UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament de Teoria del senyal i comunicacions

MULTIRESOLUTION IMAGE SEGMENTATION BASED ON COMPOUND RANDOM FIELDS: APPLICATION TO IMAGE CODING

Autor: Ferran Marqués
Director: Antoni Gasull

Barcelona, Diciembre de 1992

CHAPTER II

IMAGE MODELS

As it has been highlighted in Chapter I, a segmentation algorithm should use elaborate image models as well as multiresolution decompositions in order to improve its performance. This chapter is devoted to the discussion of the first point; that is, stochastic image modelling. The discussion about multiresolution decomposition is postponed to a posterior chapter.

Throughout the following chapters, images, which are denoted by I(i, j), are defined on an $N_1 * N_2$ rectangular lattice:

$$\mathcal{S}\,(i,j) = \{\,(i,j) : 1 \le i \le N_1,\ 1 \le j \le N_2\,\}, \qquad \text{(II.1)}$$

which is usually square. Furthermore, image models assume that images are formed by a set of regions $R = \{\,R_n\,\}$, where $n \in [1\ ..\ N_1 * N_2]$, and each region is characterised by a single model. Note that, in an image model, each region may be characterised by a different kind of model. However, when defining an

image model, the same region model is usually assumed for the whole set of regions. Therefore, a stochastic image model is defined by the kind of stochastic model that describes its regions.

Usually, regions are supposed to be statistically independent and, thus, so are the different models. Moreover, given that each region should represent a homogeneous zone within the image, each region model is assumed to be stationary. The set of picture elements (pixels) within a region R_n is denoted by

$$I_n(i, j) = \{ I(i, j) : (i, j) \in R_n \}. \qquad (II.2)$$

Since each I(i, j) is considered to be a random variable, each one of these sets I_n is assumed to be a realisation $x = \{ x_{ij} \}$ of a random field $X = \{ X_{ij} \}$. Therefore, the study of an image model is performed by analysing the kind of random field characterising its regions. Given that images are defined on a discrete, finite lattice $\mathbb{S}$ (II.1) and that they can only take discrete values, the discussion is restricted to discrete-index random fields defined on discrete, finite lattices.

II. 1.- Gaussian random fields

In this kind of model, each region R_n is characterised by a Gaussian Random Field (GRF) [41]. The statistics of a GRF are completely determined in terms of its mean value μ and its covariance matrix $\underline{C}$. For computing them, the components of the random field $\{ X_{ij} \}$ can be ordered into a vector $\mathbf{X}$. In this way, the probability density function of the GRF, $P(X = x) = P(\mathbf{X} = \mathbf{x})$, has the form

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\underline{C})^{1/2}}} \exp\left\{ -\frac{(\mathbf{x}-\mu)^T \underline{C}^{-1} (\mathbf{x}-\mu)}{2} \right\}, \qquad (II.3)$$

where N is the number of elements in R_n; that is, the dimension of vector $\mathbf{x}$.

The computation of the joint probability of a single element of a GRF demands the estimation of the elements of the covariance matrix and the computation of its inverse. Note that $\underline{C}$ can be rather large since its dimensions are NxN. In order to simplify the computation when processing the model, the covariance matrix can be assumed diagonal [42]; that is:

$$C(k,l) = \begin{cases} \sigma^2 & \text{if } k = l \\ 0 & \text{otherwise .} \end{cases} \qquad \text{(II.4)}$$

In this case, elements within a region are mutually uncorrelated and the GRF becomes a white Gaussian random field. The whiteness assumption is usually made for mathematical tractability of the problem. Indeed, image data do not really fit pure white Gaussian models very well. Moreover, note that white gaussian models do not take into account the spatial distribution of pixels within the region. For instance, a set of pixels distributed on a rectangular region has the same parameters (same mean and white covariance matrix) when distributed on a region with a very complicated shape. Thus, both regions share the same characterisation. In addition, given a region, pixels in its interior can be located in a different way without changing its mean and white covariance matrix. Therefore, in order to describe complex textures, more elaborated models (e.g.: taking into account spatial correlation) have to be used. This case is discussed in the next section.

A second problem of Gaussian models is the estimation of the model parameters. If the exact values of these parameters are not known a priori (knowing them is more an academic case than a real one), they have to be estimated from the data. As it has been emphasised in Section I.4.5, the tasks of segmenting an image and of estimating features of its regions are very related. Indeed, for obtaining correct estimations a segmentation of the scene has to be already available. Usual estimation procedures do not cope correctly with the problem of uncertainty in the contours and result in image estimates suffering from oversmoothing of edges and/or overgrowing of regions.

These drawbacks can be easily seen when applying Gaussian models to bottom-up and top-down segmentation techniques. In the first case, the process starts by analysing small regions or even assuming each single pixel as a region. Therefore, the amount of points used to estimate the parameters of the GRF is too small, which leads to very poor estimates. Note that, for estimating the covariance matrix, a minimum number of points are necessary in order to avoid $\det(\underline{C}) = 0$.

A contour image of a segmentation of the Cameraman image is shown in Figure II.1.a. This result has been obtained by applying a bottom-up technique (region growing) and an image model using white GRFs. The oversmoothing of edges can be clearly seen in Figure II.1.b, where each region

is filled with its mean gray value. This procedure for showing the segmentation results will be utilised throughout the sequel.

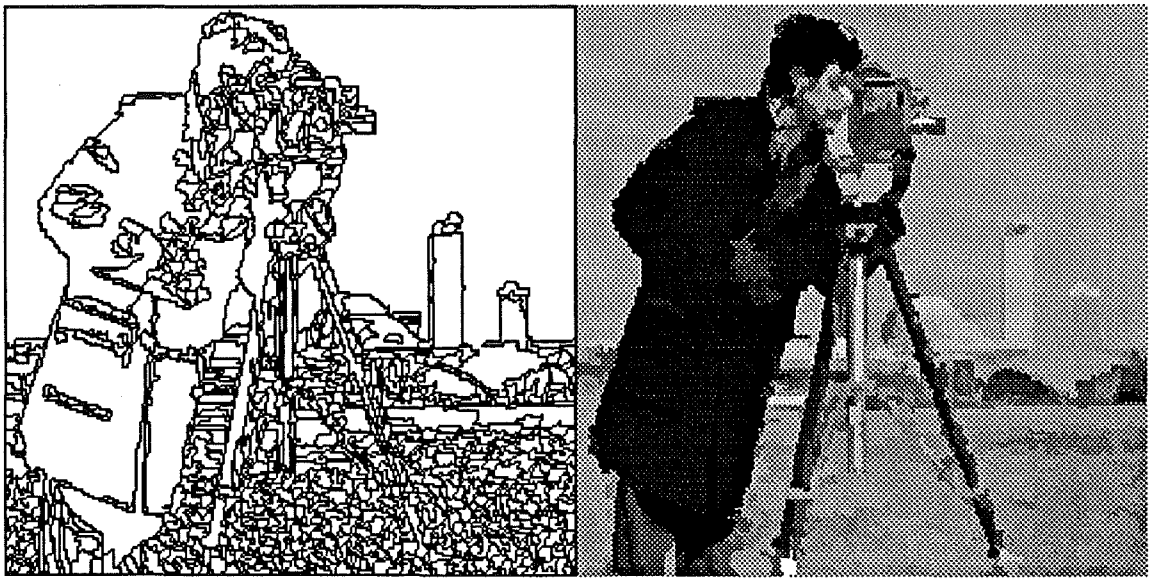

Fig. II.1.- Example of segmentation using bottom-up techniques and GRFs

In top-down techniques, the problems come from the fact that the low resolution analysis (general view of the image) may overlook some small details. This loss of detail is owing to the fact that the parameter estimate of a large area containing a small region does not differ too much from the estimate of the same area without this interior region. This effect results in an overgrowing of regions. This problem can be seen in Figure II.2 in which a top-down segmentation of the Cameraman image is shown. The segmentation has been carried out by a "split & merge" algorithm and using GRF models.

In order to solve the parameter estimation problem, different solutions have been proposed. The estimation can be helped by roughly marking the transition between zones corresponding to different region models. To this end, some approaches introduce external information in the estimation process (the so-called visibility functions). This information governs the switching between homogeneous zones. Visibility functions can be obtained by different methods: from simple approaches based on local variance estimates [43] up to more elaborated techniques dealing with directional and high frequency components [44]. The main problem with all these techniques is that the resulting functions are rather ad hoc.

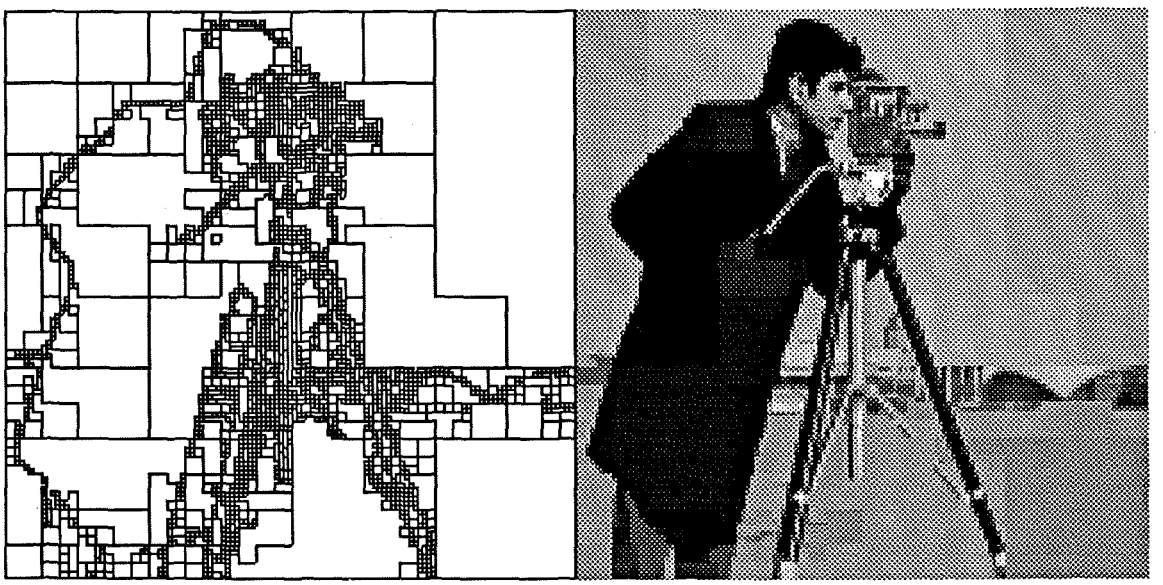

Fig. II.2.- Example of segmentation using top-down techniques and GRFs

As it has been pointed out, white Gaussian models do not help characterising complex regions. A way to improve the gaussian models performance is by withdrawing the white assumption on the random field. This translates into a covariance matrix different from diagonal. Therefore, samples within regions are no longer independent, but do depend on some other points of the image. This improvement leads to the concept of Markov random fields.

II. 2.- Markov random fields

Several 1-D and 2-D Markov processes (Markov Chains and Markov Random Fields -MRFs-, respectively) have been proposed in the literature over the years [41, 45]. In the image processing framework, MRFs provide with a well-suited tool for characterising the existing relationship among nearby pixels. In this section, image models based on MRFs are presented. Since images are defined on a discrete, finite lattice S (II.1) and can only take discrete values, the discussion is restricted to discrete-index MRFs defined on discrete, finite lattices.

II.2.1.- Neighbourhood systems

In order to define MRFs, a neighbourhood system over the lattice $\mathbb{S}$ has first to be defined:

Definition II.1: A neighbourhood system η related to a lattice $\mathbb{S}$ is defined as:

$$\eta = \{ \eta_{ij} \subset \mathbb{S} \ / (i, j) \in \mathbb{S} \} \qquad \text{(II.5.i)}$$

where η_{ij} is the neighbourhood of site (i, j). This neighbourhood is such that

$$\text{a) } (i, j) \notin \eta_{ij} \qquad \text{(II.5.ii)}$$

$$\text{b) } (k, l) \in \eta_{ij} \text{ implies } (i, j) \in \eta_{kl} \qquad \text{(II.5.iii)} \bullet$$

Since causality is not a natural assumption in image processing, a hierarchical sequence of non-causal neighbourhood systems is often used. The hierarchy of these systems is related to the maximum distance between sites within the system.

Definition II.2: The distance between site (i, j) and (k, l) on a discrete, rectangular lattice is defined as an euclidean distance

$$P = \sqrt{(i - k)^2 + (j - l)^2} \qquad \text{(II.6)} \quad \bullet$$

Given that images are usually defined on a rectangular lattice, the distance P can only take a discrete set of values, namely $\{1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}, \ldots\}$. The order in this set gives the hierarchy among systems. Therefore, a first-order system η^1 consists of those points located at most at a distance P = 1 of the central point (that is, its four nearest neighbours); a second-order system η^2 consists of these points located at most at a distance $P = \sqrt{2}$ of the central point (that is, its eight nearest neighbours); and so on. In Figure II.3 these two first systems are shown, as well as the sequence of neighbourhoods systems up to η^5.

It is worth noting that, although these are the most used neighbourhood systems in image processing, other systems can also be defined. In particular, it has to be said that non-isotropic systems satisfy conditions (II.5.ii) and (II.5.iii). Actually, causal systems (and thus, non-isotropic) have been widely used in image processing [42, 45].

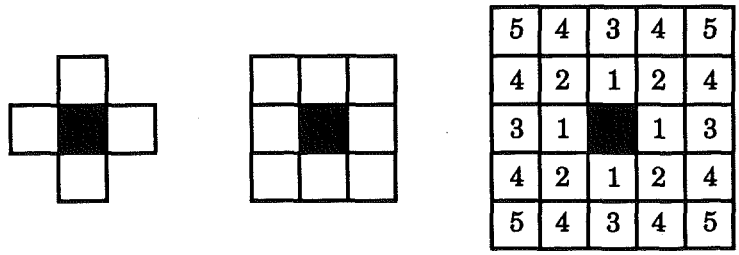

Fig. II.3.- Hierarchical neighbourhood systems

II.2.2- Definition of Markov random fields

Once the concept of neighbourhood system has been introduced, Markov random fields can be defined:

Definition II.3: Let $X = \{ X_{ij} \}$ be a random field; that is, a collection of random variables, one at each site of the lattice $\mathcal{S}$. X is said to be a Markov random field with respect to $(\mathcal{S}, \eta)$ and $P(X = x) > 0$ for all x if and only if

$$P(X_{ij} / X_{kl}, (k, l) \in \Omega) = P(X_{ij} / X_{kl}, (k, l) \in \eta_{ij}) \qquad (II.7)$$

where Ω is any subset of sites of $\mathcal{S}$ containing completely η_{ij} but not (i, j). ●

As it can be seen, the definition of a MRF is not very restrictive. Actually, any random field defined over a finite lattice is a MRF simply assuming as neighbourhood: $\eta_{ij} = \mathcal{S} \setminus (i, j)$. It has to be noticed that neither definition II.1 nor definition II.3 demand the neighbourhood system to be the same through the whole lattice. Therefore, a MRF having a neighbourhood system varying with the position, $\eta = \eta(i, j)$, can be defined. Indeed, this choice is seldom done and MRFs are chosen to be homogeneous - same η_{ij} for every (i, j) -. However, non-homogeneous MRFs provide with a theoretical way of handling the problem of boundaries in the finite lattice case. In these cases, sites near the boundary have associated systems with fewer neighbours than interior sites. This is a more natural manner of coping with the boundary problem than a toroidal lattice or any periodical assumption.

Using the concept of hierarchical neighbourhood systems, Markov-P models have been defined [46]. The general definition deals with continuous fields with global Markovian properties, but here, for the sake of simplicity, only discrete Markov-P random fields with local Markovian properties are

defined. Local properties imply that the conditional statistics of the value at a single site X_{ij} are stated. However, when working with global properties, the conditional statistics refer to the values of a set of sites $\{ X_{ij} \}$.

Definition II.4: A random field X is said to be a Markov-P random field with $P(X = x) > 0$ for all x if and only if

$$P(X_{ij} / X_{kl}, (k, l) \in \Omega) = P(X_{ij} / X_{kl}, (k, l) \in \eta^{\pi}) \qquad (II.8)$$

where Ω is defined as in (II.7) and π is the index in the neighbourhood system hierarchy related to the distance P. ●

Following this definition, a Markov-1 and a Markov-$\sqrt{2}$ random field are equivalent to a first-order and a second-order MRF, respectively. Note that the class of Markov-P models is less general than MRF models since the former is restricted to have a neighbourhood system with circular symmetry (isotropic neighbourhood), whereas the latter is not.

As it has been highlighted, although Markov-P models are widely used, other kind of MRFs relying on different neighbourhood systems have also been studied. For instance, early works, such as [47], make use of causal neighbourhoods. These models are usually referred as Markov mesh random fields or unilateral MRFs.

II.2.3.- Gibbs distributions

It is worth noticing that, until this moment, when defining an MRF no constraint has been put on the kind of conditional probability defining the random field. As it has been pointed out by several authors [45, 46], problems arise when the behaviour of a random field is characterised only in terms of its conditional probabilities $P(X_{ij} / X_{kl}, (k, l) \in \eta)$, without regarding the resulting joint probability $P(X = x)$. This set of conditional probabilities cannot be arbitrarily specified expecting that a joint probability, consistent with it, exists.

This problem is solved in the unpublished work by Hammersley and Cliffor. In it, the necessity of defining a Gibbs distribution (GD) in order to have a MRF is stated. GDs provide with a way for defining conditional probabilities consistent with the joint probability. An extensive study of the relation between Gibbs distributions and MRFs, as well as of different kinds of random fields that GDs generate, can be found in [48]. Here, the discussion is restricted to the

finite lattice case. In order to define Gibbs distributions, the concept of clique has to be first introduced.

Definition II.5: Given a lattice $\mathcal{S}$ and a neighbourhood system η, a set of sites c is said to be a clique if and only if this set is

a) a single point, or

b) if $(i, j) \neq (k, l)$, if $(i, j) \in c$ and if $(k, l) \in c$ then $(i, j) \in \eta_{kl}$.

That is, if any two elements in the set are neighbours of each other. The set of all cliques $\{ c \}$ of $(\mathcal{S}, \eta)$ is denoted by $\mathcal{C}$. ●

The set of cliques which can be defined on a first-order $\mathcal{C}^1$, and on a second-order $\mathcal{C}^2$ neighbourhood systems are shown in Figure II.4. Note that a configuration containing three sites in diagonal does not form a clique in η^2. This exclusion is owing to the fact that an extreme site in this configuration does not belong to the neighbourhood of the other extreme site. The extension to higher order neighbourhoods is straightforward. It has to be noticed that, by definition, $\mathcal{C}^n$ contains not only the set of cliques only related to η^n but also the whole cliques of $\mathcal{C}^{n-1}$.

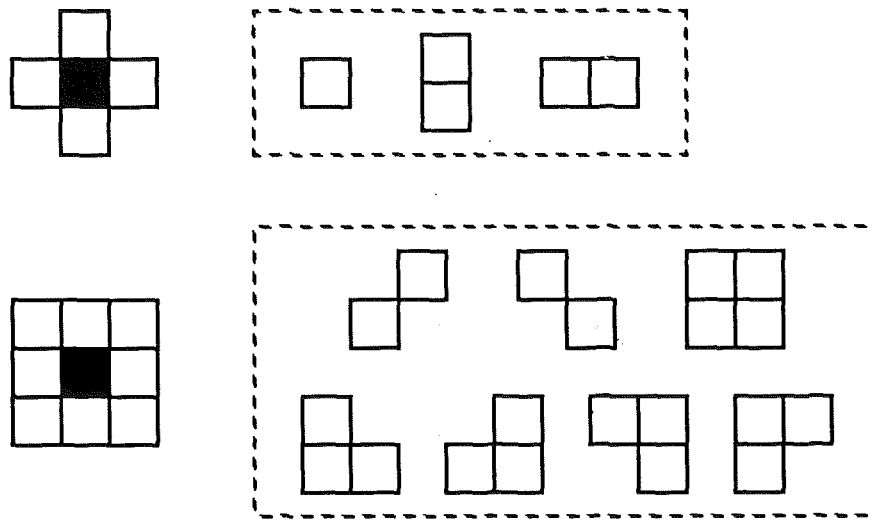

Fig. II.4.- First-order and second-order neighbourhood systems and their sets of cliques

Once the clique concept has been introduced, a Gibbs distribution can be defined as follows:

Definition II.6: A random field $X = \{ X_{ij} \}$ defined on a lattice $\mathcal{S}$ has a Gibbs distribution or is called to be a Gibbs random field with respect to

a neighbourhood system η if and only if its joint distribution is of the form

$$P(X = x) = \frac{1}{Z} \exp(-U(x)), \qquad (II.9)$$

where U(x) is the energy function, and Z is simply a normalising constant called the partition function:

$$Z = \sum_{x} \exp(-U(x)). \qquad (II.10) \bullet$$

The energy function U(x) is usually expressed as

$$U(x) = \frac{1}{T} \sum_{c \in C} V_c(x) \qquad (II.11)$$

where x is the site on which the neighbourhood is computed, $V_c(x)$ represents a potential associated with the clique c of the neighbourhood of x and T is a constant that stands for "temperature". Taking into account that the origin of GDs lies in statistical mechanics, the terminology used in their definition should not surprise. Actually, the first known work on GDs dates of 1925 and dealt with modelling the magnetic properties of ferromagnetics materials (the nowadays called Ising model [48]). For such a goal, a binary, homogeneous, isotropic field was proposed.

The physical interpretation of the joint distribution of (II.9) is that the smaller the energy of a realisation U(x), the more likely the realisation is. Moreover, paying attention to (II.11) it can be seen that, for the same clique potentials in a realisation, a variation of the temperature T results in a variation of the probability. This variation is in such a way that when T is large (the system is hot) the probability is less dependent on the potentials, and several realisations have similar high probabilities. On the contrary, when T is small (the system is cold), the influence of the potentials increases and the distribution P(X = x) becomes more peaked. Note that the behaviour of this distribution seems suitable for modelling natural systems such as lattice gas or ferromagnetics systems.

As it has been said above, the main characteristic of GDs is their one-to-one relation equivalence with MRFs. This equivalence is established in the Hammersley and Clifford's theorem.

Theorem II.1: A random field $X = \{ X_{ij} \}$ is a MRF with respect to the neighbourhood system η and $P(X = x) > 0$ for all x, if and only if X is a Gibbs random field with respect to the same system η and its associated cliques. ●

This theorem was proven in the 1970's independently by several researchers. For a proof, as well as for a detailed account of the MRF-GD correspondence, the reader is refereed to [48] and references therein. The power of this theorem is that, unlike in the simple MRF case, the GD characterisation of the joint distribution of the random field is free from consistency problems. Furthermore, the MRF-GD equivalence translates into a one-to-one relationship between a global description of a random field (provided by GDs) and a local one (provided by MRFs). Therefore, this equivalence leads to the possibility of using global characteristics when modelling random fields, but local features when processing them.

In the segmentation framework, MRFs can be used to obtain global maximum likelihood segmentations of images modelled by elaborate random fields. As it has been said in Chapter I, a complex random field can characterise textured areas by taking advantage of spatial correlation among samples. Due to the MRF-GD equivalence, algorithms can yield global maximum likelihood segmentations coping only with local computations. As it will be seen in the following sections, the main problem when seeking the maximum of likelihood functions is that these functions are neither linear nor unimodal. Therefore, maximisation procedures demand the use of elaborated techniques or suboptimal methods.

The Gibbs description of a random field is basically performed by means of an exponential distribution. Nevertheless, the definition of the clique potentials gives as much flexibility as the definition of the neighbourhood system in the case of MRFs. Thus, by correctly choosing the set of potential $V_c(x)$, very complex distributions can be formulated as special cases of GD, as above stated. A very detailed description of these possibilities can be found in [45, 48]. Here, only three cases are discussed, presenting for each one their applications to image processing problems as well as their main advantages and drawbacks.

II.2.4.- Gaussian Markov random fields

As it has been stated in Section II.1, the assumption of a Gaussian random field (GRF) with a region of support in which the samples are correlated leads to the concept of MRFs. Therefore, any GRF can be defined and studied as a specific case of MRF (even the case of white gaussian random fields if the neighbourhood system is reduced to the zero site case η^0). Actually, several authors have proposed and studied different Gaussian image models which are special cases of MRFs. In this way, Gauss-Markov Fields are discussed in [46], Auto-normal Processes in [48] and Spatial Markov Processes in [49].

The probability density function of a Gaussian random field is given in (II.3). By assuming zero mean ($\mu = 0$) for simplicity, the exponent of this expression reduces to a function of ($\mathbf{x}^T \underline{C}^{-1} \mathbf{x}$). Note that the rows of the matrix have zeros everywhere except in a few locations. Actually, in a generic i-th row, these non-zero elements correspond exactly to the locations of the neighbours of the i-th pixel in the neighbourhood system η. This analysis allows to write the conditional probability of an element in a GRF, given the value of the other sites in $\mathcal{S}$, as [45]:

$$P\left(X_{ij} / X_{kl}, (k, l) \in \Omega\right) = \frac{1}{\sqrt{(2\Pi)\,\sigma^2}} \exp\left\{ -\frac{\left(X_{ij} + \sum V_{i-k,\, j-l}\, X_{kl}\right)^2}{2\sigma^2} \right\} \quad \text{(II.12)}$$

where Ω is defined as in (II.7) and the summation is performed on $(k, l) \in \eta_{ij}$. Note that (II.12) satisfies (II.7) and, therefore, the statement saying that any GRF is a special case of MRFs is true. This kind of MRF is called a Gaussian Markov Random Field (GMRF).

II.2.5.- Auto-binomial models

From the class of MRFs called auto-models [48], auto-binomial models are one of the most utilised subclasses. It is easier to specify them in terms of their conditional probabilities, rather than in terms of the clique potentials. The model assumes that the conditional probability is binomial with parameter θ and number of tries G-1 [41]. Moreover, each random variable X_{ij} within the random field can take values in the set {0, 1, 2, ..., G-1}, where, in the image modelling case, G is the number of possible gray levels. Therefore, the expression for the conditional probability is

$$P\,(X_{ij} = m \,/\, X_{kl},\ (k, l) \in \eta_{ij}) = \binom{G-1}{m} \theta^{\,m} (1 - \theta)^{\,G-1-m}\,, \qquad \text{(II.13)}$$

where $\theta = \dfrac{\exp(T)}{1 + \exp(T)}$, T being a function of η_{ij}. For a neighbourhood of first order, $\eta^{\,1}$,

$$T = \alpha + \beta_1 (x_{i,j-1} + x_{i,\,j+1}) + \beta_2 (x_{i-1,j} + x_{i+1,\,j})\,. \qquad \text{(II.14)}$$

For higher order neighbourhood systems, the parameter T follows an analogous expression. The main interest of auto-binomial models is that the values of their parameters (α, β_1, β_2, ...) provide with an easy tool to control the main directions of the texture that they are modelling. For instance, in the first order system, β_1 controls clusterings in the vertical direction whereas β_2 performs the horizontal control.

II.2.6.- Strauss processes

First, it has to be underlined that the name given to this kind of random fields is not universally accepted. The same kind of models is sometimes referred as Multi-Level Logistic models, Generalised Ising models or used without giving any name to it. In this model, the clique potentials are defined depending on the homogeneity of the cliques; that is, given a non-single-site clique c

$$V_c(x) = \begin{cases} \beta_1 & \text{if all } x_{ij} \text{ in c are equal} \\ -\beta_2 & \text{otherwise}\,, \end{cases} \qquad \text{(II.15)}$$

where, usually, $\beta_1 = \beta_2$. Potentials of single-site cliques are defined separately: $V_c(x) = \alpha$.

In the case of binary-valued functions, Strauss processes, as well as auto-binomial models, are equivalent to the Ising model [48]. The main feature of Strauss models is that clique potentials only rely on whether all values within the clique are the same or not; that is, they do not take into account the actual values of X_{ij}s. This characteristic (sometimes referred as "colour-blindness") makes Strauss processes very suitable for modelling labelled images and Hidden Markov processes.

II.2.7.- On applying MRFs to image processing

In this section, an historical overview on the applications of MRFs within the framework of image processing is presented. Since the results obtained in different disciplines (mainly estimation, segmentation and synthesis) have interacted, this overview is not restricted to the segmentation case.

GMRFs are widely used in image processing, being first applied to the frameworks of image restoration, coding, estimation and synthesis [42]. As it has been highlighted in Section II.1, the main problem arises when trying to choose or to estimate the parameters of such models. In [50], studies on the optimum neighbourhood and estimation method for GMRFs are performed. The work is carried out by analysing several synthetic (Gaussian and non-Gaussian) texture images (each image containing a realisation of a random field). The estimates are verified by comparing the quality of the results analytically as well as visually.

New maximum likelihood estimation methods are proposed in [50] and compared with the coding estimate proposed in [48] which requires the solution of a set of non-linear equations. Better results are obtained both in computational load and in estimation accuracy by these new techniques. However, these techniques are still cumbersome and difficult to use reliably. Regarding the choice of neighbourhood system in a GMRF, an algorithm for estimating the optimum system is presented in [50]. It performs fairly well when the analysed random field is a GMRF itself, although it is shown that models with different parameters yield similar results. In the non-Gaussian case, though no results are shown, it is said that the quality of synthesised pictures is rather inferior.

A work similar to that in [50] is performed in [51] dealing, in this case, with auto-binomial models. In this study, natural and synthetic textures are analysed both in the binary and gray level case. When dealing with gray level textures, the gray scale is quantised into eight gray levels, given that the estimation algorithm proposed is very sensitive to the number of gray levels. Moreover, the largest neighbourhood system studied is of fourth order, since, the estimation of parameters of larger ones requires too many random field samples. Actually, the authors claimed that even experiments on third-order models cannot be reliably performed.

The results obtained in the estimation and synthesis of binary random fields are good. However, in the case of gray level images, the results present some drawbacks. The reason for these drawbacks is mainly twofold. First, some gray level textures may need higher than third order models to be characterised [52] and, as it has been before commented, such models cannot be reliably analysed. Secondly, binomial models have an intrinsic problem when trying to model sharp transitions. This problem results in blurry synthesised textures. As it is explained in [51], "The binomial distribution is unimodal and, as a consequence, values above and below the mean gray value are highly probable also. This results in a tapering of the gray scale around maxima and minima. Such a tapering as one moves away from black or white points has an effect similar to a neighbourhood averaging or low-pass filter."

Summarising, image estimation relying on MRFs (either GMRFs or other kinds) arises some problems. Parameter estimation techniques are rather complex even for low order models. In addition, low order models do not characterise correctly all kinds of data. Finally, when using high order models, reliable estimations can only be achieved by supplying the algorithm with large amounts of data.

It is worth noting that, in the segmentation case, the fact of not knowing the boundaries of the models should be added to this list of problems; that is, not having a priori the segmentation of the image. Taking into account the problems when estimating the MRF parameters, the first intent to produce global maximum likelihood segmentations made use of GMRFs for modelling images, assuming that the parameters of the texture fields were known, as well as the number (N) of regions [53]. That is, the problem of estimating the model parameters is avoided and only the correct location of each model within the image is sought.

The algorithm presented in [53] uses a quadtree data structure and, at each block of the quadtree representation, a segmentation into two regions is carried out as a maximum likelihood estimation. That is, each block is supposed to consist of only two different regions, out of N possible ones, modelled each one by a GMRF. The sought segmentation of each block is that which maximises the likelihood of the data. At each level of the quadtree, the connectivity of regions is preserved as much as possible.

In a posterior extension of this work [54], the case of having GMRFs with unknown parameters is discussed. This work provides with two possible solutions for estimating the parameters: if the image is simple enough, to rely

on edge detectors in order to have a first estimate of the boundaries of the random fields; for complex images, to perform a rough segmentation and to estimate the parameters from it. In the case of coloured (non-white) GMRFs, the author proposes to apply first the algorithm assuming white GMRFs and to estimate the parameters of the coloured ones on the resulting segmentation. Then, the algorithm should be re-applied with the new parameters.

Main drawbacks of these algorithms are rather clear. First, in both cases, the number of regions within the scene has to be known before hand. Moreover, in order to overcome the problem of looking for the segmentation maximising the likelihood of the whole data, a two-region partition is assumed at each block of the quadtree decomposition. This simplification, which can be acceptable at low levels of the tree, may lead to really erroneous segmentations at its high levels. Note that, due to the quadtree data structure, mistakes made when segmenting at high levels cannot be fixed at lower ones. Finally, in the first algorithm [53], the different model parameters are assumed to be known. Furthermore, in order to withdraw this assumption in the second algorithm [54], some boundary information has to be introduced in the procedure. That is, either a contour detection or a previous rough segmentation has to be carried out.

In [55], a different approach for image segmentation, relying in stochastic image model, is presented. The image model is formed by causal GMRFs (Markov mesh random fields) defined on a second-order neighbourhood. The work handles the problem of segmenting binary images or gray level images into two regions. The segmentation is performed seeking the realisation of the model maximising the a posteriori probability, given the full initial image. The assumption of causal random fields allows the use of well-known 1-D techniques for the estimation problem. In spite of this, since the actual implementation of the algorithm involves prohibitive processing, a suboptimal approach is suggested. In this suboptimal algorithm, instead of using the data from the whole image, only the data contained in a small strip (3 pixel wide) are processed. A second version of the algorithm is also proposed, in which an iterative procedure is utilised to estimate the a posteriori probability.

Although the algorithm of [55] yields good results in the case of two-class images, the computational load of the method is rather prohibitive. This load increases when the algorithm is extented to more complex problems (for instance, gray level images to be segmented into more than two regions). Furthermore, the segmentation results are not very accurate in locating the

region boundaries. This problem is alleviated by a post-processing of the data (a median filtering is suggested).

Finally, it has to be pointed out that the solution of relying on a first estimation of the boundaries or of the random fields parameters to help the algorithm has been already commented in Section II.1. The problem with such techniques is that they are either rather ad hoc (case of visibility functions or similar) or time demanding (case of iterative methods as performing twice the segmentation). Therefore, it would be better if the information about the boundaries of the random fields could be included in the model itself.

II.3.- Compound random fields

In order to introduce boundary information in the image model, a new kind of random field has been proposed. This random field not only describes the behaviour of a given texture or zone of the image, but it represents the whole image. This global characterisation is achieved by using a model with two different levels, hierarchically distributed.

II.3.1.- Definition

The model assumes that an image is composed of a set of regions, each one characterised by an independent random field. The union of these random fields forms the observed image, named the upper level of the model. The location of these random fields within the image (position and shape of the regions) is governed by an underlying random field, named the lower level of the model. This kind of models are called hierarchical models [56] or Compound Random Fields (CRFs) [57].

Definition II.7: A Compound random field $(Q, X) = \{ Q_{ij}, X_{ij} \}$ defined on a lattice S is formed by a lower level Q and a upper level X. The lower level is a discrete valued random field where each Q_{ij} can take values from the set $\{1, \ldots, M\}$. The upper level is a random field where each X_{ij} can take values from a set of M independent random fields $\{ X^k \}$, with $k = 1, \ldots, M$, following

$$X_{ij}(\cdot) = X^k_{ij}(\cdot) \quad \text{if} \quad Q_{ij}(\cdot) = k \quad \text{for} \quad k = 1, \ldots, M. \tag{II.16}$$

Processes Q and $\{ X^k \}$ are mutually independent random fields. ●

Note that each realisation q of the process Q performs a partition of the lattice $\mathcal{S}$ into M region types. Therefore, in the case of segmented images, a realisation Q = q can be seen as an image of labels, relating each point to a region, and each one of the random fields X^k as the model for the texture of region k. In Figure II.5, this concept is depicted marking only the boundaries of the partition.

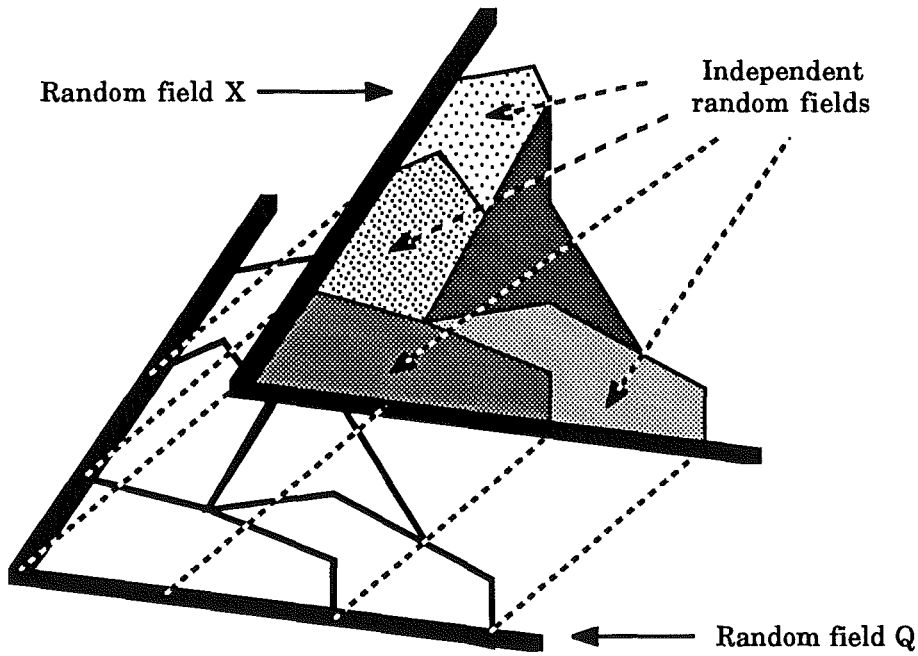

Fig. II.5.- Example of Compound random field

The usual procedure for dealing with this kind of models in the image processing framework is by assuming that the image to be processed is a realisation of the upper random field (X = x). Thus, the objective is to determine the realisation of the lower random field (Q = q) that has given rise to x. This objective can be achieved by a maximum a posteriori (MAP) estimation; that is, maximising the a posteriori distribution P (Q = q / X = x) for a given x [58]. Using Bayes' rule

$$P(Q = q / X = x) = \frac{P(X = x / Q = q)\, P(Q = q)}{P(X = x)}, \qquad \text{(II.17)}$$

where P (X = x) does not affect the maximisation procedure. Therefore, instead of maximising P (Q = q / X = x), since

$$P(Q = q, X = x) = P(X = x / Q = q)\, P(Q = q), \qquad \text{(II.18)}$$

the joint distribution P (Q , X) can be maximised, yielding the same result.

Nevertheless, functions representing these joint probabilities are very complicated, usually non-linear and multimodal. Therefore, since a brute force maximisation procedure is absolutely out of question (there are $M^{N_1 N_2}$ possible configurations), it is desired to find other kind of algorithms for maximising (II.17).

II.3.2.- Image segmentation algorithms using Compound random fields

The first work on image segmentation dealing with CRFs was presented in [59]. In it, the upper level of the model is formed by a collection of white Gaussian random fields and the lower level by a Strauss process defined on a first- or on a second-order neighbourhood system (both possibilities are studied). Whichever is the chosen system, only cliques containing one or two pixels have non-zero potential and $\beta_1 = \beta_2$. The maximisation algorithm is an extension of the "strip processing" [55] which has been outlined in Section II.2.7. That is, the maximisation procedure is performed, by means of a dynamic programming algorithm, on overlapping strips of small width. This algorithm requires knowing some boundary conditions, such as the segmentation of the first row and column.

As in [55], the restrictions of this algorithm are rather strong. The maximum feasible width for the strips is four pixels and, in order to be computationally tractable, the algorithm should perform segmentations into two regions - that is, $M = 2$ in (II.16) -. For segmenting the image into more than two regions ($M > 2$), the authors proposed to apply the algorithm for 2 regions M-1 times.

On the other hand, the behaviour of the compound model is remarkable. Disregarding the computational load, the algorithm performs fairly well: segmentations are accurate and robust to the choice of the parameter values. It has to be highlighted that, while the parameter values of the upper level are estimate from the data, these of the lower level are set before hand. Extensions of this work can be found in [60, 61], where the same algorithm is applied using other image models. However, results are very alike: same improvements with respect to non-hierarchical models and similar trade-offs in the dynamic programming algorithm.

A different approach for maximising (II.17) is presented in [56]. This work, although aimed at the restoration problem, performs an indirect segmentation of the image. Actually, it would not be worthwhile to present this work as a simple application of CRFs to image processing problems, given

that, along with [48], it set the theoretical bases for most of the posterior works in this field. In [56], the assumed model for the upper level X is formed by a set of MRFs, being the lower level Q a MRF as well (note that this defines the most general case). The interpretation of the random field Q is not only given in terms of a partition of the image, but the concept of "line process" is also introduced. That is, the process Q is said to characterise the shape of the regions in the image model. This idea opens the possibility of choosing the parameters of the model not only in order to control the switching between regions within the image but also the desired behaviour of the boundaries.

The algorithm used in [56] to perform the MAP estimation relies on a stochastic relaxation technique called simulated annealing [62]. It has been demonstrated that, with the proper annealing schedule, this technique can converge to the global optimum of a function. In [56], an annealing schedule is proposed, which is proven to lead to the optimum solution. This technique is interleaved with a maximum likelihood estimation of the random field parameters. Moreover, an algorithm for generating realisations from a given MRF, the so-called Gibbs Sampler, is also presented.

In order to test the performance of the maximising procedure, several experiments are carried out. In these experiments, Strauss processes defined on first- and second-order neighbourhoods are used for both the lower and the upper level models. Clique potentials are also taken as $\beta_1 = \beta_2$. As in the previous case, the obtained segmentations are of high quality but they required around 300 iterations of the algorithm (sometimes, even 1000). This is the main trade-off of simulated annealing techniques: in order to reach the global optimum, the schedule has to be much too slow to be practically useful. Furthermore, if a fast schedule is implemented, the algorithm is more sensitive to its initial conditions. On the other hand, it has to be emphasised the clear improvement introduced by the lower level process: some experiments are performed in the same conditions with and without the line process, yielding the former clearly better results.

Following these ideas, similar works have been presented. In this way, the iterative procedure for segmenting and estimating parameters proposed in [56] is further elaborated in [63]. Here, experiments are carried out by using white GRFs for the upper level, while keeping the Strauss process for the lower level. Results and conclusions are very similar to these of [56].

Trying to overcome the computational load of stochastic relaxation techniques such as simulated annealing, some authors have faced the

maximisation problem from a deterministic relaxation viewpoint. These methods perform the maximisation by changing the pixel classifications in a deterministic and iterative way. Therefore, although they are faster, they may get trapped in local maxima. These two techniques are compared in an image restoration framework in [57]. Image models are defined as second-order compound random fields: the "line process" of [56] for the lower level and both a set of causal and a set of non-causal GRFs for the upper level. The parameters of the lower level process are set heuristically, while the upper level ones are estimated from the data following the method described in [50].

Results show that there are no main differences between causal and non-causal models. On the contrary, the choice of compound versus simple random fields is of major importance. When comparing stochastic with deterministic techniques (simulated annealing versus iterated conditional modes), it is shown that better results are yield by stochastic ones. However, the difference in computational load is really relevant: while the order of 200 iterations are required in the first case, only around 25 are needed in order to reach a maximum in the second one. It has to be noticed that deterministic techniques usually depend very much on their initial conditions. This trade-off is not noticeable in [63] since the work deals with a restoration problem. However, when coping with segmentation, the importance of the initial estimate increases.

Finally, the work presented in [64] should be commented. In it, the image is assumed to be a compound random field consisting in a Strauss process for the lower level and a collection of auto-binomial models for the upper level. First, a supervised segmentation algorithm is proposed based on a previous knowledge of the random field parameters and a simulated annealing technique. An unsupervised algorithm is also proposed in which the value of the parameters of the lower level are set before hand. On their turn, lower level parameters are estimated from a coarse segmentation of the image performed by splitting it into a set of disjoint blocks. Although the assumptions for the unsupervised algorithm seem rather strong, both algorithms yield good quality results, being both kind of segmentations very alike.

II.4.- Summary

In this chapter, the main stochastic models used in image processing have been presented; namely, Gaussian random fields, Markov random fields and Compound random fields. GRFs and MRFs share the characteristic of being non-hierarchical. Furthermore, GRFs have been demonstrated to be a special case of MRFs. Both models have been defined and their advantages, as well as their drawbacks, pointed out. It has been shown that a random field, in order to be a MRF, must have associated a Gibbs distribution. This relation does not represent a serious constraint since, by choosing correctly the potentials defining a GD, a wide set of models can be obtained. Moreover, the MRF-GD equivalence allows algorithms to yield global maximum likelihood segmentations while dealing only with local computations.

It has been suggested that when using algorithms based on MRFs image models some problems arise. The estimation of the parameters characterising the models is not an easy task. In order to estimate them, large sets of data are required. Furthermore, the existing algorithms are rather cumbersome and little reliable. The situation worsens in the segmentation framework, since the boundaries of the random fields are not known a priori. In this case, rough information about the location of boundaries is usually introduced in the algorithms, which leads to rather ad hoc and/or time demanding techniques. Another problem comes from the fact that, in order to obtain global maximum likelihood segmentations, functions that are neither linear nor unimodal have to be maximised. This maximisation has to be carried out by means of very elaborated techniques, which imply a huge computational load, or suboptimal methods, which reduce the range of application of the methods and do not ensure global maxima.

Towards the goal of introducing boundary information in the algorithm, compound image models are defined. These models contain themselves the boundary information. Compound models consist of two different random fields, disposed in a hierarchical manner: the lower level (hidden random field) governs the upper level (observable random field) which is, on its turn, composed of several independent random fields. Given this kind of models, the segmentation problem can be formulated as a MAP estimation: assuming the original image as a realisation of the upper level of the model, the realisation of the lower model maximising the a posteriori probability is sought.

Several algorithms based on this MAP criterion and using compound random fields have been presented. The lower level of the CRF is usually

chosen to be a Strauss process whereas, for the upper level, different possibilities have been suggested (mainly GRFs, auto-binomial models and Strauss processes). The use of CRFs for modelling images have turned out to be a very powerful tool, yielding, in all cases, better results that using usual non-hierarchical models. Finally, it has been pointed out that results are fairly insensitive to the choice of the sort of random fields assumed to form the upper level.

Regarding the algorithms, mainly three techniques are used, namely: dynamic programming, stochastic relaxation (simulated annealing) and deterministic relaxation (iterated conditional modes). The first kind of techniques are very time consuming. Therefore, only suboptimal methods dealing with very simple cases and not ensuring a global maximum are feasible. Moreover, they depend on the initial conditions of the algorithm. On its turn, simulated annealing algorithms do converge to the global optimum. However, there exists a trade off between the computational load (the annealing schedule) and the sensitivity with respect to the initial conditions. In any case, simulated annealing techniques remain a computationally intensive method for optimisation. Finally, deterministic relaxation algorithms do not ensure convergence to the optimum maximum and are very dependent on the initial conditions. Nevertheless, they are computationally very efficient and may yield results of comparable quality to these obtained by stochastic relaxation techniques.