J Math Imaging Vis 25: 387–402, 2006

DOI: 10.1007/s10851-006-7251-1

Mixed-State Auto-Models and Motion Texture Modeling

P. BOUTHEMY
IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

C. HARDOUIN
SAMOS/Université de Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex 13, France

G. PIRIOU
IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

J. YAO
IRMAR/Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France

Published online: 29 August 2006

Abstract. In image motion analysis as well as for several application fields like daily pluviometry data modeling, observations contain two components of different nature. A first part is made with discrete values accounting for some symbolic information and a second part records a continuous (real-valued) measurement. We call such type of observations "mixed-state observations". In this work we introduce a generalization of Besag's auto-models to deal with mixed-state observations at each site of a lattice. A careful construction as well as important properties of the model will be given. A special class of positive Gaussian mixed-state auto-models is proposed for the analysis of motion textures from video sequences. This model is first explored via simulations. We then apply it to real images of dynamic natural scenes.

Keywords: mixed states, auto-models, Gaussian models, dynamic textures, motion analysis

1. Introduction

It is of common understanding that the type of any observation data is either continuous or discrete. The situation, where a measurement presents continuous values sometimes and discrete values at other times, is rarely considered in statistical literature. However, such situations are frequent in applications. For examples, daily pluviometry time series at a given site records many zeros when the rain is absent, followed by periods with positive rainfall values (see e.g. [1]). Similar phenomena also occur in speech recordings where, interchanges are permanent between absences and presences of the signal. Another example arises in the motion analysis problem from image sequences considered in this paper. Typically, the histograms of local motion measures present a composite picture. An important peak appears at the origin accounting for regions where no motion is present, while a large continuous component encompasses actual motion magnitudes in the images. It then raises the question to find accurate models for this type of data—we shall call them *observations with mixed states*, collected from the image lattice.

From a mathematical point of view, we are searching for models for a random field $\{X_s\}$ with the constraint

that the marginal distributions of the X_s's are composed with a discrete component and a continuous component. In its most general form and for the discrete component, we may take any distribution with support on a countable set $\{e_1, \ldots, e_k, \ldots\}$ of symbolic values, while for the continuous component any standard distribution could be considered. However in this work, we will restrict our attention to distributions with one atomic value, typically $\{0\}$, and a continuous component supported on the interval $(0, \infty)$. The state space, called a *mixed-state space*, is then $E = \{0\} \cup (0, \infty)$ with the point 0 playing a special role.

Markov random fields (MRF) models are now a standard tool in image analysis, see e.g. [4]. However, up to our knowledge, the existing models deal with either continuous variables, or discrete variables, but never with variables that can take continuous as well as discrete values. Furthermore, the discrete component could not be simply neglected, because -as it will becomes clear in the motion analysis application addressed below-, these symbolic values as well as their spatial correlations convey important pixel-wise and contextual information. On the other hand, such discrete phenomena are usually taken into account by introducing a label process L_s where, in our case, $L_s = 1$ if no motion is present at pixel s, i.e. $X_s = 0$, and $L_s = 2$ when a positive motion measurement is recorded, i.e. $X_s > 0$. However, the label process is a hidden process and the resulting statistical inference methods need in general a restoration of the hidden process (i.e., segmentation). This classical approach is then possible only upon the cost of a generally huge computation effort.

The approach we propose is different. We aim at giving a model which automatically deal with the two types of observations, without introduction (and then the inference) of a hidden process. The main idea is then to introduce mixed-state distributions in a random field set-up. More precisely, we will follow J. Besag's construction of auto-models [3] by introducing necessary adaptations for mixed-state variables.

In Section 2, we first recall some backgrounds on motion computation in an image sequence. This will also give a precise description of the data we have at hand, and in particular their mixed-state nature. In Section 3, we introduce a new class of random field models, named *mixed-state auto-models*. The construction, their basic properties as well as estimation methods will be given. A special specification based on positive Gaussian distribution for the continuous component is introduced. Then in Section 4 we explore the basic properties of the proposed auto-model via several simulation experiments. Next in Section 5, we carry out an application to the analysis of motion textures from real image sequences of dynamic natural scenes. Finally, concluding remarks are given in Section 6.

2. Motion Analysis from Image Sequences

The data we are dealing with are local motion measurements from video sequences. For the definition and the computation of these measures, we follow the approach developed in [12, 13], where several meaningful applications are also given. Here, we briefly recall the basics of their computation.

2.1. Motion Decomposition and the Residual Motion

The aim of motion computation is to obtain dense motion measures that can be easily and reliably computed from any image sequence (e.g. videos) and that inform on the dynamic content of the depicted scene. The motion information in an image sequence is completely captured by the optical flow. However, methods for estimating optical flow remain complex and time consuming if general video content has to be handled, while not always ensuring accurate and reliable measurements. As a consequence, motion measurements we will analyze in the sequel are related to normal flow. On one hand, they supply a partial motion information only since they correspond to the motion component parallel to the spatial image gradient. On the other hand, they can be locally computed in a straightforward way. In case of a static camera, these motion measures are directly related to the scene motion. More precisely, it is image motion conveyed by spatio-temporal photometric variations and related to the projection of the scene motion. If the camera is moving, we have first to cancel the camera motion. In that case, we will compute local motion features related to the so-called residual normal flow. More precisely, we first estimate the dominant image motion which can be assumed (in most cases) due to the camera motion. The following 2D-affine motion model (which is a usual choice, a 8-parameter quadratic model could be considered as well) is considered:

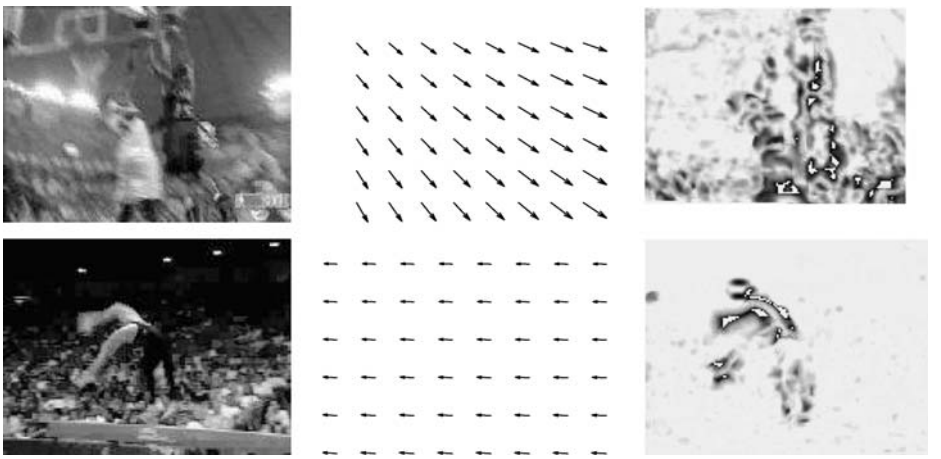


Figure 1. Two images of sports video segments (involving respectively, a zoom combined with an upward-tilt camera motion, and a right panning motion) and their corresponding maps of the estimated dominant image motion fields $w_{\widehat{\xi}}$ and of residual motion measurements v_{res} (white = 0; black = maximum value).

$$\mathbf{w}_{\xi}(p) = \begin{pmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{pmatrix}, \tag{1}$$

where $\xi = (a_i, i = 1, \ldots, 6)$ is the model parameter vector and $p = (x, y)$ is an image point. This simple motion model can correctly handle different camera motions such as panning, zooming, tracking. Different methods are available to compute such a motion model and to get the parameter estimate $\widehat{\xi}$. We use the robust real-time multi-resolution algorithm described in [13].

Then, the residual motion measurements $v_{res}(p, t)$ we consider, are defined as the weighted local mean of the normal residual flow magnitudes $|v_n|$, the weights being given by the square of the magnitude of the spatial intensity gradient. This allows us to smooth out the noise attached to the computation of the normal flow and to enforce the reliability of the motion measurements. We get the following expression:

$$v_{res}(p, t) = \frac{\sum_{q \in \mathcal{F}(p)} \|\nabla I(q, t)\|^2 . |v_n(q, t)|}{\max(\eta^2, \sum_{q \in \mathcal{F}(p)} \|\nabla I(q, t)\|^2)}, \tag{2}$$

with $v_n(q, t) = [I(q, t) - I(q + w_{\widehat{\xi}_t}(q), t + 1)]/\|\nabla I(q, t)\|$. Here, $\mathcal{F}(p)$ is a local spatial window centered in pixel p (typically a 3×3 window), $\nabla I(q, t)$ is the spatial intensity gradient of pixel q at time t and η^2 is a predetermined constant related to the noise level. This class of local motion measurements have already been proved useful for motion detection [11, 12] and for motion recognition [8]. Figure 1 displays two images of sports videos with the corresponding maps of the estimated dominant motion vectors and those of residual motion measurements v_{res}. These examples show that the camera motion is reliably captured even in case of multiple moving elements in the scene. It also indicates that the scene motion is correctly accounted by the residual motion measurements. From Eq. (2), it can be straightforwardly noted that we only get information related to motion magnitude, and consequently, we lose the motion direction. However, under the general objective of motion characterization, we aim at addressing issues such as detecting similar motion contents, grouping "qualitative" motion classes, or recognizing predefined motion classes. As demonstrated by the results reported later, these goals can be attained using this type of motion information.

2.2. *Motion Textures, Histograms and Mixed States*

The above motion computation principle has been applied to various video sequences. In this paper, we are mostly concerned with motions of *natural dynamic scenes* such as views of moving grass, moving foliage, sea waves, rivers, fire, steam, smoke, etc. As motivated in Section 5, we will call *motion textures* the resulting local motion measurements $\{v_{res}\}$. Figure 2 gives a set of sample images from six different types of videos including grass, foliage, sea-waves, trees and turbulent rivers, respectively. The corresponding motion fields $\{v_{res}\}$ are displayed in Fig. 3. As a matter of fact, these

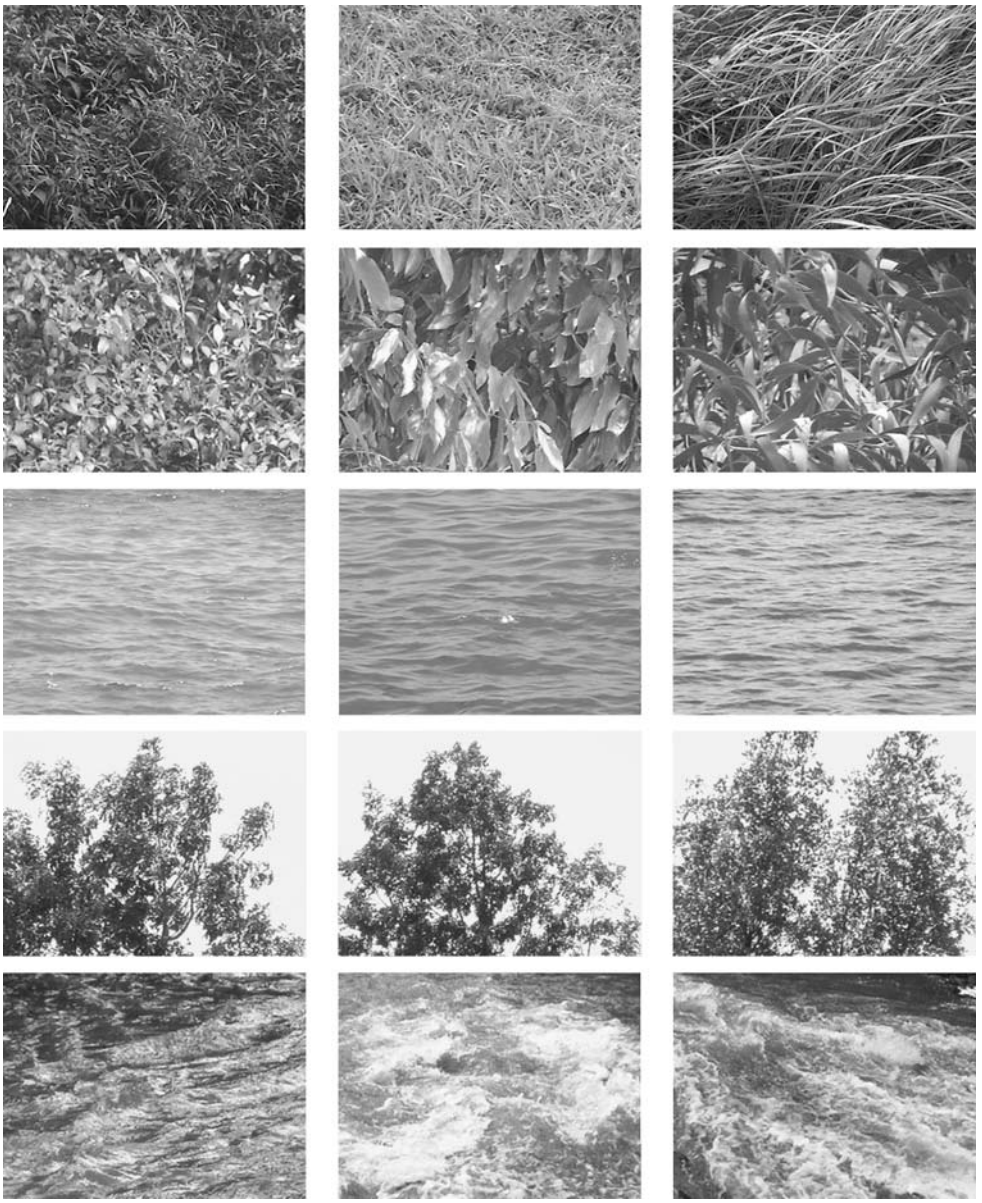


Figure 2. Sample images from different videos of natural dynamic scenes. Top to bottom: moving grass, foliage, sea-waves, trees and rivers.

videos were acquired with a static camera; therefore, the $\{v_{res}\}$ are computed with $\mathbf{w}_\xi \equiv 0$ in (2). As illustrated in Fig. 3, the normal flow magnitudes can be sufficient to reveal the intrinsic space-time arrangement of dynamic textures and then to properly model and characterize them, even if they supply a partial motion information only (local motion direction is not involved as pointed out before). An advantage is that we have to deal with scalar motion fields. Besides, local motion directions do not convey really pertinent information by themselves in case of dynamic textures.

In Fig. 4, we have displayed several typical histograms from motion textures $\{v_{res}\}$. As explained in Introduction, these histograms are of mixed-state type with a prominent peak at the origin accounting for regions where no motion is present, and a continuous component reporting the magnitudes of actual motion in the images. This can also be inferred from

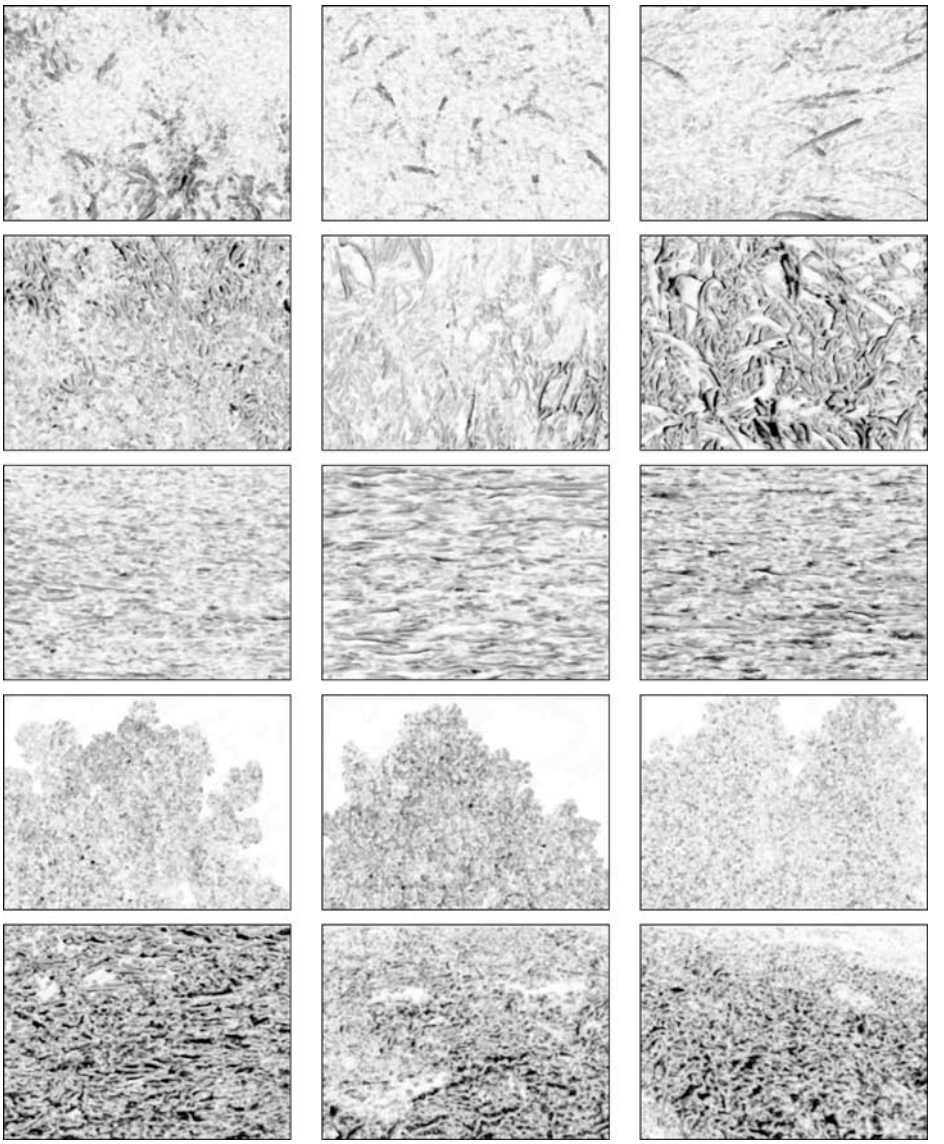


Figure 3. Sample motion measures $\{v_{res}\}$ from the videos of Fig. 2. Top to bottom: grass, foliage, sea-waves, trees and rivers (white = 0; black = maximum value).

the examples of normal flow magnitudes of Fig. 3 which typically exhibit heterogeneous content of mixed static and moving points in a more or less random way.

3. Mixed-State Auto-Models for Motion Textures $\{v_{res}\}$

Our main purpose is to construct a random field model for the (residual) normal flow magnitudes fields $\{v_{res}\}$, introduced in Section 2, which are mixed-state observations. Indeed we need to consider a general class of random field models called *multi-parameter auto-models*. Such a theory is exposed in Section 3.1. Using this theory we then introduce in Section 3.2, a *positive Gaussian auto-model* with mixed states in $\{0\} \cup (0, \infty)$ for modeling motion textures. A specification for the four nearest-neighbours system is proposed in details in Section 3.3. Such positive Gaussian auto-models will be used in Section 5 to analyze motion textures $\{v_{res}\}$.

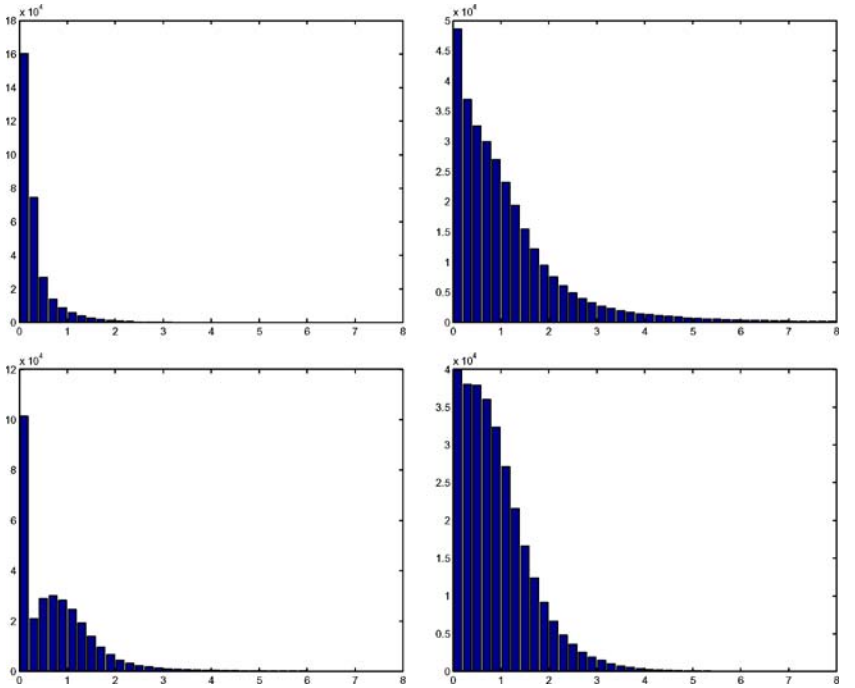


Figure 4. Sample histograms of motion measures $\{v_{res}\}$. Top to bottom, and left to right: grass, foliage, trees and sea-waves.

3.1. A General Theory of Multi-Parameter Auto-models

Consider a general system of real random vectors $\{X_i, i \in S\}$ indexed by a finite set $S = \{1, \ldots, n\}$. For a *site* i, let

$$\mu_i(x_i|\cdot) = \mu_i(x_i \mid x_j, j \neq i) \ ,$$

the probability density of X_i given the event $\{X_j = x_j, j \neq i\}$. A classical approach in stochastic modeling consists in specifying the family of all these conditional distributions $\{\mu_i\ (x_i|\cdot)\}$, and then to determine a joint distribution μ of the system, which is compatible with this family, i.e. the μ_i's are exactly the conditional distributions associated to μ. We refer to the seminar paper [3] which presents general results including a summary of earlier results about the "nearest neighbours systems" from [2, 15].

In this paper, we focus our attention on the auto-models introduced by [3]. A characteristic property of these auto-models is that the local conditional distributions belong to a one-parameter exponential family, such as exponential distribution or Poisson distribution, giving the auto-exponential model and the auto—Poisson model, respectively. Note also that these auto-models belong to the wider class of Gibbs field models (or Markov random field models).

To handle mixed-state observations addressed in this paper, we first need to extend the above *one-parameter auto-models* to a multi-parameter setting. Let us specify some notations. We are given a measurable state space $(E, \mathcal{E}, m)$ where E is a subset of $\mathbb{R}^d$. The field X is taking values in a *configuration space* $\Omega = E^S$, equipped with the product structure $(\mathcal{E}, m)^{\otimes S}$. A random field on S is specified by a probability distribution μ on Ω. We will always assume that μ has a *everywhere positive density* P with respect to the product measure $\nu = m^{\otimes S}$. In other words,

$$\mu(dx) = P(x)\nu(dx), \quad P(x) = Z^{-1} \exp Q(x), \quad (3)$$

where Z is a normalization constant. The positivity condition implies that at each site i, the conditional distribution $(X_i|X_j = x_j, j \neq i)$ has a positive density $\mu_i(x_i|\cdot)$ with respect to $m(dx_i)$.

Our construction of multi-parameter auto-models are based on the following three conditions:

[A] The potentials involve at most two points i.e.

$$Q(x) = \sum_{i \in S} G_i(x_i) + \sum_{\{i,j\}} G_{ij}(x_i, x_j) \ .$$

[B] For each site i, the conditional distribution $\mu_i(x_i|\cdot)$ belongs to a *multi-parameter exponential family*:

$$\log \mu_i(x_i|\cdot) = \langle A_i(\cdot), B_i(x_i)\rangle + C_i(x_i) + D_i(\cdot) \ ,$$

$$A_i(\cdot) \in \mathbb{R}^d, \ B_i(x_i) \in \mathbb{R}^d.$$

[C] The family of sufficient statistics $\{B(x_i)\}$ is *regular* in the sense that

$$\text{for all } i \in S, \ \ \text{Span}\{B_i(x_i), x_i \in E\} = \mathbb{R}^d \ .$$

We have then the following theorem (for a proof, see [10]).

Theorem 1. *Assume that the random field probability distribution μ of and its energy function $Q(x)$ satisfy Conditions **[A]–[B]–[C]**. Then, there are for all $i, j \in S$, $i \neq j$, a family of vectors $\alpha_i \in \mathbb{R}^d$ and a family of $d \times d$ matrices β_{ij} satisfying $\beta_{ij} = \beta_{ji}^T$, such that*

$$A_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B_j(x_j) \ . \tag{4}$$

Consequently the set of potentials is given by

$$G_i(x_i) = \langle \alpha_i, B_i(x_i)\rangle + C_i(x_i) \ , \tag{5}$$

$$G_{ij}(x_i, x_j) = B_i^T(x_i)\beta_{ij} B_j(x_j) \ . \tag{6}$$

A model satisfying the assumptions of the theorem is called a *multi-parameter auto-model*. For a practical use of Theorem 3.1 as in Section 3.2 below, one starts by specifying a family of local conditional distributions $\{\mu_i\}$ satisfying Conditions **[B]–[C]**. The extra-condition that need to be checked is that the associated "energy" function $Q(x)$ is *admissible* in the sense that:

$$\int_\Omega \exp Q(x) \nu(dx) < \infty. \tag{7}$$

3.2. Mixed-State Auto-Models for Motion Textures

We aim to construct auto-models for mixed-state observations $\{v_{res}\}$ on the state space $E = \{0\} + (0, \infty)$.

3.2.1. The Positive Gaussian Mixed-State Distribution Let us first define a family of distributions on E which will be used below for modeling local conditional distributions of motion textures $\{v_{res}\}$. A variable X of this family, called *positive Gaussian mixed-state variable*, is constructed as following: with probability $p \in (0, 1)$ we set $X = 0$, and with probability $1 - p$, X follows the distribution of the *module* of a zero-mean normal distribution with variance σ^2:

$$g_s(x) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = g_s(0) e^{-sx^2} \ .$$

Here we have set $s = (2\sigma^2)^{-1}$. Note that $g_s(0) = 2(2\pi\sigma^2)^{-1/2} = 2(s/\pi)^{1/2}$.

It is worth noticing that this positive Gaussian distribution for the continuous component is a natural choice regarding the sample histograms of motion textures as displayed in Fig. 4. We now compute the density function of this distribution. Let the space E be equipped with a "mixed" reference measure

$$m(dx) = \delta_0(dx) + \lambda(dx) \ ,$$

where δ_0 is the Dirac measure at 0 and λ the Lebesgue measure on $(0, \infty)$. Let us define the indicator function $\delta(x) = \mathbf{1}_{\{0\}}(x)$ and its complementary function $\delta^*(x) = 1 - \delta(x)$. Then, the above random variable X has the following density function, w.r.t. $m(dx)$,

$$\begin{aligned} f_\theta(x) &= p\delta(x) + (1-p)\delta^*(x) g_s(x) \\ &= \exp\left[-\delta^*(x) \log \frac{p}{(1-p)g_s(0)} - sx^2 + \log p \right] \\ &= \exp[\langle \theta, B(x)\rangle + \log p] \end{aligned} \tag{8}$$

where we have set

$$\theta = (\theta_1, \theta_2)^T = \left(\log \frac{(1-p)g_s(0)}{p}, s \right)^T \ ,$$

$$B(x) = (\delta^*(x), -x^2)^T \ .$$

In other words, this distribution belongs to an exponential family and the dimension of its parameters θ (or of its sufficient statistic $B(x)$) is two. We

also have the following one-to-one correspondence between the natural parameter θ and the original parameters s and p:

$$s = \theta_2 \ , \quad p = \frac{g_s(0)}{g_s(0) + e^{\theta_1}} = \frac{2(s/\pi)^{1/2}}{2(s/\pi)^{1/2} + e^{\theta_1}} \ .$$

This distribution, called *positive mixed-state Gaussian distribution*, will be denoted by $\mathcal{G}_m(p, s)$.

3.2.2. Mixed-State Auto-Models for Motion Textures. To construct auto-models for mixed-state observations $\{v_{res}\}$, we start by assuming that the family of conditional distributions $\mu_i(x_i|\cdot)$ belongs to the family of positive Gaussian mixed-state distribution $f_{\theta_i(\cdot)}(x_i)$ given in (8). Here the parameter $\theta_i(\cdot)$ is a function of neighbouring configuration $(\cdot) = (x_j, j \neq i)$. In other words, we assume that

$$\log \mu_i(x_i|\cdot) = \langle \theta_i(\cdot), B(x_i) \rangle + \log p_i(\cdot) \qquad (9)$$

with $B(x) = (\delta^*(x), -x^2)$.

By Theorem 1, there are a family of vectors $\alpha_i = (a_i, b_i) \in \mathbb{R}^2$ and 2×2 matrices

$$\beta_{ij} = \begin{pmatrix} c_{ij} & d_{ij} \\ d^*_{ij} & e_{ij} \end{pmatrix} \ ,$$

satisfying $\beta_{ij} = \beta^T_{ji}$, such that

$$\theta_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B(x_j) \ . \qquad (10)$$

Moreover, the associated energy function is given by

$$Q(x_1, \ldots, x_n) = \sum_{i \in S} \left[a_i \delta^*(x_i) - b_i x_i^2 \right] + \sum_{\{i,j\}} \left(\delta^*(x_i), -x_i^2 \right) \beta_{ij} \left(\delta^*(x_j), -x_j^2 \right)^T \ . \qquad (11)$$

Let us describe in more details the local conditional distributions $\{\mu_i\}$. By construction, at each site i, the conditional distribution $\mu_i(x_i|\cdot)$ is $\mathcal{G}_m(p_i(\cdot), s_i(\cdot))$ with parameters

$$\theta_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B(x_j) = \left[\log \left(\frac{[1 - p_i(\cdot)] s_i(\cdot)}{p_i(\cdot)} \right), \ s_i(\cdot) \right]^T \ .$$

More explicitly

$$\theta_{i,1}(\cdot) = a_i + \sum_{j \neq i} \left[c_{ij} \delta^*(x_j) - d_{ij} x_j^2 \right] , \qquad (12)$$

$$\theta_{i,2}(\cdot) = b_i + \sum_{j \neq i} \left[d^*_{ij} \delta^*(x_j) - e_{ij} x_j^2 \right] . \qquad (13)$$

We have in particular

$$s_i(\cdot) = \frac{1}{2\sigma_i^2(\cdot)} = \theta_{i,2}(\cdot) \ ,$$

$$p_i(\cdot) = \frac{2[s_i(\cdot)/\pi]^{1/2}}{2[s_i(\cdot)/\pi]^{1/2} + e^{\theta_{i,1}(\cdot)}} \ .$$

It follows that necessarily for all i and its possible neighbouring configuration $(\cdot) = (x_j, j \neq i)$, the variance parameter $s_i(\cdot) = 1/[2\sigma_i^2(\cdot)]$ of the Gaussian component must be positive, i.e.

$$s_i(\cdot) = \frac{1}{2\sigma_i^2(\cdot)} = b_i + \sum_{j \neq i} \left[d^*_{ij} \delta^*(x_j) - e_{ij} x_j^2 \right] > 0 \ .$$

As x is arbitrary, this is equivalent to require the

Conditions [D]:
(i) for all $\{i, j\}$, $e_{ij} \leq 0$.
(ii) for all i and any subset $A \subset S \backslash \{i\}$,
$b_i + \sum_{j \in A} d^*_{ij} > 0$ (in particular $b_i > 0$).

It turns out that these necessary conditions are also sufficient for the admissibility (7) of the energy function Q given in (11) The next proposition is important: it defines precisely the set of parameter values corresponding to a valid definition of positive Gaussian auto-models.

Proposition 1. *Under the conditions [D], the energy function Q is admissible. Consequently, the associated positive Gaussian auto-model is well-defined.*

Proof: We need only prove the admissibility of the energy function Q, since the last conclusion follows from Theorem 3.1.

For any subset $A \subset S$ and a configuration x, we denote by x_A the trace of x on A: $x_A = (x_i, i \in A)$. The configuration space Ω can be decomposed as

$$\Omega = \sum_{A \subset S} \Omega_A$$

with

$$\Omega_A = \{x : x_i > 0,\ i \in A\ ;\quad x_i = 0,\ i \notin A\}\ .$$

Moreover for $x \in \Omega_A$,

$$Q(x) = \phi_A(x_A) := \sum_{i\in A}\left[a_i - b_i x_i^2\right] + \sum_{\{i,j\}\subset A}\left(1, -x_i^2\right)\beta_{ij}(1, -x_j^2)^T\ .$$

Therefore

$$\int_\Omega \exp Q(x)\nu(dx) = \sum_{A\subset S}\int_{\Omega_A}\exp Q(x)\prod_{i=1}^{n}\{\delta_0(dx_i) + \lambda(dx_i)\} = \sum_{A\subset S}\int_{(0,\infty)^{|A|}}\exp\phi_A(x_A)\prod_{i\in A}\lambda(dx_i)\ ,$$

where we have set $|A| = \mathrm{card}(A)$. As

$$(1, -x_i^2)\beta_{ij}\left(1, -x_j^2\right)^T = c_{ij} - d_{ij}x_j^2 - d_{ij}^* x_i^2 + e_{ij}x_i^2 x_j^2\ ,$$

and $e_{ij} \le 0$, we have for some constant $C > 0$,

$$C^{-1}\phi_A(x_A) \le -\sum_{i\in A} b_i x_i^2 - \sum_{\{i,j\}\subset A}\left(d_{ij}x_j^2 + d_{ij}^* x_i^2\right)\ .$$

By the symmetry property $d_{ij}^* = d_{ji}$, the second sum in the r.h.s is exactly

$$\sum_{i\in A}\sum_{j\in A,\ j\neq i} d_{ij}^* x_i^2\ .$$

Hence

$$C^{-1}\phi_A(x_A) \le -\sum_{i\in A}\left(b_i + \sum_{j\in A,\ j\neq i} d_{ij}^*\right)x_i^2\ .$$

By Conditions **[D]**, $b_i + \sum_{j\in A,\ j\neq i} d_{ij}^* > 0$, so that

$$\int_{(0,\infty)^{|A|}}\exp\phi_A(x_A)\prod_{i\in A}\lambda(dx_i) < \infty\ .$$

The proof is then complete. □

3.3. *A Specification for the Four-Nearest-Neighbour System*

We describe in this section a particular positive Gaussian auto-model using the four-nearest-neighbour system. Recall that the set of sites is $S = \{1, \ldots, n\} = [1, M] \times [1, N]$ and the neighbour system is to be completed with usual correction on the boundary. We denote by $\{i_e = i + (1, 0), i_o = i - (1, 0), i_n = i + (0, 1), i_s = i - (0, 1)\}$ the four neighbours of i.

Furthermore, we assume that the field is homogeneous in space, i.e., the parameters are the same for all sites. Moreover we will allow possible anisotropy between the horizontal and vertical directions. Under all these considerations and by the previous results, there exist a vector $\alpha = (a, b)$ and two 2×2 matrices $\beta^{(1)}$ and $\beta^{(2)}$ such that $\forall i,\ \alpha_i = \alpha$, and for $\forall\{i, j\},\ \beta_{ij} = 0$ unless i and j are neighbours where

$$\beta_{i,i_e} = \beta^{(1)} = \begin{pmatrix} c_1 & d_1 \\ d_1^* & e_1 \end{pmatrix} = \beta_{i_o,i}^T,$$

$$\beta_{i,i_n} = \beta^{(2)} = \begin{pmatrix} c_2 & d_2 \\ d_2^* & e_2 \end{pmatrix} = \beta_{i_s,i}^T,$$

The model has then 10 parameters $(a, b, c_1, d_1, d_1^*, e_1, c_2, d_2, d_2^*, e_2)$. However, for the application developed in Section 5, we need to further constrain the parameters $d_k,\ d_k^*$ and e_k, $k = 1, 2$ to be zero, since otherwise with $d_k > 0$, $d_k^* > 0$ or $e_k < 0$, the correlation between neighbouring sites becomes negative, i.e., the field is repulsive and neighbouring sites are "in competition" (note this is also typical for classical auto-models of [3], see also [10] for more details). This is clearly not suited for homogeneous motion textures we intend to analyze here.

Finally, this auto-model with four-nearest-neighbour system has four parameters, namely $\phi = (a, b, c_1, c_2)$. The admissibility condition **[D]** is reduced in the present case to the unique simple condition

$$b > 0\ .$$

It is useful to note that the parameters of the local conditional distributions $\mu_i = \mathcal{G}_m(p_i(\cdot), s_i(\cdot))$ at a site i take the form

$$\theta_{i,1}(\cdot) = a + c_1\left[\delta^*\left(x_{i_e}\right) + \delta^*\left(x_{i_o}\right)\right] + c_2\left[\delta^*\left(x_{i_n}\right) + \delta^*\left(x_{i_s}\right)\right]\ . \qquad (14)$$

$$\theta_{i,2}(\cdot) = b$$

Notice that the effect of a non null value x_j from neighbouring sites depends on the indicator value $\delta^*(x_j)$ independently of its magnitude $|x_j|$ (the squares x_j^2 are absent in the above formula).

In case we impose the equality $c_1 = c_2 = c$, we get an isotropic model with three parameters $\phi = (a, b, c)$. This model will also be used in Section 5 to test the existence or not of a spatial isotropy.

As for the estimation of the parameter ϕ, we use the pseudo-likelihood method by maximizing the pseudo-likelihood (in fact its logarithm)

$$L(x; \phi) = \sum_{i \in S} \log \mu_i(x_i \mid x_j, \ j \neq i) \ . \qquad (15)$$

This method has good consistency properties for classical one-parameter auto-models, see e.g. [9]. We conjecture that it is still the case for multi-parameter auto-models considered here, although we are not aware of any proof of such consistency.

4. Simulation Experiments

In this section we propose some simulation experiments to explore the basic properties of the auto-model with the four-nearest-neighbour system defined in Section 3.3. Our approach is analogous to [5] for the analysis of spatial intensity textures.

First, Fig. 5 presents two realizations of the model on a 128×128 lattice with parameter values $(a, b, c_1, c_2) = (0.6, 3, 0.5, 0.5)$ and $(a, b, c_1, c_2) = (0.6, 0.5, 0.5, 0.5)$, respectively. The simulations are produced with 100 scans of the Gibbs sampler where the initial configuration is chosen at random.

Recall that our figures use a reversed gray color-map so that the white color corresponds to the null value of the field. Note also that this model has four parameters and in particular, the variance parameter $\sigma_i^2(\cdot)$ of the Gaussian component of the conditional distributions is a constant independently of neighbouring configurations. Furthermore, this conditional variance becomes smaller for larger value of b so that the values of the field is more concentrated near the origin and the figure looks whiter.

We observe also that the contrast is generally low on these figures. As a consequence and to have a better visual judgment, we will display simulations on a binary basis in the sequel: all non null values are printed with an unique black color, while the white color is kept to identify the null value of the field.

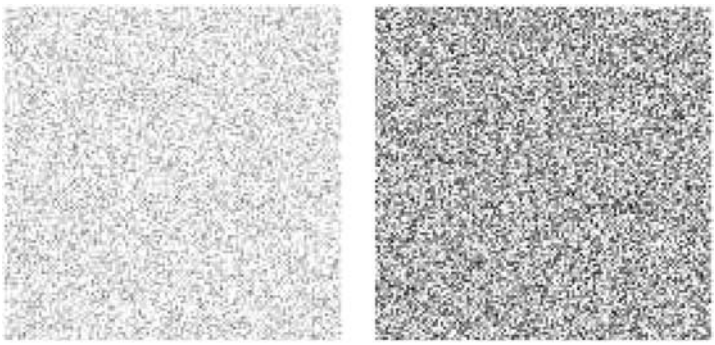
Next, we explore the different roles played by the remaining parameters a, c_1 and c_2.

4.1. *Influence of the Parameter a*

Figure 6 displays two realizations both with $(b, c_1, c_2) = (1, 1.5, 1.5)$ but with different values of $a = 2.9$ and $a = -2.9$, respectively. When a increases, the conditional probability to have a non null value of the field becomes higher. Therefore a controls the density of non null values of the field.

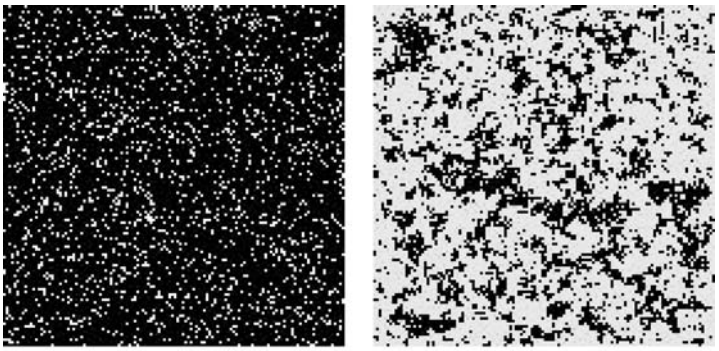
4.2. *Texture Granularity*

Figure 7 displays two realizations which share the same parameter values $(a, b) = (-2.9, 1)$ but with interaction parameters $c_1 = c_2 = 1.6$ and $c_1 = c_2 = 3$ respectively. As one can see, a higher value of the



(a) (b)

Figure 5. Realizations of size 128×128 by the Gibbs sampler of the auto-model with the four-nearest-neighbour systems. (a). $(a, b, c_1, c_2) = (0.6, 3, 0.5, 0.5)$. (b). $(a, b, c_1, c_2) = (0.6, 0.5, 0.5, 0.5)$.



(a) (b)

Figure 6. Influence of the parameter a. Common parameter values are $(b, c_1, c_2) = (1, 1.5, 1.5)$ and the different one is (a) $a = 2.9$, (b) $a = -2.9$.

c_k's produces finer white grains corresponding to points with the symbolic value 0.

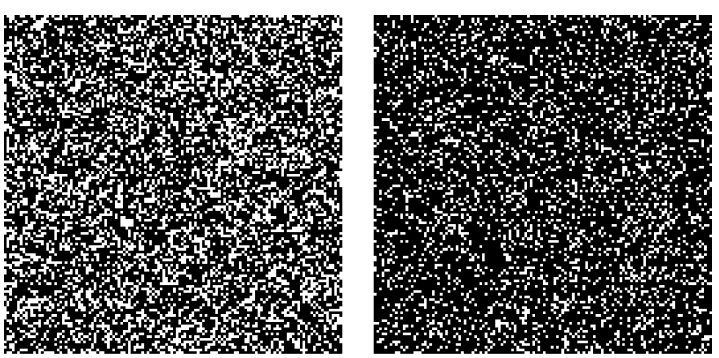
4.3. Texture Orientations

Figure 8 displays two realizations both with $(a, b) = (-2.9, 1)$ but with interaction parameters $(c_1, c_2) = (3, 0)$ and $(c_1, c_2) = (0, 3)$ respectively. The opposite signs in the c_k's produces a well-organized orientation in the textures.

In the last simulation experiment, we aim at checking the consistency of the pseudo-likelihood estimator. We then simulated 200 realizations of the model with $(a,\ b,\ c_1,\ c_2) = (0.6,\ 3,\ 0.5,\ 0.5)$ and computed the parameter estimates from these realizations. The average estimates from these 200 estimations are (0.6177, 2.9971, 0.4945, 0.4960) with respective standard deviations (0.1502, 0.0807, 0.0828, 0.0896). Therefore the pseudo-likelihood estimator is clearly consistent although the standard deviation of the parameter a seems slightly higher than the others.

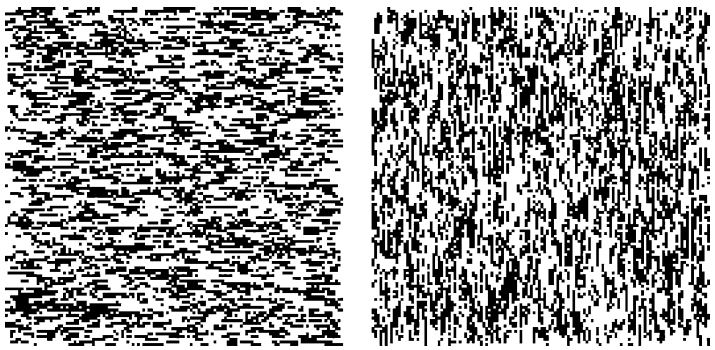
5. Applications to Motion Texture Analysis

As mentioned in Section 2, temporal textures (or dynamic textures) designate video contents involving natural (almost stationary) dynamic phenomena such as rivers, sea-waves, smokes, steams, fires, fountains, moving grass or foliage, etc. No tractable 3D kinematic models can be exhibited to account for these motions and to allow the derivation of relevant and efficient image motion models. Therefore, the analysis of dynamic textures is a challenging issue while of practical interest for various applications. This problem has been mainly investigated by considering these image



(a) (b)

Figure 7. Texture granularity. Common parameter values are $(a, b) = (2.9, 1)$ and the different ones are (a) $c_1 = c_2 = 1.6$, (b) $c_1 = c_2 = 3$.



(a) (b)

Figure 8. Orientation of the textures. Common parameter values are $(a, b) = (-2.9, 1)$ and the different ones are (a) $(c_1, c_2) = (3, 0)$, (b) $(c_1, c_2) = (0, 3)$.

sequences as visual signals only, that is by modeling the time-varying intensity function only. In the early work by [14] on temporal texture modeling, a spatio-temporal auto-regressive model was introduced which was also causal in the spatial domain and could handle a restricted range of motion contents only. A significant extension has been then designed by [6] exploiting ARMA models and system identification tools. Issues of modeling, learning, recognizing, compressing or synthesizing dynamic textures were addressed with this modeling framework. Recently, multi-scale AR models have been applied to this problem by [7], along with a closed-loop linear dynamic system by [16]. However, these methods present two main limitations: they consider linear models only and they operate on the pixel intensities.

Mixed-state auto-models allow us to specify non-linear models, to take into account the spatial context, and to introduce both symbolic information (no motion) and continuous motion values, which is of great interest to handle dynamic textures. Furthermore, we do not model the time-varying intensity function but the motion measurements themselves. Thus, the designed models are intrinsic to the motion content of the video. Consequently, we prefer to use the term "motion texture" in that context.

We report hereafter a set of preliminary results on the modeling of motion textures. All experiments of this section are made with the positive Gaussian mixed-state auto-model introduced in Section 3.3.

5.1. Isotropy and Anisotropy

We first examine how the introduced auto-models can account for a fundamental characteristic of a homogeneous texture, namely spatial isotropy or spatial anisotropy when applied to real motion textures. For the positive Gaussian auto-models used here, see Eq. 14 for instance, the isotropy occurs if (and only if) $c_1 = c_2$.

Test experiment 1: The first dynamic texture we consider represents motion from trees. Such a sample image and the associated motion texture $\{v_{res}\}$ are shown in Fig. 9. Moreover, this test motion texture from trees is believed to be spatially isotropic.

We have estimated the full positive Gaussian model using one $\{v_{res}\}$ map of the tree sequence. The pseudo-likelihood estimates of the 4 parameters $\phi = (a, b, c_1, c_2)$, as defined in (15), are reported in Table 1.

From Table 1, we see that the parameters c_1 and c_2 are almost identical (with regard to standard deviations of these estimates we have computed from similar motion texture maps computed at other instants of the same tree sequence). Therefore, spatial isotropy is well reflected here by the equality between the parameters $\{c_k\}$. This statement is further confirmed by the estimated *isotropic* positive Gaussian model with three parameters $\phi = (a, b, c)$. The corresponding estimates are given in the bottom row of Table 1. Clearly the value of c is very close to the c_k's found above for the full model.

Table 1. Parameter estimates: the full model (top row) and the isotropic model (bottom row)

Full model	a	b	c_1	c_2
	−5.8049	3.0435	3.0568	2.9541
Isotropic model	a	b	c	
	−5.7813	3.0441	3.0000	

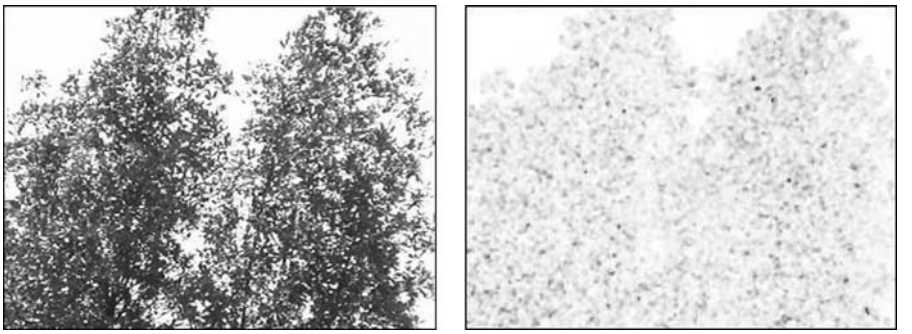


Figure 9. Sample image of "tree" sequence and the associated map of local motion measures (white = 0; black = maximum value).

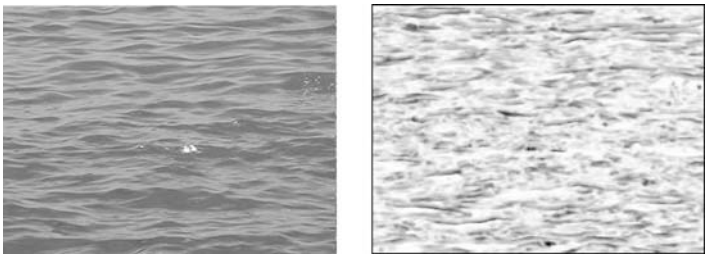


Figure 10. Sample image of the "sea-waves" sequence and the associated normal flow magnitudes map.

Test experiment 2: We next consider a motion texture from a video sequence involving sea-waves. Since these waves are clearly oriented, we should have an anisotropic motion texture in presence. A sample image and the associated motion measures $\{v_{res}\}$ are given in Fig. 10.

As in Experiment 1, we have estimated the full positive Gaussian auto-model. The parameter estimates are given in Table 2. Clearly, the difference between c_1 and c_2 is significant. Therefore, by these differences, the fitted model is able to reflect the spatial anisotropy of the considered motion texture.

5.2. *Spatial Stationarity*

Here, we propose to analyze another characteristic feature of motion textures, namely the spatial stationarity. To this end and for a given texture, we divided the motion map into 12 disjoint blocks of size 65×65 pixels each. We then fit the full positive Gaussian auto-model to each of the blocks.

Table 2. Parameters of the full model for motion from sea-waves.

Full model	a	b	c_1	c_2
	−7.9412	0.3697	5.7920	1.4219

Test experiment 3: We process here sea-waves images. Figure 11 shows the blocks $B_1 \ldots B_{12}$ extracted from a motion map $\{v_{res}\}$ at a given time instant of the video sequence.

The estimates of the parameters for these blocks are given in Table 3, where the standard deviations for these 12 sets of estimates are also computed (bottom row).

From Table 3, we can see that the 12 sets of parameters are nearly the same, taking into account the associated standard deviations. This then confirms the spatial stationarity believed in motion textures from sea-waves.

Test experiment 4: We have conducted a similar experiment with a motion texture from a video depicting a river (Fig. 2). Sample blocks $B_1 \ldots B_{12}$ of motion textures are illustrated in Fig. 12.

The estimates of the parameters for these blocks are given in Table 4, where the standard deviations from these 12 sets of estimates are also computed (bottom row).

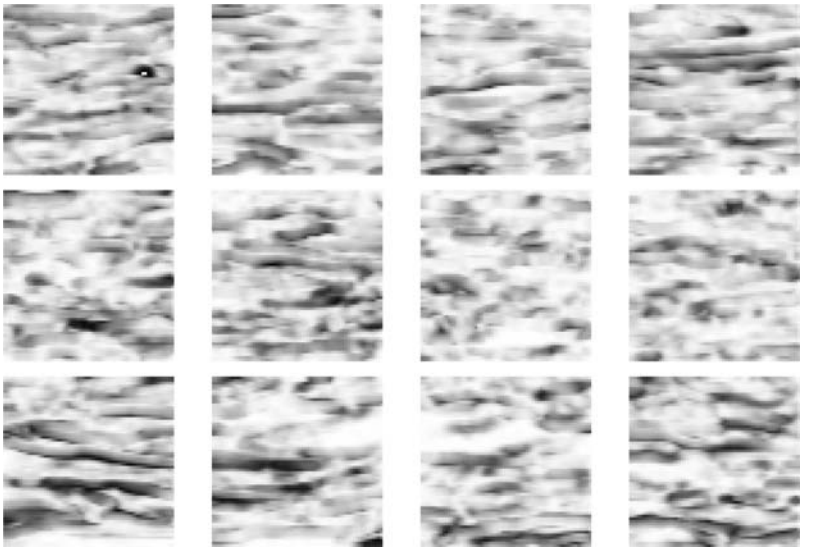


Figure 11. Sea-waves sequence: Set of disjoint blocks of local motion measures of size 65×65 at a given time instant of the sequence. Top to bottom and left to right: $B_1 \ldots B_{12}$.

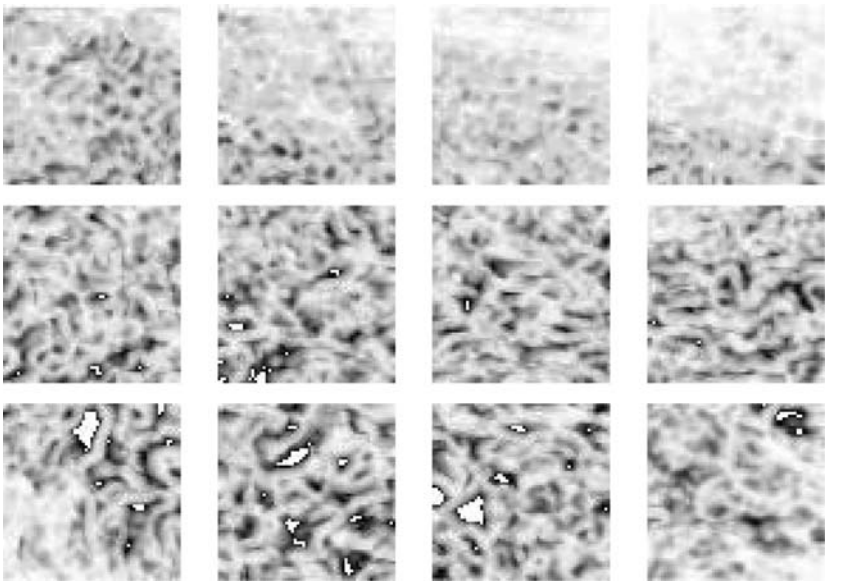


Figure 12. River sequence: Set of disjoint blocks of local motion measures of size 65×65 at a given time instant of the sequence. Top to bottom and left to right: $B_1 \ldots B_{12}$.

An overall impression from Table 4 is that these 12 sets of parameters are significantly different, resulting in much bigger standard deviation values as compared to Table 4 in the previous experiment. This difference is particularly clear between 3 block lines (of 4 blocks each). Therefore, we can reasonably assert that the non-stationarity of the river motion texture is correctly reflected by different parameter estimates of the proposed positive Gaussian mixed-state auto-model over the blocks.

Table 3. *Sea-waves*: Estimates of the full model for 12 blocks at a same time instant of a motion texture from a sea-waves video. The standard deviation of these 12 sets of estimates are given in the bottom row.

	a	b	c_1	c_2
B_1	−9.30205	0.29694	5.87898	2.13676
B_2	−9.39952	0.32878	5.46864	2.72478
B_3	−9.04816	0.34152	7.23150	1.14051
B_4	−9.60199	0.32897	7.33581	1.36095
B_5	−8.91001	0.37100	5.65410	2.04668
B_6	−7.35726	0.39963	5.74128	1.10768
B_7	−7.57434	0.43946	5.24632	1.71629
B_8	−7.47818	0.58794	5.08877	1.85792
B_9	−8.30468	0.36271	6.36993	1.18090
B_{10}	−7.61364	0.30174	6.41590	0.70918
B_{11}	−8.86299	0.28625	7.59329	0.65160
B_{12}	−8.87836	0.32870	5.83942	1.85034
st. d.	0.8220	0.0830	0.8403	0.6203

Table 4. *River-sequence:* Estimates of the full model for 12 blocks at a same time instant of a motion texture from a river video. The standard deviation of these 12 sets of estimates are given in the bottom row.

	a	b	c_1	c_2
B_1	−10.02623	0.36189	4.76893	3.48537
B_2	− 8.39303	0.44575	5.71362	1.93230
B_3	− 7.10429	0.69088	3.51732	3.46525
B_4	− 5.69048	0.87768	3.22945	2.40213
B_5	17.13015	0.11962	21.95125	10.82028
B_6	8.17958	0.11417	13.38513	7.91897
B_7	8.27053	0.10669	13.18007	8.31032
B_8	8.18484	0.12814	13.56948	8.01166
B_9	−11.78903	0.11195	9.79198	1.06964
B_{10}	8.04156	0.07512	13.68093	7.89088
B_{11}	− 3.50982	0.09938	11.51522	−4.15503
B_{12}	−12.91980	0.11302	4.74714	5.13039
st.d.	10.1118	0.2690	5.6741	4.1530

6. Conclusion

For analysis of mixed-state observations such as motion measurements in an image sequence, we have introduced a new class of random field models, namely mixed-state auto-models. This approach is made possible by extending Besag's one-parameter auto-models to the multi-parameter case. We have provided a careful construction of these models as well as a detailed discussion about their basic properties.

We have proposed a special mixed-state auto-model relying on a positive Gaussian distribution for the continuous component. This model was first explored via several simulation experiments where the model is proven to be able to reproduce important texture characteristics such as density, contrast, granularity or spatial orientation. The performance of the model is further experimentally validated by analyzing different motion textures from real images of natural scenes.

Although the adopted positive Gaussian model (with the four-nearest-neighbour system) of Section 3.3 seems somewhat crude, our experiments proved that important texture characteristics like spatial isotropy or anisotropy, as well as spatial stationarity or non-stationarity can be correctly reflected by the model parameters.

There are several unanswered questions which need further investigations. First, from a theoretic point of view, the convergence of the pseudo-likelihood estimators has to be established. Also, some efficient Monte-Carlo simulation algorithms of mixed-state auto-models have to be designed in the vein of the theory of Markov random fields.

Secondly, as for the analysis of motion textures we proposed here, it is clear that the adopted positive Gaussian model need to be improved. A first possibility is to enrich spatial correlations with a larger neighbour system, typically a eight nearest-neighbours system. A second possibility that we are currently considering is to include non zero-mean Gaussian or non-Gaussian distributions for the continuous component of motion measurements. Such richer family of distributions would probably further improve model fits for various motion textures. In particular, spatial correlation between continuous motion values will be taken into account as well. Furthermore, it is worth mentioning the possibility to adapt the mixed-state modeling to other type of motion measurements like the signed normal flow $\{v_n(p,t)\}$ or the dense optical flow$\{\vec{\mathbf{v}}(p,t)\}$.We will also exploit this non-linear modeling framework to investigate issues such as the classification or the segmentation of motion textures in videos.

References

1. D.J. Allcroft and C.A. Glasbey, "A latent gaussian markov random-field model for spatiotemporal rainfall disaggregation," *J. Roy. Statist. Soc. Ser. C*, Vol. 52, No. 4, pp. 487–498, 2003.
2. M. Bartlett, "A further note on nearest neighbour models," *J. Roy. Statist. Soc.*, Vol. A, No. 131, pp. 579–580, 1968.
3. J. Besag, "Spatial interactions and the statistical analysis of lattice systems," *J. Roy. Statist. Soc.*, Vol. B, No. 148, pp. 1–36, 1974.
4. B. Chalmond, *Modeling and Inverse Problems in Image Analysis*, Vol. 155 of *Applied Mathematical Sciences*. Springer-Verlag: New York, 2003.
5. G. Cross and A. Jain, "Markov random field texture models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 5, No. 1, pp. 25–39, 1983.

6. G. Doretto, A. Chiuso, Y. Wu, and S. Soatto, "Dynamic textures," *International Journal of Computer Vision*, Vol. 51, No. 2, pp. 91–109, 2003.
7. G. Doretto, E. Jones and S. Soatto, "Spatially homogeneous dynamic textures," in *Proc. 8th European Conf. on Computer Vision*, LNCS 3022. Springer: Prague, 2004, pp. 591–602.
8. R. Fablet and P. Bouthemy, "Motion recognition using non parametric image motion models estimated from temporal and multiscale cooccurrence statistics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 122, pp. 1619–1624, 2003.
9. X. Guyon, *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer-Verlag: New York, 1995.
10. C. Hardouin and J. Yao, "Multi parameter auto-models and application to mixed state data analysis," Technical report, IRMAR/Université de Rennes 1, 2005.
11. M. Irani, B. Rousso, and S. Peleg, "Detecting and tracking multiple moving objects using temporal integration," in *Proc of European Conf. on Computer Vision, ECCV'92*. Springer-Verlag: Santa Margherita 1992, pp. 282–287.
12. J. Odobez and P. Bouthemy, *Separation of Moving Regions from Background in an Image Sequence Acquired with a Mobile Camera*, Chapt. 8, pp. 283–311, Video Data Compression for Multimedia Computing, H.H. Li,S. Sun and H. Derin (Eds.). Kluwer Academic Publisher, 1997.
13. J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *J. of Visual Comm. and Image Repr.*, Vol. 6, No. 4, pp. 348–365, 1995.
14. M. Szummer and R. Picard, "Temporal texture modeling," in *Proc. IEEE Int. Conf. on Image Processing, ICIP'96*, Lausanne, 1996.
15. P. Whittle, "Stochastic processes in several dimensions," *Bull. Inst. Statist. Inst.*, Vol. 40, pp. 974–994, 1963.
16. L. Yuan, F. Wen, C. Liu, and H.-Y. Shum, "Synthesizing dynamic texture with closed-loop linear dynamic system," *in Proc. 8th European Conf. on Computer Vision*, LNCS 3022. Prague: Springer, 2004, pp. 603–616.

Patrick Bouthemy graduated from ENST, Paris, in 1980, and received the Ph.D degree in Computer Science from the University of Rennes, France, in 1982. From December 1982 until February 1984, he was employed by INRS-Telecommunications, Montreal, in the Department of Visual Communications. Since April 1984 he has been with INRIA, at IRISA in Rennes. He is currently "Directeur de Recherche" Inria and head of Vista group. His main research interests are: statistical approaches for image sequence processing, motion analysis, learning, recognition and classification of dynamic contents in video. He has served as member of the program committees of the major conferences in image processing and computer vision. He was Associate Editor of the IEEE Transactions on Image Processing from 1999 to 2003.

Cécile Hardouin received the Ph.D degree in Applied Mathematics from University Paris 7, France, in 1992. Since July 1992 she is lecturer at University Paris 10, Nanterre. Since her Ph.D., she is member of the research center SAMOS - MATISSE, located in University Paris 1. After some works on long memory processes, her main research interests are now Spatial statistics: Markovian random fields, standards adoption dynamics, spatial coordination.

Gwenaëlle Piriou was born in 1976. She received the Ph.D. degree in signal processing and telecommunications from the university of Rennes, France, in 2005. She is currently an assistant professor with the computer science department of university of Rennes. Her main research interests are probabilistic motion modeling in image sequence, recognition and detection of dynamic content and temporal textures.

Jian-feng Yao recieved the Ph.D. degree in Applied Mathematics from Université Paris-Sud (Orsay), France, in 1990. From January 1990 to August 2000, he was a Maître de Conference in University of Paris I. Since September 2000, he is a full professor in Statistics in University of Rennes I. His current research interests are nonlinear time series analysis, spectral theory of large-dimensional random matrices and mathematical modeling for image understanding.