). Therefore, a marked point process (MPP) of ellipses is used

to extract the boats. A MPP is a point process with marks attached to each point. The point process is defined with respect to a reference process, which is Poisson (see van Lieshout (2000) and Illian et al. (2008)) in our case. MPPs have been introduced in computer vision by mathematicians trying to solve pattern analysis problems. Different models have been developed and applied to extract various object types from remotely sensed images. For detailed information see Perrin et al. (2005), Ortner et al. (2008) and Chatelain et al. (2009).

The idea behind MPPs is to view an image as a collection of shapes, in our case ellipses. The marks of the MPP are given by the parameterization of an ellipse, as shown in Figure 1 (left). Moreover, each ellipse has an energy term assigned to it. This energy term can be divided in two : a data energy term, which controls how good the ellipse actually fits a boat in the image ; and a prior energy term, which includes end-user knowledge (i.e. constraints). The goal is to search for a set of ellipses, called a configuration, that best fits the given image. In this paper, we deal with the specific problem of boat extraction in harbors. Dramatic increases in the recreational boating fleet as well as other changes in the use of harbors has resulted in a competition for space within and along the shores of harbors. Harbor management can be conducted more easily once information on the number and location of boats is available. We will first present a previous marked point process model developed for boat extraction in harbors. Then, we will identify some of its drawbacks and propose some modifications to this model. Section 2 gives a brief description of the initial model. Section 3 presents a first improvement brought to this model. Some of the drawbacks of the approach are pointed out in section 4 and solutions to them are proposed, by relaxing the constraints imposed. This yields a new and more general model, as presented in Section 4. Section 5 describes an extension of the new model to more general images, where other objects are present (i.e. buildings, trees, etc.). Section 6 holds the conclusions and future work.

2. Initial model for boat extraction

In this study we focus on a model previously developed to extract boats in harbors (see Ben Hadj et al. (2010)). A marked point process of ellipses is considered. The object space, $\mathcal{W}$, is a bounded set in $\mathbb{R}^5$ defined as :

$$\mathcal{W} = [0, X_M] \times [0, Y_M] \times [a_m, a_M] \times [b_m, b_M] \times [0, \pi]. \quad (1)$$

Here, X_M and Y_M represent the width and height of the image, respectively, $[a_m, a_M]$ is the range for the length of the semi-major axis, $[b_m, b_M]$ is the range for the length of the semi-minor axis and $\omega \in [0, \pi]$ is the orientation of the ellipse.

A particular type of processes, called a Gibbs process, is used. The probability density function of such a process is given by :

$$f_\theta(X = \mathbf{x}|\mathbf{y}) = \frac{1}{c(\theta|\mathbf{y})} \exp^{-U_\theta(\mathbf{x},\mathbf{y})} \quad (2)$$

with

$$c(\theta|\mathbf{y}) = \int_\Omega \exp^{-U_\theta(\mathbf{x},\mathbf{y})} \mu(d\mathbf{x}), \quad (3)$$

where :
- $\mathbf{x}$ is the objects configuration ;
- $\mathbf{y}$ is the given image ;
- θ denotes a vector of parameters for the model ;
- $c(\theta|\mathbf{y})$ is the normalizing constant ;
- $\mu(\cdot)$ denotes the intensity measure of the reference Poisson process ;
- Ω is the configuration space ;
- $U_\theta(\mathbf{x}, \mathbf{y})$ denotes the energy function.

As mentioned before, the energy function is divided into two terms : a data energy term which reflects how good the model fits the actual data, denoted $U_{\theta_d}^d(\mathbf{x}, \mathbf{y})$, and a prior energy term including constraints imposed on the configuration, denoted $U_{\theta_p}^p(\mathbf{x})$. The vector θ_d contains the parameters of the data energy term that have to be estimated, while θ_p contains the parameters of the prior energy term. Hence, the parameter vector for the entire model can be written as $\theta = \{\theta_d, \theta_p\}$. The most likely configuration which allows the extraction of objects corresponds to the global minimum of the total energy $U_\theta(\mathbf{x}, \mathbf{y})$:

$$x \in \underset{\mathbf{x} \in \Omega}{Argmax}\, f_\theta(X = \mathbf{x}|\mathbf{y}) = \underset{\mathbf{x} \in \Omega}{Argmin}[U_\theta(\mathbf{x}, \mathbf{y})], \quad (4)$$

if the Maximum A Posteriori (MAP) criterion is used. Once the parameter vector has been estimated, a solution to this optimization problem can be found using simulated annealing (cf. van Laarhoven and Aarts (1987)) combined with a sampling algorithm such as Reversible Jump Markov Chain Monte Carlo (RJMCMC) (a detailed description can be found in Green (1995)) or Multiple Birth and Death (MBD) (for details see Descamps et al. (2008) and Descombes et al. (2011)). For additional information on MCMC techniques, see Gilks et al. (1995) and Robert and Casella (2005). Parameter estimation techniques are presented in detail in Chatelain et al. (2009), Ben Hadj et al. (2010) and Ben Hadj et al. (2011).
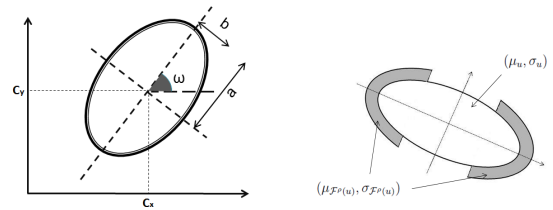

FIGURE 1: left : Parameterization of an ellipse of center $c = (c_x, c_y)$ with semi-major axis, a, semi-minor axis, b and angle, ω ; right : Border considered for the computation of the data energy term : (μ_u, σ_u) represent the mean and variance of the interior of the ellipse, $(\mu_{\mathcal{F}^\rho(u)}, \sigma_{\mathcal{F}^\rho(u)})$ represent the mean and variance of the border area.

2.1. The data energy term

In our approach, the computation of the data energy term takes place locally, for each ellipse. The data energy of

the entire configuration is computed as the sum of the individual data energies :

$$U_{\theta_d}^d(\mathbf{x}, \mathbf{y}) = \gamma_d \sum_{u \in \mathbf{x}} U^d(u, \mathbf{y}), \qquad (5)$$

where γ_d represents the data energy weight and has to be estimated. Thus, $\theta_d = \{\gamma_d\}$ is the parameter vector for the data term. The computation of the local data energy term, $U^d(u, \mathbf{y})$, relies on the use of a contrast measure $d(\cdot, \cdot)$ between the interior of the ellipse, denoted by u, and its border, denoted $\mathcal{F}^\rho(u)$. A pixel $p = (p_x, p_y)$ belongs to the neighborhood $\mathcal{F}^\rho(u)$ of ellipse u if and only if :

$$\left(\frac{(p_x - c_x)\cos\omega + (p_y - c_y)\sin\omega}{a} \right)^2 +$$
$$\left(\frac{(p_x - c_x)(-\sin\omega) + (p_y - c_y)\cos\omega}{b} \right)^2 > 1$$
$$\left(\frac{(p_x - c_x)\cos\omega + (p_y - c_y)\sin\omega}{a + \rho} \right)^2 + \qquad (6)$$
$$\left(\frac{(p_x - c_x)(-\sin\omega) + (p_y - c_y)\cos\omega}{b + \rho} \right)^2 \leq 1$$
$$||(p_x - c_x)\cos\omega + (p_y - c_y)\sin\omega|| \geq \frac{a}{2}$$

where c_x and c_y are the coordinates of the center of the ellipse and a, b and ω are its semi-major axis, semi-minor axis and orientation, respectively. Figure 1 (right) illustrates the border $\mathcal{F}^\rho(u)$ considered. Here, ρ is the width of the boundary. The contrast measure used is similar to the Bhattacharyya distance (see Goudail et al. (2004)) :

$$d_B(u, \mathcal{F}^\rho(u)) = \left[\frac{(\mu_u - \mu_{\mathcal{F}})^2}{4\sqrt{\sigma_u^2 + \sigma_{\mathcal{F}}^2}} - \frac{1}{2}\log\left(\frac{2\sqrt{\sigma_u^2 \sigma_{\mathcal{F}^2}}}{\sigma_u^2 + \sigma_{\mathcal{F}}^2} \right) \right] \qquad (7)$$

where (μ_u, σ_u^2) and $(\mu_{\mathcal{F}}, \sigma_{\mathcal{F}}^2)$ represent empirical means and variances of the object u and its ρ-border $\mathcal{F}^\rho(u)$, respectively.

A threshold, $d_0(\mathbf{y})$ for the contrast is manually determined based on the image, $\mathbf{y}$ (cf. Ben Hadj et al. (2010)). High threshold values are set for images where the objects are clearly distinguishable from the background, with a big contrast between the objects and the background. However, lower threshold values must be set for images where the contrast between the objects and the background is small. Finally, a quality function is used to compensate for errors when the output of the contrast measure is close to the threshold. Thus, the local data energy term can be defined as :

$$U^d(u, \mathbf{y}) = \mathcal{Q}\left(\frac{d(u, \mathcal{F}^\rho(u))}{d_0(\mathbf{y})} \right) \qquad (8)$$

where

$$\mathcal{Q}(x) = \begin{cases} 1 - x^{1/3} & \text{if } x < 1 \\ exp(-\frac{x-1}{3}) - 1 & \text{if } x \geq 1 \end{cases} \qquad (9)$$

The quality function, depicted in Figure 2 (left), is defined on $\mathcal{Q} : \mathbb{R}^+ \rightarrowtail [-1, 1]$. The lower the energy of one configuration, the better the configuration fits the image. Thus,

the quality function attributes a negative value to well placed objects (e.g. those objects u for which $d(u, \mathcal{F}^\rho(u))$ is higher than the threshold $d_0(\mathbf{y})$) and a positive value to misplaced objects. The use of the cubic root allows for a moderate penalization when the output of the distance measure is near the threshold.

2.2. The prior energy term

The prior energy term is decomposed into two parts and handles three constraints. The first one takes care of overlapping, meaning that objects are not allowed to overlap more than a given extent. Thus, the first part of the prior energy term corresponds to a penalization of overlapping objects, avoiding the detection of the same object several times. The proposed model uses a *hard core* process to handle object overlapping, meaning that all configurations with objects that overlap more than a given extent will be disregarded (i.e. the energy value assigned to such configurations will be infinitely high). Thus, denoting by $A(u_i, u_j) = \frac{Area(u_i \cap u_j)}{\min(Area(u_i), Area(u_j))}$ the area of intersection between the objects u_i and u_j, the prior energy can be defined as :

$$U_o^p(\mathbf{x}) = \sum_{1 \leq i \neq j \leq n(\mathbf{x})} t_s(u_i, u_j) \qquad (10)$$

with :

$$t_s(u_i, u_j) = \begin{cases} 0 & \text{if } A(u_i, u_j) < s \\ +\infty & \text{otherwise} \end{cases} \qquad (11)$$

where $s \in [0, 1]$ corresponds to the amount of overlapping allowed by the model and $n(\mathbf{x})$ is the number of objects in the configuration $\mathbf{x}$. Accordingly, all configurations containing at least two objects that overlap to a higher ratio than specified by s are prohibited.
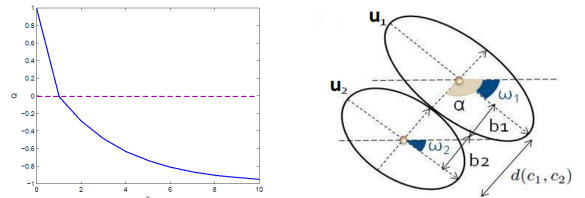

FIGURE 2: left : Quality function $Q(x)$; right : Alignment constraint between two elipses u_1 and u_2, with b_1 and b_2 being the semi-minor axis of u_1 and u_2 respectively, ω_1 and ω_2 are the orientations of the two ellipses, $d(c_1, c_2)$ is the distance between the centers of the ellipses and α is the angle between the line that connects the two centers of the ellipses and the horizontal axis.

The second part of the prior energy term incorporates two constraints. On the one hand, close and aligned ellipses are favored. The alignment interaction $\sim_{al}$ between two ellipses u_1 and u_2 is defined as :

$$u_1 \sim_{al} u_2 \Leftrightarrow \begin{cases} d_\omega(u_1, u_2) \leq d_{\omega_{max}} \\ d_\alpha(u_1, u_2) \leq d_{\alpha_{max}} \\ d_C(u_1, u_2) \leq d_{C_{max}} \end{cases} \qquad (12)$$

where :

- $d_\omega(u_1, u_2) = |\omega_1 - \omega_2|$ is the difference in orientation between the two ellipses ;
- $d_\alpha(u_1, u_2) = |\alpha - \frac{\omega_1 + \omega_2}{2} + \frac{\pi}{2}|$ is a measure that checks that the ellipses are not shifted, α being the angle between the line that unites the centers of the two ellipses and the horizontal ;
- $d_C(u_1, u_2) = |d(c_1, c_2) - (b_1 + b_2)|$, where $d(c_1, c_2)$ stands for the Euclidean distance between the centers of the ellipses.

The alignment interaction is graphically represented in Figure 2 (right). On the other hand, all ellipses in the image must have the same global orientation ω_N. This constraint was imposed for a first series of harbors considered, where all boats had the same orientation. In this way, the object space was reduced, but nevertheless, this approach makes the model very specific.

Therefore, the second part of the prior term is given by :

$$U_{al_{\omega_N}}(u) = \begin{cases} U_{al}^p(\mathbf{x} \cup u) - U_{al}^p(\mathbf{x}) & \text{if } |\omega_u - \omega_N| \leq d_{\omega_{max}} \\ 0 & \text{otherwise} \end{cases}$$
(13)

where $U_{al}^p(\mathbf{x})$ is given by :

$$U_{al}^p(\mathbf{x}) = \gamma_{al} \sum_{1 \leq i < j \leq n(\mathbf{x})} U_{al}(u_i, u_j).$$
(14)

$U_{al}(u_i, u_j)$ is the energy associated to the aligment interaction $\sim_{al}$ and is given by :

$$U_{al}(u_i, u_j) = \begin{cases} \delta\varpi(d_\alpha(u_1, u_2), d_{\alpha_{max}}) & \text{if } u_1 \sim_{al} u_2 \\ 0 & \text{otherwise} \end{cases}$$
(15)

with :

$$\varpi(x, x_{max}) = -\frac{1}{x_{max}^2}\left[\frac{1 + x_{max}^2}{1 + x^2} - 1\right], \; for \; x \leq x_{max}$$
(16)

being a reward function previously introduced in (Ortner et al., 2008). The function $\varpi(\cdot, \cdot) : \mathbb{R} \to [0, 1]$ is designed such that $\varpi(0, x_{max}) = 1$ and $\varpi(x_{max}, x_{max}) = 0$.

Finally, the total prior energy term can be written as :

$$U_{\theta_p}^p(\mathbf{x}) = U_o^p(\mathbf{x}) + \sum_{u \in \mathbf{x}} U_{al_{\omega_N}}^p(u)$$
(17)

with $\theta_p = \{\gamma_{al}\}$, the weight of the aligment interaction, present in equation 14.

This model uses the parameter ω_N for the global direction of the objects to be extracted. This parameter was extracted by trial and error in Ben Hadj et al. (2010).

3. Determining the global orientation of the boats

Our experiments on several types of harbors have revealed the importance of having information on the orientation of the boats. Lack of such information results in high extraction errors. Our first contribution to the initial model of Ben Hadj et al. (2010) is the automatic identification of the global orientation of the boats. Two ways of determining the global direction of the boats have been investigated. The first method is based on mathematical morphology and uses a technique called *rose of directions*. This technique wad introduced by Serra (1982) as a way to determine the predominant orientation of bright (or dark) structures within an image.

The second method makes use of the Fourier transform to identify the high frequencies in the image, i.e. the highest changes in pixel values. These values correspond mainly correspond to the edges between the water area and the boats. Both methods are described in more detail in this section. Extraction results are presented at the end of this section.
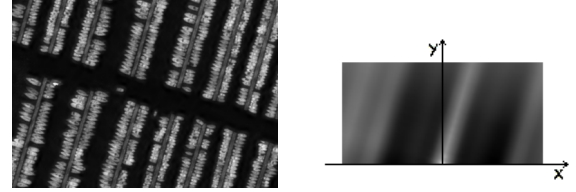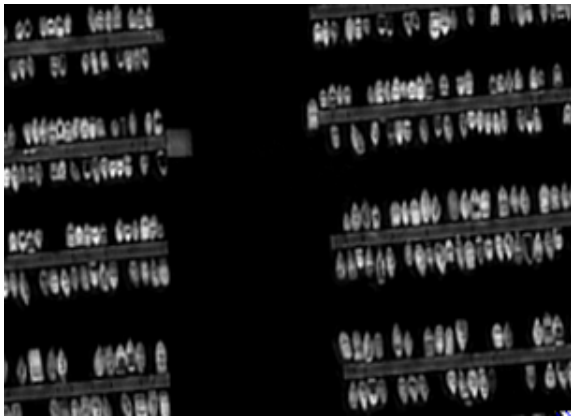

FIGURE 3: left : Boats in a Mediterranean harbor image ©CNES ; right : The corresponding rose of direction by openings (Soille and Talbot, 1998), as presented in Section 3.1
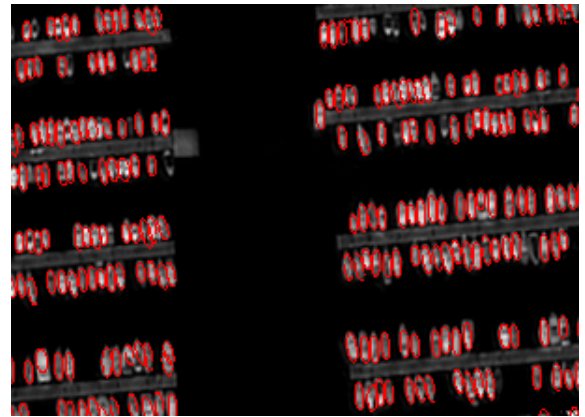
3.1. Mathematical morphology to determine the global direction

The direction of boats in an image can be identified by ploting the *rose of directions* (see Serra (1982); Soille and Talbot (1998)). The rose of directions consists in representing the polar diagram that indicates the amount of image structures in each direction considered. Opening and closing operations (opening for bright image structures and closing for dark image structures) are used for directional filtering via mathematical morphology. The method is applied to determine the direction of the docks in the harbor. The docks are represented as long lines in the images and are therefore easier to identify. We assume that the orientation of the objects to be extracted (i.e. boats) is perpendicular to the orientation of the docks.

Since the docks are brighter than the background (i.e. water in our case), opening operations are used. For 2-D gray-scale images, the value of the gray-level rose of directions by opening operations at pixel (i, j) is defined as the volume of the image opened by a discrete line segment of length $2 \max\{|i|, j\} + 1$ in pixels, called structuring element (SE), having the orientation defined by $\arctan j/i$. The volume of an image is defined as the sum of its pixel values. By rotating the SE in a discrete manner under all angles in the interval $[0, \pi)$ and by systematically modifying its length, we can identify predominant orientations for bright line structures. The output of the rose of directions is illustrated in Figure 3 (right). The bright line starting at the origin of the coordinate system in the diagram on the right represents the predominant orientation of bright line structures in the image. Identical values in all directions would indicate that there is no predominant orientation in the image (see Craciun and Zerubia (2013a) for more details).
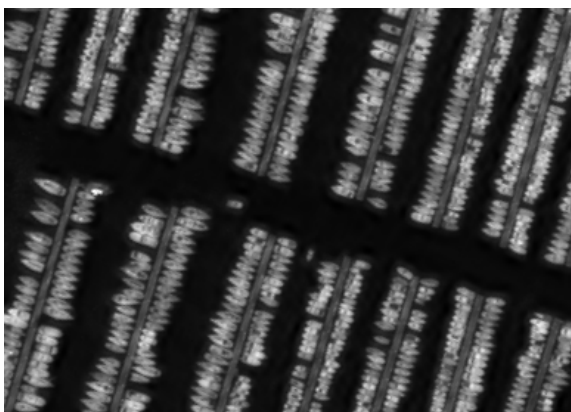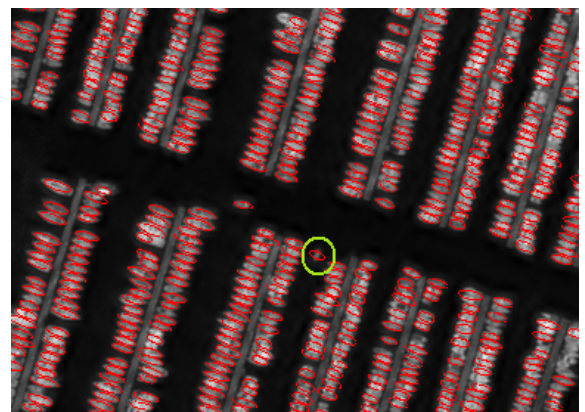
FIGURE 5: (a) - (c) Images of boats in several Mediterranean harbors ©CNES ; (d) - (f) Extraction results using the proposed model in Ben Hadj et al. (2010) with the automatic detection of the global orientation of the boats, presented in Section 3

3.2. Determining the global direction using the Fourier transform

A completely different method for extracting the global direction consists in using the Fourier transform of the image. The idea is to detect high frequencies in the image, which correspond to the high changes in pixel values. Since the boats are bright structures in the image and the water is dark, the change in intensity values between these structures is very high.
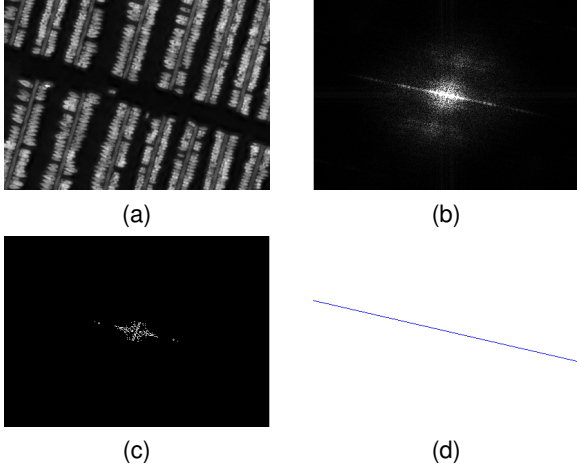

(a)   (b)

(c)   (d)

FIGURE 4: (a) Boats in Mediterranean harbor image ©CNES ; (b) Corresponding image in the Fourier domain ; (c) Edge detection applied on the transformed image ; (d) Hough transform

Figure 4 shows the corresponding image in the Fourier domain, after applying a Fast Fourier Transform (FFT) and shifting the result such that the zero-frequency is in the middle of the image. Once the corresponding image in the Fourier domain has been obtained, we then detect the edges using a classical Sobel edge detector. Finally we use Hough transform to detect the orientation of the predominant line that corresponds directly to the orientation of the boats. The Sobel edge detector yields best results in this case. Other edge detectors, such as the Canny-Deriche edge detector or Prewitt edge detector, provide either too many or too few edges, resulting in a lower performance of the Hough transform.

3.3. Results and Discussions

Both methods for computing the global direction of the boats have low computational complexity and therefore the computation time is around 1 second for each of them on images of approximately 350×250 pixels in size. Although the results are similar in complexity as well as in value, some differences have to be noted between the two approaches. First, while the output of the rose of directions depends on the length of the structuring element, which is set by hand, no such parameters are necessary for the second approach. A second difference regards the assumptions made. While using the rose of directions, we detect the docks and then assume that the boats are perpendicular to them, using the Fourier transform does not depend on such an assumption. Therefore,

we can consider that the second method is more robust than the first one in this case. The extraction results are

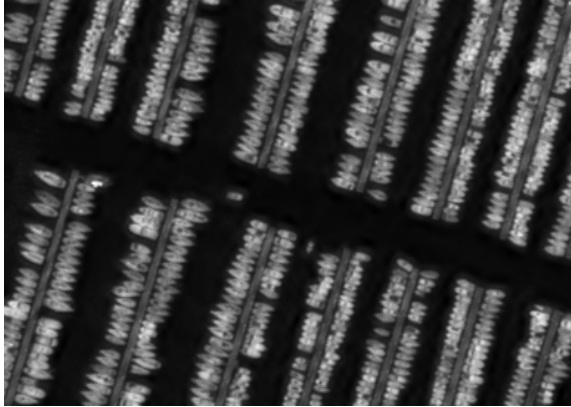| Boat Image	Ground truth	Rose of Direction	Fourier Transform
Figure 5 (a)	7.2987	7.2801	7.1000
Figure 5 (b)	7.5390	7.3333	7.6278
Figure 5 (c)	0.3000	0.2914	0.2280

TABLE 1: Global direction values

shown in Figure 5. We note that the extraction results are independent of the method used for extracting the global direction, since the model does not use the exact value of the global direction, but uses a small interval around it. We highlight errors caused by wrong orientation (i.e. boats having a different orientation than the global orientation considered) in green. The extraction lasted $32min\ 38sec$ for Figure 5(a), $29min\ 54sec$ for Figure 5(b) and $55min\ 21sec$ for Figure 5(c). The sizes of the images and the expected number of boats are 304×220 pixels for Figure 5(a) with roughly 240 boats of which 225 detected, 241×178 for Figure 5(b) with roughly 225 boats out of which 207 detected and 385×275 for Figure 5(c) with roughly 518 boats out of which 501 detected. The values for the global directions are presented in Table 1. The ground truth has been computed by hand. Nevertheless, this model can be used only when all of the boats in the image have the same orientation. This is not generally true, since the harbors have more complicated structures that result in the boats having different orientations. Therefore, we propose a new model to handle this problem.

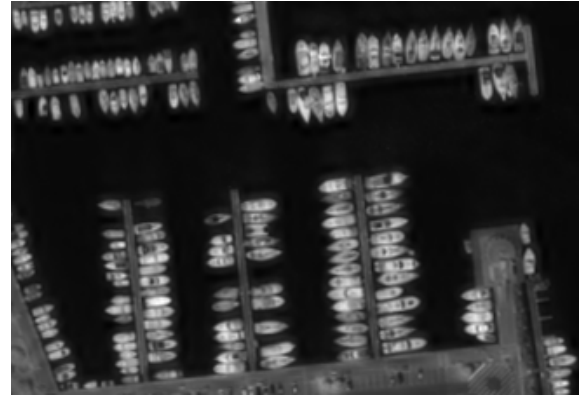## 4. Modifying the model. Relaxing constraints on the orientation

The main disadvantage of the previously described model is the strong constraint on the orientation of the boats. In general, the orientation of the boats varies within a harbor. The described model fails to handle such cases. Additionally, we want to extend the model to more general cases. In harbors, objects are not obliged to be tangent to each other and thus, we cannot put a constraint on the distance $d_C(u_1, u_2)$ between the centers of the ellipses u_1 and u_2. Furthermore, if boats of significantly different sizes are close to each other, their centers will be shifted. For this reason, we cannot put a constraint anymore on the distance $d_\alpha(u_1, u_2)$ between two ellipses. We propose to relax some of the constraints of the model of Ben Hadj et al. (2010) and limit them to the following ones (see Craciun and Zerubia (2013b)) :

– boats should not overlap more than a given extent ;
– neighboring boats should have similar orientation ;
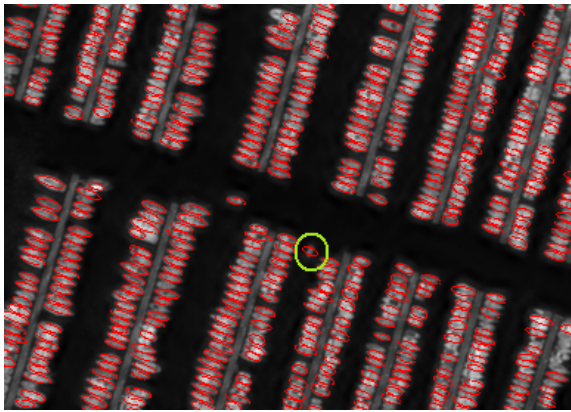– the orientation of each boat is determined locally.

The first item is exactly the non-overlaping constraint presented previously. The second item restricts neighboring boats from having completely different orientation, a threshold is set for the amount of difference in the orientation
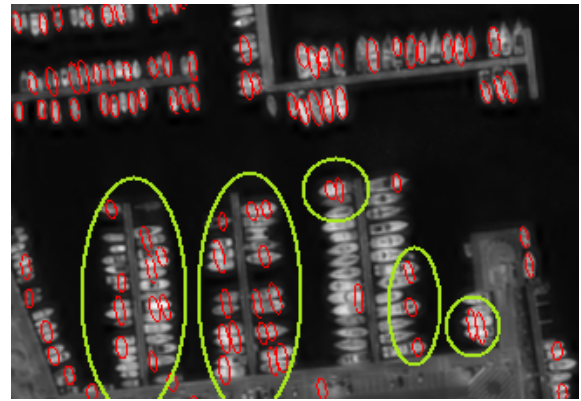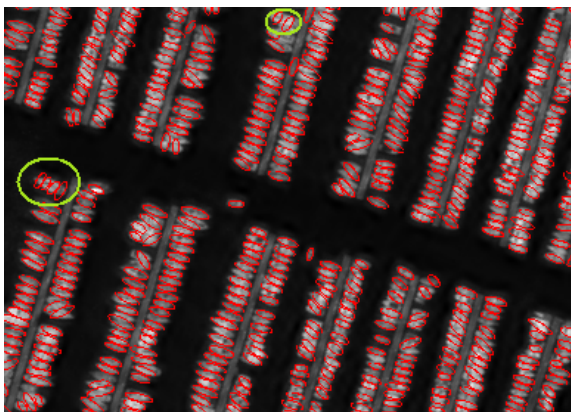
(a)

(d)

(b)

(e)

(c)

(f)

FIGURE 7: (a), (d) Image of boats in a Mediterranean harbor ©CNES ; (b),(e) Extraction results using the model proposed in Ben Hadj et al. (2010) ; (c),(f) Extraction results using the proposed model.

that is allowed. Finally, the orientation of the boats is determined locally.

Another subtle modification concerning the data energy is proposed. We observed that quite often the cabin of the boat is seen as a darker area within the boat. This usually results in smaller contrasts, and thus smaller data energy values. An additional term in the data energy has been introduced to account for this drawback. Results are presented at the end of this section.

4.1. Modification of the data energy term

Considering the entire interior of the boats when computing the contrast measure might not yield the best results, since the interior is not always homogeneous. Therefore, we introduce an additional term, by taking into account the contrast measure between the interior border of the objects and the exterior one, as shown in Figure 6. The data term becomes therefore :

$$U^d(u, \mathbf{y}) = \mathcal{Q}(\frac{d(u, \mathcal{F}^\rho(u))}{d_0(\mathbf{y})}) + \gamma_c \mathcal{Q}(\frac{d(\mathcal{I}^\rho(u), \mathcal{F}^\rho(u))}{d_0(\mathbf{y})}) \tag{18}$$

where γ_c is the weight of the contrast measure between the interior border, $\mathcal{I}^\rho(u)$, and the exterior border, $\mathcal{F}^\rho(u)$, of the object u. Border width, ρ, and contrast threshold, $d_0(\mathbf{y})$, are kept the same for both parts of $U^d(u, \mathbf{y})$ in order to keep the number of parameters as low as possible. In our experiments, $\gamma_c = 1$, resulting in both terms having the same weight in the data energy.
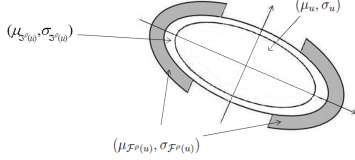

FIGURE 6: Exterior and interior border considered for the computation of the new data energy term described in equation 18.

4.2. Modification of the prior energy term

As stated previously, the alignment interaction between neighboring boats should impose them to have similar orientations. Thus, we define an alignment interaction between two neighboring ellipses $u1$ and $u2$ in the following way :

$$u_1 \sim'_{al} u_2 \Leftrightarrow \begin{cases} d_\omega(u_1, u_2) \leq d_{\omega_{max}} \\ d_C(u_1, u_2) \leq d_{C_{max}} \end{cases} \tag{19}$$

where $d_\omega(u_1, u_2)$ and $d_C(u_1, u_2)$ have been previously defined in eq. 12. Then, we append a prior energy that promotes this alignment in the following way :

$$U_{al}(u_1, u_2) = \begin{cases} \delta\varpi(d_\omega(u_1, u_2), d_{\omega_{max}}) \\ \qquad \text{if } u_1 \sim'_{al} u_2 \\ 0 \qquad \text{otherwise} \end{cases} \tag{20}$$

where $\varpi(x, x_{max})$ is a reward function that favors aligned frames and was defined in equation 16. The prior

energy of a configuration $\mathbf{x}$ corresponding to the alignment constraint is the sum over all the object pairs of the configuration $\mathbf{x}$:

$$U^p_{al}(\mathbf{x}) = \gamma_{al} \sum_{1 \leq i < j \leq n(\mathbf{x})} U_{al}(u_i, u_j). \tag{21}$$

The weight γ_{al} is also a parameter that has to be estimated.

A method proposed in Li and Briggs (2009) for road extraction, is used in order to locally determine the orientation of the docks using the water, in a preprocessing step. The idea consists in first applying an edge-detector to the input image. Afterward, at each pixel of the image containing the edges, we find the circle centered in that pixel with the largest radius that does not intersect any edge. Finally, we keep only those circles with the largest radii. These circles are mainly located over the water area since the water represents the largest structure in the image where no edges can be found. The centers of those circles form lines, commonly known as medial axes, for which the Hough transform is used to determine their orientation. The pixels which are centers of those circles and are intersected by the lines obtained through the Hough transform are called *relevant pixels*, i.e. pixels that incorporate information about the orientation of the water in that neighborhood.

Finally, we search for the local orientation of the objects, i.e. at each creation of an object, we search for the nearest relevant pixel in its spatial neighborhood. If no relevant pixel is found within a certain distance, we favor ellipses that have similar orientations with the ellipses in their neighborhood. If a relevant pixel is found, its position w.r.t. the semi-axes of the ellipse is determined. We will favor ellipses perpendicular to the orientation indicated by the relevant pixel. Thus, the final prior energy term that refers to alignment becomes :

$$U^p_{al_{\omega_l}}(\mathbf{x}) = \sum_{u \in \mathbf{x}} U^p_{al_{\omega_l}}(u) \tag{22}$$

where :

$$U^p_{al_{\omega_l}}(u) = \begin{cases} 0 \qquad \text{if a relevant pixel is found and} \\ \qquad ||\omega_u - \omega_l|| > d_{\omega_{max}} \\ \gamma_{al} \sum_{v \in \mathbf{x}} U_{al}(u, v) \quad \text{otherwise} \end{cases} \tag{23}$$

where ω_u is the orientation of object u and ω_l is the orientation perpendicular to the one retained in the relevant pixel.

The total prior energy term can be written as :

$$U^p_{\theta_p}(\mathbf{x}) = U^p_o(\mathbf{x}) + U^p_{al_{\omega_l}}(\mathbf{x}) \tag{24}$$

with $\theta_p = \{\gamma_{al}\}$, the weight of the aligment interaction, present in equation 23.

4.3. Results and Discussions

The modification brought to the data term increases the detection accuracy slightly when compared to the data energy specified in Ben Hadj et al. (2010). After running several tests by trial and error, we have concluded

FIGURE 8: (a) Image of boats in Melbourne harbor ©Airbus D&S ; (b), (c) Zoom in on particular parts of the harbor ; (d),(e),(f) Extraction results using the proposed model.

Boat Image	Detection error for old model Ben Hadj et al. (2010)	Detection error for new model
Figure 7 (a)	$\sim 1\%$	$\sim 10\%$
Figure 7 (d)	$> 45\%$	$< 30\%$

TABLE 2: Comparison between detection errors of BenHadj et al. model vs. the one proposed

that equal weights between the two terms in the right hand side of equation 18 gives the best results. The increase in detection accuracy is maintained both before and after modifying the prior term. Figure 7 shows the results obtained with this model. Figure 7 (a) and (d) represent two test images of boats in harbors. Figure 7 (b) and (e) show the results obtained by applying the model presented in Ben Hadj et al. (2010). Figure 7 (c) and (f) show the results obtained by applying the model we propose in this paper. The detection results are a little less accurate than in Ben Hadj et al. (2010) resulting in a total number of 501 ellipses, but the computation time decreased to $45min\ 3sec$ for parameter estimation and $55min\ 21sec$ for the detection, on a $2.20GHz$ processor, compared to $1h\ 38min$ only for the parameter estimation, on a $1.86GHz$ processor obtained in Ben Hadj et al. (2010).

The strength of the modified model is visible in Figure 7 (f). Where the object extraction of the model presented in Ben Hadj et al. (2010) fails due to the different orientations of the boats in the image (see Figure 7 (e)), the modified model proves to be more efficient. Qualitative detection results are presented in Table 2. The detection error (DE) is computed as the ratio between the sum of false positives (FP) and false negatives (FN), and the total number of detected objects (DO) : DE $= \frac{\text{FP}+\text{FN}}{\text{DO}}$. If the difference between the actual orientation of the object on which one ellipse is placed and the orientation of that ellipse is larger than a threshold, we consider that ellipse to be a false positive. Problematic areas are shown in green in Figure 7. The overall performance of the modified model is very good.

5. Extracting boats in larger scenes

A natural extension of this framework is its use in a more general context. Therefore, we consider an image containing more objects than just the harbor itself. Such a scene is presented in Figure 8 (a). The image contains the surrounding area of the harbor, which includes buildings, roads, trees, etc. with a size of 1620×1450 pixels.

Since our purpose is to detect boats in the harbor, a first step is to delimit the harbor area and ignore the rest of the scene, when searching for objects. One can easily identify that the pixel intensity is lower for the water area. Various methods can be applied to differenciate water and land ranging from image classification to region growing segmentation (see Gonzalez and Woods (1992)) or

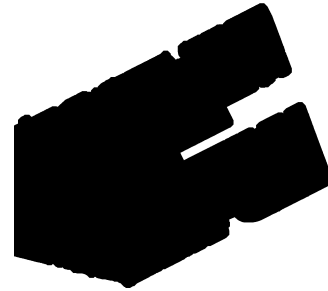three dimensional modeling (see Rau et al. (2002)). A

FIGURE 9: The extracted water area for Figure 8(a)

simple, fast and efficient method to extract the water area is to apply thresholding. As shown in Dare (2005), bimodal histogram splitting is a robust method for the selection of the threshold value. Two main classes are obtained after thresholding. One class represents the water, the second one represents the shadows created by tall buildings. We follow the proposed method in Dare (2005) and compute the variance for each region obtained after thresholding. This approach is based on the idea that the variance of the radiometric values of pixels within shadow areas is high, when compared to water areas. The higher values for the variance are caused by the heterogeneity of the underlying structures on which the shadows are cast. This is why we retain the regions with the lowest variance, thus representing the water.

Erosion operation is used to delete the bright structures within the water. The structuring element is a disk. Its size should be big enough to encapsulate the boats in the harbor and was therefore chosen to be 25×25 for the image considered. Figure 9 shows the resulting water area corresponding to Figure 8 (a). We use this image as a mask and apply the previously described model only within the remaining area of interest. Figure 8 (d) shows the extraction results on the considered image, as well as two zoom-ins on parts of that image (Figure 8 (e) and Figure 8 (f)). The extraction results are very good. The extraction of the water area is a necessary preprocessing step for two main reasons :

- it limits the search space during the optimization process, thus reducing the computation time ;
- it avoids the detection of erroneous objects on land that have similar characteristics to boats.

However, although only a delimited area is considered for boat extration, the size of this area is about 1350×950 pixels and therefore the computation time was large (i.e. $1h\ 38min\ 57sec$).

6. Conclusions and future work

In this paper we have shown how the global direction, a key parameter of a MPP for extracting boats in harbors, can be automatically computed using two different approaches. The first approach is based on a combination of morphological operators yielding a polar diagram called Rose of Directions, while the later is based on the

Fourier transform. Both approaches can be successfully introduced as a preprocessing step and have low computational complexities. On the one hand, the output of the Rose of Directions is more precise, but is highly influenced by the size of the structuring element used. On the other hand, the Fourier transform does not depend on any parameter, but the output is less precise. Although the model proposed in the work of Ben Hadj et al. (2010) yields very good results when all boats have the same orientation, it does not cope with images where the boats have different orientations. To handle this drawback, modifications to the model have been proposed in this paper. This resulted in a new, more general model for boat extraction in harbor images. We have extended our model to fit larger scenes, by incorporating a preprocessing step to extract first the region of interest.

A major contribution of this paper is the computation of the orientation of the boats locally, which is done using the water areas in the image, and the extention to larger scenes. Future work might include identifying the docks directly, thus having more precise information about the orientation of the boats. The models presented in this paper have been applied only to static images. An extension of this framework to dynamic images to implement tracking techniques can be considered. Furthermore, the computation times are generally large. State of the art high performance computing techniques can be used to develop efficient parallel samplers that would significantly speed up the optimization process.

7. Acknowledgements

Références

Ben Hadj, S., Chatelain, F., Descombes, X., Zerubia, J., 2010. Parameter estimation for a marked point process within a framework of multidimensional shape extraction from remote sensing images. Proc. of ISPRS XXXVIII.

Ben Hadj, S., Chatelain, F., Descombes, X., Zerubia, J., 2011. Approche non supervisée par processus ponctuels marqués pour l'extraction d'objets à partir d'images aériennes et satellitaires. Revue Française de Photogrammétrie et Télédétection 194, 2–15.

Chatelain, F., Descombes, X., Zerubia, J., 2009. Parameter estimation for a marked point process. application to object extraction from remote sensing images. Proc. of EMMCVPR 5681, 221–234.

Craciun, P., Zerubia, J., 2013a. Boat extraction in harbors from high resolution satellite images using mathematical morphology and marked point processes. Proc. of GRETSI.

Craciun, P., Zerubia, J., 2013b. Unsupervised marked point process model for boat extraction in harbors from high resolution optical remotely sensed image. Proc. of ICIP.

Dare, P. M., 2005. Shadow analysis in high-resolution satellite imagery of urban areas. Photogrammetric Engineering and Remote Sensing 71, 169–177.

Descamps, S., Descombes, X., Bèchet, A., Zerubia, J., 2008. Automatic flamingo detection using multiple birth and death process. Proc.of ICASSP.

Descombes, X., Chatelain, F., Lafarge, F., Lantuejoul, C., Mallet, C., Minlos, M., Schmitt, M., Sigelle, M., Stoica, R., Zhizhina, E., 2011. Stochastic Geometry for Image Analysis. John Wiley and Sons.

Descombes, X., Zerubia, J., 2002. Marked point processes in image analysis. IEEE Signal Processing Magazine 19 (5), 77–84.

Desolneux, A., Moisan, L., Morel, J. M., 2008. From Gestalt Theory to Image Analysis : A Probabilistic Approach. Springer Verlag.

Gilks, W., Richardson, S., Spiegelhalter, D., 1995. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC.

Gonzalez, R., Woods, R., 1992. Digital Image Processing. Addison-Wiley Press.

Goudail, F., Réfrégier, P., Delyon, G., 2004. Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. Journal of Optical Science of America A 21 (7), 1231–1240.

Green, P., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons.

Li, Y., Briggs, R., 2009. Automatic extraction of roads from high resolution aerial and satellite images with heavy noise. Proc. of the 6th International Conference on Geographic Information Systems.

Máttyus, G., 2013. Near real-time automatic marine vessel detection on optical satellite images. ISPRS Hannover Workshop, 233–237.

Ortner, M., Descombes, X., Zerubia, J., 2008. A marked point process of rectangles and segments for automatic analysis of Digital Elevation Models. IEEE Transactions of Pattern Analysis and Machine Intelligence 30, 105–119.

Perrin, G., Descombes, X., Zerubia, J., 2005. A marked point process model for tree crown extraction in plantations. Proc. ICIP 1, 661–664.

Proia, N., Pagé, V., 2011. Détection de bateaux dans des images satellitaires optiques panchromatiques. Revue Fraçaise de Photogrammétrie et Télédétection 194.

Rau, J., Chen, N., Chen, L., 2002. True orthophoto generation of built-iin areas using multi-view images. Photogrammetric Engeneering and Remote Sensing 68 (6), 581–588.

Robert, C., Casella, G., 2005. Monte Carlo Statistical Methods. Springer.

Serra, J., 1982. Image Analysis and Mathematical Morphology. Academic Press.

Soille, P., Talbot, H., 1998. Image structure orientation using mathematical morphology. Proc. of ICPR 2, 1467–1469.

Stoyan, D., Kendall, W. S., Mecke, J., 1987. Stochastic Geometry and its Applications. John Wiley and Sons.

Stoyan, D., Stoyan, H., 1994. Fractals, Random Shapes and Point Fields : Methods of Geometrical Statistics. John Wiley and Sons.

van Laarhoven, P., Aarts, E., 1987. Simulated Annealing : Theory and Applications. Kluwer Academic Publishers.

van Lieshout, M., 2000. Markov Point Processes and Their Applications. Imperial College Press.